

TÉCNICAS E INSTRUMENTOS DE RECOGIDA Y ANÁLISIS DE DATOS

Catalina Martínez Mediano
(*Coordinadora*)

Educación
Social



Catalina Martínez Mediano
(Coordinadora)

TÉCNICAS E INSTRUMENTOS DE RECOGIDA Y ANÁLISIS DE DATOS



UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

TÉCNICAS E INSTRUMENTOS DE RECOGIDA Y ANÁLISIS DE DATOS

Quedan rigurosamente prohibidas, sin la autorización escrita de los titulares del Copyright, bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler o préstamos públicos.

© UNIVERSIDAD NACIONAL
DE EDUCACIÓN A DISTANCIA - Madrid, 2014

www.uned.es/publicaciones

© Catalina Martínez Mediano (de toda la obra menos del tema 9) y
Arturo González Galán (tema 9)

ISBN electrónico: 978-84-362-6822-5

Edición digital: junio de 2014

ÍNDICE

PRESENTACIÓN

1. La asignatura	19
1.1. Los objetivos.....	20
1.2. El temario de contenidos.....	22
1.3. Los medios y recursos	23
1.4. La evaluación	24

PRIMERA UNIDAD DIDÁCTICA

LOS MÉTODOS DE INVESTIGACIÓN EN EDUCACIÓN SOCIAL

Tema 1. La investigación en educación social. Tipos y enfoques.....

.....	31
1.1. El conocimiento científico y las ciencias	33
1.2., Tipos, de, investigación, en, educación, y, en, Ciencias, Sociales	
1.2.1. Tipos de investigación según el grado de abstracción: básica y aplicada	35
1.2.2. Tipos de investigación según el proceso formal: deductivo, inductivo, hipotético deductivo	36
1.2.3. Tipos de investigación según la intención: descriptiva, comparativa, evaluativa, explicativa	37
1.2.4. Tipos de investigación según el momento: histórica y actual	38
1.3. Los diferentes paradigmas de investigación	39
1.3.1. El paradigma positivista y neopositivista	40
1.3.2. El paradigma crítico	41

1.3.3. El paradigma interpretativo o naturalista	42
1.4. El debate metodológico y la complementariedad metodológica	
1.5. Ejercicios de autoevaluación	44
1.6., Solución,a,los,ejercicios,de,autoevaluación	
1.7. Referencias bibliográficas	45
Tema 2. La investigación descriptiva	47
2.1., Introducción	
2.2. La investigación tipo encuesta.....	49
2.3. El proceso de realización de una investigación tipo encuesta...	50
2.3.1., Definición,del,problema,de,investigación	
2.3.2. Revisión de la literatura relevante al problema de estudio	51
2.3.3., Selección,de,la,muestra	
2.3.4. Selección de la técnica de encuesta	52
2.3.5., Construcción,del,instrumento,de,la,encuesta	
2.3.6. Los procedimientos del trabajo de campo.....	53
2.3.7., Realización,de,una,aplicación,piloto	
2.3.8. El análisis de datos descriptivos.....	54
2.3.9. La generalización de los resultados	56
2.3.10. El informe de los resultados de investigación	57
2.4., La,investigación,exploratoria	
2.4.1. Estudios de relación y basados en el análisis factorial....	58
2.4.2. Estudios de predicción	59
2.4.3. Estudios causales comparativos o investigación «ex-post-facto»	59
2.5., Ejercicios,de,autoevaluación	
2.5.1., Conceptos,fundamentales,del,tema,	
2.5.2., Síntesis,del,tema	
2.5.3., Cuestiones,sobre,los,contenidos,del,tema,	
2.6., Soluciones,a,los,ejercicios,de,autoevaluación	
2.7. Referencias bibliográficas	60
Tema 3. La investigación experimental	61
3.1., Introducción	
3.2. Funciones y objetivos de los diseños de investigación	63
3.3. Clasificación de los diseños.....	65
3.4. Criterios para valorar los diseños: su validez.....	67
3.4.1., La,validez,interna	
3.4.2. La validez de constructo	68
3.4.3. La validez estadística	69

3.4.4. La validez externa	69
3.5. El control de las variables extrañas	70
3.6. El proceso general de investigación experimental.....	72
3.6.1., Formulación,del,problema	
3.6.2. Revisión de la literatura sobre el tema de investigación .	73
3.6.3. Formulación de las hipótesis.....	74
3.6.4. Constitución del plan experimental	76
3.6.5., Realización,del,experimento	
3.6.6. Procedimientos para el análisis de los datos	78
3.6.7. Interpretación de los resultados y conclusiones	79
3.6.8. El informe de investigación.....	80
3.7. Ejercicios de autoevaluación	83
3.8., Solución,a,los,ejercicios,de,autoevaluación	
3.9. Referencias bibliográficas	84
Tema 4. La investigación cualitativa	85
4.1., Introducción	
4.2. Características de la investigación cualitativa	87
4.3. La investigación cualitativa aplicada a la Educación.....	89
4.4. La investigación cualitativa y la investigación experimental.	
Criterios para su comparación	90
4.4.1.,Teoría,primero,o,teoría,después	
4.4.2. Análisis inductivo vs. deductivo.....	91
4.4.3. Perspectiva holística	92
4.4.4.,Preconcepciones,o,posconcepciones	
4.4.5. Selección y muestreo.....	93
4.4.6. Contacto personal.....	94
4.4.7.,Triangulación,y,convergencia,	
4.4.8. Comparabilidad o generalización.....	95
4.5. El diseño etnográfico.....	98
4.6. El análisis de datos en la investigación cualitativa.....	100
4.6.1. Descubrimiento en progreso	101
4.6.2., Codificación,de,los,datos	
4.6.3. Relativización de los datos y conclusiones.....	102
4.7., Criterios,para,valorar,las,investigaciones,cualitativas,	
4.7.1. Criterios de credibilidad, seguridad, confirmabilidad y	
transferibilidad.....	104
4.7.2., Criterios,hermenéuticos,	
4.7.3. Criterios de autenticidad	106
4.8. Elaboración del informe de investigación.....	108
4.9., Ejercicios,de,autoevaluación,	
4.9.1. Conceptos fundamentales del tema	111

4.9.2., Cuestiones,de,síntesis	
4.9.3. Cuestiones sobre los contenidos del tema	111
4.10., Soluciones,a,las,actividades,de,autoevaluación	
4.11. Referencias bibliográficas.....	112

Tema 5. La investigación evaluativa. Evaluación de programas e instituciones.....	113
5.1. La evaluación de programas: Concepto, finalidad y funciones .	115
5.2. Cuestiones a tener en cuenta en el proceso de realización de la evaluación,	
5.2.1. Las características de los programas y de las instituciones	117
5.2.2. Para quién se hace la evaluación.....	118
5.2.3., Quién,debe,hacer,la,evaluación	
5.2.4. Sensibilización hacia la evaluación	119
5.2.5. Planificación de la evaluación	120
5.2.6. Qué información debe ser recogida	122
5.2.7. Los métodos en la evaluación.....	124
5.3. Los informes de la evaluación y la utilización de los resultados..	125
5.4. La realización de proyectos innovativos de mejora.....	126
5.5. Los criterios y normas para juzgar la evaluación	129
5.6., El,enfoque,de,calidad,total,y,el,modelo,de,excelencia,europeo	
5.6.1., El,concepto,de,la,calidad	
5.6.2. Los principios de la Calidad Total.....	131
5.6.3. El Modelo de Excelencia Europeo: Estructura y Criterios	134
5.6.4. La autoevaluación y el Modelo de Excelencia Europeo ..	137
5.6.5. La elaboración de los planes de mejora.....	138
5.7., Ejercicios,de,autoevaluación	
5.8. Referencias bibliográficas.....	140

SEGUNDA UNIDAD DIDÁCTICA

LA MEDIDA EN EDUCACIÓN Y LAS TÉCNICAS PARA LA RECOGIDA DE INFORMACIÓN

Tema 6. La medida en Educación y en Ciencias Sociales	145
6.1. El concepto de medida	147
6.2. Elementos a considerar en la medición	148
6.2.1., El,objeto,a,medir	
6.2.2. La elaboración del instrumento de medida.....	149
6.2.3. El acto de la medición	150

6.2.4. El tratamiento de la información.....	150
6.3., Problemática,de,la,medida,en,Educación,y,en,Ciencias,Sociales	
6.3.1. Del fenómeno a los conceptos, a las dimensiones o variables, a los indicadores y a los índices.....	151
6.3.2. Las variables y sus clases.....	154
6.3.3. Los indicadores	156
6.3.4. Los índices.....	157
6.4. Los niveles de medición: escalas nominal, ordinal, de intervalo y,de,razón	
6.4.1. Nivel de medida nominal.....	159
6.4.2. Nivel de medida ordinal.....	161
6.4.3. Niveles de medida de intervalo y de razón	163
6.5. Relación entre los indicadores, los índices, los niveles de medida y las pruebas estadísticas	164
6.6., Ejercicios,de,autoevaluación,	
6.6.1., Conceptos,fundamentales,del,tema	
6.6.2. Cuestiones de síntesis	165
6.6.3., Cuestiones,sobre,los,contenidos,del,tema	
6.7. Soluciones a los ejercicios de autoevaluación	166
6.8. Referencias bibliográficas	167

Tema 7. Las escalas para la medida de las actitudes. Las escalas de medida de Thurstone	169
7.1. Las escalas para la medida de las actitudes	171
7.1.1. Las «gradaciones ordinales».....	172
7.2. Las escalas de medida de Thurstone	175
7.2.1. La ley del Juicio Comparativo de Thurstone.....	177
7.2.2. La ley del juicio categórico de Thurstone.....	179
7.3. Construcción de una escala de intervalos de Thurstone	180
7.3.1., Definición,del,rasgo,a,medir	
7.3.2. Elaboración de los elementos iniciales.....	181
7.3.3. Clasificación de las proposiciones: prueba de los jueces.	182
7.3.4. Criterios para la selección de los elementos.....	183
7.3.5. Elaboración de la escala	187
7.3.6. Calificación de la escala.....	188
7.4. Fiabilidad y validez de la escala.....	189
7.5. Ejercicios de autoevaluación	190
7.6., Soluciones,a,los,ejercicios,de,autoevaluación	
7.7. Referencias bibliográficas	192

Tema 8. Las escalas para la medida de las actitudes de Likert

8.1. Las escalas de medida de Likert	
8.2. Construcción de una escala de estimaciones sumadas de Likert.	195
8.2.1. Elaboración de las proposiciones	
8.2.2. Valoración de las proposiciones	196
8.2.3. Asignación de puntuaciones a las proposiciones	197
8.2.4. Análisis de los ítems iniciales	198
8.2.5. La escala definitiva. Aplicación e interpretación de las puntuaciones	206
8.3. Fiabilidad y validez de la escala	207
8.4. Ejercicios de autoevaluación	210
8.5. Soluciones a los ejercicios de autoevaluación	
8.6. Referencias bibliográficas	211

Tema 9. Características técnicas de los instrumentos de medida desde la teoría clásica de los tests. Fiabilidad y validez (Arturo González Galán)

9.1. El concepto de fiabilidad	215
9.2. Tipos de fiabilidad	217
9.2.1. Fiabilidad como estabilidad	218
9.2.2. Fiabilidad como equivalencia	219
9.2.3. Fiabilidad como consistencia interna	220
9.3. Aspectos a tener en cuenta al valorar un coeficiente de fiabilidad	
9.3.1. Variabilidad y tipo de muestra	
9.3.2. Tipo de instrumento	222
9.3.3. Elección del coeficiente de fiabilidad y diferencias entre sus valores	223
9.3.4. Valores aceptables para el coeficiente « α » de Crombach	
9.3.5. Fiabilidad y unidimensionalidad: suma de elementos	224
9.4. Un ejemplo del cálculo de la fiabilidad de una escala con SPSS	225
9.5. El concepto de validez	227
9.6. Tipos de validez	229
9.6.1. Validez de contenido	230
9.6.2. Validez de criterio	231
9.7. Procedimientos para hallar la validez	232
9.8. Relación entre validez, fiabilidad y longitud de la prueba	237
9.9. Ejercicios de autoevaluación	238
9.10. Referencias bibliográficas	240

Tema 10. Procedimientos para la recogida de información I. La observación sistemática, el cuestionario la entrevista, y las pruebas de rendimiento.....	241
10.1. Los instrumentos para recogida de información en el proceso general,de,investigación	
10.2. La observación como procedimiento de representación de la realidad.....	243
10.2.1. Los sistemas de recogida de información en la observación sistemática	248
10.2.2. Las unidades de observación.....	250
10.2.3. Cuestiones sobre el muestreo en la observación	251
10.2.4. Técnicas de registro	252
10.2.5. Fuentes de error en las observaciones	253
10.3. El cuestionario y sus características.....	255
10.3.1. Construcción de los cuestionarios.....	256
10.3.2. Tipos de preguntas en los cuestionarios	258
10.4. La entrevista. Objetivos y características.....	262
10.4.1. Modalidades de la entrevista	263
10.5. Las pruebas de rendimiento. Evaluación normativa y evaluación,criterial	
10.5.1. Las pruebas de examen en la evaluación normativa....	265
10.5.2. Las pruebas de evaluación criterial.....	271
10.6., Ejercicios,de,autoevaluación	
10.7. Referencias bibliográficas	276
 Tema 11. Procedimientos para la recogida de información II. La observación participante, el análisis de contenido, la técnica Delphi y los grupos de discusión.....	 277
11.1. La observación participante. Orígenes y características	279
11.1.1. La función del observador en la observación participante.....	280
11.1.2. El diseño para la realización de la observación participante.....	283
11.1.3. El proceso de recogida de datos en la observación participante	284
11.1.4. El análisis de la información en la observación participante.....	288
11.1.5. La elaboración de informe de la observación participante.....	289
11.2. El análisis de contenido. Definición y características	290
11.2.1. La fases del análisis de contenido	292

11.2.2. Las unidades de análisis y su contexto	293
11.2.3. La inferencia como objetivo intelectual básico del análisis de contenido y la validez como criterio de éxito...	294
11.3. El análisis y la evaluación de necesidades	295
11.3.1. Las fases seguidas en la valoración de necesidades	296
11.4. La técnica Delphi. Características y posibilidad	297
11.4.1. El proceso de recogida de información y datos mediante Delphi	298
11.5. Los grupos de discusión	301
11.5.1. Aspectos a considerar para la formación y funcionamiento de los grupos de discusión	302
11.6. Ejercicios de autoevaluación	303
11.7. Referencias bibliográficas	304

TERCERA UNIDAD DIDÁCTICA

LA ESTADÍSTICA DESCRIPTIVA E INFERENCIAL PARA EL ANÁLISIS DE LA INFORMACIÓN. LOS DISEÑOS DE INVESTIGACIÓN

Tema 12. Distribuciones de variables y representaciones gráficas	309
12.1., Distribuciones, de, variables	
12.1.1. Distribuciones numéricas y reducción de los datos	311
12.1.2., La, distribución, binomial	
12.1.3. La distribución continua	314
12.1.4. La distribución normal de probabilidad o curva normal.....	315
12.2., Representaciones, gráficas	
12.2.1. Diagrama de barras	316
12.2.2. Histograma	318
12.2.3. Diagrama circular de sectores	321
12.2.4. Gráficas de frecuencias	322
12.3., Ejercicios, de, autoevaluación	
12.4. Solución a los ejercicios de autoevaluación	323
12.5. Referencias bibliográficas	325
Tema 13. Características de las distribuciones de frecuencias: medidas de tendencia central, de posición y de dispersión. Puntuaciones individuales	327
13.1. Medidas de tendencia central: la moda, la mediana y la media	329

13.2. Medidas de posición. Cuantiles	331
13.2.1. Percentiles, cuartiles, deciles y amplitud intercuartílica	332
13.3., Medidas, de, variabilidad, y, de, dispersión	
13.3.1., La, amplitud	
13.3.2. La desviación media	335
13.3.3. La varianza, la desviación típica y el coeficiente de variación	336
13.4., Las, puntuaciones, individuales	
13.4.1., Las, puntuaciones, diferenciales	
13.4.2. Las puntuaciones típicas	341
13.4.3. La curva normal	342
13.4.4. Normalización de las distribuciones	344
13.4.5. Asimetría y curtosis	347
13.5. Ejercicios de autoevaluación	349
13.6. Solución a los ejercicios de autoevaluación	352
13.7. Referencias bibliográficas	355
Tema 14. Correlaciones bivariadas	357
14.1. Las distribuciones bidimensionales y las relaciones	359
14.2. Las tablas de correlación y las tablas de contingencia	360
14.3. La covarianza	363
14.4. El diagrama de dispersión	368
14.5. Las medidas correlacionales bivariadas y su interpretación ...	370
14.5.1. El coeficiente de correlación producto momento de Pearson	375
14.5.2. Interpretación de la correlación producto momento de Pearson	377
14.5.3. Cálculo de la correlación producto momento de Pearson mediante el programa informático SPSS	378
14.5.4. El coeficiente de correlación ordinal de Spearman	380
14.5.5. Interpretación del coeficiente de correlación ordinal de Spearman	381
14.5.6. Cálculo de la correlación ordinal de Spearman mediante el programa informático SPSS	382
14.5.7. El coeficiente de contingencia o de asociación C de Pearson	383
14.5.8. Interpretación del coeficiente de asociación C de Pearson	385
14.6. Ejercicios de autoevaluación	386
14.7. Solución a los ejercicios de autoevaluación	389
14.8. Referencias bibliográficas	395

Tema 15. Probabilidad y muestreo. Distribución muestral y estimación de parámetros	397
15.1. La probabilidad.....	399
15.2. El muestreo	400
15.2.1., Población,y,muestra	
15.2.2., Aleatoriedad,y,aleatorización	
15.2.3. El muestreo y sus tipos	401
15.2.4. Representatividad y tamaño de la muestra	406
15.3., La,distribución,muestral	
15.3.1. Suma, media, varianza y distribución muestral.....	407
15.3.2. El error típico	408
15.3.3. El teorema del límite central	409
15.4. La estimación de parámetros.....	411
15.4.1. Estimación puntual y estimación por intervalos.....	412
15.4.2. El intervalo de confianza y los límites de confianza para la estimación de parámetros. Estimación de la media poblacional	413
15.4.3. Propiedades de los estimadores	416
15.5., Ejercicios,de,autoevaluación	
15.5.1. Conceptos fundamentales del tema	417
15.6. Solución a los ejercicios de autoevaluación.....	418
15.7. Referencias bibliográficas	419
 Tema 16. Estimación de la diferencia de dos medias. La prueba de hipótesis	
16.1. Estimación de la diferencia de dos medias	423
16.2. La prueba de hipótesis	427
16.2.1. El valor de probabilidad o valor crítico para la prueba de hipótesis	428
16.2.2. El valor de probabilidad o valor crítico unilateral para la,prueba,de,hipótesis	
16.3. Errores posibles en los que se pueden incurrir en las decisiones estadísticas en el contraste de hipótesis	433
16.4., Distribuciones,muestrales,asociadas,a,la,prueba,de,hipótesis	
16.4.1. La distribución del estadístico z y la del estadístico t de Student.....	435
16.4.2. La distribución del estadístico de χ^2	436
16.4.3. La distribución del estadístico F de Snedecor y la distribución,de,z	
16.5. El modelo estadístico y las pruebas para el contraste de hipótesis	437

16.5.1. Los modelos estadísticos paramétricos	437
16.5.2. Los modelos estadísticos no paramétricos	438
16.6. Ejercicios de autoevaluación	440
16.7. Solución a los ejercicios de autoevaluación.....	441
16.8. Referencias bibliográficas	442
Tema 17. Resolución de diseños de investigación de dos grupos	443
17.1., Los, diseños, de, investigación, de, dos, grupos	
17.2. Tipos de diseños de dos grupos y sus pruebas estadísticas para el contraste de hipótesis.....	445
17.3. Resolución de un diseño de investigación de dos grupos independientes y nivel de medida en la variable dependiente de intervalo mediante la prueba t de Student.....	446
17.4. Resolución de un diseño de investigación de dos grupos relacionados y nivel de medida en la variable dependiente de intervalo mediante la prueba t de Student.....	451
17.5. Resolución de un diseño de investigación de dos grupos independientes y nivel de medida en la variable dependiente ordinal mediante la prueba U de Mann-Whitney	455
17.6. Resolución de un diseño de investigación de dos grupos relacionados y nivel de medida en la variable dependiente ordinal mediante la prueba T de Wilcoxon.....	461
17.7. Ejercicios de autoevaluación	468
17.8., Solución, a, los, ejercicios, de, autoevaluación	
17.9. Referencias bibliográficas	474
Tema 18. Resolución de diseños de investigación de dos y más grupos. El análisis de la varianza	475
18.1., El, concepto, de, varianza, y, sus, tipos	
18.1.1., Definición, de, varianza	
18.1.2. La varianza de la población y la varianza de las muestras.....	477
18.1.3. Tipos de varianza.....	478
18.2. La resolución del análisis de la varianza	483
18.2.1. Ecuaciones para la resolución del análisis de varianza .	484
18.2.2. Interpretación del valor de «F»	488
18.2.3. Las pruebas de Tukey y Scheffé para el contraste a posteriori.....	489

18.3. Resolución de un diseño de investigación de tres grupos independientes y nivel de medida en la variable dependiente de intervalo mediante el análisis de la varianza	490
18.4. Resolución de un diseño de investigación de tres grupos independientes y nivel de medida en la variable dependiente de ordinal, mediante la prueba de Kruskal-Wallis	497
18.5. Ejercicios de autoevaluación	502
18.6. Solución a los ejercicios de autoevaluación	503
18.7. Referencias bibliográficas.....	504

Apéndice

Tablas estadísticas	505
---------------------------	-----

PRESENTACIÓN

Estas **Unidades didácticas** de *TÉCNICAS E INSTRUMENTOS DE RECOGIDA Y ANÁLISIS DE DATOS* tienen como finalidad facilitar al alumno el estudio de los contenidos de la materia, ayudándole en el proceso de aprendizaje de los contenidos teóricos y práctica de la asignatura del mismo nombre, perteneciente a la titulación de la **Diplomatura en Educación Social** de la Facultad de Educación de la Universidad Nacional de Educación a Distancia.

Las **Unidades didácticas** de la asignatura están estructuradas en torno a tres unidades didácticas integradas por 16 temas que conforman el programa.

Los autores de estas **Unidades didácticas** son:

Catalina Martínez Mediano, de los temas 1 al 8 y 10 al 18, Dirección y Coordinación.

Arturo González Galán, del tema 9.

1. LA ASIGNATURA

Esta asignatura de *TÉCNICAS E INSTRUMENTOS DE RECOGIDA Y ANÁLISIS DE DATOS*, de carácter optativo, cuatrimestral, de cinco créditos, se cursa en el Segundo curso de la Diplomatura de Educación Social. El alumno ya ha tenido la oportunidad de estudiar la asignatura de *MÉTODOS DE INVESTIGACIÓN EN EDUCACIÓN SOCIAL*, materia obligatoria, cuatrimestral de 5 créditos. En dicha asignatura se han estudiado contenidos relacionados con «*Las características de la investigación en educación social*», las «*Técnicas e instrumentos de recogida de datos*» y «*El análisis de los datos cuantitativos*». Lo cierto es que los descriptores de ambas asignaturas son muy similares siendo, inevitablemente,

gran parte de sus contenidos coincidentes. El alumno debe ver esta circunstancia como una oportunidad de profundizar en el conocimiento y manejo de las materias de contenidos metodológicos, que sin duda le prestarán un gran servicio para su formación aplicando sus contenidos a otras materias de la carrera, y muy especialmente, a su desarrollo profesional.

Estamos ante una materia de carácter metodológico que le aportará conocimientos específicos para poder diseñar, realizar, evaluar y mejorar sus propias investigaciones aplicadas a su ámbito profesional, y asimismo le capacitará para comprender, interpretar y valorar informes de investigaciones realizadas por otros, aportándole conocimientos tanto metodológicos como sustantivos necesarios para su desarrollo profesional y el consiguiente ejercicio profesional, en el que tendrá que elaborar planes y proyectos educativos, en los ámbitos educativo y social, que deberán ser realizados con el necesario rigor y sistematización de modo que las acciones que de ellos se deriven contribuyan, de modo eficaz, a mejorar las situaciones objeto de la intervención educativa y su evaluación así lo confirmen.

Esta materia metodológica pretende aportarle conocimientos y contribuir al desarrollo de habilidades, destrezas, técnicas y actitudes de gran importancia para su formación profesional como Educador Social y en la medida en que lo consiga le permitirá acceder con la suficiente capacidad técnica y, por tanto, crítica, a publicaciones sobre investigaciones, valorarlas, decidir sobre su calidad y apreciar la conveniencia o no de aprovechar los conocimientos allí recogidos para su propio hacer profesional y para su actualización profesional.

1.1. Los objetivos

El objetivo fundamental de la asignatura de *TÉCNICAS E INSTRUMENTOS DE RECOGIDA Y ANÁLISIS DE DATOS* dentro del título de **Diplomado en Educación Social** es el de contribuir, junto con las demás asignaturas metodológicas de la carrera, a la formación del alumno para que sea capaz de:

- Enfrentar los problemas del ámbito de la educación con una actitud científica, rigurosa, que le permita buscar soluciones adecuadas a las situaciones que observa.
- Comprender y valorar investigaciones en el ámbito de la educación.

- Diseñar, implementar y evaluar proyectos y programas para la mejora de las condiciones educativas y sociales.

Para ello, será necesario que el alumno conozca las características de la metodología científica, sus procedimientos, sus técnicas e instrumentos que utiliza para planear una investigación, seleccionar unas muestras poblacionales y un campo de aplicación, las acciones para elaborar un plan fundamentado para tratar de resolver el problema que le preocupa, unas estrategias para la aplicación de plan, unas técnicas e instrumentos para recoger la información como consecuencia del plan, y unas técnicas para analizar la información y extraer conclusiones que le permitan comprobar si sus intentos por resolver el problema han actuado en la dirección pretendida, valorar lo adecuado y mejorable del proceso y las recomendaciones que se deriven de ello, expuesto en formato de informe en tal modo que facilite su comprensión y contribuya al desarrollo de la comunidad educativa y social.

En fin, esta materia, junto con otras similares de metodología de la ciencia, deberá contribuir al doble objetivo de que el Diplomado en Educación Social sea capaz de valorar, seleccionar y transmitir, a la comunidad educativa y social, los resultados de investigaciones ya consolidadas en Educación y Ciencias Sociales, y también a que sea capaz de formar equipos, liderar y colaborar en investigaciones analizando los problemas de su aula y de su entorno educativo y social para realizar investigaciones, que realizadas con el suficiente rigor, aporte soluciones y propuestas de cambio y transmita los resultados con el compromiso puesto en la mejora de la calidad de la educación y de la sociedad.

Los objetivos específicos de esta materia, que pretendemos que domine el alumno al terminar el curso, son los siguientes:

- Conocer, comprender y valorar las características de la metodología científica.
- Comprender los procedimientos para la realización de investigaciones.
- Comprender y seleccionar los diferentes enfoques de investigación, positivista, cualitativa o evaluativa, y su adecuación para cada problema objeto de estudio.
- Interpretar con criterios científicos —rigurosos y sistemáticos—, la realidad educativa susceptible de investigación para la búsqueda de mejoras.
- Formular adecuadamente los problemas.

- Documentarse sobre el estado de la cuestión de los problemas a investigar consultando fuentes bibliográficas y documentales.
- Seleccionar el diseño.
- Seleccionar y / o elaborar las técnicas e instrumentos de recogida de información y datos.
- Seleccionar las técnicas adecuadas para el tratamiento y análisis de datos.
- Interpretar los resultados y extraer o deducir las conclusiones que se deriven.
- Plantear las decisiones en coherencia con las conclusiones que se deriven de los resultados de las investigaciones.
- Elaborar informes y planes de mejora, informar su aplicación y evaluación.
- Analizar críticamente las investigaciones, ya sean propias ya de otros investigadores.

1.2. El temario de contenidos

Los contenidos de la asignatura se estructuran en torno a tres unidades didácticas básicas. La primera se refiere a «*Los métodos de investigación científica en Educación social*», la segunda a «*La medida en educación y los instrumentos para la recogida de información*» y la tercera a «*La estadística descriptiva e inferencial para el análisis de la información*». Incluyen los siguientes temas:

PRIMERA UNIDAD DIDÁCTICA: LOS MÉTODOS DE INVESTIGACIÓN EN EDUCACIÓN SOCIAL

- Tema 1. La investigación en educación social. Tipos y enfoques.
- Tema 2. La investigación descriptiva.
- Tema 3. Los diseños de grupo. La investigación experimental.
- Tema 4. La investigación cualitativa.
- Tema 5. La investigación evaluativa. Evaluación de programas e instituciones.

SEGUNDA UNIDAD DIDÁCTICA. LA MEDIDA EN EDUCACIÓN Y LAS TÉCNICAS PARA LA RECOGIDA DE INFORMACIÓN

- Tema 6. La medida en Educación y en Ciencias Sociales.
- Tema 7. Las escalas para la medida de las actitudes. Escalas de medida de Thurstone.
- Tema 8. Las escalas para la medida de las actitudes. Las escalas de medida de Likert.
- Tema 9. Características técnicas de los instrumentos de medida desde la teoría clásica de los tests. Fiabilidad y validez.
- Tema 10. Procedimientos para la recogida de información I. La observación sistemática, el cuestionario, la entrevista y las pruebas de rendimiento.
- Tema 11. Procedimientos para la recogida de información I. La observación participante, el análisis de contenido, la técnica Delphi y los grupos de discusión.

TERCERA UNIDAD DIDÁCTICA. LA ESTADÍSTICA. DESCRIPTIVA E INFERENCIAL PARA EL ANÁLISIS DE LA INFORMACIÓN. LOS DISEÑOS DE INVESTIGACIÓN

- Tema 12. Distribuciones de variables y representaciones gráficas
- Tema 13. Características de las distribuciones de frecuencias: medidas de tendencia central, de posición y de dispersión. Puntuaciones individuales.
- Tema 14. Correlaciones bivariadas.
- Tema 15. Probabilidad y muestreo. Distribución muestral y estimación de parámetros.
- Tema 16. Estimación de la diferencia de dos medias. La prueba de hipótesis.
- Tema 17. Resolución de diseños de investigación de dos grupos.
- Tema 18. Estadística inferencial IV. Resolución de diseños de investigación de dos y más grupos. El análisis de la varianza.

1.3. Los medios y recursos

Las presentes **Unidades didácticas** pretenden ser un *texto de estudio* que oriente el aprendizaje del alumno que estudia en una Universidad a Distancia, como es la UNED, en la medida en que desarrolla contenidos teóricos guiados por los objetivos de aprendizaje, mostrando ejemplos y ejercicios prácticos, que ayuden a la comprensión de la teoría y de su apli-

cación, así como con la inclusión de ejercicios de autoevaluación. La posibilidad de consultar al profesorado de la asignatura mediante el teléfono, el correo electrónico y postal, según consta en la Guía de la Carrera, ayudarán al alumno a la comprensión de los contenidos de la materia.

1.4. La evaluación

La evaluación se concreta en la realización de un *trabajo* y en la *prueba presencial*, en la que el alumno deberá poner de manifiesto el grado de dominio de los objetivos de la materia.

La evaluación cumple una doble función: la función formativa y la sumativa o de certificación. La evaluación formativa, término aportado por Scriven, tiene como meta la mejora, e implica su consideración como medio, lo que supone una mayor continuidad en el proceso de recogida de información sobre el proceso y progreso de aprendizaje del alumno, y su información al alumno para ayudarle a tomar conciencia de sus progresos, de sus logros, de sus errores y de sus carencias, dado el carácter motivador que suele representar este hecho; y, por último, su utilización para la toma inmediata y a tiempo de decisiones de mejora sobre cada alumno y sobre el programa en cada una de las unidades en que esté estructurado.

La función sumativa, como medio para asegurar que el estudiante ha alcanzado las metas u objetivos previstos por el profesor en relación con los temas o contenidos de la disciplina que integran el programa. Es además una actividad más en el proceso de enseñanza, dentro de las funciones específicas del profesor, junto con la actividad de planificación, diseño y realización de la enseñanza, siendo un elemento esencial para guiar la cantidad y calidad de los objetivos y contenidos de aprendizaje.

Para contribuir a la función formativa de la evaluación tenemos los ejemplos y ejercicios que se incluyen al final de cada tema. Y en relación con la función sumativa, el alumno deberá realizar una prueba escrita durante los exámenes de mayo-junio, en su convocatoria ordinaria, o en septiembre, en la convocatoria extraordinaria, que se celebran en los Centros Asociados de la UNED.

En la GUÍA DE LA CARRERA de la Diplomatura de Educación Social, el alumno encontrará información más precisa sobre estos aspectos de la evaluación. Este es el medio fundamental para que el profesorado de la asignatura comunique a los alumnos aspectos relacionados con los objetivos, contenidos, y procedimientos y criterios de la evaluación. Asimismo, se espera que en breve esta asignatura tenga una página web

con el fin de facilitar y agilizar las orientaciones precisas a los alumnos para su mejor aprovechamiento, así como para la realización de preguntas y debates.

Los alumnos que cuenten con profesor tutor en los Centros Asociados de la UNED tendrán la oportunidad de trabajar sobre las cuestiones relacionadas con la asignatura en las tutorías. Los demás, deberán consultar al profesor de la Sede central.

1.4.1. *Las pruebas presenciales*

Dado que la asignatura es cuatrimestral, cuenta con una única prueba, con dos convocatorias, y como se cursa en el segundo cuatrimestre del año académico, estas pruebas se realizarán en junio, en convocatoria ordinaria, y en septiembre, en convocatoria extraordinaria. Con una duración de dos horas, las pruebas constará de dos partes claramente diferenciadas, pero coherentes entre sí y de similar importancia a la hora de valorar: una parte *teórica* y otra de carácter *práctico*.

Las cuestiones sobre *teoría* se formularán de forma que puedan ser respondidas con *brevedad* (de 10 a 20 líneas cada una), *precisión y claridad*. También cabe la posibilidad de que se incorpore una breve *prueba objetiva*, e incluso una cuestión de desarrollo, en la que se valorará el *nivel de dominio*, la *comprensión* y la *asimilación personal*, además de otros criterios genéricos, como *la correcta expresión y la adecuada organización en la realización de las cuestiones*.

La *práctica* se concreta en *problemas* y en *cuestiones*. Entendemos por *problema* un enunciado amplio en el que el alumno deberá evidenciar tanto el dominio de la teoría correspondiente como el de su aplicación. Y es que el dominio de la teoría es la base para la realización de una buena práctica.

Los *criterios* generalmente utilizados para valorar las pruebas presenciales son los siguientes:

- Adecuación y precisión en las respuestas dadas a las preguntas relacionadas con la teoría.
- Corrección en el planteamiento de la resolución del problema práctico.
- Justificación de los pasos a dar en la resolución del problema.
- Exactitud y precisión en los cálculos.
- Interpretación correcta.
- Valoración crítica.

Los meros errores de cálculo tienen baja penalización, pero si tal error de cálculo supone una carencia de carácter teórico, entonces si penalizarán. Por ejemplo, si resuelve unos cálculos y obtiene un valor de r (correlación de Pearson) > 1 , el error es sólo de cálculo si hace saber que ese valor no es posible, pero se convierte en un error conceptual grave si sigue adelante como si tal valor fuera posible. Algo similar podemos afirmar si obtiene sumas de cuadrados (SC) o medias cuadráticas (MC) negativas en un análisis de varianza.

1.4.2. *El trabajo*

El alumno deberá realizar un trabajo práctico obligatorio consistente, en esencia, en el informe sobre una investigación, real o simulada, siguiendo las orientaciones del equipo docente de la asignatura y del profesor tutor. Dado que la materia recoge diferentes enfoques de investigación, el trabajo puede realizarse desde la perspectiva de investigación que más interese al alumno, ya sea desde el enfoque experimental, como del cualitativo o del evaluativo.

Si el alumno se decidiera por el enfoque experimental, la estructura básica del trabajo sería la siguiente:

- El problema de investigación.
- Fundamentación: consulta de fuentes y estado de la cuestión.
- La hipótesis.
- Las variables.
- El diseño.
- Desarrollo de la investigación.
- Recogida de datos.
- Tratamiento estadístico de datos. Contraste de hipótesis.
- Discusión y conclusiones.
- Significación pedagógica de los resultados.
- Valoración crítica. Análisis del trabajo, de la validez interna y externa.

Si el enfoque elegido fuera el cualitativo, el informe debe seguir la siguiente estructura:

- El problema de investigación.
- El marco teórico. Experiencia, enfoque y roles del investigador.
- El propósito de la investigación. Su diseño. Las modificaciones.

- La naturaleza y el número de los escenarios e informantes.
- El modo de acceso al campo.
- Las diversas fuentes de información: personas, bibliografía, documentos.
- Las relaciones con los informantes.
- La metodología: observación, entrevistas, lectura de informes.
- Las estrategias de recogida de información y datos y su control.
- El tiempo y extensión del estudio y su distribución.
- Las técnicas de análisis de la información y datos.
- Los hallazgos, las interpretaciones, las conclusiones.
- Las aplicaciones.
- Las recomendaciones.

Si se tratara de una investigación evaluativa, el informe deberá incluir:

- Los propósitos de la evaluación.
- Las cuestiones que se van a revisar.
- Los criterios que se van a utilizar en la valoración del programa o institución.
- Los métodos y los procedimientos que se van a utilizar para la recogida de información.
- Los métodos y técnicas seguidos para el análisis de datos, tanto numéricos como de otra naturaleza.
- Los criterios seguidos para enjuiciar o valorar la evaluación.
- La síntesis descriptiva objetiva y valorativa que relacione los propósitos del programa con los recursos y los resultados, que recoja los aspectos más positivos del programa y aquellos que hay que modificar y mejorar.
- La elaboración de un plan de mejora del programa fundamentado en las necesidades de cambio y mejora del programa, que incluya estrategias y tiempos para su realización.

La valoración del trabajo se llevará a cabo atendiendo a los siguientes criterios adaptados al enfoque de investigación utilizado:

- Relevancia, adecuación y correcta formulación del problema.
- Rigor, amplitud y profundidad de la fundamentación.
- Formulación adecuada de la hipótesis.
- Corrección en la definición de las variables.

- Adecuación entre el problema, la hipótesis y el diseño.
- Adecuación, corrección y exactitud del análisis y tratamiento de los datos.
- Interpretación adecuada de los resultados de la prueba estadística.
- Rigor en la discusión y adecuación de las conclusiones.
- Análisis adecuado de la relevancia pedagógica de los resultados.
- Profundidad en la valoración crítica del propio trabajo.

Primera Unidad Didáctica

LOS MÉTODOS DE INVESTIGACIÓN EN EDUCACIÓN SOCIAL

TEMA 1

La investigación en educación social. Tipos y enfoques

ESQUEMA

- 1.1.** El conocimiento científico y la ciencia
- 1.2.** Tipos de investigación en educación y en ciencias sociales
 - 1.2.1. Tipos de investigación según el grado de abstracción: básica y aplicada
 - 1.2.2. Tipos de investigación según el proceso formal: deductivo, inductivo, hipotético-deductivo
 - 1.2.3. Tipos de investigación según la intención: descriptiva, comparativa, evaluativa, explicativa
 - 1.2.4. Tipos de investigación según el momento: histórica y actual
- 1.3.** Los diferentes paradigmas de investigación
 - 1.3.1. El paradigma positivista y neopositivista
 - 1.3.2. El paradigma crítico
 - 1.3.3. El paradigma interpretativo o naturalista
- 1.4.** El debate metodológico y la complementariedad metodológica
- 1.5.** Ejercicios de autoevaluación
- 1.6.** Solución a los ejercicios de autoevaluación
- 1.7.** Referencias bibliográficas

OBJETIVOS

- ✓ Comprender las características y los fines del conocimiento científico
- ✓ Conocer, comprender y valorar la importancia de los métodos de investigación como procedimiento para alcanzar el conocimiento científico
- ✓ Comprender las distintas clasificaciones de investigación atendiendo al grado de abstracción, proceso formal, intención y momento.
- ✓ Conocer y comprender la naturaleza de los diferentes paradigmas o enfoques de investigación
- ✓ Comprender y valorar las posibilidades que ofrecen los diferentes enfoques de investigación y las condiciones para la complementariedad metodológica

1.1. El conocimiento científico y las ciencias

El conocimiento humano se sirve de diversos métodos como base del progreso social. Un primer método utilizado por la Humanidad se fundamenta en *las «verdades» que se repiten* siempre, siendo esta repetición una manera de confirmar su validez. Un segundo método se apoya en *la autoridad* de la tradición o de los expertos, para comprender y orientar los comportamientos. Ambos modos de conocimiento deben ser confirmados o rechazados por la observación de los acontecimientos y mediante la investigación. Un tercer método de conocimiento se fundamenta en apriorismos basados en proposiciones aceptadas como evidentes por sí mismas; Es el modo de conocer del *racionalismo*. Las ideas apriorísticas pueden concordar con la razón pero no necesariamente con la experiencia. El modo de conocer del racionalismo consiste en añadir sobre los dos modos de conocimiento, anteriormente mencionados, pensamiento, deducción y análisis lógico sobre los acontecimientos, pudiendo llegar a generalizaciones y predicciones útiles. Este modo de investigación ayuda a ampliar la teoría y origina preguntas. Otro método de conocimiento es aquél que incorpora las observaciones y datos recogidos sistemáticamente para ser analizados con el fin de obtener una comprensión de los fenómenos. Las nuevas evidencias son constantemente comprobadas y las teorías existentes modificadas y corregidas a la luz de los nuevos hallazgos. Esta forma de conocimiento es conocida como *método científico* y se construye sobre la autocorrección.

La utilización del método científico en la investigación educativa y social supone un intento sistemático dirigido a cuestionar y/o resolver problemas a través de la recogida y análisis de datos primarios con el propósito de describir, explicar, generalizar y predecir, siendo las características propias del método científico la repetibilidad, la estabilidad y la comunicabilidad. El método científico es esencial para la investigación científica y la construcción de la ciencia, consistiendo aquélla en una actividad que combina experiencia y razonamiento, siendo su principal característica el uso sistemático y consciente del aspecto autocorrectivo. Lo verdaderamente importante del método científico es la racionalidad controlada de la investigación científica como un proceso de indagación reflexiva integrado por diversas partes interdependientes, en donde el problema y su expresión representa el valor supremo. (Kerlinger, 1987:4).

El vocablo «método», del griego, hace referencia a los procedimientos ordenados que se siguen para alcanzar una meta u objetivo determinado. alguna de las características del método científico son:

- Contrastación con la evidencia empírica disponible.
- Interacción entre teoría y realidad en la construcción de la primera y el estudio de la segunda.
- Exclusión deliberada de manipulación de la información y su interpretación.
- Justificación de la verdad en términos probabilísticos, nunca en términos absolutos.

«La peculiaridad de la ciencia consiste en el modo como opera para alcanzar un objetivo determinado, o sea, en el método científico»... «Donde no hay método científico, no hay ciencia... Puede utilizarse en todos los campos del conocimiento... la naturaleza del objeto en estudio dicta los posibles métodos especiales del tema o campo de investigación correspondiente» (Bunge, 1976:29).

La ciencia pretende ser un sistema ordenado y coherente de relaciones que mutuamente se apoyan y sostienen. Persigue un conocimiento relacionado y sistemático que debe incardinarse con la totalidad. Lo propio de la ciencia, desde esta perspectiva, es contribuir al desarrollo del conocimiento teórico con el fin de explicar los fenómenos observados, formular hipótesis y comprobarlas, relacionar fenómenos, explicar la realidad, generalizar y establecer leyes.

Algunos autores mantienen que las ciencias sociales no pueden ser científicas porque no se puede aplicar el método experimental al ser su objeto de estudio el hombre, y sus comportamientos difíciles de medir; y por tanto supone una gran dificultad operar y aplicar cálculos matemáticos a sus acciones, y deducir conclusiones a partir de ellas. Sin embargo, con todas las limitaciones y precauciones, desde estos enfoques, el método científico ha contribuido al desarrollo de la Educación y de las Ciencias Sociales. Dicho esto, conviene añadir que las ciencias humanas no deben mirarse sólo en el espejo de las ciencias físico-naturales, tomándolas como único modo de estudiar y explicar los fenómenos y desarrollar el conocimiento. Dadas las muchas particularidades que presentan, y la complejidad de su objeto de estudio, el hombre, lo humano, requiere de otros enfoques de aproximación al estudio de su comportamiento. Y es que no siempre se puede pretender conseguir el objetivo de alcanzar una teoría con significado nomotético, con normas fijas, inamovibles, cosa que, por otra parte, tampoco se puede conseguir siempre en las ciencias físico-naturales.

Otro enfoque de investigación, el de la corriente naturalista, concibe la ciencia desde un punto de vista dinámico, cuyo objeto de estudio es la

persona en sus contextos. La ciencia, desde esta concepción, es más inclusiva que el positivismo, recoge lo idiográfico, lo personal, lo experiencial, lo transcendental, desde una perspectiva global, holística. Para Maslow (1966), que defiende este enfoque desde la corriente humanista, la ciencia no debe estar centrada en un único método sino en los problemas y buscar los métodos apropiados para resolverlos. El énfasis lo centran en la comprensión profunda de las personas en las situaciones en las que se desenvuelven, más que en la explicación, situándose en el polo opuesto del positivismo. La ciencia, desde este punto de vista, tiene como finalidad comprender y transformar para mejorar las condiciones humanas; es idiográfica, es decir, se interesa por lo individual y lo particular sin pretender llegar al establecimiento de leyes generales.

1.2. Tipos de investigación en Educación y en Ciencias Sociales

La investigación en Educación y en Ciencias Sociales ha sido clasificada de muy diferentes maneras atendiendo a diversos criterios no siempre excluyentes. Algunas veces, el criterio de clasificación de los métodos de investigación ha sido el proceso formal que utiliza; otras el grado de abstracción; otras el de manipulación y control de las variables y el de generalización y la naturaleza de los datos en otros casos.

A continuación presentamos los tipos de investigación de acuerdo con el siguiente esquema:

- Según el grado de abstracción: básica y aplicada.
- Según el proceso formal: deductivo, inductivo, hipotético-deductivo.
- Según la intención: descriptiva, comparativa, explicativa.
- Según el momento: histórica y actual.
- Según la cuestión a investigar, el control y el momento temporal.

1.2.1. *La investigación según el grado de abstracción: básica y aplicada*

Atendiendo al grado de abstracción nos encontramos con la investigación básica y la investigación aplicada. La distinción entre ambas debe ser entendida más como los extremos de un continuo que como una dicotomía. Ambos tipos de investigación convergen en su finalidad: hacer progresar el saber y simultáneamente transformar la realidad. **La investigación básica** tiene como objetivo el incremento de conocimiento por sí

mismo sin pretensiones de aplicación inmediata de sus hallazgos. De otro lado, la **investigación aplicada** pretende la aplicación del conocimiento para la resolución de problemas, y cuando ésta tiene éxito se traduce en planes o directrices para su expansión; el conocimiento producido suele tener utilidad casi inmediata, mientras que no ocurre lo mismo en la investigación básica.

Desde la perspectiva positivista de la ciencia, a la investigación básica le preocupan las relaciones de las variables y la causa por las que se relacionan de una determinada forma, cuidando, especialmente, la validez interna de la investigación. Y en la investigación aplicada, además de lo anterior, se da una gran importancia a la generalización de los hallazgos, pues se desea extender los resultados a otras personas y situaciones. En educación la mayor parte de la investigación es aplicada, tendente a resolver problemas inmediatos para introducir mejoras en la práctica educativa, y cuando estas soluciones son consistentes y generalizables contribuye al desarrollo del corpus teórico necesario para comprender los fenómenos en la práctica.

1.2.2. La investigación según el proceso formal: deductivo, inductivo, hipotético-deductivo

Según el proceso formal seguido por la investigación nos encontramos con los procedimientos de investigación siguientes: deductiva, inductiva e hipotético-deductiva.

El método deductivo parte de una premisa general, o universal, para sacar conclusiones de un caso particular. Partiendo de enunciados generales formula enunciados particulares. Sigue el modelo aristotélico esquematizado en el silogismo; enlazando las premisas mayor y menor llega a la conclusión, siendo el silogismo correcto si las premisas de que parte son ciertas. El científico que utiliza este método pone el énfasis en la teoría, en la explicación, en los modelos teóricos, en la abstracción, no en recoger datos empíricos o en la observación y en la experimentación. Para Van Dalen y Meyer (1979:38), la conclusión del razonamiento deductivo no proporciona conocimiento más allá del que está en las premisas. Bacon y Mill señalaron que las principales aportaciones del método deductivo descansan en la coherencia lógica de su procedimiento.

El método inductivo procede de lo particular a lo universal. Consiste en analizar casos particulares a partir de los cuales se derivan conclusiones de carácter general. Fue introducido por Bacon y desarrollado por Mill. La inducción es un tipo de inferencia que parte de enunciados sin-

gulares para llegar a enunciados generales. Se basa en la observación y en la experimentación. Su objetivo es descubrir generalizaciones y teorías a partir de observaciones sistemáticas de la realidad. Pero, salvo que se trate de una *inducción completa o perfecta*, en la que se estudian todos los casos particulares para llegar al universal, se da siempre un riesgo en una conclusión general sobre la base de una premisa en que únicamente se han estudiado un cierto número de todos los casos. Para que la inducción sea aceptada es preciso que cumpla las reglas de la inducción, como son la selección adecuada de los casos mediante el muestreo aleatorio y suficiente, el control de las condiciones para evitar sesgos, y aceptar que las generalizaciones nunca se pueden hacer en términos de certeza absoluta sino de probabilidad. El método inductivo se utiliza en los estudios descriptivos y experimentales, desde distintas concepciones de muestreo, control y diferentes pretensiones de generalización de sus resultados.

El método hipotético-deductivo parte de la observación de casos particulares para formular un problema y a través de un proceso inductivo remite a una teoría. Después, apoyándose en la teoría, formula unas hipótesis de un modo deductivo, que intentará validar empíricamente. Este ciclo inductivo-deductivo se conoce como proceso hipotético-deductivo. Los estudios experimentales utilizan el método hipotético-deductivo.

1.2.3. Tipos de investigación según la intención: descriptiva, comparativa, evaluativa, explicativa

Cuando el investigador realiza una investigación puede pretender *describir* el comportamiento de un determinado acontecimiento, sujeto o grupo de sujetos, y posiblemente hacer alguna valoración o juicio a partir de su descripción, realizando **investigación descriptiva**. Pero también puede pretender *comparar* los resultados de dos estudios descriptivos sobre un mismo sujeto o grupo en dos momentos a lo largo del tiempo, o entre dos grupos diferentes, buscando conocer sus semejanzas y sus diferencias, comparando los sujetos y las situaciones realizando una **investigación comparativa**. Y cuando para comparar los resultados de observación del grupo utiliza unos criterios que le sirven de referencia, se denomina **investigación evaluativa**. La ciencia también busca la **explicación**, mediante la puesta a prueba de sus teorías, a través de la prueba de hipótesis, y contrasta sus resultados con las teorías previamente construidas, que como sabemos pueden refutarse. La ciencia busca describir situaciones, conjuntos, grupos, establecer comparaciones buscando semejanzas y diferencias, y poner a prueba hipótesis para solucionar problemas, contribuir al desarrollo de teorías y explicar los fenómenos que le preocupan.

En *la investigación explicativa*, una vez formulado el problema, revisada la literatura e investigaciones relevantes y planteadas las hipótesis, se diseña el plan de ejecución, o experimento, para poner a prueba dichas hipótesis, y deducidas las consecuencias particulares comprobables de las mismas, se aplica el plan, y mediante instrumentos —observación, cuestionarios, tests—, se miden sus efectos y se analiza la información y los datos mediante pruebas estadísticas, que permitirán confirmar o rechazar las hipótesis al nivel de confianza fijado, puesto que se trabaja en términos probabilísticos, pudiendo, a partir de esto, realizar inferencias de la muestra estudiada, a grupos poblacionales, contextos y problemas similares.

1.2.4. Tipos de investigación según el momento: histórica y actual

La *investigación histórica* puede utilizar el método de investigación descriptivo, cuando su finalidad es describir los hechos ocurridos en sus contextos —para una mayor y ajustada comprensión—, el método de investigación exploratorio, cuando busca la posible relación entre variables, y asimismo el método de investigación explicativo, no en sentido estricto, pero sí en lo que se refiere a la utilización de la lógica del proceso de la investigación explicativa, que busca confirmar hipótesis formuladas por el investigador para aportar luz a un problema.

La investigación que se centra en los hechos pasados también es llamada *investigación ex post facto*, para indicar que los acontecimientos que se estudian ya han acontecido.

La historia es una ciencia normativa que utiliza métodos descriptivos, exploratorios y en ocasiones aplica la lógica de la investigación científica desde el enfoque positivista. El método experimental no puede aplicarse en sentido estricto a la investigación histórica ya que su objeto de estudio son los acontecimientos del pasado, y el método experimental busca probar hipótesis poniendo a prueba un experimento, comprobando posteriormente sus efectos, es decir, en el futuro.

El método tradicionalmente usado por los historiadores se ha basado en la búsqueda y reunión de fuentes, sirviéndose de disciplinas auxiliares como la diplomática, numismática, filología, sigilografía, paleografía, criptografía, epigrafía, papirografía, genealogía, heráldica y cronología. Tradicionalmente los historiadores se apoyan en el *análisis de documentos* escritos, o de otro tipo, en contraste con los métodos del estudio del momento actual, del presente, que utiliza la observación directa de los

acontecimientos que estudia, sean provocados, como ocurre en la *investigación experimental* mediante el experimento o no, como en la *investigación naturalista*, en la cual observan los acontecimientos naturales en su propio contexto. Los documentos escritos son analizados con profundidad porque pueden no ser auténticos, tener errores, sesgos, por lo que para obtener una información lo más objetiva y rigurosa de los hechos se deben contrastar los textos de diferentes autores y utilizar diferentes fuentes de información. Hay que desconfiar sistemáticamente de lo que no esté probado mediante diversidad de fuentes.

La lógica del método científico puede aplicarse al estudio de la historia llegándose a plantear hipótesis que serán puestas a prueba mediante procedimientos diferentes a los utilizados en la comprobación de hipótesis de hechos presentes y hacia el futuro. Como en cualquier disciplina científica, consiste básicamente en seguir los pasos del proceso de investigación científica, con las peculiaridades propias de esta ciencia.

Por otro lado, en la *investigación actual*, dependiendo del objetivo se puede realizar investigación descriptiva, exploratoria o explicativa, tal como hemos indicado en el epígrafe anterior.

1.3. Los diferentes paradigmas de investigación

La persuasiva tesis de Kuhn (1962) de que ciertas concepciones del conocimiento —o paradigmas— dominan el mundo de la ciencia durante un tiempo hasta que después de una crisis surge otra nueva concepción, vino a apoyar a los investigadores que, como Maslow (1966) abrigaban ciertas dudas sobre la concepción dominante de la ciencia. Ciertamente, los investigadores sociales —los científicos sociales— preocupados por los sentimientos o por la comprensión subjetiva de la realidad, no contaban con un método tan fuerte, coherente y dominante como los procedimientos de medida utilizados por la investigación experimental contruidos apoyándose en teorías matemáticas.

Un rasgo característico de la investigación en Educación y en las Ciencias Sociales es el espectacular desarrollo de su metodología y la versatilidad de la misma. Como ejemplos la generalización de los diseños multivariantes, el análisis causal, los diseños cuasi-experimentales y el creciente uso de los métodos cualitativos. La complejidad de los fenómenos educativos sólo podrá ser captada y explicada mediante enfoques flexibles y complejos que los avances en la metodología de investigación y su previsible desarrollo nos ofrecen. Una vez más en la historia de la revolución científica, esta será un efecto de la revolución metodológica.

Las diferencias entre los métodos de investigación no radican en el hecho de que puedan ser concebidos como vías alternativas para alcanzar el mismo fin o responder de manera distinta a la misma pregunta o cuestión. Shulman (1981) subraya que lo que distingue radicalmente a los métodos entre sí, en virtud de sus diversos orígenes disciplinares, es, no sólo los procedimientos que aplican, sino fundamentalmente, el tipo de cuestiones que pretenden o pueden resolver, así como los presupuestos y postulados básicos que determinan y especifican los criterios de evidencia y la interpretación de las respuestas obtenidas. En resumen, la diversidad metodológica se ilumina en el contexto de la concepción paradigmática de la ciencia en el sentido kuhniano. Sin embargo, la tesis de Kuhn no se ha confirmado puesto que más que sustituir un paradigma a otro, en realidad, tienden a coexistir e o incluso, en ocasiones, a complementarse.

Soltis (1984, en De la Orden, 1985:XII) clasifica la investigación pedagógica en tres grandes líneas: empírica, interpretativa y normativo-crítica. Estas líneas se vinculan a las grandes tradiciones filosóficas del siglo XIX, el empirismo lógico neopositivista, la filosofía analítica y la fenomenológica y la teoría crítica del marxismo, en las que se halla la base del pensamiento epistemológico de las ciencias sociales. Guba (1990) clasifica los paradigmas en el positivista, el neopositivista, el crítico o ideológico y el interpretativo o constructivista. Veamos una síntesis de los mismos.

1.3.1. El paradigma positivista y neopositivista

La creencia básica del *sistema positivista* está enraizada en un realismo ontológico: la realidad existe y procede de acuerdo con leyes naturales y mecanismos inmutables. El conocimiento de estas entidades, sus leyes y mecanismos, es formalmente comprobable y generalizable, adoptando la forma de leyes de causa-efecto. El rol del investigador en relación con lo que investiga es distante, con la finalidad de no influir en lo que investiga. Los valores considerados como poco precisos, son excluidos con el fin de que no interactúen, junto con las variables de estudio, medidas con instrumentos precisos, en los resultados. El objetivo de este enfoque de investigación es conocer los fenómenos naturales para poder predecirlos y controlarlos. La metodología más adecuada es el experimento. Las preguntas e hipótesis son enunciadas en forma proposicional a partir de la observación de los hechos y su interacción con el cuerpo teórico existente para, posteriormente, ser comprobadas empíricamente bajo un cuidadoso control de las condiciones en las cuales se realiza el experimento. Formula hipótesis que trata de probar para posteriormente reali-

zar inferencias sobre el fenómeno que observa desde la muestra a la población definida.

El *neopositivismo* es considerado como una versión modificada del positivismo. La predicción y el control continúan siendo sus objetivos pero desde un realismo crítico. La esencia de esta posición es que, si bien el mundo real se mueve por las causas naturales existentes, para los humanos es imposible percibirlo verdaderamente debido a sus mecanismos sensoriales e intelectuales imperfectos. Los investigadores necesitan ser críticos sobre su trabajo precisamente a causa de la fragilidad humana. La objetividad permanece como un ideal regulador pero sólo puede ser aproximativa poniendo un énfasis especial sobre los guardianes externos, tales como la tradición y la comunidad crítica. Metodológicamente destaca la multiplicidad crítica. Reajusta los desequilibrios haciendo una investigación en situaciones más naturales, usando más métodos cualitativos, dependiendo de la teoría positivista pero reintroduciendo el descubrimiento en el proceso de investigación.

1.3.2. *El paradigma crítico*

Desde la *teoría crítica* no se busca incrementar el conocimiento como principal objetivo, sino que su objetivo principal es la emancipación del hombre de la esclavitud. Aunque las personas pueden actuar de forma consciente para cambiar la sociedad, la economía y la historia, los investigadores críticos reconocen que las posibilidades de cambio están supeditadas a diversas formas de dominación social, cultural y política.

El objetivo fundamental de la investigación crítica es poner en evidencia las situaciones discriminatorias hacia ciertos grupos de la sociedad. Se centra en señalar los conflictos y las contradicciones sociales, para liberar y ayudar a eliminar las situaciones de dominación y opresión. La línea de investigación pedagógica crítica, orientada ideológicamente, está representada por los filósofos Habermas (1971) y Popkewitz (1990), e integra el freirismo, el neomarxismo, el feminismo, la investigación participativa y otros movimientos similares, así como la propia teoría crítica. Considera al *neopositivismo* como una ideología que consagra la eficiencia técnica, el hallazgo de los medios más adecuados para conseguir unos fines no cuestionados que, sin embargo, preservan y refuerzan la situación privilegiada de las clases dominantes.

La corriente crítica intenta desmitificar las instituciones y prácticas educativas hacia una mayor conciencia colectiva y transformarlas en instrumento de emancipación individual y social. No acepta la idea de

una investigación social, o educativa, neutra en la que los valores queden al margen de los fenómenos que estudia. La naturaleza no puede ser vista como realmente es ni como realmente funciona si no es a través de los valores. Si los valores ciertamente entran en toda investigación, la pregunta que inmediatamente aflora es qué valores son esos y a quiénes gobierna. Si los hallazgos del estudio pueden variar dependiendo de los valores elegidos, entonces la elección de un particular sistema de valores tiende a fortalecer a ciertos grupos frente a otros. La investigación se convierte en un acto político. La tarea del investigador es, por definición, elevar a la gente al nivel de verdad consciente. Una vez que se aprecian como son, pueden actuar para transformar el mundo.

El enfoque crítico pretende proporcionar a los miembros de la comunidad educativa una mayor comprensión sobre sí mismos y sobre sus situaciones. Es defendido por Carr y Kemmis (1986:162), con el nombre de investigación-acción emancipatoria, considerada como una forma de investigación autorreflexiva realizada por los participantes en las situaciones sociales, con el fin de mejorar la racionalidad y la justicia de sus propias prácticas desde la comprensión de esas prácticas y las situaciones en donde se realizan.

En el aspecto metodológico, los teóricos críticos adoptan un enfoque dialogante, más que manipulativo e intervencionista, buscando entre todos los participantes un punto de vista común ¿la verdad?, el consenso. El resultado concerniente con la acción es la transformación.

El enfoque crítico es conocido, también, con el nombre de investigación acción colaborativa o cooperativa, aunque cada investigador matiza, desde su sistema de valores, finalidades investigadoras distintas.

1.3.3. El paradigma interpretativo o naturalista

La corriente de *investigación interpretativa*, naturalista o constructivista, parte de la idea de que para acceder al conocimiento de la realidad sólo puede hacerse a través de las construcciones sociales como el lenguaje, como expresión de las construcciones mentales y los significados compartidos. Sus fundamentos filosóficos están en la fenomenología y la hermenéutica. La investigación interpretativa estudia las manifestaciones humanas como totalidad sin previamente definir las variables que intervienen en los fenómenos, abordando directamente la complejidad y pretendiendo comprender los fenómenos a partir del sentido que para las personas tienen y les atribuyen. Presuponen que la investigación positivista no puede dar una respuesta a los problemas más significativos de la

educación, concebida como una actividad humana. Abandonando la obsesión por lo observable y mensurable buscan la objetividad en el ámbito del significado intersubjetivo, basado en el consenso como criterio de evidencia. (En De la Orden, 1985a:XII). El enfoque fenomenológico desarrollado en educación por Langeveld (1983:7), tiene como objetivo el estudio del mundo tal como aparece en nuestra conciencia subjetiva. Pretende que podamos comprendernos a nosotros mismos y a aquellos que están bajo nuestra responsabilidad educativa.

Los constructivistas se declaran relativistas y defienden que la distinción entre ontología y epistemología es anulada por la naturaleza interactiva de la díada investigador / investigado. Las realidades son múltiples y existen en las mentes de la gente, fundamentadas social y empíricamente, locales y específicas, dependiendo sus formas y contenidos de las personas que las sustentan. El constructivismo opta por una posición subjetivista argumentando que es el único medio de comprender las construcciones sostenidas por los individuos. Si las realidades existen sólo en las mentes, las interacciones subjetivas parecen ser el único modo de acceder a ellas. Investigador e investigado se funden en una entidad singular, monista. Los hallazgos son, literalmente, la creación del proceso de interacción entre los dos.

El objetivo de los constructivistas es identificar la variedad de construcciones que existen entre las personas y lograr el mayor consenso posible. Metodológicamente, este proceso tiene dos aspectos: hermenéutico y dialéctico. El aspecto hermenéutico consiste en descubrir las construcciones individuales tan exactamente como sea posible, mientras que el aspecto dialéctico consiste en comparar y contrastar estas construcciones individuales, incluyendo la del investigador. El objetivo de la metodología hermenéutica-dialéctica es producir una construcción mental tan informada y elaborada como sea posible. Su metodología consiste en mantener vías de comunicación abiertas de modo que la información y elaboración puedan ser continuamente mejoradas. El constructivismo no intenta predecir ni controlar el mundo real, tampoco transformarlo, sino reconstruir el mundo en el único punto en el que existe: en las mentes de los constructores. Es la mente la que debe ser transformada, no el mundo real.

Guba (1990:17-27), después de analizar las distintas corrientes paradigmáticas, afirma que hemos de situarnos en un nivel distinto y reemplazar todos los paradigmas actuales por otro que él denomina el paradigma dialogante. Dialogante con la problemática y realidad a la que pretende comprender y mejorar.

1.4. El debate metodológico y la complementariedad metodológica

La tesis kunhniana de que los paradigmas se suceden después de un período de crisis del viejo y la emergencia del nuevo paradigma que le sustituye, hasta hoy no se ha cumplido. Más que suceder un paradigma a otro tienden a coexistir y hasta a complementarse.

Toda investigación comparte, al menos, el intento de dar respuesta a un problema sentido, a un problema real, detectado en la práctica. Después un enfoque metodológico, el que busca una explicación para predecir y generalizar, se acerca a la teoría que puede aportar luz para aclararlo y ayudar a formular conjeturas posibles de solución, informando, también, la teoría, la elaboración de los instrumentos de recogida de datos, confeccionados éstos de un modo cualitativo, y buscando siempre la representatividad en la confección de los mismos y en su aplicación. El enfoque naturalista, de otro lado, trata de resolver el problema acercándose a la situación donde se produce ese problema, con el afán de comprenderlo desde el punto de vista de los propios protagonistas, mediante métodos y técnicas de observación y análisis de la realidad, como son la observación participante, los grupos de discusión o las entrevistas. Busca la profundidad más que la representatividad, selecciona las muestras intencionadamente, frente al muestreo al azar del enfoque explicativo. Pero lo cierto es que la explicación se enriquece con la comprensión.

La coexistencia de los paradigmas de investigación en las Ciencias Sociales y la búsqueda de la complementariedad metodológica, cuando ello es posible, caracteriza hoy la investigación en educación y en ciencias sociales.

1.5. Ejercicios de autoevaluación

1. Características esenciales del concepto de ciencia desde la perspectiva positivista.
2. Concepción de la ciencia desde la perspectiva humanista.
3. Características definitorias del método científico.
4. Mencione las características esenciales del paradigma positivista.
5. Mencione las características esenciales del paradigma crítico.
6. Mencione las características esenciales del paradigma interpretativo.

7. Elabore un cuadro de los diferentes tipos de investigación con sus características definitorias esenciales.
8. Defina el concepto de paradigma en el contexto de la investigación.
9. Haga una valoración crítica de las posibilidades reales de la complementariedad metodológica.

1.6. Solución a los ejercicios de autoevaluación

Consulte los epígrafes señalados entre paréntesis al final de cada pregunta.

1.7. Referencias bibliográficas

- BELTRÁN, M. (1994). «Cinco vías de acceso a la realidad social». *Revista Española de Investigaciones Sociológicas*. n.º 29, 1985. En García Ferrando, J., Ibáñez, J. y Alvira, F. (1994). *Análisis de la realidad social. Métodos y técnicas de investigación*. Madrid: Alianza Universidad.
- BUNGE, M. (1976). *La investigación científica*. Barcelona: Ariel.
- CARR, W. y KEMMIS, S. (1986). *Becoming critical: Education, knowledge and action research*. Lewes: Falmer Press.
- FEYERABEN, P. K. (1975). *Against method: Outline of an anarchistic theory of knowledge*, Atlantic Highlands, NJ: Humanities Press.
- GUBA, E. G. (1990). «The alternative paradigm dialog». En E. G. Guba (Ed). *The paradigm dialog*. Newbury, CA: Sage.
- HABERMAS, J. (1971). *Knowledge and human interests*. Boston: Beacon Press.
- KERLINGER, F. N. (1987). *Investigación del comportamiento*. México: Interamericana.
- KUHN, T. S. (1962). *The structure of scientific revolutions*. Chicago, Ill.: UCP.
- LANGEVEL D. M. (1983). *Reflections on phenomenology and pedagogy*, *Phenomenology and Pedagogy*, 1, 1 (5-7).
- MASLOW, A. H. (1966). *The psychology of science*. New York: Harper & Row.
- MIGUEL, M. DE (1985). «Investigación aplicada». «Investigación básica». En A. de la Orden (Coord.). *Investigación educativa. Diccionario de Ciencias de la Educación*. Madrid: Anaya. (130-31).
- ORDEN, A. DE LA (1985). «Introducción». En A. de la Orden (Coord.): *Investigación educativa. Diccionario de Ciencias de la Educación*. Madrid: Anaya.

- PÉREZ JUSTE, R. (1987). «El proceso y la lógica de la investigación científico-experimental». En: C. Jiménez Fernández, E. López-Barajas Zayas y R. Pérez Juste. *Pedagogía Experimental II*. Unidades Didácticas. Madrid: UNED.
- POPKIEWITZ, T. S. (1990). «Whose future? Whose past? Notes on critical theory and methodology». En Egon Guba (Ed.): O. c.
- SHULMAN, L.S. (1981). «Disciplines of inquiry in education: An overview». *Educational Researcher*, 10, 6 (5-12).
- SOLTIS, J. F. (1984). «On the nature of educational research» *Educational Researcher*, 13, 10 (5-10).
- VAN DALEN, D. B. y MEYER, W. J. (1979). *Manual de técnicas de investigación educativa*. Buenos Aires: Paidós.

TEMA 2

La investigación descriptiva

ESQUEMA

- 2.1.** Introducción
- 2.2.** La investigación tipo encuesta
- 2.3.** El proceso de realización de una investigación tipo encuesta
 - 2.3.1. La definición del problema de investigación
 - 2.3.2. La revisión de la literatura relevante al problema de estudio
 - 2.3.3. La identificación de la población objeto de estudio y la selección de la muestra
 - 2.3.4. La selección de la técnica de encuesta
 - 2.3.5. La construcción del instrumento de la encuesta
 - 2.3.6. Los procedimientos del trabajo de campo
 - 2.3.7. La realización de una aplicación piloto
 - 2.3.8. El análisis de datos descriptivos
 - 2.3.9. La generalización de los resultados
 - 2.3.10. El informe de los resultados de investigación
- 2.4.** La investigación exploratoria
 - 2.4.1. Estudios de relación y basados en el análisis factorial
 - 2.4.2. Estudios de predicción
 - 2.4.3. Estudios causales comparativos o investigación «ex-post-facto»
- 2.5.** Ejercicios de autoevaluación
- 2.6.** Soluciones a los ejercicios de autoevaluación
- 2.7.** Referencias bibliográficas

OBJETIVOS

- ✓ Conocer y comprender los fines de la investigación tipo encuesta.
- ✓ Comprender y valorar las características de la investigación tipo encuesta.
- ✓ Conocer y comprender los pasos para la realización de una investigación tipo encuesta.
- ✓ Conocer y comprender los fines de la investigación exploratoria y sus ámbitos de aplicación.

2.1. Introducción

La investigación descriptiva se caracteriza y diferencia de la investigación experimental porque el investigador no hace nada sobre los objetos o sujetos que investiga, excepto observarlos o encuestarlos con el fin de obtener información sobre los objetos o personas tal como son, para describir los hechos de la realidad objeto del estudio sin alterar nada. Son, por tanto, estudios exploratorios previos a otros posibles de relación o de causalidad entre variables. Introducen al lector en los acontecimientos tal cual ocurren, aportando una primera aproximación a la realidad tal como es. Puede estudiar tanto acontecimientos actuales como pasados, históricos. Cuanto más intrusiva es la investigación menos son las oportunidades de que refleje los hechos reales tal como son en la realidad.

Dentro de los estudios descriptivos se incluyen los estudios etnográficos y los estudios tipo encuesta. La investigación tipo encuesta va a ser el objeto de estudio de este tema. La investigación etnográfica se estudiará en el tema siguiente dedicado a la investigación cualitativa.

2.2. La investigación tipo encuesta

El propósito de la investigación tipo encuesta es describir las características específicas de un amplio grupo de personas, objetos o instituciones, mediante la utilización de diferentes técnicas de recogida de datos en un grupo más reducido.

Así, una investigación tipo encuesta podría tener el propósito de conocer el tiempo que tardan los universitarios en concluir sus estudios. Otra podría tener como objeto conocer las expectativas de los alumnos ante las nuevas titulaciones cortas. Otra podría centrarse en conocer las razones que llevan a los alumnos a matricularse en una Universidad a Distancia.

Estas posibles investigaciones comparten varias características como son:

- 1.º Conocer ciertos hechos específicos que describen a un grupo.
- 2.º Estudiar grupos bien definidos.
- 3.º Conocer cuestiones específicas supuestas en un grupo.
- 4.º El modo de asegurar que se recoge la información deseada es preguntar a la gente adecuada mediante la formulación de la pregunta correcta de acuerdo con la cuestión a investigar.

Si un investigador estuviera interesado en conocer el número de licenciados en Pedagogía que forman parte de los Equipos de Orientación en los Centros de Educación Secundaria Obligatoria del Estado Español, y el número de años que llevan trabajando como tales con dedicación completa, podría dirigirse a todos estos centros para recabar dicha información, y una vez que los centros aporten dicha información podría calcular la media de años que llevan dedicándose a tiempo completo a esta profesión, y realizar diversos análisis estadísticos de los datos.

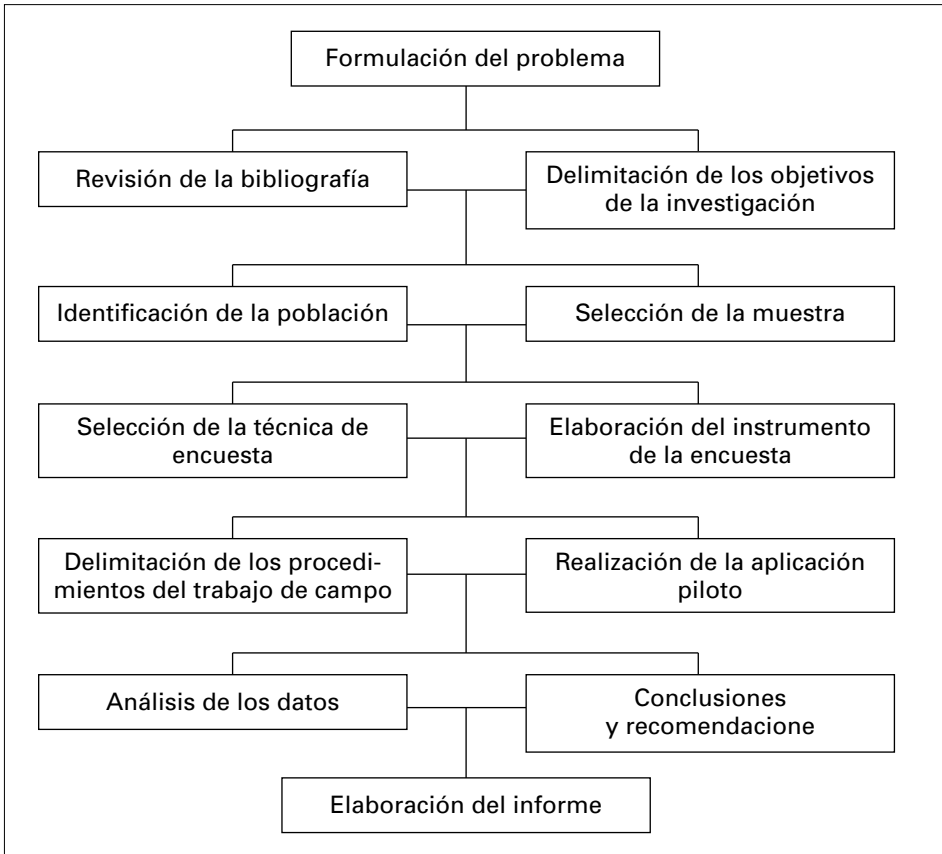
En las investigaciones tipo encuesta los datos se recogen de algunos miembros del grupo de interés, más que de todos los miembros del grupo, siempre que al menos tengan una característica en común –en nuestro caso ser Licenciados en Pedagogía, trabajar en Centros de Educación Secundaria Obligatoria, y realizar funciones de Orientación con dedicación completa–, conforman la *población*. El grupo reducido procedente de una población utilizado para estudiar las cuestiones que interesan al investigador se denomina *muestra*.

Si pretendemos que los valores obtenidos del estudio de las muestras nos proporcionen información útil para estimar los parámetros de la población, las muestras deben ser elegidas de acuerdo con unas reglas bien definidas, debiendo ser muestras aleatorias y de tamaño suficiente de modo que sean representativas de la población a la que pretendemos trasladar nuestros resultados mediante reglas de inferencia. Los resultados obtenidos de dichas muestras representativas se denominan *estadísticos muestrales*, y además de describir la muestra estudiada nos van a permitir estimar los parámetros de la población a la que pertenece la muestra. Podemos concluir que la investigación tipo encuesta es un estudio de investigación en la cual se recoge información procedente de los miembros de una muestra con el propósito de estimar uno o más parámetros de la población.

2.3. El proceso de realización de una investigación tipo encuesta

Para realizar una investigación tipo encuesta, se deben realizar diferentes pasos que explicamos a continuación.

LA INVESTIGACIÓN POR ENCUESTA



2.3.1. Definición del problema de investigación

El problema de investigación debe definirse con claridad y exhaustividad, asegurando que el problema contiene las cuestiones esenciales por las que se realiza el estudio, y eliminando las cuestiones no esenciales. Se puede comenzar por las cuestiones más generales y descender hasta las más específicas y concretas.

2.3.2. Revisión de la literatura relevante al problema de estudio

Los problemas objeto de estudio en investigación suelen contar con estudios ya realizados sobre problemas similares que deben ser consultados. Estos nos servirán para comprobar cómo estamos planificando

nuestra investigación, si habíamos ignorado alguna cuestión importante, comparar nuestros instrumentos de recogida de datos y el modo de acercarnos al campo y su análisis posterior y las conclusiones aportadas, o incluso si se han encontrado soluciones que hacen innecesaria nuestra investigación. Siempre nos ayudará a resolver algunos problemas y a evitar otros. Las revistas de investigación ofrecen fuentes para profundizar y comparar nuestra investigación con las que en ellas se publican. De cualquier modo no podemos olvidar nuestra intención de investigación y el contexto en el que se desenvuelve.

2.3.3. Selección de la muestra

La población objeto del estudio la componen aquél grupo de personas, objetos o instituciones que definen el objeto de la investigación. La definición inequívoca de la población objeto de estudio es una exigencia esencial de la investigación tipo encuesta. Debe estar definida la población de tal modo que quede claro si una persona, objeto o institución pertenecen o no a dicha población. La población estudiada debe ser aquella que sirva a los objetivos pretendidos por la investigación. Si esto no sucede así los datos obtenidos de miembros que no pertenecen a la población objeto del estudio carecerán de valor para ser analizados, y asimismo los resultados y conclusiones de la investigación.

2.3.4. Selección de la técnica de encuesta

Las principales técnicas para recoger información de la población objeto de nuestro estudio son los cuestionarios, la entrevista personal o la entrevista telefónica. Los cuestionarios por correo tienen las ventajas de su economía y amplio alcance. Sin embargo, el no tener la certeza sobre quién y cómo son contestados, la valoración de sus resultados siempre se debe hacer con muchas precauciones. Las encuestas por teléfono reducen los costos de traslado personal pero son también costosas. Las entrevistas personales es la modalidad de encuesta más costosa, pero también de las más eficaces, dado que consigue con mayor facilidad la cooperación de los encuestados y aporta oportunidades de asegurar la interpretación de la pregunta y la exactitud de la respuesta; la observación del encuestado durante la entrevista aporta información adicional importante, y asegura quién es la persona que responde. La entrevista, dependiendo de los objetivos de la investigación, puede ser la modalidad de encuesta más recomendable.

Por motivos de reducción de gastos, se suele utilizar el cuestionario mediante su envío por correo, aunque a veces, en función del objetivo de investigación, se combinan las tres alternativas, o al menos en las diferentes fases de la realización de la investigación; así en la etapa piloto se suelen aplicar los cuestionarios en grupos reducidos para aclarar dudas, y en etapas finales se puede seleccionar una muestra reducida para profundizar en algunas cuestiones; también, después de enviar las encuestas por correo, se puede realizar llamadas telefónicas para conseguir una mayor colaboración. Cada procedimiento requiere su propio análisis.

2.3.5. Construcción de los instrumentos de la encuesta

El instrumento de recogida de la información, ya sea cuestionario, encuesta telefónica o entrevista, o varios, es uno de los elementos más importantes, sino el que más, en una investigación tipo encuesta; sus preguntas deben estar formuladas de modo claro, unívoco, adecuado al objetivo de la investigación. La claridad en las cuestiones se entiende en términos de que sea interpretada del mismo modo por todos los encuestados, es decir, en términos de fiabilidad. Otra cosa es como cada cual la responda. La longitud del cuestionario debe ser tal que evite preguntas que no aporten información esencial para aquello que se estudia. El vocabulario en un nivel que pueda ser entendido por la mayoría de la población con la que se trabaja. El tipo de preguntas será diferente en función de que estén dirigidas a conocer hechos, opiniones, juicios de valor. Su formato puede ser de respuesta de sí-no, de respuesta múltiple, de indicación de grado de acuerdo en una escala de diferente amplitud, de libre respuesta, etc.

La calidad técnica con la que se construyan los instrumentos de recogida de información será determinante para recoger exactamente la información pretendida, debiendo estar estructurados en torno a las categorías de contenidos que conforman el tema investigado, debiendo dichas categorías estar bien definidas, ser completas en sí mismas, y no debiéndose superponer unas con otras; Asimismo, los ítems que integran cada categoría deben estar bien definidos y contruidos, de modo que sean unívocos, dirigidos a recoger la información en coherencia con la categoría a la que pertenecen.

2.3.6. Los procedimientos del trabajo de campo

La etapa piloto nos habrá aportado información para revisar aspectos tales como las instrucciones de aplicación, la suficiencia de la muestra

dada la ratio de encuestas completadas y devueltas, las cuestiones ambiguas, las cuestiones innecesarias y las cuestiones que deberán añadirse, los tiempos necesarios para cada fase del proceso completo. A partir de esto debemos planificar con una mayor precisión los procedimientos para la realización de la investigación definitiva; sus objetivos y sus plazos deben estar previstos y fijados y se debe perseguir su cumplimiento, aunque los plazos de devolución y las ratios de respuesta pueden diferir de los aportados por la aplicación piloto. Se deben fijar los plazos de devolución de los cuestionarios contestados, y llevar un registro de los cuestionarios completados recibidos. En una investigación mediante entrevista, una vez elaborado el protocolo de las cuestiones incluidas en la entrevista el plan de trabajo de campo debe incluir entrenamiento de los entrevistadores, seguimiento del proceso de las entrevistas, recogida de los protocolos completados de las entrevistas y normas por si fuera necesario completar la información de los entrevistados. Si el plan de trabajo no está bien definido la calidad de la investigación se puede resentir.

2.3.7. Realización de una aplicación piloto

La aplicación piloto supone una puesta a prueba de los procedimientos de toda la investigación a escala reducida, desde la definición de la población meta, a la delimitación de la muestra por el procedimiento elegido, el envío de los cuestionarios o la realización de las entrevistas, la recogida de los mismos comprobando los tiempos y las ratios de respuestas, la reducción de los datos aportados por las encuestas en bases de datos, el análisis de los mismos y su interpretación, lo que nos llevará a detectar errores y fallos que nos dará la oportunidad de evitarlos en la realización de la investigación a la escala definitiva.

2.3.8. El análisis de los datos descriptivos

Cuando los datos en una investigación son recogidos mediante cuestionario o entrevista, en sus formatos originales no permiten el análisis inmediato de la información que contienen. Hay que pasar las puntuaciones e información de los cuestionarios a una base de datos al ordenador para su análisis mediante programas informáticos de análisis de información y datos.

El proceso de transformar los datos para su análisis se denomina *reducción de datos*, y puede hacerse por el procedimiento manual de introducir cada dato numérico en su fila y columna, o mediante un procedimiento mecánico con hojas de respuesta de lectura óptica que son

introducidas en un lector óptico que automáticamente transfiere los datos al ordenador, o incluso mediante un terminal de ordenador en el que directamente el encuestado introduce sus respuestas. Los datos de las respuestas se sitúan en filas y columnas: en las filas se ordenan los sujetos, de acuerdo con un número de orden otorgado a cada cuestionario individual y en las columnas se sitúa el valor numérico que otorga dicho sujeto a cada una de las preguntas o ítems, de modo que la totalidad de la fila corresponde a las respuestas dadas por un único y mismo sujeto, y cada columna corresponde a la respuesta otorgada a una y misma pregunta del cuestionario por la totalidad de los sujetos que la contestan.

Las cuestiones de respuesta abierta deben ser transcritas directamente al ordenador para su análisis textual posterior, o ser transformadas en códigos numéricos para su análisis numérico. Casi siempre hay algunos cuestionarios incompletos, con fallos, con contradicciones, y para detectarlos la base de datos obtenida tras el proceso de reducción de los datos debe ser revisada para separar los datos procedentes de aquellos cuestionarios incompletos. A veces los errores pueden deberse a preguntas mal formuladas, ambiguas o equívocas, de modo que pueden inducir a respuestas ambiguas y contradictorias, por lo que se recomienda una cuidadosa revisión previa del cuestionario o del protocolo de las entrevistas, e incluso es recomendable realizar análisis de datos previos con el fin de detectar posibles errores en la formulación de las preguntas. Normalmente los propios programas de análisis de datos detectan los casos incompletos y contradictorios, los contabilizan y los excluye del análisis general.

Una vez organizada la información en una base de datos, se procede a su *análisis*. Todo proyecto de investigación debe especificar su propuesta de análisis de información, aunque no siempre se puede prever todos los análisis de datos posibles en una investigación tipo encuesta. Sin embargo, concretar el tipo de análisis de datos que se pretende realizar determinará, en parte, el tipo de preguntas que deben ser incluidas en la encuesta.

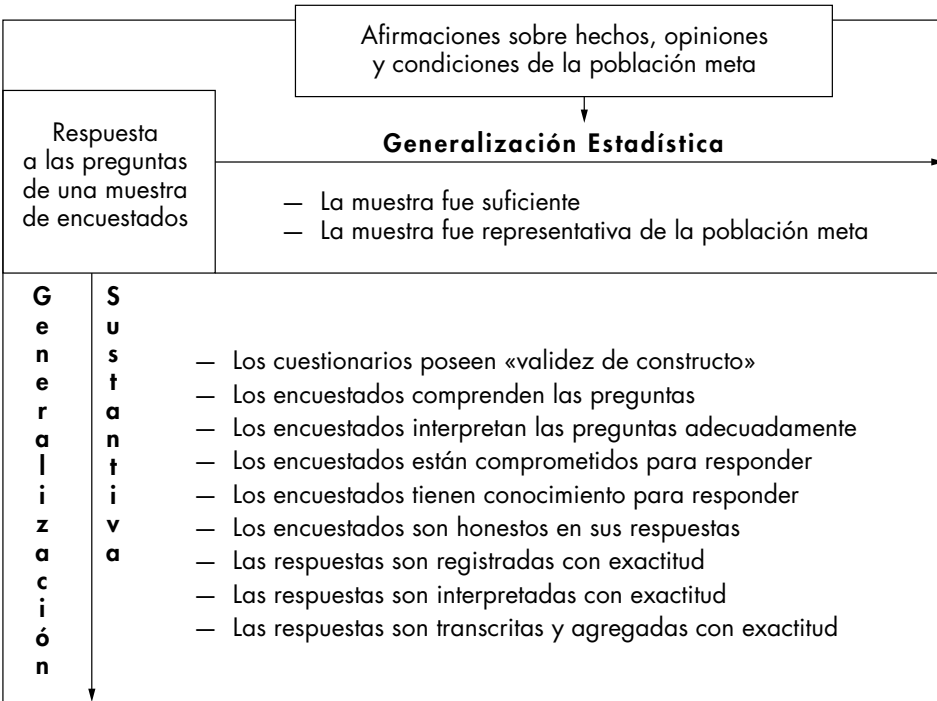
La presentación de los datos mediante tablas y gráficos facilita un análisis e interpretación intuitiva de la información recogida en la encuesta. Los análisis relacionales, como por ejemplo los análisis factoriales exploratorios, ayudan a detectar si los ítems están presentes en más de una categoría, o si algunas categorías o variables se solapan o superponen.

2.3.9. *La generalización de los resultados*

La generalización de los resultados interesa, de un modo u otro, a todas las investigaciones, pero es esencial en las de tipo encuesta. Se trabaja con grupos reducidos representativos de la población y sobre información recogida a partir de una cierta cantidad de preguntas integradas en un cuestionario, con el fin de conocer las opiniones, condiciones, estatus, valoraciones de instituciones, de políticas, de programas, etc, para, a partir de este conocimiento de los hechos y de sus valoraciones, elaborar proyectos, planes, programas, reformas.

La importancia de la finalidad de la investigación tipo encuesta hace absolutamente necesario que la credibilidad de los informes y sus conclusiones estén fundamentados en la teoría, la lógica y la metodología rigurosas seguidas en el proceso de toda la investigación.

La generalización de los resultados de una investigación es parcialmente estadística y parcialmente sustantiva como hemos señalado al hablar de la representatividad de las muestras de la población, de la representatividad de las preguntas introducidas en los instrumentos de las preguntas de investigación, y de la adecuación de los análisis de datos al tipo de preguntas. Las condiciones de respuesta también deben tenerse en cuenta, recomendándose controles para asegurar cómo y quién responde a las cuestiones introducidas en las encuestas de modo que altere lo menos posible las características de la población con la que trabajamos. Las generalizaciones, no obstante, deben ser interpretadas, dependiendo del tipo de investigación, con las precauciones que la ciencia estadística y el sentido práctico nos aconsejan.



Generalización estadística y sustantiva en la investigación por encuesta.
(Jaeger, 1988:326)

2.3.10. El informe de los resultados de la investigación

El informe de los resultados de la investigación es el documento final de la misma y debe garantizar la consistencia con los objetivos de la investigación. El informe debe dar respuesta a las preguntas iniciales tales como las siguientes: ¿Por qué se querían recoger esa información? ¿Cómo se pretendía recoger? Y ¿Para qué se va a utilizar? Esto ayudará a asegurar la utilidad de la información recogida y la coherencia entre los objetivos de investigación y la investigación realmente realizada. El informe deberá seguir una estructura que refleje todo el proceso seguido en la investigación, y los resultados deben ser expuestos utilizando las diferentes categorías que integran el problema de investigación. Se deben ir presentado las tablas de resultados, los gráficos y la interpretación literal, de un modo secuenciado, destacando lo más relevante.

El informe de investigación se puede presentar en diferentes formatos y extensión en función de las diferentes audiencias que los leerá. Todo esto deberá ser previsto en el plan de la investigación, incluyendo las

fechas previstas en la entrega del informe y el modo de distribución a las diferentes audiencias o grupos interesados en conocer los resultados, junto con aquellos con los que nos habíamos comprometidos a hacerlos llegar.

En un apartado final del informe se debe incluir una revisión crítica valorando los aspectos positivos y negativos de la investigación y realizando recomendaciones para posteriores investigaciones. El informe debe concluir con una síntesis final de interpretación y valoración global de la información aportada por los datos.

2.4. La investigación exploratoria

La investigación exploratoria tiene tres objetivos fundamentales:

- descubrir las variables significativas en la situación de «campo»,
- detectar las relaciones entre las variables y
- poner los cimientos para investigaciones posteriores. (Kerlinger, 1985:286).

Dentro de los estudios de campo exploratorios cabe distinguir entre aquellos que buscan describir variables significativas de las situaciones y los correlacionales, que pretenden descubrir el grado de relación o asociación existente entre las variables que intervienen en un fenómeno previamente estudiado mediante estudios descriptivos y que aportan indicios sobre posibles relaciones entre las variables. Suele utilizar cuestionario y pruebas objetivas para recoger datos acerca de las variables que estudia, para posteriormente calcular las correlaciones existentes entre las variables estudiadas mediante pruebas estadística correlacionales. Son exploratorios porque tratan de ver lo que hay en las situaciones que estudian. Los estudios correlacionales, como cualquier estudio exploratorio, son previos a los estudios causales confirmatorios.

Dentro de los estudios correlacionales se agrupan de acuerdo con la siguiente tipología:

- Estudios de relación y los basados en el análisis factorial
- Estudios de predicción
- Estudios causales comparativos o investigación «ex-post-facto»

2.4.1. Estudios de relación y basados en el análisis factorial

Los estudios de relación tratan de señalar las relaciones actuales existentes entre las variables a estudiar, lo que nos va a permitir eliminar el

efecto enturbiador de otra u otras variables, así como ver el grado de relación entre una variable y un grupo de variables conjuntamente. Suelen configurar un paso previo a estudios de tipo causal.

Los estudios basados en el análisis factorial pretenden descubrir los factores o constructos que hacen que las puntuaciones obtenidas a través de un determinado instrumento de recogida de datos se relacionen. Su finalidad es la de simplificar y organizar un gran número de correlaciones, para ello identifican los factores comunes de todas las medidas. Debemos distinguir entre el análisis factorial exploratorio y el confirmatorio en línea de los estudios causales.

Tanto en los estudios descriptivos como los correlacionales, desde una perspectiva temporal se realizan estudios transversales —en el tiempo— y longitudinales —a lo largo del tiempo.

2.4.2. Estudios de predicción

Los estudios de predicción pretenden conocer el efecto y magnitud de una o más variables que sirven para predecir otra variable dependiente o criterio. Las variables independientes serían las predictoras. Estos estudios se basan en la correlación y la ecuación de regresión. La utilidad de los estudios de predicción puede verse en el campo del diagnóstico, la orientación, la evaluación, así como en el estudio de las características técnicas de los instrumentos de medida, para conocer su validez predictiva.

2.4.3. Estudios causales comparativos o investigación «ex-post-facto»

Dentro de los estudios correlacionales, Fox (1972) agrupa los estudios causales comparativos, también denominados investigación «ex-post-facto» (Kerlinger, 1985: 268). La investigación «ex-post-facto» es definida como «una búsqueda sistemática empírica, en la cual el científico no tiene control directo sobre las variables independientes, porque ya acontecieron sus manifestaciones o por ser intrínsecamente no manipulables. Se hacen inferencias a partir de la variación concomitante de las variables independientes y dependientes». En la investigación ex-post-facto no se puede afirmar que tal variable independiente sea la causa de otra variable dependiente, debido a que los acontecimientos ya han ocurrido y es, por tanto, imposible controlar si otras posibles variables independientes han podido incidir en la variable dependiente.

La finalidad de los estudios retrospectivos es la de conocer los antecedentes de alguna situación actual mediante la aclaración de aspectos del pasado, y se basan en el análisis de datos recogidos cierto tiempo después del acontecimiento que se estudia.

2.5. Ejercicios de autoevaluación

2.5.1. Conceptos fundamentales del tema

1. Características definitorias de la investigación descriptiva.
2. Características esenciales de la investigación tipo encuesta.
3. Definición de la investigación exploratoria

2.5.2. Síntesis del tema

1. Elabore una síntesis personal sobre los contenidos desarrollados en el tema.
2. Haga una síntesis del proceso de realización de una investigación tipo encuesta.

2.5.3. Cuestiones sobre los contenidos del tema

1. Defina un posible problema a investigar mediante la investigación tipo encuesta.
2. Indique el procedimiento que seguiría para resolverlo. Argúmentelo.

2.6. Soluciones a los ejercicios de autoevaluación

Consulte los diferentes epígrafes del tema.

2.7. Referencias bibliográficas

FOX, D. (1980). *El proceso de investigación en educación*. Navarra: EUNSA.

JAEGER, R. M. (Ed) (1988). *Complementary methods for research in education*. Washington: American Educational Research Association.

KERLINGER, F. N. (1985). *Investigación del comportamiento*. México: Interamericana.

TEMA 3

La investigación experimental

ESQUEMA

- 3.1.** Introducción
- 3.2.** Funciones y objetivos de los diseños
- 3.3.** Clasificación de los diseños
- 3.4.** Criterios para valorar los diseños: su validez
- 3.5.** El control de las variables extrañas
- 3.6.** El proceso general de la investigación experimental
 - 3.6.1. Formulación del problema
 - 3.6.2. Revisión de la literatura sobre el tema de investigación
 - 3.6.3. Formulación de las hipótesis
 - 3.6.4. Constitución del plan experimental
 - 3.6.5. Realización del experimento
 - 3.6.6. Procedimientos para el análisis de los datos
 - 3.6.7. Interpretación de los resultados y conclusiones
 - 3.6.8. El informe de investigación. Finalidad y estructura del informe
- 3.7.** Ejercicios de autoevaluación
- 3.8.** Solución a los ejercicios de autoevaluación
- 3.9.** Referencias bibliográficas

OBJETIVOS

- ✓ Conocer y comprender los fines de la investigación experimental
- ✓ Conocer y comprender las características definitorias de la investigación experimental
- ✓ Comprender y valorar el proceso general de investigación experimental.
- ✓ Comprender y valorar la importancia de la definición del problema.
- ✓ Comprender y valorar las hipótesis y su formulación.
- ✓ Valorar las características de la constitución del plan experimental.
- ✓ Comprender la importancia del análisis estadístico de los datos
- ✓ Comprender la importancia de la interpretación de los resultados y la elaboración de las conclusiones
- ✓ Comprender y valorar la importancia del informe de investigación para la difusión de sus resultados.

3.1. Introducción

El experimento, consiste en modificar deliberadamente y de manera controlada las condiciones que determinan un hecho y en observar e interpretar los cambios que ocurren en este último. El objetivo esencial de la experimentación consiste en la comprobación de hipótesis, ya sea para aumentar el conocimiento científico en campos todavía poco conocidos, bien para estudiar cuál es la explicación correcta de un fenómeno, bien para confirmar la validez de una determinada teoría. Pretende lograr la evidencia empírica mediante la provocación de cambios en una variable llamada independiente y el registro de los posibles cambios, o la ausencia de ellos, en otra variable llamada dependiente, manteniendo controladas el resto de variables que aparecen en el proceso intentando neutralizarlas. Con el oportuno control, podemos garantizar la existencia de una relación válida entre esas variables, pudiéndose establecer relaciones de causa-efecto.

Mediante la experimentación se observan y estudian los hechos, se predicen determinadas relaciones, se formulan hipótesis para confirmar dichas predicciones, y su confirmación permite el establecimiento de leyes. Sus rasgos distintivos son la manipulación de la variable experimental y el control de las variables intervinientes en el fenómeno estudiado.

3.2. Funciones y objetivos de los diseños de experimentos

Kerlinger define el diseño experimental como «el plan, la estructura y la estrategia de investigación concebidos para obtener respuestas a preguntas de investigación y controlar la varianza». El plan es el esbozo general del proyecto de investigación e incluye desde la formulación de la hipótesis como tentativa para solucionar el problema de investigación, hasta el análisis de datos. La estructura del diseño experimental se refiere al esquema del papel que van a jugar las variables de estudio. La estrategia hace referencia a los métodos de recogida y análisis de los datos.

Arnau (1981:10) le da un sentido más general al diseño experimental, incluyendo en él todos los procedimientos requeridos en una investigación experimental, que van, desde la formulación de las hipótesis, determinación de las variables independientes y las dependientes, al control de las variables experimentales, la manipulación de las variables independientes y el registro de los valores de la variable dependiente, el análisis de la varianza producida en la variable dependiente y la inferencia de las relaciones entre las variables en estudio.

Siguiendo a Kerlinger y a Arnau, las metas de un diseño experimental son responder a las preguntas de investigación, controlar la variabilidad, o la varianza, de la variable dependiente y la generalización de la investigación a sujetos, tratamientos y situaciones a la población.

Responder a la o las preguntas de investigación es el primer objetivo de todo diseño de investigación. Un problema de investigación admite varias hipótesis como soluciones posibles, siendo algunas de ellas más fáciles de poner a prueba empíricamente que otras. Por ello, podría decirse que hay tantos tipos de diseños como posibilidades tengan las hipótesis de ser puestas a prueba.

El diseño sugiere las direcciones de la observación y del análisis. Según sea el enunciado de una hipótesis el diseño sugiere si se necesita uno, dos o más grupos; también nos sugiere el número de sujetos por grupo para utilizar una determinada prueba estadística; el tipo de observaciones que para utilizar dicha prueba tiene que obtener; las características de los instrumentos para recoger los datos, el nivel de medición exigible para aplicar determinada prueba estadística, las variables extrañas que puede controlar y las que no, cuales son las variables activas y cuales las atributivas, y finalmente, permite aportar posibles soluciones extraídas de los resultados de los análisis estadísticos de los datos.

Controlar la variabilidad es el segundo objetivo de los diseños experimentales. Se trata de controlar que la variabilidad (varianza) de la variable dependiente entre los grupos estudiados sea debida al efecto de la variable independiente experimental y no a otras variables extrañas. Para ello hay que comprobar si realmente la variable experimental activa influye realmente en la variable dependiente, de forma significativa, y no sólo como efecto del azar. Este segundo objetivo de los diseños de investigación experimental lo expresa Kerlinger con el principio estadístico *max-min-con*, que significa *maximizar* la varianza debida a la variable de investigación sustantiva —la variable independiente o tratamiento—, *minimizar* la varianza de error o varianza aleatoria debida a las variables extrañas que puedan intervenir con la variable independiente en su influencia en la variable dependiente, *controlar* la parte de la varianza debida a las variables extrañas que pueden afectar a los resultados experimentales.

La varianza experimental es la modificación sufrida en la variable dependiente debida, presumiblemente, al tratamiento experimental y/o a su interacción; la varianza de error es la variabilidad de las medidas debido a fluctuaciones aleatorias cuya característica es su auto-compensación; la varianza de error se minimiza cuando más controladas están

las condiciones del experimento, cuando se aumenta la fiabilidad de los instrumentos de medida.

La generalización es el tercer objetivo o meta de los diseños experimentales. A la generalización también se la denomina *validez externa* de la investigación, lo que significa su poder de generalización. En sentido general se trata de la posibilidad de extender los resultados de la investigación a otros sujetos, tratamientos, situaciones. La generalización pasa por dos etapas diferentes: *a)* la definición clara de las poblaciones presentes en el experimento y *b)* la creación de muestras representativas.

La generalización de los resultados de una investigación es parcialmente estadística y parcialmente sustantiva en relación con el tema investigado. La primera guarda relación con la representatividad de las muestras, y la segunda con la representatividad y adecuación de las preguntas introducidas en los instrumentos de recogida de información —medición—, y con la adecuación de los análisis de datos al tipo de preguntas y características de los grupos.

En definitiva, el diseño de investigación nos proporciona un modo de poner a prueba la relación causal que entre las variables independientes y dependientes se postula en la hipótesis. Todo ello nos permitirá establecer inferencias válidas sobre la relación entre las variables.

3.3. Clasificación de los diseños

Los criterios más comúnmente empleados para la clasificación de los diseños estadísticos son los siguientes:

- La línea de investigación.
- El grado de control.
- La técnica de control.
- El número de variables dependientes.

Según **la línea de investigación**, cabe hablar de *diseños estadísticos* y *diseños no estadísticos*. A los diseños estadísticos también se les conoce con el nombre de diseños de grupos; siguen la línea de investigación hipotético-deductiva, recurren a la aleatorización como técnica de control y emplean técnicas estadísticas para el contraste de hipótesis. Los segundos, los diseños no estadísticos, siguen la línea de investigación inductiva. Estudian a un sólo sujeto, o muestras muy reducidas, y no utilizan pruebas estadísticas. Para el control de los efectos del tratamiento o incidencia de la variable independiente sobre la dependiente utilizan la constancia y la eliminación o aparición de lo que pretendían con el tratamiento.

Otro criterio de clasificación de los diseños de investigación es **el grado de control**. Atendiendo a este criterio se clasifican en diseños *pre-experimentales*, *cuasi-experimentales* y *experimentales* (Campbell y Stanley, 1974). Con el término control nos estamos refiriendo al de las variables experimentales que pueden interferir con la variable independiente o tratamiento experimental e influir, sin saber cómo, en los resultados observados en el experimento. Son diseños de grupo y utilizan técnicas de control como la aleatorización y el emparejamiento al formar los grupos. Los diseños pre-experimentales no utilizan grupo de control por lo que no se consideran diseños experimentales, propiamente dicho.

Los diseños de investigación cuasi-experimental son aquellos en los que, como en la investigación experimental, también se manipula el tratamiento pero la asignación de los sujetos al azar en los distintos grupos no es posible. Si los grupos no han sido seleccionados al azar los participantes pueden ser diferentes y por tanto las diferencias finales no pueden suponerse causadas exclusivamente por el tratamiento. Esto se debe a las imposiciones de la muestra disponible por trabajar con sujetos en un medio natural, en una situación real, por lo que debemos ser conscientes de que se trata de una situación más proclive a la contaminación debida a la presencia de variables independientes extrañas.

Los experimentos de campo suponen un menor control sobre la investigación que los experimentos de laboratorio, pues el investigador no interviene en el proceso excepto al decidir qué fenómeno hay que observar o medir y en el acto mismo de la medición o evaluación. Sin embargo, debemos señalar las ventajas de los experimentos de campo sobre los experimentos de laboratorio para su aplicación a las situaciones naturales, lo que posibilita una mayor validez externa permitiendo transferir los resultados a otros grupos y situaciones con problemáticas similares a las estudiadas, siempre que la validez interna esté garantizada.

Los diseños experimentales se caracterizan por realizar un control óptimo sobre las variables, tanto las experimentales independientes como sobre otras variables que en situaciones naturales podrían interferir entre las variables experimentales y las dependientes contaminando los resultados de la investigación. El investigador, de este modo, produce las condiciones bajo las que va a observar los efectos de la variable independiente sobre la variable dependiente, alcanzando una gran validez interna y una mínima varianza de error.

Kerlinger señala los siguientes objetivos de los diseños experimentales:

- Descubrir relaciones entre variables en condiciones no contaminadas, es decir, de riguroso control.

- Probar hipótesis derivadas de la teoría.
- Perfeccionar teorías e hipótesis formulando hipótesis relacionadas con otras y ayudando a la elaboración de sistemas teóricos complejos.

Las diferencias existentes entre ambos radican en el grado de control de la varianza debida a la incidencia de variables independientes extrañas —y esto lo posibilita el medio en el que se realiza la investigación: laboratorio, aislado de factores contaminantes—. Posteriormente se pueden poner a prueba estas hipótesis en situaciones normales.

Otro criterio de clasificación de los diseños es el referido a **la técnica de control de las variables**. Si se emplea la técnica de control de la aleatorización tenemos los diseños de dos o más grupos seleccionados al azar. Asimismo se emplea esta técnica para asignar los tratamientos a los grupos. También se puede realizar el control de los grupos mediante la equivalencia inicial de éstos en los valores de una determinada variable —técnica de grupos apareados—, o mediante la técnica de bloqueo consistente en dividir la muestra en grupos atendiendo a alguna característica o variable de bloqueo. Tanto la variable de apareamiento como la de bloqueo deben guardar relación con la variable dependiente que queremos estudiar. Si el sujeto es su propio control tenemos el diseño intrasujetos, y en él se somete a cada uno de los sujetos a todos los valores de la variable independiente o condiciones experimentales, midiéndose en cada caso los valores que alcanza la variable dependiente. La comparación de ambas medidas permite determinar los efectos de los distintos valores de la variable independiente.

Finalmente, otro de los criterios utilizados para la clasificación de los diseños de investigación es el **número de variables independientes** objeto de estudio. De ese modo nos podemos referir a los diseños *univariados*, *bivariados* y *multivariados*. A estos últimos, los multivariados, se les conoce con el término de diseños factoriales, que admiten nuevas clasificaciones dependiendo de la técnica de control utilizada.

3.4. Criterios para valorar los diseños: su validez

Hay cuatro grandes manifestaciones de la validez interna, validez de constructo, validez estadística y validez externa, agrupadas en las ya clásicas *validez interna* y *externa*.

3.4.1. *La validez interna*

La validez interna es definida como el requisito mínimo y necesario para hacer que los resultados del experimento sean interpretables en términos de control, medición, análisis y procedimiento. Su objetivo es asegurar que los valores obtenidos en la variable dependiente se deben a los efectos de la variable independiente. El problema de la validez interna radica en si una tercera o terceras variables pueden ser la causa de tales efectos por introducirse entre ambas modificando su relación. Los factores que pueden amenazar la validez interna de las investigaciones son los siguientes:

- La historia, o hechos externos diferentes de la VI que pueden incidir en los valores de la VD, especialmente en las investigaciones en situaciones o contextos naturales y que son de larga duración.
- La maduración, definida como los cambios debidos a factores como el crecimiento biológico y psicológico de los sujetos que pueden confundirse con los cambios producibles por la VI.
- La prueba, refiriéndonos al nivel de exigencia de «suelo» y de «techo», es decir, mínimo y máximo que impiden apreciar el efecto diferencial en alguno de los sujetos.
- La instrumentación, referida a la habilidad y experiencia de los observadores.
- La regresión estadística, tendencia de los valores extremos hacia la media, con lo que el efecto de la intervención puede quedar desfigurado.
- La selección posiblemente sesgada de los sujetos, si son grupos pequeños, o grupos naturales.
- La mortalidad experimental o el abandono de sujetos de la investigación que puede producir desequilibrios en los resultados.
- La intercomunicación de los sujetos de los grupos de tratamiento y de control.

3.4.2. *La validez de constructo*

La validez de construcción o de constructo incide en la validez interna y hace referencia a la definición y operativización de las variables que integran los instrumentos de medida para recoger la información pertinente en la variable dependiente. Las definiciones inadecuadas o incom-

pletas de las variables se convierten en una de las principales amenazas de la validez de construcción y por tanto de las relaciones establecidas entre las variables.

3.4.3. *La validez estadística*

La validez estadística, relacionada con la validez interna de los diseños, pretende dar respuesta a la pregunta de si se da, o no, relación entre las variables relacionadas en la hipótesis. La validez estadística también se relaciona con la validez externa o posibilidad de inferir las conclusiones sobre relaciones a partir de los datos obtenidos en la muestra a la población, dependiendo de la adecuada extracción aleatoria de la muestra y la asignación aleatoria de los tratamientos a los grupos del experimento.

La validez estadística va a depender de:

- La potencia de la prueba, que en parte viene determinada por el tamaño de la muestra.
- Los supuestos del modelo estadístico (paramétrico o no paramétrico).
- La calidad de las medidas (fiabilidad y validez de los instrumentos de recogida de datos).
- La aplicación de los tratamientos. (Constancia y uniformidad en la aplicación de los tratamientos).
- El contexto en el que se realiza la investigación.

3.4.4. *La validez externa*

El segundo tipo de validez general de los diseños es el de la *validez externa*. El interés por asegurar la validez externa de los diseños radica en que posibilita la extensión, o generalización, de los resultados de la investigación a otras situaciones, poblaciones y contextos. Ello exige una definición clara de las poblaciones presentes en el experimento y la creación de muestras representativas de dichas poblaciones. La generalización, ya lo hemos mencionado, es el tercer objetivo o meta de los diseños experimentales.

3.5. El control de las variables extrañas

Las dos condiciones esenciales de la investigación experimental son la separación de la variable independiente experimental de las otras variables extrañas y la igualación de los grupos a estudio, para que estén en la misma posición de partida en relación con el tratamiento o variable independiente que queremos estudiar, mediante técnicas experimentales o estadísticas para controlar la posible incidencia de las variables independientes extrañas con el fin de neutralizarlas. Pues bien, para controlar las variables extrañas de un modo experimental tenemos que:

- Elegir los sujetos de modo que sean lo más homogéneos posibles en determinadas variables independientes no activas.
- Asignar los sujetos al azar a los grupos y condiciones experimentales.
- Controlar una variable extraña introduciéndola en el diseño de investigación como una variable atributo, alcanzando así control sobre ella y recogiendo información adicional acerca de su efecto sobre la variable dependiente además de información de su posible interacción con otras variables independientes experimentales.
- Cuando una variable de emparejamiento está substancialmente correlacionada con la variable dependiente, el emparejamiento como forma de control de la varianza puede ser provechoso y deseable (Kerlinger, 1985:322 y ss.).

Uno de los más importantes avances en la metodología científica ha sido el reconocimiento de la necesidad de las condiciones de control, denominadas condiciones «normales» contra las que se evalúan los tratamientos experimentales (McGuigan, 1977:147).

El *control* es la palabra clave en los diseños experimentales. Control para comprobar que las variaciones medidas en la variable dependiente se deben a los efectos introducidos por la variable independiente, y control y constancia de las demás variables independientes que de no hacerlo actuarían como extrañas. Si la asignación de los sujetos a los grupos se hace al azar o por técnicas de equiparación, como por ejemplo el emparejamiento, el diseño posibilita el control como medio para asegurar que inicialmente no se dan diferencias significativas entre los dos grupos en las variables relevantes a la variable que se quiere estudiar. El control se debe ejercer sobre:

- Las variables independientes.

- La asignación de los sujetos a los grupos.
- La asignación de los tratamientos a los grupos.
- Las variables extrañas.

El primer paso para controlar las variables extrañas será identificarlas, centrándonos en las que pudieran afectar a la variable dependiente. Un modo de conocerlas es estudiando la literatura relevante a nuestra variable dependiente, y sus investigaciones experimentales, para ver qué variables identificaron y demostraron que afectan a la variable dependiente.

Existen diferentes *técnicas de control*. La utilización de una técnica de control determinada condiciona la elección de la técnica estadística para el contraste de hipótesis consecuente con ella. Pasamos a mencionar alguna de ellas:

- *Control por eliminación*. Es la mejor manera de controlar las variables extrañas, eliminándolas. Por ejemplo, las condiciones climatológicas se eliminan desarrollando el experimento en un lugar cerrado en el que se mantienen igualadas las condiciones de temperatura, luminosidad, viento, etc.
- *Constancia en las condiciones*. Cuando no se pueden eliminar las variables extrañas, hay que tratar de mantenerlas constantes, es decir, que presenten el mismo valor para todos los sujetos de los grupos que estamos estudiando. Por ejemplo, si para medir una variable dependiente sabemos que la concentración en la tarea favorecería su ejecución, las condiciones de ruido, si no se pueden eliminar, al menos deben ser igualadas en los distintos grupos. Otro ejemplo de variable extraña a mantener constante sería las instrucciones que se les da a los sujetos cuando se les aplica una prueba para ver los posibles efectos diferenciales resultantes tras el tratamiento en la variable dependiente. Para eliminar este posible efecto, se pueden presentar las instrucciones grabadas y el experimentador se limita a conectar el reproductor de voz en cada grupo. Los instrumentos y aparatos para registrar los valores de la variable dependiente, también deben ser los mismos para todos los sujetos.
- *Selección al azar*: El azar se utiliza como técnica de control cuando de una misma población definida para el estudio de una determinada variable (universitarios de segundo año de la carrera de físicas, universitarios de último curso), que se supone que la variable que queremos estudiar se distribuye en dicha población siguiendo el modelo de la curva normal, se selecciona una muestra al azar, espe-

rando, en virtud de las leyes del azar, que los valores estadísticos (varianza, media) de los sujetos en dicha variables, sean similares, y explicables sus diferencias por el azar, tendiendo estas diferencias a autocompensarse, permitiendo asegurar a un nivel de confianza fijado, que las diferencias tras el diseño de grupo se deban a los efectos del tratamiento, y no sean fruto del azar. No obstante, existe el riesgo de que las muestras aleatorias sean inicialmente estadísticamente diferentes, por lo que los resultados estarían viciados desde el principio. La aleatorización completa, para el control de las variables extrañas, supone:

- Seleccionar al azar el número de casos de la población definida, que permita la representatividad (exigencia del muestreo aleatorio).
 - Asignar los sujetos a los grupos al azar.
 - Asignar los tratamientos a los grupos al azar.
- *El emparejamiento*: La aleatorización resulta efectiva en la mayoría de los casos pero presenta ciertos riesgos de que los grupos sean inicialmente diferentes por lo que se aconseja la realización de mediciones iniciales en las variables relevantes a aquella que queremos estudiar, o el emparejamiento de los sujetos en alguna variable relevante a la que estudiamos.
- *Balanceo*. Cuando no se pueden ni eliminar ni mantener constantes las variables, se recurre a la técnica de balanceo, consistente en equilibrar el efecto de las variables extrañas mediante la selección de los grupos a partir de una variable de bloqueo en la que son medidos todos los sujetos, formando parejas entre aquellos que han obtenido una puntuación similar y asignando un sujeto de cada pareja, al azar, a cada uno de los grupos. Tanto la variable de emparejamiento como la de bloqueo deben mantener una alta correlación con la variable dependiente.
- *Contrabalanceo*. Mediante la técnica de contrabalanceo se pretende que todos los sujetos pasen por las mismas condiciones en orden alternativo.

3.6. El proceso general de la investigación experimental

La investigación experimental supone un proceso que se estructura en torno a las siguientes fases que pasamos a comentar.

3.6.1. *Formulación del problema*

La formulación del problema debe hacerse con la mayor precisión posible y en términos que manifiesten su resolubilidad. Si el problema es complejo deberá dividirse en los diferentes subproblemas que lo integran, y tratar de ver si se puede abordar el estudio de los diferentes subproblemas en la investigación o hay que priorizar alguno de éstos para estudiarlos primero. En relación con el problema de investigación, hay que tener en cuenta que ha de expresar una relación entre dos o más variables.

Veamos un ejemplo que comentaremos a lo largo de todo el proceso:

Los profesores del Departamento de Física de un Instituto de Educación Secundaria están preocupados por el bajo rendimiento de los alumnos en los contenidos teóricos y prácticos en la materia de Física.

3.6.2. *Revisión de la literatura sobre el tema de investigación*

Una vez formulado el problema en términos que esboza su posible resolución, se debe realizar una revisión de la literatura y de las investigaciones relacionadas con el tema que nos preocupa.

El problema de investigación, y las variables que lo integran, se fundamentará en la literatura general y específica experimental. La primera fundamentación, la general, es la más importante: consiste en esclarecer la base teórica del problema. De este modo el investigador refleja la teoría relevante, fruto de investigaciones previas, el estado de la cuestión, lo que le permitirá formular con mayor precisión el problema y una mejor identificación de los elementos necesarios para resolver el problema.

En nuestro ejemplo, el equipo docente, después de revisar la literatura e investigaciones relacionadas sobre el problema y revisar programas de realización de prácticas mediante el uso de equipos informáticos, y estudiar su eficacia, decide que ha de reforzar el aprendizaje de contenidos teóricos y prácticos mediante la realización de prácticas simuladas en el ordenador. Y formulan el problema en forma de pregunta: ¿La enseñanza de prácticas de física simuladas con el ordenador mejorará el aprendizaje de los contenidos teóricos y prácticos en dicha materia?

El equipo docente ha formulado el problema de modo operativo, dado que relaciona la variable independiente (VI): *la enseñanza de prác-*

ticas de física simuladas con el ordenador, con es la variable dependiente (VD): *la mejora del aprendizaje de los contenidos teóricos y prácticos en dicha materia*. El investigador espera que VI incida en VD y debe probar dicha incidencia.

La formulación del problema de forma operativa, tal como se acaba de expresar, debe permitir su verificación empírica y las variables de la relación deben poder medirse de algún modo. Y así es:

La formulación del problema expresa una relación real entre los hechos (el bajo rendimiento) y el modo de abordarlos (probar la eficacia de un programa de prácticas). Y las variables podrán medirse: la variable independiente va a ser dosificada en sesiones de utilización del programa informático para la simulación de prácticas de física durante dos sesiones de 30 minutos introducidas en sendas clases de física a la semana, durante el primer trimestre del curso. Pasado este tiempo medirá la variable dependiente mediante un test de rendimiento con preguntas de contenidos teóricos y resolución de prácticas.

3.6.3. Formulación de las hipótesis

Una hipótesis es una relación conjetural de la relación que existe entre dos o más variables, debiendo quedar perfectamente relacionadas la variable independiente —experimental—, o tratamiento, y la variable dependiente en la que se espera influir mediante el tratamiento. La deducción de las consecuencias y la definición de los términos básicos de las variables deben quedar explícitas en la hipótesis. La hipótesis se formula en forma afirmativa, y relaciona, de manera general o específica, una variable con otra.

La hipótesis de nuestro ejemplo podría formularse del siguiente modo:

«Si utilizamos un programa informático de prácticas de física entonces mejorará el rendimiento de los alumnos en los contenidos de física tanto teóricos como prácticos».

Los criterios para identificar hipótesis científicas correctas son:

- Deben expresar las relaciones que se postulan existen entre las variables independiente y dependiente.
- Deben señalar la necesidad de verificar las relaciones expresadas, por lo que deben contener dos o más variables medibles y especificar su relación entre ellas:

— *Hipótesis unilaterales*: el investigador se decanta por la influencia de la VI (A) sobre la VD (B) en una determinada dirección:

H: «Si A entonces «mejorará», «será superior», «será inferior», B. (H. unilateral derecha).

H: «Si A entonces «empeorará», «será superior», «será inferior», B. (H. unilateral izquierda).

— *Hipótesis Bilaterales*: el investigador no se decanta por la dirección en que influirá la VI (A) sobre la VD (B) en una determinada

H: «A incidirá en B. (no se sabe si para mejor o para peor).

- Junto con la hipótesis que pretende probar el investigador en la situación experimental, se debe formular la hipótesis nula que afirma que los efectos del tratamiento no son debidos al tratamiento experimental sino que se deben al azar. Por ello se debe fijar un nivel de significación, que en Ciencias Sociales suele ser del 0,05, o más exigente del 0,01. Los datos obtenidos en la medición de la variable dependiente mediante la prueba de rendimiento se analizan utilizando la prueba estadística de hipótesis adecuada, y si la diferencia entre el valor empírico obtenido es igual o menor al valor teórico reflejado en las tablas de los valores probabilísticos de la distribución del estadístico correspondiente, con el nivel de significación fijado, podremos rechazar la hipótesis nula y dar por válida la hipótesis del investigador.
- Las hipótesis deben formularse en términos textuales, como hemos hecho más arriba, y en términos estadísticos, del siguiente modo:

$$H_1 = \mu_A > \mu_B$$

$$H_0 = \mu_A = \mu_B$$

$$\alpha = 0,05$$

Siendo μ_A la media del grupo experimental, μ_B la media del grupo de control, y α el nivel de significación fijado, en este caso, de 0,05.

Las relaciones expresadas en la hipótesis se verificarán mediante la medición de los efectos de la variable independiente o tratamiento sobre la variable dependiente. Pero, para llegar a confirmar o rechazar la hipótesis de investigación antes debemos diseñar el plan experimental, llevarlo a la práctica, medir sus resultados, organizar los datos y analizarlos mediante las pruebas estadísticas de contraste de hipótesis adecuadas.

3.6.4. Constitución del plan experimental

El plan experimental, o plan de actuación, es lo que nos va a permitir poner a prueba las hipótesis. El plan debe contemplar todos los elementos necesarios para probar las hipótesis, lo cual supone:

- Identificar todas las variables no experimentales que pudieran ejercer influencia sobre el experimento y determinar de qué manera pueden ser controladas. Por ejemplo preguntando a los alumnos si alguno de ellos está recibiendo clases particulares o complementos al estudio mediante el uso de programas informáticos para el estudio de la asignatura. Estos alumnos no podrían ser tenidos en cuenta en el experimento a la hora de utilizar sus puntuaciones en las pruebas de rendimiento para probar las hipótesis, dado que podría contaminar los resultados debidos a nuestro plan experimental.
- Definir una población acerca de la cual se desea hacer una afirmación. En nuestro ejemplo, los estudiantes de bachillerato de un determinado curso, 2.º de bachillerato.
- Seleccionar una muestra de sujetos representativos de la población, a quienes se les va a aplicar el método que queremos probar. Sólo si la muestra se selecciona al azar puede suponerse que es representativa de la población que hemos definido para nuestro estudio. Un modo de selección al azar, en este caso, sería si tenemos en el instituto cuatro grupos de 2.º de bachillerato, meteríamos en un sombrero cuatro papeletas en las cuales estuviera escrito en cada una de ellas un identificativo: curso A, curso B, curso C y curso D. Elegiríamos dos de ellas, que constituirían nuestros grupos: uno experimental y otro de control. En este caso los sujetos no son elegidos al azar, por tratarse de grupos naturales ya formados en el centro. El diseño sería cuasi-experimental, no experimental. Los grupos van a estar formados por 25 sujetos cada uno.

Aunque podíamos haber previsto el plan en el curso anterior y haber distribuido a los alumnos al azar a los cuatro distintos grupos, y después haber echado cuatro papeletas en el sombrero y elegido dos.

- Después de la selección de los grupos, para garantizar en mayor medida la aleatorización de nuestra investigación, asignamos el tratamiento experimental a uno de los dos grupos al azar, por ejemplo, lanzando una moneda al aire que determinara que si

sale cara, el primer grupo seleccionado será el experimental y el otro el control.

Nuestro procedimiento ha permitido distribuir los sujetos en los grupos al azar y asignar el tratamiento experimental a dichos grupos de un modo aleatorio, pidiendo decir cual será el grupo experimental y cual el de control. Existe un presupuesto básico importante que aconseja este procedimiento, y es que las medias aritméticas de los grupos en las variables que vamos a estudiar, o en otras similares, no difieren significativamente al iniciarse el experimento. Es decir, los grupos deben pertenecer a la misma población, esperando que su varianza no difiera significativamente. La selección al azar se hace precisamente para esto, igualar a los sujetos antes de comenzar el experimento, lo que supone una forma de control de las variables extrañas que pudieran incidir e interactuar junto con la variable experimental, o independiente, en la variable dependiente. Si no estuviéramos seguros podríamos aplicar a esta muestra una prueba que midiera alguna característica claramente relacionada con la variable dependiente, y comparar las medias de cada uno de los grupos. Y si no difieren significativamente confirmaría nuestro supuesto de igualdad de varianzas.

- Para conocer los efectos de nuestro experimento debemos realizar observaciones, consistentes en la recogida de datos empíricos, mediante la medición de la variable dependiente. También hemos debido observar todo el proceso para controlar las variables extrañas, neutralizándolas mediante igualación en los dos grupos, o evitándolas en los dos grupos, experimental y control.

El instrumento para la recogida de información, en nuestro caso una prueba de rendimiento, debe garantizar que mide lo que pretende medir: conocimientos teóricos, resolución de problemas, procedimientos de prácticas, que es válido, especialmente debe estar garantizada la validez de constructo —que mide la variable que pretende medir— y que es fiable, fundamentalmente la fiabilidad como consistencia interna que significa que todos los ítems de la prueba miden el constructo que pretende medir, y que es consistente en el tiempo con el mismo tipo de sujetos.

- A partir de la información que disponemos elegiremos la prueba estadística para el contraste de hipótesis, en función de la selección de los sujetos, el número de grupos, el tamaño de los grupos, el modo de asignación del tratamiento a los grupos, el número de

variables, el nivel de medición de las variables mediante la prueba de rendimiento.

- Es recomendable elaborar procedimientos de recogida de datos y, si es posible, llevar a cabo una investigación previa para perfeccionar los instrumentos y el procedimiento general.

3.6.5. Realización del experimento

La realización del experimental supone la aplicación del tratamiento o variable independiente y la observación y registro de los valores obtenidos en la variable dependiente.

- Se administran los dos valores de la variable independiente a ambos grupos; en nuestro caso a uno se aplica el tratamiento, o una cierta cantidad de la variable independiente, y al otro no se le aplica el tratamiento, o cantidad cero de esa variable; al primero se le denomina grupo experimental y al segundo grupo de control.
- Pasado el tiempo fijado para realizar la experiencia, un trimestre, se registran los valores obtenidos por todos los sujetos en la variable dependiente mediante las pruebas de rendimiento elaboradas para tal fin, con el requisito de que estas pruebas sean válidas —representativas de la variable que se quiere medir— y fiables —que tengan consistencia interna y que midan en sucesivas ocasiones lo mismo—, y que sean adecuadas para las características la muestra con la que trabajamos. De no hacerlo así nuestras conclusiones serían dudosas, afectando a la validez interna del diseño y por tanto, a las conclusiones.

3.6.6. Procedimientos para el análisis de los datos

Si los datos han sido recogidos mediante una prueba de rendimiento que alcanza el nivel de medida de intervalo, se pueden introducir directamente en la hoja del editor de datos del programa estadísticos informático, y seleccionar la prueba estadística adecuada para analizar los datos, basada en las diferencias entre medias,. Si no se dispone de un programa estadístico para el análisis de datos se pueden realizar, cuando los datos son reducidos, el análisis utilizando las fórmulas de las pruebas estadísticas adecuadas, y haciendo los cálculos necesarios con una calculadora.

En el caso que estamos estudiando, de dos grupos independientes, seleccionados al azar, y asignación al azar, que alcanza en la prueba de

medida de la variable independiente el nivel de intervalo, y supuestas la homogeneidad de las varianzas de los grupos, se aplicaría una prueba estadística paramétrica para grupos independientes. Dependiendo del tamaño de la muestra, si los grupos son menores de 30, en este caso son 25, se puede utilizar la prueba *t* de Student o la prueba del Análisis de la Varianza.

Si al comparar el valor empírico obtenido tras la aplicación de la prueba estadística con el de las tablas del estadístico, nos da una probabilidad asociada igual o inferior al nivel de significación fijado, nos permite rechazar la hipótesis nula y aceptar la del investigador que decía que la utilización de un programa de simulación de prácticas de física mejoraría el rendimiento en física de los alumnos. Las diferencias obtenidas entre los grupos se deben al efecto de la variable independiente o tratamiento. Si no se dieran dichas diferencias, el investigador, o equipo, deberá revisar todo el proceso para detectar a qué puede deberse el hecho de que su programa no consiga las mejoras pretendidas. Deberá revisar tanto aspectos relacionados con el propio programa, como el modo de aplicarlo, las condiciones de colaboración entre los estudiantes, el clima de aula, así como revisar otros elementos que hayan podido afectar a la validez interna del diseño, como pueden ser variables extrañas que interactúen con la variable independiente o cualquier otra variable que esté interfiriendo en el grupo de control, y muy especialmente la instrumentación refiriéndonos a las características del instrumento de medida, en este caso, la prueba de rendimiento para valorar el aprendizaje de los alumnos en los contenidos de teoría y práctica en la materia de física.

3.6.7. Interpretación de los resultados y conclusiones

Los resultados se deben interpretar en términos de validez interna y externa. Desde la validez interna, si el programa ha respondido de acuerdo con las expectativas del investigador, obteniendo diferencias significativas a favor del grupo experimental, y se ha comprobado que ha sido aplicado de acuerdo con el plan previsto, que se han controlado las variables que pudieran contaminar la investigación, y que las pruebas reunían las características técnicas de validez y fiabilidad requeridas, asegurada por tanto la validez interna del diseño, podemos pasar a realizar recomendaciones de utilización del programa a todos los grupos de bachillerato, adaptados al nivel de exigencia del programa de la materia, e incluso la recomendación de utilización de programas similares para el aprendizaje de otras materias, por ejemplo, de idiomas, donde la interacción simulada que ofrece la enseñanza multimedia, podría favorecer el

aprendizaje de los alumnos en estas áreas. Si, por otro lado, la hipótesis del investigador hubiera sido rechazada, habría analizar todo el proceso para buscar las posibles causas que han contribuido a no poder confirmar la hipótesis del investigador, pudiendo aconsejar que los programas estén mejor elaborados, o que su aplicación sea realizada de un modo más riguroso.

Hemos dicho en el primer tema de este libro que una de las características del conocimiento científico es que es autocorrectivo. Si un experimento nos llevara a una conclusión errónea, que manifiesta una inconsistencia con la teoría existente, con lo esperable. La repetición de los experimentos, subsanando las deficiencias detectadas y mejorando el control para evitar interferencias, aportaría información necesaria para apoyar nuestras conclusiones. La difusión de los resultados es recomendable que se haga mediante comunicación de la experiencia para contrastar los resultados en seminarios, congresos, revistas de difusión científica, etc.

3.6.8. *El informe de investigación*

El informe de una investigación tiene como *propósito y finalidad* indicar a los lectores cuál es el problema investigado, los métodos usados para resolverlo, los resultados de la investigación y las conclusiones inferidas a partir de aquellos. (En Kerlinger, 1987:488 y ss.). El informe debe ser escrito de tal modo que quien lo lea pueda llegar a sus propias conclusiones sobre la adecuación de la investigación y la validez de sus resultados y conclusiones.

El criterio de calidad del informe se apoya en la posibilidad de que otro investigador pueda repetir la investigación apoyándose en el informe; si los métodos utilizados para resolver el problema de investigación no han sido correctos, o no quedan claramente reflejados en el informe, la investigación no podrá ser valorada.

La *estructura del informe de investigación* es sencilla; sigue los mismos pasos que la propia investigación:

- El problema
- La metodología para su resolución
- Los resultados

3.6.8.1. *El problema*

En el primer apartado del informe, referido al ***problema y su fundamentación***, se debe mostrar el problema, la revisión de la teoría, las hipótesis que fueron sometidas a prueba, la definición de las variables y su significado, así como los informes de investigaciones previas y documentos de referencia.

La fundamentación del problema, en la literatura general y específica experimental, es importante para explicar y esclarecer las bases teóricas de la investigación, e informar al lector del informe acerca de lo que se ha investigado y lo que no se ha hecho en relación con el problema. El fin último es situar la investigación dentro del marco más general de las investigaciones y teorías ya elaboradas sobre el tema, y destacar su aportación.

3.6.8.2. *La metodología para su resolución*

El segundo apartado del informe, referido a la ***metodología y obtención de datos***, se centra en las cuestiones siguientes:

- Muestra y métodos de muestreo.
- Metodología seguida para verificar las hipótesis, tipo de diseño de investigación.
- Instrumentos para la medición de las variables.
- Métodos de análisis estadístico.
- Pruebas previas y estudios piloto si los hubiera.

En el apartado de metodología y obtención de datos el investigador tratará de exponer del modo más claro posible lo que hizo para resolver el problema. El informe debe mostrar que el criterio de «repetibilidad» se ha cumplido, es decir, que se puede repetir la investigación, analizar de nuevo los datos, y llegar a conclusiones claras sobre la adecuación de los métodos utilizados para la realización del experimento y para la obtención de los datos.

En esta sección se debe informar de *las muestras* que se emplearon, cómo se seleccionaron, y sus razones. El modo de asignación de los alumnos y de los tratamientos a los grupos, aleatorio o no, y se explicará por qué se hizo una cosa u otra.

El método mediante el cual se probaron las hipótesis debe exponerse con detalle. Si el estudio es de tipo experimental ha de describirse la

forma en la cual la variable o variables independientes fueron manipuladas, los instrumentos utilizados como material audiovisual, refuerzos adicionales, metodología de enseñanza específica, puesta a prueba, precauciones de control, etc. Si el estudio fuera ex-post-facto, se informará de los métodos utilizados para la reunión de datos.

También debe describir el informe cómo se midieron las variables dependientes. Si se han utilizado instrumentos específicos para medir la variable dependiente y aportar descripciones y justificaciones del empleo de dichos instrumentos, adjuntando pruebas de su validez, fiabilidad y exactitud. Asimismo se incluirán explicaciones y justificaciones de los procedimientos seguidos para el análisis de los datos. Si se ha realizado un estudio piloto será conveniente informar cómo se hizo y los resultados obtenidos.

3.6.8.3. Los resultados y conclusiones

La tercera y última parte del informe es la de los **resultados, su interpretación y las conclusiones**. Los resultados o los datos de una investigación son los «materiales brutos» para la solución del problema bajo estudio. Los datos y sus análisis constituyen la base de la prueba de hipótesis en la investigación. La metodología y la obtención de datos son las herramientas que permiten obtener el material «directo» de la prueba de la hipótesis y los datos.

La pregunta básica es: ¿Los datos confirman o invalidan la hipótesis de investigación? Nunca se hará bastante hincapié en la idea de que la metodología, la obtención de datos y el análisis se eligen y usan con el fin de verificar las hipótesis operativas, deducidas a partir de las preguntas generales de investigación, del problema a resolver. Este apartado se debe cuidar especialmente, presentándolo de modo preciso y completo, reflejando la forma en que los resultados guardan relación con las hipótesis. Los datos deben ser mostrados en forma reducida mediante tablas, gráficos, diagramas, de modo que ayuden a interpretar los resultados antes de pasar a comentarlos verbalmente, teniendo en cuenta la pregunta: ¿Los datos confirman o invalidan la hipótesis?

En este apartado se deben comentar las limitaciones y puntos débiles de la investigación. Si no se hubiera hecho un muestreo aleatorio, los resultados habría que circunscribirlos a la muestra particular utilizada y a otras muestras similares a ésta. Si la asignación de los grupos y de los tratamientos no se hizo de manera aleatoria, podría cuestionarse el control de variables independientes extrañas que pudieran interactuar con la

variable independiente experimental, generando dudas en las conclusiones.

Las decisiones estadísticas también pueden producir conclusiones incorrectas. Las deficiencias en la medición, en la calidad del instrumento de medida de la variable dependiente, también afectarán a las conclusiones. Si un instrumento es fiable a medias, los hallazgos serán ambiguos y no concluyentes. Si se cuestiona la validez del instrumento se cuestionará seriamente las conclusiones. De modo que, al escribir las conclusiones se debe tener en cuenta las limitaciones y las deficiencias reconocidas, siendo responsabilidad del investigador consignarlas por escrito para que el lector del informe las conozca.

Y como recomendación general, para escribir un informe, hay que decir que la primera redacción no puede ser la definitiva. Todo escrito mejora con su revisión, teniendo como meta una mayor sencillez, claridad y brevedad. Su relectura pasados unos días, y la aportación de los comentarios de los compañeros mejorarán el informe.

3.7. Ejercicios de autoevaluación

1. Características definitorias de la investigación experimental.
2. Defina la validez interna de los diseños de investigación.
3. Defina la validez externa de los diseños.
4. Definición y notas características del problema de investigación.
5. El concepto de hipótesis.
6. Las variables de la investigación. Defínalas.
7. Concepto de significación estadística para la prueba de hipótesis.
8. Objetivos del informe.
10. Haga una síntesis sobre los objetivos de los diseños de investigación de experimentos.
11. Haga una síntesis sobre la clasificación de los diseños.
12. Elabore un cuadro sinóptico sobre el proceso de investigación experimental.
13. Estructura del informe de investigación.
14. ¿Qué tipo de diseño utilizaría para probar la eficacia del aprendizaje de inglés por ordenador (aprendizaje on-line o e-learning) sobre el aprendizaje en el aula tradicional?

15. Elementos del plan experimental.
16. Importancia de la validez interna de los diseños en la investigación experimental.
17. Importancia de la validez externa en la investigación experimental.
18. Cuestiones a tener en cuenta en la interpretación de los resultados de una investigación.

3.8. Soluciones a los ejercicios de autoevaluación

Consulte los diferentes epígrafes del tema.

3.9. Referencias bibliográficas

- ARNAU, J. (1981). *Psicología experimental. Un enfoque metodológico*. México: Trillas.
- CAMPBELL, S. y STANLEY, J. (1974). *Diseños experimentales y cuasi-experimentales en la investigación social*. Buenos Aires, Amorrortu.
- KERLINGER, F. N. (1987). *Investigación del comportamiento*. México: Interamericana.
- PÉREZ JUSTE: *Pedagogía Experimental I*. Madrid: UNED (6588).
- MCGUIGAN, F. J. (1976). *Psicología Experimental*. México: Trillas.

TEMA 4

La investigación cualitativa

ESQUEMA

- 4.1.** Introducción
- 4.2.** Características de la investigación cualitativa
- 4.3.** La investigación cualitativa aplicada a la educación
- 4.4.** La investigación cualitativa y la investigación experimental. Criterios para su comparación
 - 4.4.1. Teoría primero o teoría después
 - 4.4.2. Análisis inductivo vs. deductivo
 - 4.4.3. Perspectiva holística
 - 4.4.4. Preconcepciones o posconcepciones
 - 4.4.5. Selección y muestreo
 - 4.4.6. Contacto personal
 - 4.4.7. Triangulación y convergencia
 - 4.4.8. Comparabilidad y traducibilidad vs. generalización
- 4.5.** El diseño etnográfico
- 4.6.** El análisis de datos en la investigación cualitativa
 - 4.6.1. Descubrimiento en progreso
 - 4.6.2. Codificación de los datos
 - 4.6.3. Relativización de los datos
- 4.7.** Criterios para valorar las investigaciones cualitativas
 - 4.7.1. Criterios de credibilidad, seguridad, confirmabilidad y transferibilidad
 - 4.7.2. Criterios hermenéuticos
 - 4.7.3. Criterios de autenticidad
- 4.8.** Elaboración del informe de investigación
- 4.9.** Ejercicios de autoevaluación
- 4.10.** Soluciones a los ejercicios de autoevaluación
- 4.11.** Referencias bibliográficas

OBJETIVOS

Tras el estudio de este tema, se pretende que el alumno sea capaz de:

- ✓ Conocer y comprender la naturaleza y los fines de la investigación cualitativa o naturalista.
- ✓ Conocer y comprender las notas características de la investigación naturalista.
- ✓ Comprender y valorar los supuestos teóricos en los que se fundamenta la investigación naturalista.
- ✓ Conocer, comprender las estrategias del trabajo de campo.
- ✓ Conocer y comprender los pasos para la realización de una investigación naturalista o cualitativa.
- ✓ Conocer y los ámbitos de aplicación de la investigación cualitativa.
- ✓ Comprender y valorar los criterios de valoración de las investigaciones cualitativas
- ✓ Valorar las aportaciones de los métodos de investigación cualitativa a la investigación y trabajo en Educación y Ciencias Sociales.

4.1. Introducción

En este tema vamos a estudiar otro modo de acercamiento para el estudio y búsqueda de soluciones en los contextos educativos y sociales. La investigación cualitativa tiene sus raíces en la etnografía y en la fenomenología y es utilizada en Sociología, en Psicología y en Educación. El término «investigación cualitativa» tiene un significado paralelo al de «investigación naturalista», e integra la investigación que no está relacionada con las variables y su medida. Erickson (1989:198) defiende como nombre genérico el de «métodos cualitativos» que incluye los métodos etnográficos, el interaccionista simbólico, el fenomenológico, el constructivista y el interpretativo. Las investigadoras cualitativistas Goetz y LeCompte (1988:28) prefieren utilizar el nombre genérico de «investigación etnográfica» integrando, bajo esta rúbrica, la investigación cualitativa, el trabajo de campo y la etnografía misma. Otros defienden el término de investigación naturalista e incluso interpretativo. Pero todos estos términos se utilizan con significados similares.

Las corrientes humanistas ponen el énfasis en la comprensión profunda de las personas y de las situaciones en las que se desenvuelven, más que en la explicación; ya en 1966, Maslow defendía una concepción de ciencia más inclusiva que la del paradigma positivista de modo que recoja lo idiográfico, lo experiencial, lo transcendental, lo holístico y lo personal. De acuerdo con Maslow la ciencia no debe estar centrada en un único método sino en los problemas y buscar los métodos apropiados para investigar cómo resolver la complicación.

La ciencia, desde el punto de vista heurístico y dinámico, tiene como finalidad comprender y transformar para mejorar las condiciones humanas. La ciencia, desde este enfoque, es idiográfica, es decir, se interesa por lo individual y lo particular sin pretender llegar al establecimiento de leyes generales, encontrándose, en estos aspectos, en el extremo opuesto del positivismo.

4.2. Características de la investigación naturalista

Los diferentes modos de contribuir al desarrollo de conocimiento científico se relacionan con los diferentes métodos de acercamiento a estudio de la realidad. Dicha diferenciación parte de la adaptación del método a su objeto de estudio. Por ello, lo que verdaderamente caracteriza a la investigación cualitativa es su objeto de estudio y su intención y a ellos se deben adaptar los procedimientos y las técnicas de investigación que utilicen.

La investigación naturalista pone su énfasis en el hombre como instrumento, basándose en el clásico concepto del trabajo de campo en el que el investigador usaba sus ojos y sus oídos y tomaba notas de un modo inestructurado para luego reflexionar sobre las mismas y describir los resultados en un estilo narrativo. En la investigación naturalista esta imagen se redefine: el hombre es el principal instrumento para la recogida de datos.

El objetivo del investigador naturalista es estudiar los problemas en sus propios contextos, intentando integrarse para comprender las múltiples variables que interfieren aunque sin dejar de ser un observador «extranjero». Tomará notas de campo y añadirá sus propias interpretaciones diferenciándolas. Realizará un informe de su estudio etnográfico en un estilo narrativo de realismo etnográfico, vívido, con citas anecdóticas relevantes. Después tratará de construir teoría. La finalidad es estudiar los hechos y los problemas que preocupan, no tanto para generalizar los resultados sino para buscar las soluciones más adecuadas, consensuadas, para llevarlas a la acción.

La mayoría de las investigaciones cualitativas, de acuerdo con Lincoln y Guba (1985:39-43), comparten alguna de las siguientes características que la definen:

- El objetivo es comprender el fenómeno del modo natural como ocurre en su propio contexto, en las situaciones naturales.
- El diseño es naturalista, emergente, construido en el contexto, orientado al descubrimiento.
- El investigador no intenta manipular las situaciones de investigación, procurando no introducir ninguna concepción previa sobre los resultados de la investigación.
- Atiende los distintos significados que tiene para los actores específicos, localizando puntos de contraste y de acuerdo, sobre los acontecimientos.
- Utiliza instrumentos humanos: el hombre.
- Se basa en el muestreo intencional frente a muestreo al azar de la investigación experimental.
- Utiliza métodos inductivos, generativos. Analiza los datos de modo inductivo, de lo particular a lo general.
- Se sirve de una combinación equilibrada de datos objetivos y subjetivos, siendo esencial el análisis objetivo y sistemático del significado subjetivo.

- Busca identificar vínculos causales específicos que no son identificables mediante métodos experimentales.
- Los resultados, las conclusiones, son negociados, consensuados, con los participantes en el estudio.
- Predomina una intención de búsqueda de soluciones de aplicación, en contextos particulares, determinados, frente a la de generalización.
- Utiliza criterios especiales para probar su valor.
- El informe se realiza en estilo narrativo, descriptivo.

Los métodos de investigación naturalistas tienen como finalidad describir lo que la gente hace, sabe, piensa o siente, mediante el uso de técnicas tales como la observación, la entrevista, los grupos de discusión o el análisis de documentos. Los métodos de investigación cualitativa son empíricos sin ser positivistas, en el sentido de experimentalistas que introducen una modificación para comprobar sus efectos, y rigurosos y sistemáticos al investigar los complejos fenómenos de la interacción cotidiana, utilizando como principales elementos el significado subjetivo. Así mismo se puede decir que los métodos inductivos que proceden de lo particular a lo general, y los deductivos, de lo general a lo particular, en la investigación etnográfica, están en continuo diálogo porque, si bien no se determinan de antemano ciertas categorías específicas a observar, se inducen, el investigador siempre identifica los conceptos de interés antes de comenzar el trabajo de campo, los deduce o se deducen. (Erickson, 1989:199).

La investigación cualitativa constituye una síntesis interdisciplinar por la influencia de tradiciones de investigación procedentes de la antropología, la sociología, la psicología y la educación.

4.3. La investigación cualitativa aplicada a la Educación

La Educación se interesa por la investigación cualitativa dado que contribuye al estudio de cuestiones en los propios contextos educativos, por las siguientes razones:

- 1.^a Por su peculiar modo de trabajo, que utiliza las características propias del trabajo de campo, y las técnicas de recogida de información tales como la observación, la entrevista, los grupos de discusión, para conocer lo que acontece en los centros educativos y en sus aulas.

- 2.^a Por su carácter reflexivo, que ayuda a los investigadores y a los docentes a hacer que lo familiar se vuelva extraño e interesante nuevamente. Lo que está sucediendo, y pasa desapercibido por su cotidianidad, puede hacerse visible a los ojos del extraño, el investigador. Lo que se estudia debe documentarse sistemáticamente, facilitando tomar conciencia de ello, lo que servirá para diseñar o elaborar y adoptar soluciones para mejorar la enseñanza.
- 3.^a Por la necesidad de adquirir un conocimiento específico a través de la documentación detallada de la práctica concreta.
- 4.^a Por la necesidad de considerar los significados particulares que tienen los acontecimientos para las personas que participan en ellos.
- 5.^a Por la necesidad de tener un conocimiento comparativo de diferentes ámbitos sociales, más allá de las circunstancias inmediatas del contexto local y particular. La consideración de las relaciones entre un contexto dado y su ambiente social más amplio ayuda a esclarecer lo que está sucediendo en el propio contexto. La perspectiva comparativa sirve de respaldo a los intentos de cambio (En Erickson, 1989: 201).

4.4. La investigación cualitativa y la experimental. Criterios de comparación

Tanto el investigador cualitativo como el experimental comparten las siguientes decisiones relativas al proceso de investigación:

- Formula un problema situado dentro de una perspectiva teórica, o independiente respecto de cualquier teoría.
- Escoge las fuentes de datos que posibiliten el examen del problema.
- Asume una determinada posición o rol hacia dichas fuentes de datos.
- Desarrolla procedimientos de recogida de datos.
- Analiza los datos obtenidos según su relevancia para el problema que estudia.

Sin embargo, existen otros elementos que claramente los diferencian. De acuerdo con los criterios de comparación aceptados por la mayoría de los investigadores, entre otros Goetz y LeCompte (1988) y Patton (1990), presentamos un análisis comparativo entre la investigación cua-

litativa y la investigación experimental en torno a las siguientes cuestiones:

1. Teoría primero, o teoría después
2. Análisis inductivo vs. deductivo
3. Perspectiva holística
4. Preconcepciones o posconcepciones
5. Selección y muestreo
6. Contacto personal
7. Triangulación y convergencia
8. Comparabilidad o generalización.

4.4.1. Teoría primero o teoría después

Una distinción fundamental entre los enfoques de investigación cualitativa y experimental parte de la naturaleza de los fines de sus investigaciones. Ambas investigaciones parten de un problema a resolver. Después, el investigador experimentalista se acerca a las investigaciones y teorías previas, que le ayudan a formular hipótesis hipotético deductivas como tentativa de solución al problema, y que trata de confirmar empíricamente para después relacionar sus hallazgos con las teorías, tratando de confirmarlas o refutarlas.

El investigador naturalista se acerca a la realidad para tratar de comprenderla y analizarla libre de teorías. Después tratará de describir sistemáticamente las características de los fenómenos y variables que estudia con el fin de generar categorías conceptuales, descubrir y validar relaciones entre fenómenos o comparar sus hallazgos con los constructos y postulados teóricos generados a partir de los fenómenos observados en escenarios distintos. Las construcciones teóricas creadas inductivamente pueden posteriormente tratar de confirmarse. A veces, el investigador cualitativista manipula deliberadamente el entorno siendo esta manipulación experimental un complemento de la investigación cualitativa.

4.4.2. Análisis inductivo vs. deductivo

El análisis inductivo supone una inmersión en los detalles y especificidades de los datos para descubrir interrelaciones y generar categorías de significado importantes. El investigador inductivo comienza explo-

rando de un modo amplio, con cuestiones abiertas que se van generando a partir del estudio y observación de lo que acontece en la situación, más que comprobando hipótesis derivadas de la teoría. El análisis inductivo parte de las observaciones específicas para construir modelos generales.

En contraste, en el enfoque hipotético-deductivo de los diseños experimentales se decide por adelantado, de modo deductivo, basándose en la teoría, qué variables son importantes y qué relaciones pueden esperarse entre ellas. La especificación de las principales variables y la formulación de hipótesis de investigación se realiza antes de que comience la recogida de datos, basándose en un marco teórico explícito que proporciona el marco para comprender las observaciones específicas de la realidad.

Mediante el enfoque inductivo la investigación cualitativa posibilita el análisis de las dimensiones importantes que emergen durante el estudio, sin presuposiciones previas. El investigador cualitativo intenta comprender las múltiples interrelaciones que emergen de las observaciones, sin anticipar hipótesis sobre posibles relaciones entre unas variables previamente definidas. El enfoque de aproximación inductivo facilita la comprensión de las actividades de un programa y sus resultados como fruto de la experiencia y de la permanencia del investigador observando la situación en su contexto.

4.4.3. Perspectiva holística

El investigador cualitativo pretende comprender el fenómeno que estudia o programa educativo que se desarrolla en un centro, como una totalidad. La perspectiva holística en la investigación implica la reunión de datos sobre múltiples aspectos de la situación bajo estudio, con el fin de elaborar una síntesis completa y comprehensiva de la dinámica social. Cada caso es tratado como una entidad única con su propio significado y su particular constelación de relaciones emergiendo en el contexto en el que se enmarca. La perspectiva holística asume que la totalidad es entendida como un sistema complejo superior a la suma de sus partes.

La investigación cualitativa pone el acento en la interacción entre las variables empíricas en el contexto natural no manipulado. La interrelación global de todos los factores constituye generalmente el punto de interés en la investigación cualitativa. Para una mejor comprensión de lo observado es esencial la descripción del medio social, del contexto educativo o sociopolítico de la organización en donde se produce el fenómeno.

El énfasis del enfoque holístico, de totalidad, contrasta con los procedimientos de la metodología experimental, centrada en el examen de los efectos causados por un tratamiento específico, donde la validez del diseño y la potencia del efecto del tratamiento se determinan manteniendo constantes o eliminando todos los factores extraños y contextuales posibles y operativizando las variables independientes, observando los efectos del tratamiento y recogidos a través de la medición de la variable dependiente de modo que puedan ser analizados sus efectos estadísticamente.

La crítica principal de los investigadores cualitativistas a la lógica experimentalista se fundamenta en que desde este enfoque se pierden factores importantes que no son fácilmente cuantificables y fracasa en la descripción del fenómeno que estudia como totalidad. El enfoque experimentalista simplifica la complejidad de las experiencias. La investigación experimental, en el estudio de problemas diferentes, y no poco importantes, gana en validez, precisión y poder de generalización.

4.4.4. Preconcepciones o posconcepciones

La investigación cualitativa admite las experiencias subjetivas del investigador y de los participantes, ofreciendo una profundidad en la comprensión superior a la de otros enfoques, lo que facilita el control consciente de los sesgos del observador y de la reactividad de los participantes.

Aunque los investigadores naturalistas estén familiarizados con las investigaciones empíricas relacionadas con su trabajo y utilicen marcos teóricos generales para iniciar sus estudios, asumen que se consigue una descripción más exacta si no se da por supuesto ningún aspecto de la escena social. El dejar en suspenso sus preconcepciones permite a los etnógrafos centrarse en las construcciones de los participantes, subjetivas y objetivas.

El investigador experimental se preocupa del efecto de sus preconcepciones mediante la cuidada estructura del diseño y la formulación de la hipótesis de nulidad.

4.4.5. Selección y muestreo

Quizá nada capta mejor la diferencia entre los métodos cualitativos y cuantitativos que las diferencias lógicas que subyacen en los enfoques de

muestreo. La investigación cualitativa se centra en el estudio exhaustivo de muestras relativamente pequeñas e incluso de caso único, seleccionadas determinísticamente, no probabilísticamente.

La lógica del muestreo determinístico, en la investigación cualitativa, se basa en la selección de casos ricos en información sobre un tema de especial importancia para su estudio en profundidad, de acuerdo con el propósito de la investigación, de los cuales se espera obtener gran cantidad de información.

Entre los posibles muestreos determinísticos se encuentran los siguientes:

- El de caso extremo
- El de variaciones máximas
- El homogéneo
- El de caso típico
- En función de una teoría.

En investigación cualitativa no hay normas sobre el tamaño de las muestras; el muestreo se realiza en función del propósito de la investigación, cuidando que aporte credibilidad, que sea útil y que pueda hacerse con el tiempo y recursos disponible. La validez y significatividad de la información generada por la investigación cualitativa tiene más que ver con la riqueza de información de los casos seleccionados y la capacidad de observación y análisis del investigador que con el tamaño de las muestras.

La investigación experimental suele depender de muestras amplias seleccionadas al azar. La lógica y el poder del muestreo probabilístico depende de la selección verdaderamente al azar y de la representatividad estadística de la muestra que permitirá la generalización de los hallazgos desde la muestra a la población, dado que su propósito es la generalización.

4.4.6. Contacto personal

En ambos enfoques los investigadores deben conocer el contexto y la realidad que estudian. El trabajo de campo significa introducirse de lleno participando, hasta donde sea posible, en las situaciones y los fenómenos. El investigador cualitativo combina la función de observador y de participante, con la finalidad de comprender lo que estudia desde el punto de vista interno, mientras lo describe para los que están fuera.

El trabajo de campo es la principal actividad de la investigación cualitativa. «Introducirse en el campo» significa tener un contacto personal directo con las personas en su propio medio con el fin de conseguir una comprensión de la realidad que estudia, necesitando de un periodo de tiempo suficiente para comprenderlo para hacerse familiar al grupo pero sin dejar de ver lo que acontece desde otra perspectiva. Se trata de interpretar la realidad desde el propio contexto pero con los ojos del extraño que hacen visible lo cotidiano.

El investigador experimental permanece en la escena de estudio lo imprescindible, con el objeto de evitar que su presencia interactúe con la situación que trata de modificar y comprobar sus efectos.

4.4.7. Triangulación y convergencia

La triangulación es una técnica que utiliza diferentes tipos de fuentes para asegurar las evidencias. El principio que subyace es el de recoger observaciones de una situación desde una variedad de perspectivas para después compararlas y contrastarlas. Por ejemplo, para conocer el funcionamiento de la dinámica de enseñanza y aprendizaje en el aula se pueden contrastar las observaciones realizadas por el propio profesor, las realizadas por los alumnos, y las realizadas por un observador externo, pudiendo ser contrastadas también con la revisión de una grabación de vídeo. La investigación cualitativa utiliza diversidad de fuentes y técnicas de recogida de datos para evitar sesgos y asegurar la exactitud.

En el contraste de fuentes y opiniones, es conveniente fijarse en qué difieren y en qué concuerdan. Estas diferencias y acuerdos pueden matizarse, o profundizar en ellas, mediante el debate entre todos los participantes. Sin embargo, hay que decir que el uso de la triangulación no es exclusivo de la investigación cualitativa siendo utilizada también en la investigación experimental.

4.4.8. Comparabilidad o generalización

En la investigación cualitativa se busca la comparabilidad y traducibilidad de los descubrimientos a otros problemas y contextos similares, en lugar de pretender la generalización o la transferencia directa de los resultados a grupos no investigados. La comparabilidad exige que el investigador utilice una terminología y unos procedimientos analíticos normalizados, aceptados por la comunidad científica, no únicos para cada contexto. La traducibilidad supone un nivel de explicitación de los

métodos de investigación, de las categorías analíticas y de los fenómenos y grupos estudiados, de modo que permita la realización de comparaciones con confianza (fiabilidad) y sentido entre los distintos grupos y entre los resultados del estudio con los de otras disciplinas. La comparabilidad y la traducibilidad cumplen en investigación cualitativa una función similar a la generalización de las investigaciones más controladas.

La investigación experimental pretende generalizar sus hallazgos, y por eso trabaja con muestras representativas de la población, el diseño está determinado y las condiciones controladas de modo que la validez interna esté garantizada, condición «sine qua non» para su generalización.

Tanto el enfoque cualitativo como el experimental comparten diversas técnicas y pueden llegar a complementarse. Las técnicas cualitativas pueden servir para incrementar la fiabilidad de un diseño experimental al aumentar la replicabilidad del tratamiento, y puede reforzar la validez de los resultados confirmando la relevancia de los constructos para situaciones reales. Y, de otro lado, las técnicas experimentales pueden potenciar la credibilidad de los diseños cualitativos al favorecer la posibilidad de hacer generalizaciones cuando la investigación se realiza en distintos escenarios; asimismo, se contribuye a la fiabilidad de los resultados cuando se emplean medidas estandarizadas para describir las variables de un contexto natural.

En la tabla que presentamos a continuación, mostramos una síntesis de la comparación entre la investigación cualitativa y experimental.

COMPARACIÓN ENTRE LA INVESTIGACIÓN CUALITATIVA Y LA EXPERIMENTAL		
Criterio	Elementos comunes	Elementos diferenciales
Teoría primero o teoría después	Parten de un problema a resolver.	- El I.C. prefiere ir libre de teorías preconcebidas. Se acerca a la realidad para analizarla, y después generará teorías que puede tratar de confirmar. - El I.E. se acerca a las teorías previas, para formular hipótesis que trata de confirmar empíricamente.
Análisis inductivo vs. deductivo	Las hipótesis inductivas tratan de confirmarse en investigaciones sucesivas.	- La I.C. intenta comprender las interrelaciones múltiples que emergen. Luego describe los fenómenos que estudia generando categorías conceptuales. - La I.E. se orienta a la verificación o comprobación deductiva de proposiciones causales: si... entonces...
Perspectiva holística	Los problemas de estudio se insertan en un sistema como totalidad.	- El I.C. se introduce en el campo para comprender el fenómeno como una totalidad. - El I.E. estudia el problema a la luz de las teorías previas, se centra en los efectos de un tratamiento particular y comprueba la interrelación de sus hallazgos con las teorías.
Pre-concepciones o pos-concepciones	Se preocupan del efecto de sus concepciones y expectativas hacia los resultados.	- La I.C. asume que se consigue una descripción más exacta si no se da por supuesto ningún aspecto de la escena social, centrándose en las construcciones mentales de los participantes. - El I.E. se preocupa del efecto de sus preconcepciones cuidando la estructura del diseño.
Selección y muestreo	La selección de sujetos y fuentes de información, es imprescindible	- En la I.C. no hay normas sobre el tamaño de las muestras. Selecciona casos ricos en información. - La I.E. suele depender de muestras que garanticen la representatividad estadística, que permitirá generalizar los hallazgos a la población.
Contacto personal	El investigador debe conocer el contexto y la realidad que estudia.	- La I.C. asume que es mediante la permanencia en el campo se conoce mejor la realidad sin modificarla. Se interpreta desde el propio contexto con la intención de hacer visible lo cotidiano. - El I.E. permanece en la escena de investigación sólo lo imprescindible, para evitar su influencia.
Triangulación y convergencia	Utiliza diferentes fuentes y técnicas para la recogida de información.	- La I.C. utiliza diversidad de fuentes y técnicas de recogida de datos para evitar sesgos y asegurar la exactitud. Interpreta la realidad desde el propio contexto pero con los ojos de un extraño. - El I.E. permanece en la escena sólo lo imprescindible, y trata de evitar interactuar con lo que estudia.
Comparabilidad y generalización	Buscan la traducibilidad o generalización	- En I.C. se busca la comparabilidad y traducibilidad de los descubrimientos a otros problemas y contextos similares. - En I.E. se pretende generalizar los hallazgos, por eso trabaja con muestras representativas de la población y controla la validez interna.

4.5. El diseño etnográfico

El diseño de investigación, que en la concepción de la investigación positivista o experimental supone una estructura de acción, un esquema de trabajo fijo y determinado, en la investigación cualitativa es cíclico y flexible, adaptable a las situaciones de trabajo y sin una estructura lineal fija. Emerge y se modifica durante el estudio de los problemas en su propio contexto.

En el diseño de la investigación experimental, las hipótesis iniciales marcan su desarrollo secuencial, pero en el diseño cualitativo trabajar con unas hipótesis prefijadas no haría sino constreñir el análisis de lo que acontece en la situación de estudio. En la investigación cualitativa son los objetivos los que marcan el proceso de investigación. El mundo simbólico capturado mediante el discurso no se circunscribe, en modo alguno, a premisas previamente formalizadas para su ulterior verificación.

Frente al carácter cerrado del diseño experimental, en la medida en que seguir un procedimiento determinado es parte de su estructura formal con la finalidad de hacer repetible todo el proceso de acuerdo con un modelo algorítmico, en el sentido de que admite la existencia de una serie de instrucciones no equívocas, susceptibles de ser formuladas, transmitidas y seguidas correctamente, con el control de todo el algoritmo, permitiendo la replicación casi exacta de la experiencia, el diseño cualitativo se caracteriza por la flexibilidad e incluso la posibilidad de invención, en el sentido de que da cabida a lo inesperado, propio de una realidad en continuo cambio.

El diseño en investigación cualitativa, en su etapa previa a la realización de la investigación, tiene que ver fundamentalmente con la composición de los grupos que intervendrán en la misma y su tamaño. Los criterios de la selección de los grupos son de pertinencia y no de representatividad estadística, tratando de incluir todos los componentes que aporten información relevante. La selección de los grupos en la investigación cualitativa es una cuestión de foco: cuanto más enfocada esté la selección en función de los objetivos de la investigación, más ajustada será la información que obtengamos, tratando de localizar y saturar el espacio simbólico, discursivo, sobre el tema a investigar. Se trata por tanto de un muestreo estructural no estadístico. La muestra no tiene tanto que ver con la cantidad sino con la composición adecuada de los grupos y su tamaño.

De modo sintético, podemos decir que las cuestiones que preocupan al diseño etnográfico son las siguientes:

1. Determinar el propósito del estudio y las cuestiones que se pretenden estudiar.
2. Elegir el modo de acercarse al estudio de los problemas y justificarlo.
3. Determinar las funciones que van a asumir el investigador y los participantes.
4. Acceder al campo.
5. Seleccionar los informantes claves.
6. Fijar los procedimientos para la recogida y análisis de la información.
7. Extraer explicaciones y contrastarlas con los participantes.
8. Escribir el informe etnográfico.

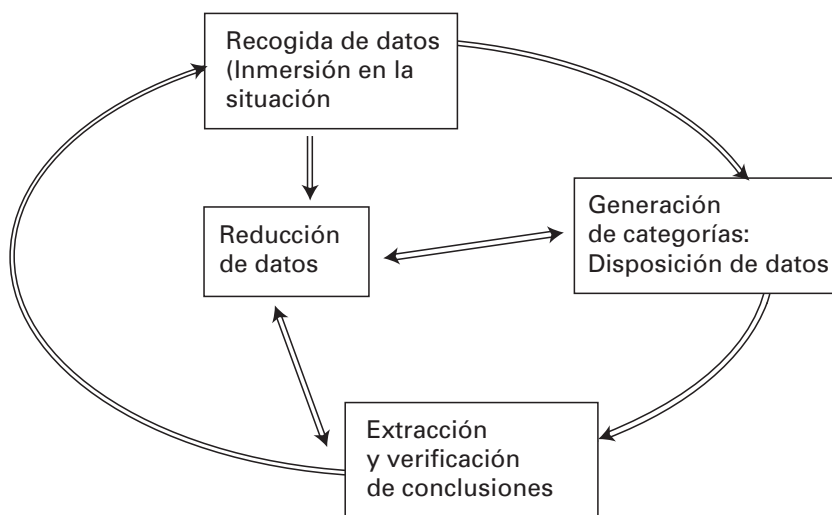
La investigación cualitativa, interpretativa, se interesa por conocer la relación entre las perspectivas de significado de los actores y las acciones en las circunstancias situacionales en donde se producen. Se interesa por conocer la ecología social, sus procesos y sus estructuras. El objeto principal de la investigación interpretativa radica en descubrir los modos específicos en los que las formas locales, o particulares, y extralocales de organización social y cultural se relacionan con las actividades de las personas concretas que efectúan opciones y se relacionan en la acción social. Esto no quiere decir que se busquen factores universales abstractos, a los que se pretende llegar en la investigación positivista a través de una generalización estadística de una muestra a una población entera determinada, sino que pretende encontrar factores universales concretos, a los cuales se llega estudiando un caso específico en detalle y luego comparándolo con otros casos estudiados en forma igualmente detallada. La paradoja es que, para lograr el descubrimiento válido de las propiedades universales es necesario mantenerse muy cerca de los casos concretos. (Erickson, 1989:221-224).

Si bien la postura del investigador de campo en los contextos educativos no es directamente evaluativa y las preguntas de su investigación no presentan la forma de ¿Qué actividades de enseñanza son las más eficaces?, sin embargo, los aspectos relativos a la eficacia son fundamentales en la investigación interpretativa en Educación. El análisis objetivo, sistemático, del significado subjetivo es esencial en la investigación social, incluida la investigación sobre la enseñanza, desde la perspectiva de la investigación interpretativa.

4.6. El análisis de los datos en la investigación cualitativa

El enfoque cualitativo se orienta hacia el desarrollo de una comprensión profunda de los escenarios o personas que se estudian. El dato cualitativo nos ofrece una información básica, primaria, acerca de la realidad, con un bajo nivel de estructuración teórica. Estos datos no permiten, de una manera directa, establecer al investigador significados para resolver el problema de estudio. Sin embargo, junto con las notas de campo aparecen juicios, opiniones, sospechas, reflexiones e interpretaciones del investigador haciendo que el análisis y la recogida de datos se alternen o superpongan en el proceso de investigación. La consideración de los procesos de recogida de datos y análisis de los mismos como fases separadas es poco productiva, defendiendo la concomitancia de dichas tareas en un proceso interactivo de recogida y análisis de datos que prosigue una vez finalizada la recogida de datos.

El análisis de datos cualitativos es una actividad que implica reflexiones, transformaciones, comprobaciones realizadas a partir de la información obtenida con el fin de obtener un significado relevante para el problema de investigación que se estudia. Miles y Huberman (1984: 23) representan del siguiente modo los componentes del análisis de datos:



A las cuatro fases propiamente dichas, se le añade una previa, de delimitación del área de estudio, y otra al final, de presentación de las conclusiones y teorías de apoyo.

FASES DEL PROCESO ANALÍTICO DE DATOS

- 1.º Inmersión en la situación y descubrimiento en progreso de la información y datos. Generación inicial de categorías. Identificación de temas claves y desarrollo de conceptos.
- 2.º Con la información recogida, codificación de los datos y el refinamiento de la comprensión del tema de estudio. Selección y definición los conceptos; comparación de los incidentes aplicables a cada categoría.
- 3.º Validación de las categorías: Frecuencia de aparición y distribución de los conceptos, se integran las categorías y sus fenómenos.
- 4.º Organización, interpretación y explicación de las categorías. Relativización de los descubrimientos, de acuerdo con el contexto en que fueron tomados.

Veamos el análisis de los datos, siguiendo a Taylor y Bogdan (1986):

4.6.1. Descubrimiento en progreso

Los investigadores cualitativos van dando sentido a lo que estudian combinando intuición y perspicacia y familiaridad con los datos. La clave está en buscar temas examinando los datos de todas las maneras posibles, de acuerdo con las siguientes pautas:

1. Lea varias veces su información.
2. Siga la pista de temas, intuiciones, interpretaciones e ideas.
3. Busque los temas claves. Elabore listas posibles de temas claves.
4. Elabore tipologías o esquemas de clasificación útiles para identificar temas y desarrollar conceptos y teorías.
5. Desarrolle conceptos y proposiciones teóricas. Avance de la descripción a la interpretación y a la elaboración de teorías a través de los conceptos y proposiciones claves. En investigación cualitativa los conceptos claves son instrumentos que proporcionan una referencia fundamental, y sugieren direcciones para la observación.
6. Lea bibliografía sobre el tema de estudio. Ayudará a contrastar conceptos, proposiciones y marcos teóricos relevantes a la investigación.

7. Desarrolle una guía del proceso de la investigación. Describa el trabajo en términos generales de modo que oriente el análisis de los lectores.

4.6.2. Codificación de los datos

En la investigación cualitativa, la codificación es un modo sistemático de desarrollar y refinar las interpretaciones de los datos. El proceso de codificación incluye la reunión y análisis de todos los datos que se refieren a temas, ideas, conceptos, interpretaciones y proposiciones. Taylor y Bogdan sugieren los siguientes pasos para la codificación de los datos cualitativos:

1. Desarrollo de categorías de codificación. Elabore una lista con todos los temas, conceptos, interpretaciones, proposiciones identificados durante el análisis inicial. Asigne un número o letra a cada categoría de codificación.
2. Codifique todos los datos. En las notas de campo indique en el margen el número asignado a cada categoría cuando aparezca en el texto. A medida que se codifican los datos hay que refinar el esquema de la codificación. En algunos casos habrá datos que pertenezcan a más de una categoría, asignándolos a cada una de ellas.
3. Separe los datos pertenecientes a las diversas categorías de codificación conservando intacto el texto original.
4. Vea los datos que han sobrado, que no entran en ninguna de las categorías existentes. Trate de crear nuevas categorías o de situar en las ya existentes. Si no se ajustan no lo fuerce.
5. Refine su análisis volviendo a leer el texto y las categorías. Hay que estar preparado para desarrollar nuevas categorías y prescindir de otras si no las consideramos adecuadas.

4.6.3. Relativización de los datos y conclusiones

Es la fase final del análisis cualitativo. Se trata de interpretar los datos en el contexto en el que fueron recogidos. Hay que tener en cuenta las siguientes consideraciones:

1. Datos solicitados o no solicitados. Tratar de ver la información que ha emergido de un modo natural y la que ha forzado el investigador.
2. Influencia del observador sobre el escenario.
3. Influencia de otras personas en el escenario.

4. Datos directos e indirectos. Literales o interpretados.
5. Fuentes. Grupo, informantes claves.
6. Nuestros propios supuestos.

En el trabajo cualitativo el investigador comienza con un mínimo de supuestos aunque nuestros preconceptos son difíciles de evitar. Los datos nunca se explican a sí mismos, sino que son interpretados en el marco de unos supuestos teóricos.

PROCESO GENERAL DE ANÁLISIS DE DATOS CUALITATIVOS		
Tareas	Actividades	Operaciones
REDUCCIÓN DE DATOS	Separación de unidades	Criterios de separación físicos, temáticos, gramaticales, sociales.
	Identificación y clasificación	Categorización y codificación.
	Síntesis y agrupamiento	Agrupamiento físico, creación de metacategorías, obtención de estadísticos, de agrupamiento y síntesis.
DISPOSICIÓN Y TRANSFORMACIÓN DE DATOS	Disposición	Elaboración de tablas, gráficos, modelos, matrices, sistemas de redes
	Transformación	Expresión de los datos en otros lenguajes (numérico, gráfico).
OBTENCIÓN DE RESULTADOS Y VERIFICACIÓN DE CONCLUSIONES	Proceso para obtener resultados	Datos textuales: descripción e interpretación; recuento, concurrencia de resultados códigos, comparación y contextualización.
	Proceso para alcanzar conclusiones	Datos textuales: consolidación teórica aplicación de otras teorías, uso de metáforas y analogías, síntesis con resultados de otros investigadores. Datos numéricos: uso de reglas de decisión (comparación de los resultados con modelos teóricos y experiencia del analista).
	Verificación de conclusiones	Comprobación o incremento de mediante la presencia prolongada en la validez el campo, intercambio de opiniones con otros investigadores, triangulación comprobación con los participantes, establecimiento de adecuación referencial, ponderación de la evidencia, Comprobación de la coherencia estructural.

(García Jimenez y otros, 1994).

4.7. Criterios para valorar las investigaciones cualitativas

Los criterios de bondad tienen sus raíces en la concepción del paradigma para el cual se designan, por lo tanto, no se puede esperar aplicar los criterios positivistas a los estudios naturalistas. ¿Qué criterios podrían servir para el paradigma naturalista?

Guba y Lincoln (1989:236), proponen tres grupos de criterios para valorar las investigaciones naturalistas o constructivistas. El primero de ellos lo consideran paralelo a los criterios clásicos del paradigma positivista de *validez interna*, *validez externa*, *fiabilidad* y *objetividad*. El segundo lo relacionan con los procesos hermenéuticos-dialécticos y el tercero con criterios de autenticidad.

4.7.1. Criterios de credibilidad, seguridad, confirmabilidad y transferibilidad

Así explican Guba y Lincoln los criterios del paradigma positivista de *validez interna*, *validez externa*, *fiabilidad* y *objetividad* y los paralelos en el paradigma naturalista, interpretativo:

- **La validez interna** es la valoración del grado de isomorfismo entre los hallazgos del estudio y el mundo real, y ello no puede estar presente en un paradigma que rechaza el realismo ontológico. Si la realidad sólo existe en las mentes de sus constructores –subjetividad-, ¿Qué sentido tiene ese isomorfismo?
- **La validez externa**, concepto que contiene la esencia de la generalización de los hallazgos de la muestra a la población, igualmente tiene poco significado si las realidades que podríamos dese- ar generalizar existen en forma diferente en las mentes de cada una de las personas, condicionadas por sus historias y circunstancias diferentes, basadas en sus experiencias particulares e interpretadas desde sus propios sistemas de valores.
- **La fiabilidad** es esencialmente una valoración de la estabilidad del fenómeno que ha sido valorado y del instrumento utilizado para registrarlo, lo que supone que el fenómeno es invariable, al menos en un corto tiempo. Pero, si el fenómeno cambia, y esto es central en el desarrollo y mejora de las construcciones mentales de las personas, entonces la fiabilidad es inútil como criterio de bondad en este paradigma.
- Finalmente **la objetividad** refleja la posición positivista de la dualidad objeto-sujeto, pero si el paradigma rival, el interpretativo,

afirma que la interacción (monismo) es inevitable, ¿Qué puede significar la objetividad?

Después de reflexionar sobre los posibles criterios que podrían servir para el paradigma naturalista, proponen un grupo de criterios paralelos a los del paradigma positivista utilizando los términos alternativos de *credibilidad*, *seguridad*, o *formalidad*, *confirmabilidad* y *transferibilidad*. Así los explican:

- **Credibilidad.** Es un criterio paralelo al de *validez interna*. La atención se centra en que las construcciones de las personas implicadas y las representaciones del propio investigador sean equivalentes. Se utilizan varias técnicas para incrementar la probabilidad de que tal isomorfismo se produzca; son las siguientes:
 - Compromiso o encuentro prolongado para evitar efectos de distorsión por la falta de información y falta de confianza para descubrir las construcciones y comprender la cultura del contexto.
 - Observación continua para identificar los elementos y las características más relevantes de la situación.
 - Informes de colegas a través del compromiso con compañeros desinteresados para la discusión y análisis de las conclusiones de los hallazgos.
 - Análisis de los casos negativos y opuestos.
 - Seguimiento de la subjetividad con el fin de controlar las construcciones del investigador.
 - Comprobaciones de los miembros y participantes de la investigación, que pueden ser formales e informales, cumpliendo un importante número de funciones:
 - Valoración de la intención de las acciones concretas del investigador.
 - Corrección de errores de hechos o de interpretación.
 - Posibilidad de información adicional, estimulando la reflexión de los implicados con el fin de confirmar evidencias y acuerdos.
 - Síntesis como un primer paso hacia el análisis.
 - Valoración de la adecuación de la información posibilitando la confirmación de los datos individuales por parte de los implicados.
- **Seguridad o formalidad.** Se corresponde con el concepto de *fiabilidad* y está relacionado con la estabilidad de los datos en el tiempo.

Frecuentemente, la inestabilidad se produce como consecuencia del cansancio de los investigadores. En el paradigma naturalista los cambios producidos por modificaciones metodológicas pueden contribuir a mejorar la seguridad de la información en un proceso de triangulación.

- **Confirmabilidad.** Es un criterio paralelo al de *objetividad*, interesado en asegurar que los datos, interpretaciones y logros de la investigación tengan sus raíces en los contextos y en las personas, además de en el investigador. Esto significa que los datos, y la lógica usada en su interpretación, deben estar unidos a sus fuentes corroborando la totalidad explícita e implícita en un informe narrativo de estudio de caso. La técnica utilizada frecuentemente para confirmar los datos y su interpretación es la auditoria de confirmabilidad, cuyas raíces se encuentran en la auditoria fiscal.
- **Transferibilidad.** Corresponde al criterio de *validez externa o generalización*. Consiste en un proceso empírico para evaluar el grado de similitud entre los contextos. A diferencia de la generalización en el paradigma positivista, aquí se trata de una cuestión relativa y la técnica adecuada consiste en hacer una descripción detallada del tiempo, lugar, contexto y cultura en los que se realiza la investigación.

4.7.2. Criterios hermenéuticos

Los criterios hermenéuticos son utilizados en el paradigma naturalista para enjuiciar sus investigaciones. El proceso hermenéutico-dialéctico se utiliza a modo de un control de calidad. Los datos son analizados inmediatamente a su producción con la finalidad de realizar una comprobación mediante comentarios, revisiones, ampliaciones, o cualquier otro modo que los participantes proporcionen durante las reuniones. Se buscan los diferentes niveles de significado en el significado aparente, los diferentes significados subyacentes al significado literal y son incorporados a medida que emergen, en un proceso de reconstrucción colaborativa. Cómo método de análisis la hermenéutica sugiere una manera de entender los datos textuales, dado que se centra en el significado del texto, ya sea oral o escrito.

4.7.3. Criterios de autenticidad

El tercer grupo de criterios que utiliza la investigación constructivista para enjuiciar su calidad es el de la *autenticidad*. Los dos criterios ante-

rios no son de entera satisfacción para los constructivistas; los primeros son, principalmente, criterios metodológicos; las metas, los resultados y la negociación son criterios igualmente importantes para juzgar una investigación. De modo que, basarse sólo en criterios metodológicos sería insuficiente para garantizar el esfuerzo investigador para lograr algo, —resolver un problema.

El segundo grupo de criterios, el proceso hermenéutico, tiene más que ver con los procesos constructivistas pero es débil por estar implícito en el proceso mismo. Por ello, han ideado los criterios de autenticidad que emergen de la concepción básica del constructivismo. Los criterios de autenticidad, representan el tercer grupo de criterios para valorar las investigaciones, e incluyen los siguientes elementos (Lincoln y Guba 1986:81):

- **Justicia.** Se refiere a la estructura de valores subyacentes en el proceso de investigación. El papel del evaluador es el de indagar para comunicar dichos valores y ver los que están en conflicto a través de la diferencia entre las creencias y el sistema de valores, mediante la realización de una negociación de recomendaciones abierta y una agenda de las acciones subsiguientes.
- **Autenticidad ontológica.** Tiene que ver con como maduran, mejoran, se expanden y se elaboran las construcciones de los participantes, dado que han alcanzado mayor información y han logrado un uso más sofisticado de ella. Es, literalmente, «una mejora en el individuo —o grupo— de la conciencia experiencial del mundo». La autenticidad ontológica se refuerza mediante la experiencia vicaria y se comprueba por el testimonio de informantes seleccionados, además de por una auditoria judicial para el estudio del caso determinado.
- **Autenticidad educativa.** Representa el grado en el que los participantes individuales comprenden y aprecian las construcciones de los otros fuera de su grupo particular.
- **Autenticidad catalítica.** Consiste en el grado en el cual la acción es estimulada y facilitada mediante el proceso de la investigación. El propósito de la investigación es la acción y la toma de decisiones, y no sería completa sin una implicación en la acción por parte de los participantes.
- **Autenticidad táctica.** Se refiere al poder de los participantes para actuar.

La autenticidad debería ser el criterio más importante para valorar la investigación naturalista. Y es que, como ya hemos visto, la autenticidad integra conceptos como el de justicia, conciencia, enriquecimiento, apreciación de los puntos de vista y necesidades de los otros, estimulación hacia la acción, y generación de cambios deseables.

4.8. Elaboración del informe de investigación

Todo informe de investigación, ya sea breve o extensa, compleja o sencilla, debe reflejar el trabajo realizado por el investigador, sus procedimientos, sus reflexiones y conclusiones. Goetz y Lecompte (1984) mencionan los siguientes apartados que deben quedar reflejados en el informe etnográfico:

1. El propósito del estudio
2. El enfoque de investigación
3. Los sujetos, marcos y contextos investigados
4. La experiencia y roles del investigador
5. Las estrategias de recopilación de datos
6. Las técnicas de análisis de datos
7. Los hallazgos, interpretaciones y aplicaciones

Una idea fundamental para la elaboración de informes consiste en tener claro en la mente lo que se quiere presentar: una lógica analítica del estudio que sigue un hilo conductor lógico desde el principio hasta el fin. La elaboración del informe debe iniciarse cuando comienza la investigación siendo una de las tareas más difíciles que exige una continua reelaboración hasta que se considera que es claramente comprensible, y su estructura y expresión son correctas. Cuando el investigador no está del todo convencido de que esto sea así, debe pedir a sus compañeros colaboración para que se lo lean, le proporcionen consejos y discutirlos.

El informe de una investigación cualitativa debe recoger todo el proceso seguido para resolver el problema de estudio, aunque en la investigación cualitativa el problema se redefine continuamente. Debe reflejar las acciones para acceder al campo, y una vez en el contexto, el proceso seguido para llegar a los acuerdos sobre el problema. Así mismo deben reflejarse el proceso seguido para concretar el diseño y las estrategias de acción, condicionadas con este descubrimiento en proceso, que hace que emerjan nuevos aspectos del problema que llevan a idear nuevas estrate-

gias para intentar solucionar el problema, por lo que a este tipo de diseño se le denomina «diseño emergente».

En la investigación cualitativa lo importante, una vez demarcado el problema, es adentrarse en el campo, conocer el problema en su propio contexto, con sus propios actores y sus condicionantes. En la investigación cualitativa el diseño, la metodología, los instrumentos -guiones de entrevistas, pautas que guíen la observación, los guiones que guíen la discusión grupal, si es que fuera necesario utilizarlos, se deciden durante el desarrollo de la investigación, si bien el investigador experimentado tiene y conoce las estrategias que previsiblemente serán necesarias utilizar. Esto significa que a medida que se va desarrollando la investigación, el investigador, en plena interacción con los informantes, el grupo, decide lo que se necesita para seguir realizando la investigación, para seguir profundizando en la búsqueda de soluciones al problema, tras comprenderlo.

Todas las investigaciones, sea cual sea el método que se utilice, requieren recoger datos. Los datos son recogidos unas veces de fuentes primarias, otras a través de la lectura, la observación, la entrevista o la administración de cuestionarios o tests. El informe debe describir las estrategias seguidas durante el proceso de recogida y análisis de los datos, la elaboración y/o la selección de los instrumentos a utilizar, la selección de los informantes claves, y sus razones, los tiempos, la recogida de los datos, su análisis, la formulación de hipótesis, la validación de los datos y la interpretación de datos, de los resultados y la elaboración de conclusiones y su confirmación contrastadas con los informantes.

Los procedimientos seguidos para su tratamiento y análisis de datos y la presentación de los mismos comentados deben justificarse y quedar reflejados en los informes. El análisis de los datos varía desde las tabulaciones y clasificaciones o categorizaciones manuales hasta el realizado por ordenador con ayuda de los programas informáticos para el análisis de datos cualitativos. El análisis de datos es uno de los aspectos más interesantes del proceso de investigación dado que ofrece posibilidades de descubrimiento y reconceptualización de las cuestiones de investigación. Los análisis de los datos frecuentemente originan la necesidad de utilizar diferentes tratamientos para capitalizar, aprovechar, la riqueza particularizada de los datos que han sido recogidos.

Una vez analizados los datos se deben mostrar de modo que comunique a los lectores lo que se ha descubierto. Los datos deben agruparse y representarse mediante tablas y figuras coherentemente con las cuestiones de estudio (preguntas) y sus componentes (variables). Las conclusio-

nes y recomendaciones deben basarse, fidedignamente en los resultados, realizando las interpretaciones pertinentes, coherentes con los datos reales y apoyados por las teorías relevantes al problema de estudio.

Las características que adopte el informe final dependerán de las características del propio trabajo y de las audiencias a las que se dirige. Se recomienda el uso de esquemas o diagramas que faciliten la integración de toda la información que debe aparecer en el informe. También deben recogerse las referencias bibliográficas y los anexos o fuentes originales de datos. Los apéndices deben ir en formato separado del resto del informe.

EL INFORME DE INVESTIGACIÓN ETNOGRÁFICA

En el informe debe explicarse todo lo que los lectores necesitan saber sobre:

1. El problema de investigación.
2. El marco teórico. Experiencia, enfoque y roles del investigador.
3. El propósito de la investigación. Su diseño. Las modificaciones.
4. La naturaleza y el número de los escenarios e informantes.
5. El modo de acceso al campo.
6. Las diversas fuentes de información: personas, bibliografía, documentos.
7. Las relaciones con los informantes.
8. La metodología: observación, entrevistas, lectura de informes.
9. Las estrategias de recogida de información y datos y su control.
10. El tiempo y extensión del estudio y su distribución.
11. Las técnicas de análisis de la información y datos.
12. Los hallazgos, las interpretaciones, las conclusiones.
13. Las aplicaciones.
14. Las recomendaciones.

La extensión del informe también dependerá de las características del trabajo, de la personalidad de quien lo escribe y de la audiencia a quien se

dirige. La longitud, debe ser la suficiente para recoger lo que se ha realizado desde el inicio del trabajo hasta el final, de modo que se comprenda bien, con las frases y párrafos justos y adecuados. La claridad del lenguaje debe ser un objetivo a conseguir.

Escribir un informe de un trabajo para colegas del mismo ámbito puede no crear tanta ansiedad que escribir el informe para una audiencia que no se conoce. Tampoco es lo mismo escribir un informe de un trabajo de tesis doctoral para ser defendido ante el tribunal que lo va a juzgar que para presentarlo en un congreso. La elaboración de informes breves para su publicación en revistas científicas también facilita la práctica y refuerza la confianza en la elaboración de informes. La ansiedad de los primeros informes se mitiga cuando se escribe el tercero o el cuarto. La calidad en el estilo expositivo mejora con la práctica.

4.9. Actividades de autoevaluación

4.9.1. *Conceptos fundamentales del tema*

1. Notas definitorias de la investigación cualitativa.
2. Defina los siguientes conceptos: Triangulación. Muestreo intencional. Credibilidad. Transferibilidad. Autenticidad.

4.9.2. *Cuestiones de síntesis*

1. Elabore una síntesis sobre los principales contenidos desarrollados en el tema.
2. Haga una síntesis sobre las características definitorias del enfoque de investigación cualitativa y sus diferencias con el enfoque de investigación experimental.

4.9.3. *Cuestiones sobre los contenidos del tema*

1. Describa el proceso de la investigación etnográfica.
2. Compare los criterios de valoración de la investigación cualitativa y la experimental.
3. ¿Es aplicable la metodología de investigación cualitativa al estudio de los procesos de enseñanza y aprendizaje desarrollados en un aula? Justifique su respuesta y argumente cómo aplicaría esta

metodología concretándola en un problema específico. Por ejemplo: El contenido desarrollado en una clase de Lengua. O de Matemáticas. Otro caso: El contenido desarrollado en una sesión grupal de tutoría.

4. ¿Cómo aplicaría la metodología de investigación cualitativa para organizar los temas prioritarios en una reunión con padres de los alumnos al comienzo del curso? Haga un esquema indicando las fases y contenidos que desarrollaría, desde el principio hasta la síntesis y conclusiones y recomendaciones para posteriores reuniones.
5. Mencione las aportaciones de la investigación cualitativa aplicada a los centros educativos. Justifique sus comentarios.

4.10. Soluciones a las actividades de autoevaluación

Consulte los diferentes epígrafes del tema.

4.11. Referencias bibliográficas

- ERICKSON, F. (1989). «Métodos cualitativos de investigación sobre la enseñanza». En M. C. Wittrock (Ed.): *La investigación de la enseñanza, II*. Madrid: Paidós, MEC.
- GARCÍA JIMÉNEZ, E., GIL, J. y RODRÍGUEZ, G. (1994). «Análisis de datos cualitativos en la investigación sobre la diferenciación educativa». *Revista de Investigación Educativa*, 23 (179-213).
- GOETZ, J. P. y LECOMPTE, M. D. (1988). *Etnografía y diseño cualitativo en investigación educativa*. Madrid: Morata.
- GUBA, E. G. (1978). *Toward a methodology of naturalistic inquiry in educational evaluation*. Los Angeles, CA: Center for the Study of Evaluation.
- GUBA, E. G. y LINCOLN, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- LINCOLN, Y. S. y GUBA, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- LINCOLN, Y. S. y GUBA, E. G. (1986). «But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation», (73-78), En D. D. Williams (Ed.): «Naturalistic evaluation». *New directions for program evaluation*, n.º 30, San Francisco: Jossey-Bass.
- MILES, M. B. y HUBERMAN, A. M. (1994). *Qualitative data analysis: an expanded sourcebook*. Newbury Park, CA: Sage.
- PATTON, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park, CA: Sage.

TEMA 5

La investigación evaluativa. Evaluación de programas e instituciones

ESQUEMA

- 5.1.** La evaluación de programas: Concepto, finalidad y funciones
- 5.2.** Cuestiones a tener en cuenta en el proceso de realización de la evaluación
 - 5.2.1. Las características de los programas y de las instituciones
 - 5.2.2. Para quién se hace la evaluación
 - 5.2.3. Quién debe hacer la evaluación
 - 5.2.4. Sensibilización hacia la evaluación
 - 5.2.5. Planificación de la evaluación
 - 5.2.6. Qué información debe ser recogida
 - 5.2.7. Los métodos en la evaluación
- 5.3.** Los informes de la evaluación y la utilización de los resultados
- 5.4.** La realización de proyectos innovativos de mejora
- 5.5.** Los criterios y normas para juzgar la evaluación
- 5.6.** El enfoque de calidad total. El modelo de excelencia europeo
- 5.7.** Realización de una práctica evaluativa
- 5.8.** Referencias bibliográficas

OBJETIVOS

- ✓ Comprender y valorar el concepto de la evaluación de programas e instituciones.
- ✓ Conocer, comprender y valorar las diferentes funciones de la evaluación.
- ✓ Valorar los posibles objetos de evaluación.
- ✓ Valorar los diferentes tipos de información requeridos para realizar la evaluación de un programa.
- ✓ Valorar los criterios que deben seguirse para realizar una evaluación.
- ✓ Conocer y comprender el proceso a seguir para realizar una evaluación.
- ✓ Valorar los métodos de investigación aplicables a la evaluación.
- ✓ Conocer, comprender y valorar diversos modelos de evaluación de programas e instituciones.
- ✓ Conocer y valorar las normas por las que se deben regir las evaluaciones.

5.1. La evaluación de programas. Concepto, finalidad y funciones

La evaluación es un proceso y un resultado. «La evaluación es el proceso de determinar el valor y calidad de algo y las evaluaciones son los resultados de este proceso» (Scriven, 1980:18). La evaluación consiste en un proceso para identificar, obtener y proporcionar información acerca de un programa valorado en sus metas, en su planificación, en su realización y en su impacto, con el propósito de guiar la toma de decisiones, proporcionar información y contribuir a su comprensión y con el criterio de su valor (como respuesta a las necesidades) y su mérito (calidad con la que lo consigue). (Stufflebeam y Shinkfield, 1987). Evaluar implica determinar los criterios del mérito de lo que se pretende evaluar, los cuales nos van a permitir comprobar la realización. Realizada dicha comparación entre los indicadores, representativos de los criterios, y la práctica, se obtienen unos resultados, y con la valoración de lo conseguido podemos concluir en un juicio evaluativo. La evaluación aporta datos objetivos y valoraciones que deben ayudar a la toma de decisiones de cambios y mejoras de lo evaluado.

La evaluación tiene como finalidad comprobar si un plan, un proyecto o un programa, elaborado para dar respuesta a unas necesidades determinadas y producir el cambio pretendido lo consigue y la calidad con que lo logra. La nota distintiva de la evaluación es la emisión de juicios valorativos fundamentados en hechos y datos que orienten la toma de decisiones para la realización de acciones de mejora de los programas, de las condiciones y de los contextos, de los grupos y de los individuos. La evaluación se ha constituido en una disciplina necesaria para conocer el funcionamiento de los programas de cualquier ámbito de la actividad humana: político, económico, social, sanitario, educativo..., de las instituciones y de las organizaciones que los desarrollan.

El potencial de la evaluación para la mejora de la calidad de los programas y de las organizaciones es de una gran trascendencia al identificar los aspectos positivos y negativos con los que opera un programa y las organizaciones y los diferentes elementos que contribuyen al logro de sus objetivos. La evaluación de programas se central en:

- El estudio de necesidades que originan programas.
- La evaluación de los programas, la calidad de sus metas, sus contenidos, su estructura, la planificación para la puesta en acción, las estrategias de aplicación, incluidos los recursos materiales necesarios y la formación del personal.

- El seguimiento de los resultados del programa y de su impacto, en términos de eficacia y de satisfacción del personal, incluidos los aplicadores y beneficiarios.

La evaluación se convierte, de este modo, en una herramienta para la mejora continua de la calidad de planes, proyectos, programas y organizaciones, de los individuos y de los grupos.

En cuanto a las funciones de la evaluación, Scriven (1967) formuló dos importantes funciones: la *formativa*, para la mejora del programa durante su desarrollo, y la *sumativa*, valorando la eficacia del programa en su totalidad, cuando ha terminado. Los términos *evaluación formativa* y *evaluación sumativa*, acuñados por Scriven, están relacionados con dos modos de sensibilidad hacia las necesidades de evaluación de los programas para su mejora.

La *evaluación formativa* es realizada desde el inicio del programa y durante su desarrollo con el fin de recoger información útil, y a tiempo, de los efectos del programa y de las necesidades de introducir cambios de mejora. Informa a la evaluación sumativa sobre el funcionamiento del programa.

La *evaluación sumativa* es realizada una vez que ha terminado el programa, o en determinados momentos claves, y su objetivo principal es recoger información sobre los resultados finales, y sobre el valor y mérito del programa. Sirve al rendimiento de cuentas.

Podemos ver la distinción entre evaluación formativa y sumativa en el siguiente cuadro:

CUADRO 1

Dimensiones de la evaluación	Evaluación formativa	Evaluación sumativa
Propósito	Mejorar el programa	Certificar la calidad del programa
Audiencia	Administradores y personal del programa	Consumidores potenciales
Quién la hace	Evaluador interno	Evaluador externo
Cuestiones que hace	¿Cómo funciona? ¿Qué necesita para mejorar?	¿Qué resultados produce? ¿A qué costo? ¿Qué objeciones debe hacer?
Exigencias del diseño	Medidas válidas y fiables Muestreo suficiente Recogida de información frecuente Resultados a tiempo	Medidas válidas y fiables Muestreo suficiente Recogida de información al final Resultados a tiempo

Los términos de evaluación interna y evaluación externa se asimilan a la evaluación formativa, la primera, al servicio de los que están aplicando el programa y para que conozcan cómo está funcionando y puedan introducir cambios de mejora, y a la evaluación sumativa, la segunda, para conocer los resultados del programa y rinda cuentas ante la sociedad. La evaluación interna también se encuentra asimilada al término de autoevaluación, para significar que es realizada por los agentes internos de la institución.

Otras dos funciones de la evaluación son la *psicológica o sociopolítica*, para motivar e incentivar una conducta deseada, promover apoyo público e incrementar el conocimiento de unas actividades especiales, y la *administrativa*, para el ejercicio de la autoridad.

Cuando se inicia un proyecto de evaluación de una institución se habla de *evaluación institucional*, y presupone las dos funciones de la evaluación, *la evaluación interna, o autoevaluación, y la evaluación externa* como rendimiento de cuentas. Desde el enfoque de evaluación institucional, se cubre el doble objetivo de conocer cómo funciona la institución, con un primer objetivo de *evaluación diagnóstica*, para introducir los ajustes necesarios en la conceptualización del programa, en los recursos materiales, la formación del personal y los medios económicos necesarios para realizar las mejoras. El segundo objetivo sería el de informar a la sociedad sobre cómo trabaja la institución para conseguir los objetivos que se ha marcado, y los hace explícitos, y qué logros consigue, fundamentada esta información en la evaluación interna, o autoevaluación, y en evaluación externa, o certificación por evaluadores externos, lo que garantiza la credibilidad de los informes de la evaluación.

5.2. Cuestiones a tener en cuenta en el proceso de realización de la evaluación

5.2.1. Las características de los programas y de las instituciones

De acuerdo con Pérez Juste (1995), todo programa es un plan elaborado y organizado, al servicio de metas consideradas valiosas, en respuesta a unas necesidades detectadas. Los programas se suelen caracterizar por las siguientes notas:

- **Complejidad:** de elementos organizativos, orientados a dar una respuesta adecuada a las necesidades a las que sirve, normalmente integrado por diferentes saberes procedentes de ámbitos multi-

disciplinarios, y por múltiples personas responsables de realizarlos y ponerlos en marcha, y diversidad de medios y recursos para asegurar su éxito.

- **Duración:** para el logro de las metas de los programas, requiere tiempo para desarrollar y afianzar los cambios de mejora perseguidos, ya sea de habilidades, actitudes o comportamientos, contando, siempre, con la colaboración de las personas implicadas en el mismo.
- **Procesualidad:** traducibilidad de las metas de los programas para su aplicación, en coherencia con los fines pretendidos.
- **Mantenimiento y transferencia:** de los logros y su transferencia a otros ámbitos y actividades del sujeto.

La estructura organizativa de una institución es el reflejo de los valores y actitudes de sus miembros. La mejora de los centros educativos es más factible cuando los proyectos a emprender son factibles de realizar, con una buena concepción y elaborado diseño, planificación, concreción de las estrategias y recursos y formación de las personas que lo van a llevar a cabo. Una cultura de consenso, colaborativa, en la que los propósitos de la institución, su misión y objetivos, concretados en la formación de alta calidad de sus alumnos y profesores, favorece el desarrollo de un clima favorable a la consecución de la misión de la institución.

5.2.2. Para quién se hace la evaluación

La evaluación está al servicio de los diferentes grupos implicados en el programa a evaluar:

- Los responsables del programa.
- Los que lo llevan a la práctica.
- Los que se benefician de los servicios del programa.
- Las empresas que emplean a los formados en el programa.
- La sociedad.

Cuando se decide realizar la evaluación de un programa, o de una institución, hay que identificar los diferentes grupos implicados y sus necesidades de información.

El evaluador, o evaluadores, debe atender las necesidades de información de quien debe tomar las decisiones, pero también a las necesidades de información de las personas que están implicadas en el funciona-

miento de los programas, informándoles de los resultados de la evaluación. Los evaluadores, tras analizar la finalidad del programa, su funcionamiento, y contrastar las diferencias entre el programa en su concepción y el programa en uso, y las causas de las diferencias, deben informar en qué debe cambiar y mejorar y cómo mejorarlo en coherencia con los cometidos iniciales del programa, siendo recomendable que incorpore las nuevas necesidades actuales y las previstas en el futuro.

5.2.3. Quién debe hacer la evaluación

La evaluación debe ser realizada por un equipo de expertos, pudiendo ser externo o interno a la institución, comprometido con la idea de la evaluación para la mejora, siendo conveniente, en este último caso, si fuera necesario, que recibieran asesoramiento específico sobre cómo actuar para realizar la evaluación.

El papel del evaluador es proporcionar a los responsables de la organización y aplicadores de los programas un plan preciso de evaluación, que incluya estrategias de información y formación para lograr la colaboración y la implicación de todos en la realización de la evaluación, y asimismo, en la colaboración de la puesta en marcha de las acciones que se deriven de los resultados de la evaluación.

5.2.4. Sensibilización hacia la evaluación

Antes de iniciar una evaluación se debe informar a todos los grupos implicados en la institución y en el programa de la intención de la evaluación, de sus objetivos y de los ámbitos en los que se va a centrar. Las intenciones de evaluar una institución, o un proyecto o un programa que se está aplicando, deben de ser conocidas por todos, y en ella deben estar comprometidos tanto la dirección de la institución como los que aplican el programa y los beneficiarios del mismo. De ese modo estarán dispuestos a asumir sus resultados y a emprender los cambios que se deriven de las conclusiones y recomendaciones de la evaluación. El evaluador debe explicar su concepción de la evaluación, las funciones que cumplirá, el tipo de información que debe ser recogida, los métodos e instrumentos de recogida de datos e información, y cómo se analizarán. Aclarar estas cuestiones al inicio, ayudará a evitar conflictos durante el proceso y al final de la evaluación.

Las evaluaciones deben ser sensibles a los posibles temas políticos implicados, por lo que el evaluador, o equipo, debe consultar al personal

que toma las decisiones, entrevistarse con los grupos que aplican el programa y con sus beneficiarios, y, como consecuencia de la información recogida, debe realizar una valoración de la adecuación, en ese momento y contexto, de la realización de la evaluación. El impulso que desde las administraciones educativas se da a la realización de evaluaciones en las instituciones educativas, debe ser aprovechado, recabando su ayuda en recursos económicos y formación del personal, para llevarlas a cabo en beneficio de la mejora de la calidad educativa.

5.2.5. Planificación de la evaluación

Para elaborar un plan de evaluación se deben tener en cuenta los siguientes elementos:

- Identificación de las metas de la organización, de los programas y de las estrategias para su aplicación, incluidas las funciones de los que lo aplican.
- Identificación de las condiciones reales en las que el programa se aplica y su incidencia en el logro de las metas.
- Desarrollo de un marco general de estrategias para la aplicación del programa incluidas la modificación de las condiciones si fuera necesario.
- Asignación de responsabilidades individuales para la realización de las actividades de intervención para conseguir los objetivos del programa.
- Plan, instrumentos y estrategias para valorar los resultados en términos de eficacia en la consecución de las metas y de satisfacción del personal y beneficiarios del programa.

Es imprescindible llegar a acuerdos en los enfoques seguidos para la realización de la evaluación, incluso cuando las propuestas partan desde las administraciones públicas. El proceso seguido para realizar una evaluación puede diferir dependiendo de la concepción evaluativa que la guíe, pero en líneas generales sigue siendo válida la propuesta del primer enfoque sistemático de «Evaluación orientada al logro de los objetivos» de Tyler, desarrollado en las décadas de los 30, 40 y 50 del siglo XX, que concebía la evaluación como la actividad dirigida a determinar si las metas del programa se habían cumplido, recomendando las siguientes fases:

- 1.º Identifique y enuncie las metas del programa en términos de conducta observable.

- 2.º Elabore instrumentos de recogida de datos sobre lo que se espera que realice el programa.
- 3.º Recoja los datos de la realidad.
- 4.º Interprete los hallazgos.
- 5.º Haga recomendaciones.

En la década de los 70, con el resurgir de los métodos de investigación naturalista, o etnográfica, interesados por el estudio en profundidad de los acontecimientos, y de los programas, en sus contextos, llevaron al desarrollo de enfoques de evaluación naturalistas como el de Guba, o el de «Evaluación Repondiente», de Stake, en el que especificaba doce pasos interactivos, en un modo característico del enfoque de investigación naturalista, que puede adoptar una secuencia circular. Son las siguientes:

1. Hable con las personas implicadas en el programa.
2. Identifique las metas del programa.
3. Revise las actividades del programa.
4. Descubra los intereses y preocupaciones de los grupos implicados.
5. Determine las preocupaciones claves.
6. Identifique las necesidades de profundizar en determinados temas y recoger más información sobre ellos.
7. Seleccione los observadores, los jueces o comités y los instrumentos de recogida de datos.
8. Observe y analice los antecedentes, los procesos y los resultados.
9. Recoja información sobre los temas claves, categorice y ordene por su importancia, prepare informes descriptivos.
10. Valide la información contrastándola con los distintos grupos implicados.
11. Presente los resultados a los grupos o audiencias.
12. Elabore informes para los diferentes grupos de audiencias.

En estos doce puntos pueden observarse cuatro fases:

- 1.^a Acercamiento al programa, a sus escenarios, y a las personas implicadas, para identificar las metas y actividades del programa, y los intereses y preocupaciones de los grupos implicados, en relación con el programa y su contexto.

- 2.^a Acuerdos en la que se determinan las preocupaciones y los intereses sobre los que profundizar.
- 3.^a Recogida y análisis de la información acerca de los antecedentes, los procesos y los resultados sobre los temas claves detectados en la fase anterior; en una forma de hacer característica de la investigación naturalista, en la que se toma información de numerosas fuentes e informantes, se categorizan y ordenan los temas relevantes, se contrasta la información y se amplía si fuera necesario.
- 4.^a Información a los diferentes grupos de los resultados, incluyendo los acuerdos para acciones futuras.

La colaboración entre los promotores de la evaluación, los participantes, los beneficiarios del programa y los evaluadores es defendida en la totalidad de los enfoques de evaluación. Incluso cuando el evaluador sea externo a la institución debe trabajar en estrecha colaboración con las personas implicadas en el programa, con el fin de conocer cómo funciona el programa y las dificultades con las que se encuentran para aplicarlo. Y es que el conocimiento de las diferencias entre el programa en su concepción y en su realización, y a las causas que generan estas diferencias, sólo se logrará mediante la colaboración entre los promotores del programa, los que lo aplican y los que lo evalúan.

5.2.6. La información que debe ser recogida

La evaluación alcanza a casi todo: desde los estudios de necesidades, las metas de los programas, el diseño de los mismos, los contextos en los que se desarrolla, su planificación, su realización en su proceso y en sus resultados. La evaluación es aplicada a proyectos, a programas, a materiales instruccionales, a estudiantes, a profesores y a instituciones, siendo fundamental, para que la evaluación tenga éxito, delimitar e identificar, de modo claro y preciso, en la fase de planificación de la evaluación, lo que se va a evaluar. Las cuestiones principales que deben ser atendidas en todas las evaluaciones son las siguientes:

- Análisis relacionados con los contextos y el estudio de necesidades, con el fin de planificar y elaborar programas con el objetivo de resolverlas.
- Análisis de posibles programas alternativos, de su conceptualización, estructura y diseño y planes más adecuados sobre cómo proceder a su implantación.

- Análisis de la situación de partida, de los recursos necesarios y disponibles, de las estrategias para la implementación de los programas, incluida la formación de los que deben aplicarlos.
- Análisis de los procesos para comprobar cómo se están aplicando los programas, en coherencia con los objetivos y planes previstos, mediante revisiones continuas para comprobar los resultados que va obteniendo en los diversos momentos claves.
- Análisis de los resultados obtenidos una vez aplicado el programa, para comprobar si las necesidades han sido satisfechas o reducidas.
- Valoración del impacto del programa, incluyendo sus efectos en relación con las necesidades que pretendía satisfacer, estimando su eficacia y la relación entre costos y eficacia y costos y beneficio.
- Decisiones sobre el programa: Propuestas para la elaboración de planes de mejora, y estrategias para su aplicación y seguimiento.

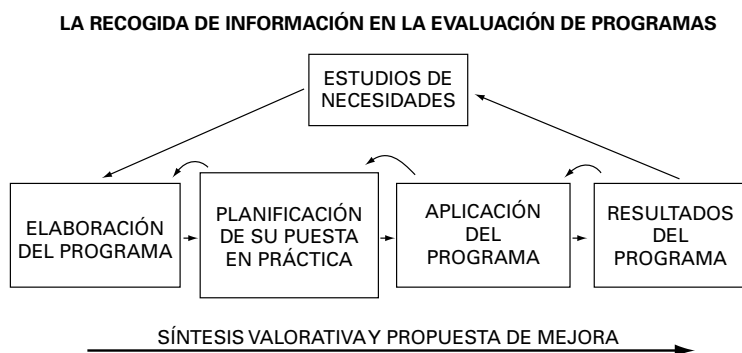


FIGURA 1.

Los prerequisites para valorar los impactos de una intervención evaluativa —sus resultados últimos—, son los siguientes:

- El programa que pretendemos evaluar debe tener sus objetivos suficientemente bien articulados para hacer posible identificarlos y elaborar a partir de ellos indicadores que permitan medir las metas logradas, o, de otro modo, el evaluador deberá ser capaz de delimitar, de un modo razonable, dichos objetivos.
- El programa ha debido ser aplicado en su mayoría, no sólo en parte, de modo que no haya dudas para poder evaluar si los elementos críticos han sido realizados adecuadamente a sus objetivos.

Por ello es importante la función de seguimiento de la aplicación del programa con el fin de resolver dos interrogantes principales: 1) si la realización de los procesos es congruente o no con las especificaciones del programa, y 2) si el programa alcanza adecuadamente, o no, a la población objeto para quien se ha elaborado.

5.2.7. Los métodos en la evaluación

La evaluación de programas es una modalidad de investigación aplicada que utiliza las metodologías de las Ciencias Sociales. Dada la complejidad que las organizaciones y los programas encierran, tanto para la evaluación de necesidades iniciales, como del diseño del programa para contribuir a la solución de las mismas, como la planificación de su puesta en funcionamiento y su aplicación, y de sus resultados, las metodologías de recogida de información y análisis de la misma que utiliza la evaluación, pueden incluir todos los enfoques de investigación, dependiendo de sus objetivos, y dimensión, y por lo tanto, diversidad métodos, técnicas e instrumentos, imponiéndose el principio de complementariedad metodológica.

Veamos el siguiente cuadro que muestra los posibles métodos y técnicas para la recogida de información en la realización de la evaluación de programas:

LOS MÉTODOS EN LA EVALUACIÓN DE PROGRAMAS			
Finalidad	Dimensiones	Objeto	Métodos y técnicas
Análisis de necesidades	• El contexto	• Las personas y sus contextos	• Resultados de evaluaciones previas • Lectura de informes, memorias • Realización de encuestas, entrevistas, observaciones
Diseño del programa	• Calidad del programa y de sus metas	• Metas • Contenido • Estructura • Viabilidad	• Análisis de la estructura y contenido del programa por expertos • Análisis de los recursos del contexto
Implantación	• Adecuación a las metas y al	• Estrategias • Actividades • Clima	• Evaluación de la planificación • Evaluación de los recursos
Análisis de los procesos	• Adecuación a la situación de partida	• Constatación • Criterios • Referencias	• Evaluación de los ritmos del programa, de la participación y satisfacción (Observación, cuestionarios...)
Análisis de los resultados	• Medida logros • Valoración	• Constatación • Criterios • Referencias	• Efectos de programa, positivos y negativos • Evaluación de logros en coherencia con los criterios • Evaluación de la satisfacción
Propuestas de mejora	• Decisiones	• Continuidad • Cambio y mejoras	• Informe y propuestas de mejoras • Planificación y seguimiento

FIGURA 2.

5.3. Los informes de la evaluación y la utilización de sus resultados

Los evaluadores deben informar a todos los grupos interesados en saber, por lo que no debe limitarse a un único estilo de informe. La información evaluativa debe aportar nuevas luces sobre las concepciones de los problemas sociales sobre la teoría implícita de los programas. Por eso, los informes de evaluación deben mostrar:

- Las cuestiones que han aconsejado la evaluación.
- Los procesos seguidos.
- Las fuentes de donde toman sus datos.

- Los métodos, técnicas e instrumentos utilizados.
- Los criterios que llevan a emitir los juicios valorativos.
- Las recomendaciones.

La evaluación no es sólo una actividad técnica sino que es, también, una investigación diseñada para dirigir el proceso de formación política. Por ello, el evaluador debe ser consciente de que la información de una evaluación debe llegar allí dónde se toman las decisiones, de un modo adecuado, y a tiempo.

Algunas de las variables que afectan a la utilización de los resultados de las evaluaciones son las siguientes:

- La pertinencia de la realización de la evaluación.
- La comunicación entre los promotores de la evaluación, los evaluadores, aplicadores de los programas y su beneficiarios.
- La participación de los implicados en el programa,.
- La información a los implicados en el programa.
- La factibilidad de la utilización de los resultados.

Queda claro, en esta ordenación, que una evaluación aportará una información utilizable cuando ha sido posible y pertinente hacerla, y si se ha establecido una relación adecuada y transparente, informando de las intenciones de la evaluación a los participantes y beneficiarios, ya que de otro modo la información que aporta habría que valorarla con ciertas dudas de verosimilitud, lo que afectaría al uso de los resultados. Si los grupos y personas implicadas en el programa han sido informados y han colaborado en la evaluación, será más factible que acepten los resultados y colaboren en las propuestas de cambio que se deriven de las conclusiones de la evaluación.

5.4. La realización de proyectos innovativos de mejora

A partir de los resultados de la evaluación, e una dinámica de mejora continua, se deben realizar proyectos de mejora. Los principios que deben regir la realización de los proyectos, de acuerdo con Aisncow y otros (2001:14), son los siguientes:

- Concreción de la «visión» de la institución o centro que queremos, que debe ser el resultado de la contribución de todos los miembros de la comunidad escolar.

- Consideración de las presiones externas de cambio que demanda la sociedad, las familias, los estudiantes, como oportunidades para asegurar las prioridades internas.
- Creación de condiciones de aprendizaje necesarias para todos los miembros de la comunidad educativa.
- Promoción de la evaluación y el control de la calidad de la institución como responsabilidad de todo el profesorado.

Estos principios generan sinergia: el todo es más que la suma de las partes. El marco de la acción de formación debe centrarse en el proyecto y en las estrategias necesarias para llevarlo a cabo, contextualizadas en el centro educativo y su entorno.

La lógica de la mejora en la escuela para por las siguientes fases:

1. Análisis de las necesidades de mejora, y desarrollo y priorización de las mismas.
2. Elaboración de proyectos en coherencia con las necesidades.
3. Concreción de las estrategias necesarias para llevarlos a cabo.
4. Establecimiento de las condiciones necesarias para realizar los proyectos.

Estos cuatro elementos se ven afectados por la historia de la organización del centro, la implicación de sus miembros en los procesos de cambio y los resultados obtenidos, por lo que es importante mejorarlas.

Para conseguir la institución que queremos, Ainscow, Hopkins, Southworth, West, (2001:125 y ss.) utilizan la metáfora del viaje. Se viaja en una dirección guiada por objetivos, por procedimientos y por una visión que refleja el conjunto básico de valores de la organización. Y el viaje no lo realiza una sola persona, sino el grupo de personas que integran la institución, un centro educativo en este caso.

Un proyecto innovativo de mejora tendrá más seguro su éxito si se apoya en un proceso simultáneo de mejora de las condiciones de gestión del centro y además se enmarca en la dinámica global del centro, dado que la mejora institucional se concibe como un proceso holístico, global, no fragmentado, transversal, que implica a la totalidad de la institución. Estos esfuerzos innovativos de mejora se verán beneficiados si se promueve la gestión por procesos transversales, interdepartamentales, en equipos y liderazgo extenso que respalden los proyectos de mejora de la institución, éstas beneficiarán a todos. Para lograrlo hay que priorizar las

necesidades de cambio de las condiciones de gestión del centro a través del:

- Liderazgo participativo.
- Búsqueda del consenso en la visión, misión, proyectos.
- Compromiso en la formación del profesorado, de los empleados.
- Establecimiento de alianzas con instituciones externas y la comunidad próxima para que apoyen y contribuyan al cambio.
- Trabajo en equipo.
- Gestión por proyectos.
- Revisión y evaluación continua.
- Informes basados en evidencias recogidas con rigurosidad.
- Propuestas de mejora basadas en necesidades sentidas y expresadas.

Los proyectos deben vincularse a las condiciones internas para poderlos llevar a cabo, en los siguientes aspectos:

- Tiempo de preparación, planificación, formación.
- Organización del trabajo.
- Liderazgo para el mantenimiento del impulso hacia el cambio. Generar compromiso e involucración de las personas implicadas en el cambio: profesores, alumnos y familias. Conseguir apoyos de formación externos. El líder de los proyectos es un factor clave para su éxito.
- Enfoque del cambio: «piensa en lo grande y empieza por lo pequeño». Ten clara la meta, pero consíguela mediante pasos estructurados, ordenados y secuenciados para conseguir el fin, revisados y valorados de forma continua.

Las preguntas claves para la mejora escolar y las acciones para resolverlas, son:

1.^a *¿Dónde estamos ahora?*

- Analizar las condiciones internas del centro educativo.
- Fomentar la participación en el análisis de las condiciones actuales.

2.^a *¿Dónde quisiéramos estar?*

- Dialogar sobre la escuela que queremos: la visión.

→ Identificar las prioridades de mejora.

3.^a *¿Cómo llegaremos?*

→ Identificar las necesidades para realizar los proyectos: necesidades de formación, recursos.

→ Planificación de los proyectos, concreción de la estrategias, descripción de los recursos, determinación de los criterios de valoración de la realización y su calidad.

4.^a *¿Qué debemos hacer para conseguirlo?*

→ Dialogar sobre los procesos, sobre las dificultades, sobre los avances.

→ Revisar frecuentemente el progreso conseguido, en contraste con el proyecto planeado.

5.^a *¿A dónde iremos después?*

→ Nuevos proyectos de mejora en coherencia con la información aportada del proyecto, su realización y sus logros.

→ Consolidar los procesos de mejora: diseño, planificación, realización, revisión, mejora.

Lo importante para que la visión y la misión tengan éxito es involucrarse e implicar a todos los miembros en la «construcción de la visión». Los métodos de investigación, y las técnicas de recogida y análisis de la información, junto con los conocimientos sobre organización de centros educativos, y diseño y aplicación de programas educativos, nos ayudarán a responder a las cuestiones planteadas.

5.5. Los criterios y normas para juzgar la evaluación

De acuerdo con Stufflebeam y Shinkfield (1987) las normas por las que deben ser juzgadas las evaluaciones son las siguientes:

- **Normas de utilidad:** Para asegurar que sirve a las necesidades de información práctica, centrada en cuestiones importantes. Los estándares de utilidad guiarán la evaluación de tal forma que sean relevantes, oportunos y de influencia. Deben facilitar informes basados en evidencias, sobre aspectos positivos y negativos y debe aportar soluciones para mejorar.
- **Normas de viabilidad:** Para asegurar que la evaluación es realista y prudente. Debe utilizar procedimientos eficientes que no ocasionen demasiados problemas. Los estándares de viabilidad reco-

nocen que las evaluaciones se realizan en contextos naturales, sus diseños deben ser operativos y sus gastos, en material, personal o tiempo, no deben exceder al requerido por las cuestiones de la evaluación.

- **Normas de honradez:** Para asegurar que la evaluación es conducida legal y éticamente, basada en compromisos explícitos, que aseguren la necesaria cooperación, la protección de los grupos implicados, y la honestidad de los resultados.
- **Normas de precisión:** Para asegurar que la evaluación revela y comunica la información encontrada; describe con claridad el objeto evaluado en su evolución y en su contexto; revela las virtudes y defectos del plan de evaluación, de los procedimientos y de las conclusiones (metaevaluación) y proporciona conclusiones válidas y fidedignas. Las conclusiones y los juicios deben ser coherentes con los datos.

Estas normas sirven de guía para la realización de evaluaciones con el fin de que sean útiles, factibles, éticas y exactas, requisitos fundamentales para que la evaluación cumpla con su objetivo primordial: la toma de decisiones informada que permita introducir mejoras ajustadas a las necesidades reales que motivaron el diseño del programa y las nuevas necesidades detectadas. La aplicación de estos cuatro grupos de normas aseguran:

- **Una evaluación integral** que afecta a la organización y a las personas implicadas en el programa, desde sus metas y objetivos, hasta la planificación, estrategias, secuenciación y recursos para conseguirlos, los resultados y sus conclusiones y valoraciones.
- **Una evaluación sistemática**, estructurada siguiendo un plan consensuado, asegurando la objetividad y honradez de los resultados.
- **Una evaluación innovadora**, que permite ir comprobando el progreso, así como los elementos que inciden en la mejora continua del programa.
- **Una evaluación para la mejora**, que al ser exacta por fundamentarse en evidencias contrastadas, permite establecer planes de mejora a corto y medio plazo y la valoración progresiva de los avances.

5.6. El enfoque de calidad total y el modelo de excelencia europeo

Todo modelo es una descripción teórica de una realidad que trata de comprender, analizar y, en su caso, modificar. Tras definir las características teóricas del modelo se elaboran criterios de calidad para compararlos con su realización.

5.6.1. *El concepto de la calidad*

El concepto actual de «calidad de la educación», como meta de toda institución educativa, asume un enfoque global de gestión de la institución hacia la consecución de metas de calidad para todos y con la colaboración de todos los implicados en la institución educativa. La calidad conseguida tiene que ver con la calidad de las metas pretendidas, concretadas en objetivos y criterios y estándares que guían su realización y comprobación de su consecución.

La calidad abarca a todas las funciones y actividades de la institución y debe estar vinculada a las necesidades relevantes de la sociedad en un ámbito y contexto dado. Las necesidades para la realización de los procesos, de los productos, de los servicios deben estar establecidas y especificadas, aunque a veces existen necesidades que no están explícitas siendo conveniente identificarlas y definirlas para poder satisfacerlas con lo que se sorprenderá al cliente con servicios en los que no había pensado, aportando valor añadido.

5.6.2. *Los principios de la Calidad Total*

Los Sistemas de gestión de calidad, se fundamenta en los principios que quedan reflejados en el siguiente diagrama relacional adaptado a partir de www.iso.quema (2000):

Esquema conceptual de la mejora continua
en los sistemas de gestión de la Calidad Total

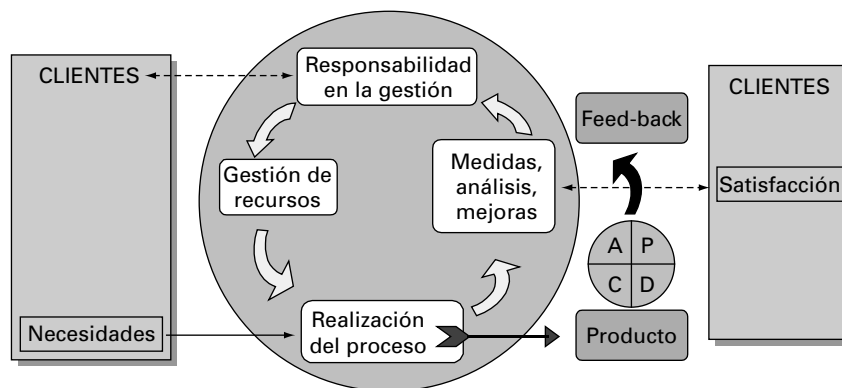


FIGURA 4.

Dichos principios han sido definidos dentro del enfoque de Calidad Total para la gestión de las organizaciones, y adoptados por el Modelo Europeo de Excelencia, en los términos en que a continuación se indican:

- **Orientación al usuario.** El usuario es la persona que se beneficia directamente de las actividades de la organización. En el caso de los centros educativos es el alumno, su familia y la sociedad. El centro ha de identificar a sus clientes, sus necesidades y expectativas para satisfacerlas. La excelencia depende del equilibrio y satisfacción de las necesidades de todos los clientes relevantes, incluidas las personas que trabajan en la organización, los que solicitan sus servicios, los proveedores y la sociedad en general, así como todos aquellos con intereses, de un tipo u otro, en la organización.
- **Liderazgo y coherencia con los objetivos.** El comportamiento de los líderes de una organización comprometida con la calidad debe contribuir a la definición de la visión, misión y estrategias para conseguirla, para satisfacer las necesidades del cliente, basándose en la claridad y unidad de propósito, favoreciendo la creación de un entorno en el cual la organización y las personas que la integran puedan alcanzar la excelencia.
- **Desarrollo, implicación y reconocimiento de las personas.** El Modelo de Excelencia Europeo otorga un papel importante a las personas que trabajan en una organización, y considera decisiva su participación e implicación en los procesos de mejora, median-

te el trabajo cooperativo. El potencial de cada una de las personas que trabaja en una organización se potencia y aflora cuando se comparten valores en una cultura, y existe confianza mutua, lo cual anima a la involucración de todos hacia el logro de los objetivos de la organización.

- **Enfoque por proyectos.** La institución consigue mejores rendimientos cuando trabaja en torno a proyectos de mejora tendentes a dar respuesta a las necesidades detectadas. Los proyectos son planificados y solventados con los recursos necesarios para llevarlos a cabo mediante una cadena sucesiva de procesos en torno a la organización general del centro, el clima escolar, concretado en las relaciones entre todos los miembros de la comunidad, en torno a los proyectos educativos de centro y aula, los proyectos de desarrollo de áreas de contenidos curriculares, incluyendo las estrategias de enseñanza-aprendizaje, la orientación y tutoría, con la participación e implicación en y de la comunidad.
- **Gestión por sistemas y hechos para la toma de decisiones.** Las mejoras propuestas deben estar fundamentadas en la información fiable que incluye las percepciones de los miembros de la institución y de los clientes. Todas las actividades deben estar interrelacionadas y ser gestionadas sistemáticamente, con la comprensión e implicación de todos los miembros de la institución. El conjunto, el todo, es más que la suma de sus partes. Las decisiones a tomar y los proyectos consecuentes deben derivarse del conocimiento sistemático y riguroso del funcionamiento de la organización.
- **Relaciones con los proveedores mutuamente beneficiosas y desarrollo de alianzas.** Referido a las personas, instituciones y organizaciones que suministran productos, servicios o conocimientos al centro educativo. La organización trabaja de un modo más efectivo cuando establece con sus proveedores unas relaciones mutuamente beneficiosas basadas en la confianza, en compartir el conocimiento y en la integración, buscando alianzas externas y provecho de los recursos internos en apoyo de su política y estrategia y del eficaz funcionamiento de sus procesos y de su personal. Basadas en la utilización de sistemas de certificación de la calidad.
- **Aprendizaje, innovación y mejora continua.** El rendimiento de una institución se maximiza cuando se fundamenta en la gestión compartida del conocimiento dentro de una cultura de aprendizaje, innovación y mejora continuas. Y este se basa en la revisión

de los resultados, en coherencia con los criterios previos utilizados en la definición de los proyectos y programas, como son la misión, la política y la estrategia, y los elementos que la determinan: la formación del personal, las colaboraciones y utilización de los recursos, la determinación de los procesos, la planificación y realización de los mismos y su evaluación, subrayando la importancia de la retroalimentación, basada en la recogida de información fidedigna que permita introducir cambios en pos de una mejora continua mediante el aprendizaje organizacional, la innovación y la mejora permanente.

- **Ética y responsabilidad.** Los intereses de una organización, incluidos los de su personal, son mejor servidos si se adoptan enfoques éticos que excedan las expectativas y las normativas de la comunidad. Los miembros de cualquier organización han de comportarse de acuerdo con una ética que otorgue significado a las acciones individuales más allá del que pueda derivarse de las normas o requisitos legales. En el ámbito de lo educativo este planteamiento adquiere especial importancia: la ética de servicio ha de impulsar a todo el personal de un centro educativo a esforzarse por mejorar continuamente, en un ámbito cuya repercusión en el futuro de las personas y de la sociedad es fundamental.

5.6.3. *El Modelo de Excelencia Europeo: Estructura y Criterios*

El Modelo de Excelencia Europeo, de carácter no normativo, tiene como finalidad ayudar a las instituciones a conocer su situación en relación con sus fines. Fue anunciado por la EFQM en 1991, bajo el patrocinio de la Comisión Europea, con el nombre de «Modelo europeo de gestión de la calidad». Desde su actualización en 1999 se llama Modelo de Excelencia Europeo, y se ha constituido en una herramienta valiosa para ayudar a las organizaciones a conseguir la satisfacción de sus clientes, tanto internos como externos, atendiendo a sus necesidades presentes y futuras, siguiendo los principios de la Calidad Total. El modelo cumple el doble objetivo de servir como herramienta para el autodiagnóstico y para la certificación de la Calidad.

Las administraciones educativas públicas, los sistemas educativos y las instituciones educativas públicas y privadas participan de este movimiento para el logro de la excelencia, a través de la implicación de todos los miembros de la institución, en sus distintos niveles de responsabilidad, en los procesos de gestión y autoevaluación para la mejora y rendimiento de cuentas a la sociedad.

La finalidad del modelo aplicado al ámbito educativo es conocer la realidad del centro, de un modo riguroso, y reflexionar sobre ella, lo que posibilitará elaborar planes y estrategias de mejora. Estas dos características, junto con la implicación de todos en el proceso de autoevaluación y toma de decisiones, se consideran factores que favorecerán la implicación de los miembros de la institución en la mejora de la calidad del centro. El modelo define unos factores críticos para el buen funcionamiento de una organización a los que denomina criterios, y que están integrados por indicadores. La comprobación sistemática de la existencia o no de dichos indicadores en la institución, nos permite tener un conocimiento objetivo del funcionamiento del centro, para de este modo poder elaborar proyectos y planes de mejora.

La calidad de una institución educativa se consigue mejorando el conocimiento real del funcionamiento de la propia institución, recogiendo información continua sobre los logros de los objetivos de aprendizaje y sobre los recursos y condiciones para el desarrollo del aprendizaje, con el fin de introducir modificaciones de mejora adecuadas y a tiempo y así mejorar el rendimiento de los alumnos y reducir el fracaso escolar. Está demostrado que uno de los aspectos clave para la mejora de la escuela son los procesos de enseñanza y aprendizaje y el clima del aula en el que se desarrollan, viéndose estos aspectos favorecidos por la organización y el clima general del centro, lo que contribuye a prevenir los fallos antes de que ocurran.

El Modelo de Excelencia está formado por dos bloques: el de los agentes o facilitadores y el de los resultados. Con el objetivo puesto en los resultados que es la satisfacción de las necesidades del cliente, es a través del bloque de los agentes, concretados en el liderazgo de la institución, en su política y estrategia, en el personal de la institución y en los colaboradores y recursos que contribuyen en la concreción de la realización de los procesos para consecución de los logros en coherencia con las necesidades del cliente.

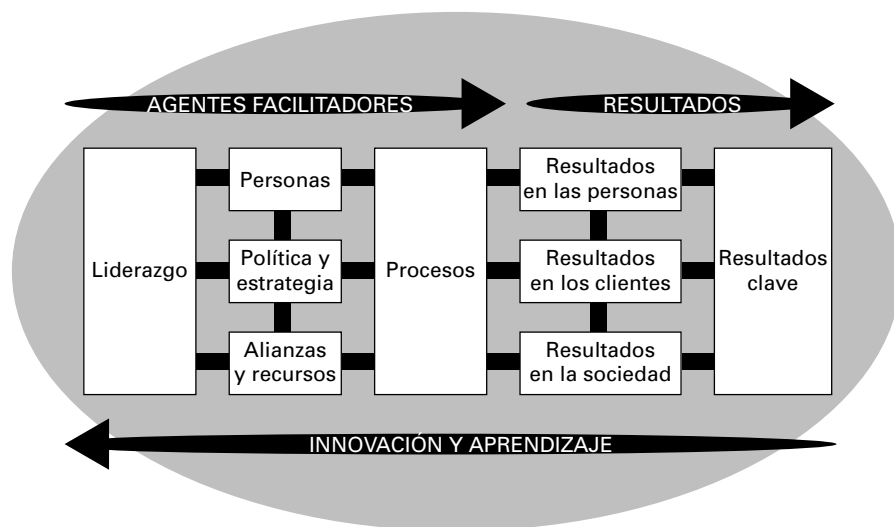


FIGURA 4. (tomada de: www.efqm.org, 2002).

5.6.3.1. Criterios agentes o facilitadores

Los criterios agentes reflejan como el centro educativo enfoca cada uno de sus subcriterios o áreas que integran los diferentes criterios. Son los siguientes:

- *Liderazgo*: comportamiento y actuación de los dirigentes y responsables de la institución, encaminado a conducir el centro escolar hacia la gestión de la calidad.
- *Planificación y Estrategia*: referida a la misión, los valores y la dirección estratégica del centro, y al modo en que se implantan los proyectos de la organización.
- *Personas*: centrada en la forma en que la institución utiliza al máximo el potencial de su personal, para lograr la mejora continua de la institución.
- *Colaboradores y Recursos*: referido al establecimiento de alianzas con colaboradores y a la gestión, uso y conservación de los medios materiales al servicio de las funciones de la organización.
- *Procesos*: referido a la forma en cómo se identifican los procesos, se planifican, se gestionan, se aplican, se revisan y se evalúan y se mejoran, en coherencia con los estándares representativos de los proyectos de la institución educativa.

5.6.3.2. Criterios de los resultados

Los criterios del bloque de los resultados tienen como finalidad conocer que ha obtenidos el centro educativo. También, visto a priori, señalan los resultados que desea obtener el centro. Son los siguientes:

- *Resultados en los clientes:* el alumno, su familia, la sociedad, las empresas, en relación con lo que se espera que consiga el centro en relación con su misión.
- *Resultados en el personal:* grado en que se da la adecuada respuesta a las necesidades y expectativas de las personas que trabajan en la organización. Esta atención integra campos como el ambiente laboral, la formación continua, la salud o la seguridad.
- *Resultados en la sociedad:* qué es lo que consigue el centro en relación con las necesidades y expectativas de la sociedad, y en el entorno próximo.
- *Resultados clave:* qué consigue el centro, en relación con la Política, su estrategia y planificación, y con respecto a la satisfacción de las necesidades y expectativas de los alumnos, de las familias, y en general de la sociedad.

5.6.4. La autoevaluación y el Modelo de Excelencia Europeo

La autoevaluación es la herramienta fundamental del Modelo de Excelencia Europeo, entendida como un examen global y sistemático de las actividades y resultados de una organización que se compara con un modelo de excelencia. Mediante el proceso de la autoevaluación se pretende conseguir una comprensión detallada del centro educativo, buscando realizar un diagnóstico sobre cual es la situación del centro, señalando los puntos fuertes y las áreas de mejora. A partir de este diagnóstico se elaboran propuestas de mejora, objetivas, concretas y conseguibles en un plazo determinado, arbitrando un seguimiento de dicho plan de mejora, para ayudar a su implantación y valoración de logro mediante la autoevaluación, introduciéndose en el centro una dinámica de mejora continua, o lo que es lo mismo, la gestión de la calidad. Aunque la autoevaluación suele ser aplicada al conjunto de la organización, también puede evaluarse un departamento o cualquier otra unidad.

LA AUTOEVALUACIÓN

La autoevaluación es un proceso por el cual se recoge información precisa y rigurosa, de modo sistemático, sobre la calidad de los programas en el contexto del centro, siguiendo, en este caso, el modelo de excelencia europeo.

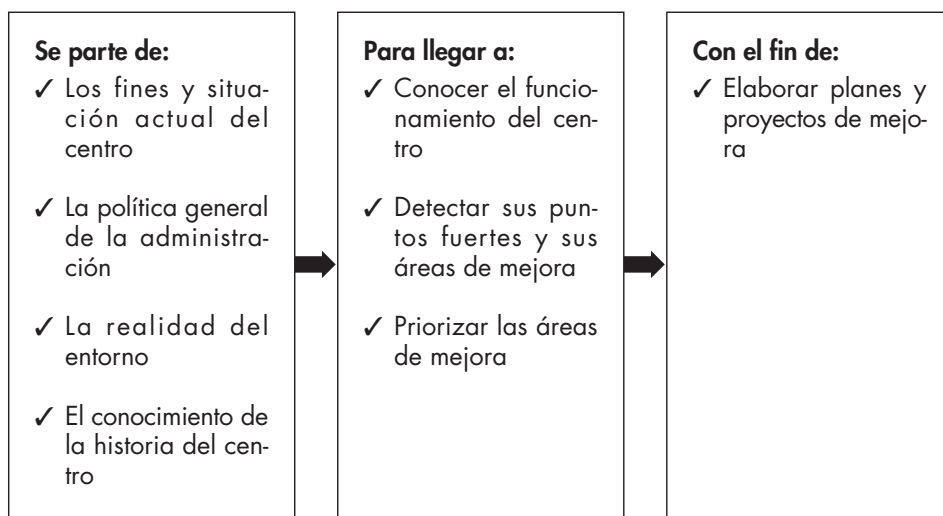


FIGURA 5.

5.6.5. La elaboración de los planes de mejora

Un plan de mejora parte de las aportaciones de la autoevaluación que ha mostrado las fortalezas y las debilidades de los objetivos del centro, de los enfoques, de los despliegues y de los resultados. La autoevaluación señala, por tanto, los puntos fuertes y las áreas de mejora. La evaluación de este modo realiza una función diagnóstica indicando las necesidades de mejora. Los criterios para priorizar las áreas de mejora, son los siguientes:

- La gravedad de las carencias detectadas
- La factibilidad de las acciones a emprender
- La rentabilidad prevista en sus resultados
- El grado de consenso entre los afectados por las decisiones a tomar

Una vez consensuadas y priorizadas las necesidades, se elaboran en coherencia los Planes de mejora. Un plan de mejora es un instrumento

que parte de la evaluación de necesidades, que ha marcado la línea base en la que se encuentra el centro, y sobre la que se deben promover las mejoras. El plan especifica los objetivos a conseguir de modo realista, concretos y evaluables, siendo necesario para conseguir las mejoras, la concreción de un plan de seguimiento para constatar sus logros. El seguimiento sirve a la institución como instrumento para el aprendizaje organizativo.

EL PLAN DE MEJORA

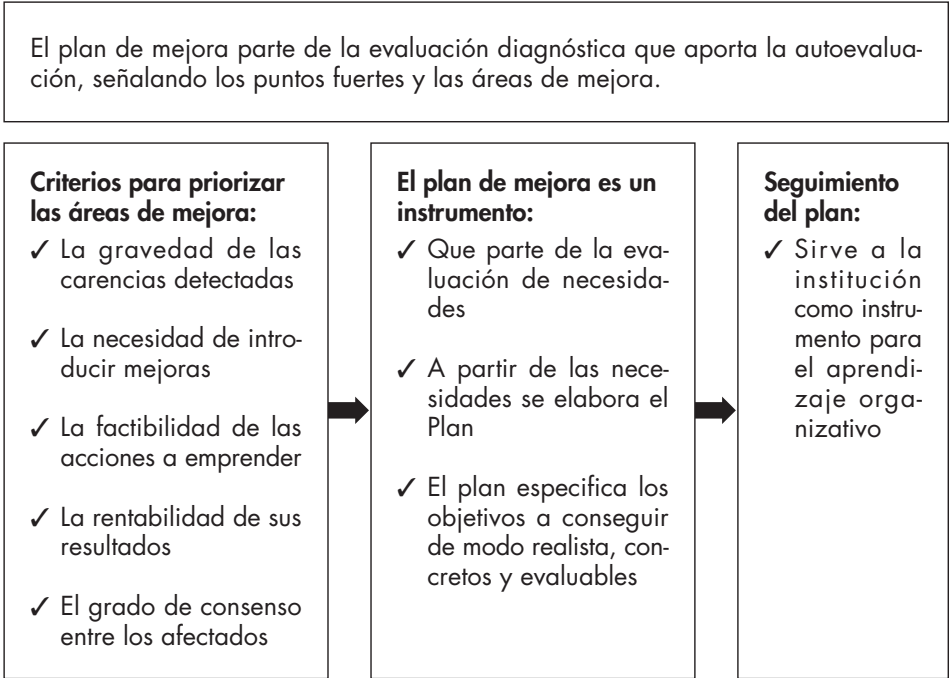


FIGURA 6.

El Modelo Europeo de Excelencia está abierto a mejoras, por lo que sus indicadores, tras las valoraciones de los usuarios, se adaptarán cada vez más a las características del sistema educativo, buscando con ello lograr una mayor participación y compromiso de la comunidad educativa en tareas de evaluación para la mejora de la calidad de los Centros Educativos.

5.7. Ejercicios de autoevaluación

Piense en un objeto posible de evaluación, con el que se sienta relacionado, o esté interesado, y realice un plan de evaluación en torno a las siguientes indicaciones:

1. Describa el objeto a evaluar.
2. Defina sus metas.
3. Describa, brevemente, cómo debe trabajar dicho objeto para conseguir sus metas.
4. Qué tipo de evaluación, o mejor, qué funciones aplicaría y por qué.
5. Qué tipo de información recogería.
6. En qué momentos la recogería.
7. De qué modo y qué instrumentos utilizaría.
8. Cuáles son los criterios de evaluación.
9. Cómo los aplicaría para evaluar el programa.
10. Importancia de la utilidad de los hallazgos.
11. Reflexione sobre los posibles modos de potenciar el uso de los resultados de la evaluación de programas. Indíquelo.

Valore la adecuación de la evaluación interna o externa en el objeto o programa que Vd. viene evaluando.

5.8. Referencias bibliográficas

- AINSCOW, M. HOPKINS, D., SOUTHWORT, G. y WEST, M. (2001). *Hacia escuelas eficaces para todos. Manual para la formación de equipos docentes*. Madrid, Narcea. (Primera edición en inglés en 1994, London, Fulton).
- MARTÍNEZ MEDIANO, C. (1997). *Evaluación de programas educativos. Modelos de evaluación de programas. Investigación evaluativa*. Madrid, UNED, Cuadernos de la UNED, 159.
- MARTÍNEZ MEDIANO, C. (1998). *La teoría de la evaluación de programas. Educación XXI. 1*. Madrid, UNED, Revista de la Facultad de Educación.
- MARTÍNEZ MEDIANO, C. y GONZÁLEZ GALÁN, A. (2001). *La evaluación para la mejora de la calidad de los centros educativos*. Madrid, UNED, Textos de Educación Permanente.

MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE: (2001). *El Modelo Europeo de Excelencia*. Madrid. Secretaría General Técnica.

STUFFLEBEAM, D. and SHINKFIELD, D. (1987). *Evaluación sistemática*. Madrid: Paidós, MEC.

www.efqm.org

www.iso.quema

Segunda Unidad Didáctica

LA MEDIDA EN EDUCACIÓN
Y LAS TÉCNICAS
PARA LA RECOGIDA DE INFORMACIÓN

TEMA 6

La medida en Educación y en Ciencias Sociales

ESQUEMA

- 6.1.** El concepto de medida.
- 6.2.** Elementos a considerar en la medición
 - 6.2.1. El objeto a medir.
 - 6.2.2. La elaboración del instrumento de medida
 - 6.2.3. El acto de la medición
 - 6.2.4. El tratamiento de la información
- 6.3.** Problemática de la medida en Educación y en Ciencias Sociales
 - 6.3.1. Del fenómeno a los conceptos, a las dimensiones o variables, a los indicadores y a los índices
 - 6.3.2. Las variables y sus clases
 - 6.3.3. Los indicadores
 - 6.3.4. Los índices
- 6.4.** Los niveles de medición nominal, ordinal, de intervalo y de razón
 - 6.4.1. Nivel de medida nominal
 - 6.4.2. Nivel de medida ordinal
 - 6.4.3. Niveles de medida de intervalo y de razón
- 6.5.** Relación entre los indicadores, los índices, los niveles de medida y las pruebas estadísticas
- 6.6.** Ejercicios de autoevaluación
- 6.7.** Soluciones a los ejercicios de autoevaluación
- 6.8.** Referencias bibliográficas

OBJETIVOS

Tras el estudio de este tema, se espera que el alumno sea capaz de:

- ✓ Comprender el concepto de medida.
- ✓ Comprender las características de la medición.
- ✓ Comprender y valorar la problemática de la medición en la Educación y Ciencias Sociales.
- ✓ Conocer y comprender los diferentes niveles de medida.
- ✓ Comprender la relación entre los niveles de medida y las pruebas estadísticas.

6.1. El concepto de medida

En el ámbito de la Educación y las Ciencias Sociales la medición es el establecimiento de correspondencias que se atribuye entre números y propiedades de los objetos o hechos como las actitudes, los valores, los intereses, las aptitudes, que constituyen su campo de interés y estudio. Se entiende por medir la asignación de símbolos numéricos a alguna magnitud, atributo, acontecimientos o hechos.

La Pedagogía como ciencia requiere que los procedimientos seguidos para la construcción de conocimiento científico sobre la educación sean sistemáticos y demostrables sirviéndose de un método. Dentro del enfoque de investigación empírica, desde la perspectiva positivista, es fundamental la medición objetiva de los fenómenos estudiados para que sea posible la comprobación de los resultados de nuestros estudios o investigaciones.

La medición de una propiedad lleva consigo una asignación de números a objetos como una forma de representación de esa propiedad. El proceso comprende un sistema formal y un sistema lógico. El sistema de números reales (símbolos) deberá corresponderse con un sistema empírico considerado como el conjunto de particularidades de la propiedad. Para poder establecer esa correspondencia entre un sistema conceptual y un sistema cifrado, es decir entre unas propiedades y unos números, y poder realizar por tanto medidas válidas, ambos sistemas —el de cifras y el de conceptos— deben ser isomórficos, las estructuras de ambos sistemas deben ser internamente semejantes.

Medir significa comparar una magnitud con otra de su misma especie, que se considera como unidad, con el fin de conocer su magnitud, extensión o cantidad. La medición se define como la asignación de numerales a objetos o hechos de acuerdo con reglas. Como podemos comprender, esta definición muestra ya las dificultades que el término medir tienen en el ámbito de las ciencias humanas para abarcar en su justa extensión aspectos como por ejemplo las habilidades cognitivas.

Bunge describe el acto de la medición del siguiente modo: «Cuantificar una propiedad de un sistema concreto es proyectar el conjunto de grados de la propiedad... sobre un conjunto de números; la serie de números se caracteriza por presentarse en un orden determinado y con una distancia cuantitativa que varía según el tipo de número; este hecho debe compararse con los objetos, hechos o propiedades que se desean medir y comprobar hasta qué punto se da el isomorfismo a que aludimos. La información que nos den los números así asignados será el

referente a tales propiedades en el sistema de los objetos y no en el de los números. (1976:787, en Pérez Juste, 1984: 134).

En las Ciencias Sociales las unidades de medida suelen ser sujetos o grupos humanos, cuya variación resulta más imprevisible que los objetos de medida sobre los que operan las ciencias físico-matemáticas. Como consecuencia de esto hay una cierta imprecisión entre las leyes que nos permitirían fijar como constantes los factores que intervienen en la medición de una variable relacionada con otra. Las medidas en ciencias sociales suelen ser más imprecisas que en las ciencias físico-matemáticas, lo cual no significa que en éstas otras ciencias sean exactas, son sólo menos imprecisas. (González Blanco, 1986: 277).

Las reflexiones teóricas sobre nuestros objetos de estudio en las ciencias humanas, pueden y deben ser contrastadas con la realidad concreta, y no sólo con el objetivo de acumular información y datos, sino con el objetivo de actuar sobre la realidad y contrastar hipótesis con la ayuda de las metodologías que utilizan las demás ciencias incluidas las físico-matemáticas. Y para ello, las mediciones nos permiten ir más allá de la mera descripción de los fenómenos observables; nos permiten concretar las observaciones, comprobar la interrelación entre distintos fenómenos sociales. El proceso de la medición, y los resultados que por su medio obtenemos, nos permiten revisar los conceptos, las hipótesis y teorías contempladas en los trabajos en las ciencias humanas. Por todo ello, el medir en educación y en ciencias sociales es una parte sustantiva del quehacer en estas ciencias humanas.

6.2. Elementos a considerar en la medición

La medición es una tarea compleja. Antes de pasar al acto concreto de la medición es necesario delimitar aquello que ha de ser medido, definir la regla de medida, e incluso construir el instrumento para la medida o recogida de la información. Posterior a todo esto, cuando se está realizando la medida en la situación concreta que tratamos estudiar, pueden surgir circunstancias que interfieren en la medición y confunden la propia medición. Una vez que se ha recogido la información mediante los instrumentos de medida, el proceso no ha terminado puesto que tenemos que realizar los tratamientos estadísticos y el análisis de los datos, la interpretación y valoración de los mismos, y la deducción de conclusiones y recomendaciones de aplicación derivadas del estudio. Si todos estos elementos no se integran de forma coherente y sin contradicciones e incompatibilidades, el esfuerzo habrá sido eficaz, de otro modo las posibles con-

clusiones que aporte el estudio no serán provechosas, sino todo lo contrario, habrá supuesto una pérdida de tiempo y esfuerzo.

6.2.1. *El objeto a medir*

Lo que pretendemos medir puede encerrar en sí mismo una enorme complejidad. Por ejemplo, si pretendemos conocer las relaciones entre el fracaso escolar y la implicación de los padres en la actividad escolar de sus hijos, a simple vista podemos apreciar la gran cantidad de variables que intervienen en este enunciado. Podríamos tratar de enumerar algunas de las cuestiones a tener en cuenta para su medición, y podríamos mencionar las siguientes: 1. Características de las relaciones entre padres e hijos: *a)* estilo afectivo de la relación, *b)* tiempo que dedica el padre a hablar con su hijo, *c)* tiempo que dedica la madre a hablar con su hijo. *d)* tiempo dedicado por el estudiante a la realización de tareas escolares en casa, *e)* frecuencia del seguimiento del progreso escolar del estudiante por parte del padre y de la madre, *f)* características del seguimiento del aprendizaje de los alumnos por parte de sus padres. Y podríamos hablar de la relación entre padres e hijos en torno al estilo de diversión de sus hijos, actividades que realizan en su tiempo libre, amistades de los hijos... Como podemos ver hay diversidad de variables, tales como las habilidades relacionales entre padres e hijos, con el interés de los padres en las actividades de aprendizaje de sus hijos, con las actividades recreativas de los hijos, con el seguimiento de la actividad escolar en el centro educativo, el conocimiento de las capacidades de su hijo, de sus intereses.

6.2.2. *La elaboración del instrumento de medida*

La dificultad de la medición se refiere a la dificultad de definir los elementos que subyacen en las construcciones teóricas que queremos estudiar, por lo que previamente a cualquier medición deberemos informarnos acerca de las características de aquello que queremos medir. La complejidad de las construcciones teóricas como la personalidad, los intereses, las relaciones interpersonales, hace de la medida en Educación y en Ciencias Sociales, una tarea no exenta de complejidad que se debe abordar con rigurosidad, por que de lo contrario nuestros estudios aportarían poco para el conocimiento y la mejora de las situaciones humanas.

La elaboración de un instrumento para la recogida de información sobre una determinada construcción teórica, exige una fundamentación teórica sobre aquello que queremos medir, y una construcción de instru-

mento contrastado con la opinión de expertos y con la plausibilidad de que pueda ser aplicable realmente para la recogida de información pretendida, comprobable mediante su aplicación y su valoración posterior para constatar su constancia, o fiabilidad, y precisión, o validez, en lo que pretende medir.

6.2.3. *El acto de la medición*

El acto de la medición supone la aplicación del instrumento de recogida de datos en la situación y objeto que queremos medir, y este hecho se encuentra afectado por:

- Las habilidades de la persona que aplica el instrumento para crear un clima favorable para la aplicación del mismo y lograr la colaboración de los sujetos en la recogida de información.
- Las circunstancias espacio-temporales en las que se produce la aplicación, procurando que favorezcan la colaboración en la respuesta. Por ejemplo se tratará de evitar aplicar el instrumento en una situación de cansancio de los sujetos, lo que podría influir negativamente en el modo como se responde; evitar condiciones ambientales de mucho frío, o calor, o poca luminosidad, o ruidos, pueden también incidir negativamente en la recogida de información sobre las variables que se pretenden medir.
- Las características del instrumento, tanto referidas a la calidad de su construcción, como a la presentación formal, contribuirán a una recogida de información óptima.

6.2.4. *El tratamiento de la información*

La información recogida a través de los diferentes instrumentos contruidos para tal fin es transformada y codificada en números para su análisis de datos e interpretación de los resultados, siempre vinculados al contexto para el que se han elaborado y en el que se han aplicado. La estadística descriptiva permite estudiar la información bruta mediante su transformación en tablas, gráficos, medidas de posición, de variabilidad o de asimetría, así como el estudio de relaciones entre dos o más variables, nos aporta una primera aproximación a la interpretación del estudio del constructo o construcciones teóricas en el grupo. La estadística inferencial nos va a permitir proyectar los estudios realizados en un grupo a la población, siempre que cumpla determinados supuestos.

6.3. Problemática de la medida en Educación y en Ciencias Sociales

Cualquier estudio, o trabajo de investigación, suele comenzar por un fenómeno, un hecho social, o educativo, o un problema, sobre el que recae nuestra atención y que queremos analizar para conocer mejor y aportar mejoras. En un principio, nuestras observaciones las realizamos de un modo vago sobre las unidades, situaciones, grupos o personas, en donde se presenta el objeto de nuestro interés, tratando de identificar una serie de conceptos que en principio son genéricos, pero que presentan una serie de características, propiedades, atributos, formas externas de manifestación del comportamiento, y por lo tanto, observables.

El problema surge al tratar de pretender captar de un modo más preciso esa realidad observada y transformar las observaciones en datos, a través de una medición de esa realidad. Pues bien, para hacer medible dicha realidad es por lo que debemos concretar las unidades de observación, los conceptos, y transformar las observaciones a valores numéricos, o a unidades de significado o categorías a las que también se les puede asignar valores numéricos o ser representadas con números. (González Blanco, 1986: 278).

6.3.1. *Del fenómeno a los conceptos, a las dimensiones o variables, a los indicadores y a los índices*

Para llegar a las unidades de medición partimos del fenómeno objeto de nuestro interés, de su concepto, y de éste a las dimensiones o variables, y de éstas a los indicadores, y de éstos a los índices. El investigador actúa entre la realidad y la teoría, y va de uno a otro. Si parte de la observación de la realidad y pretende resolver un problema, se acerca a la teoría para construir hipótesis que necesitará probar recogiendo información de sus observaciones. Los conceptos son abstracciones formadas por generalizaciones de casos particulares, por ejemplo, «dureza». Las construcciones hipotéticas suponen una invención adaptada a un propósito científico, por ejemplo, «rendimiento escolar». El término variable se refiere a las «construcciones hipotéticas» o «propiedades» que estudian los investigadores como por ejemplo «autoestima», «atención», «clase social», razonamiento numérico». La variable es una propiedad que adquiere distintos valores. La variable es algo que varía.

Lo primero que hay que distinguir son las diferentes dimensiones del concepto; por ejemplo, si queremos estudiar las competencias de un

profesional, podemos analizarlas a partir de su: *a)* formación inicial: titulación académica, *b)* formación complementaria; *c)* experiencia profesional, en empresas nacionales e internacionales; *d)* capacidad para trabajar en equipo; *e)* capacidad para comunicar resultados de proyectos; *f)* habilidades para las relaciones interpersonales, *g)* concepción personal de su profesión; aficiones.

Al analizar y desagregar los atributos y dimensiones de los conceptos se gana en concreción y favorece la precisión en la medición. A las dimensiones se las suele denominar con el término variable, o factor, y nos proporcionan una aproximación al fenómeno que queremos estudiar. A veces una determinada variable está integrada por otras variables que podemos llamar de segundo orden o subvariables. Una característica de estas variables es que son mensurables, es decir, que al contrastar la variable y sus indicadores con la realidad a la que nos queremos aproximar, obtengamos una información que pueda ser medible. En el ejemplo de las «Competencias profesionales» podemos obtener información de una persona, de entre el grupo al que le hemos recogido información a partir de las dimensiones e indicadores señaladas:

- Tiene la titulación de:
 - licenciado en Pedagogía,
 - master en Recursos Humanos,
- Conocimiento de idiomas:
 - dominio del Inglés,
 - dominio del Francés,
- Habilidades profesionales especiales:
 - habilidades para el trabajo en equipo,
- Experiencia profesional:
 - experiencia de 2 años en una empresa nacional,
 - experiencia de 6 meses en una empresa norteamericana,
- Aficiones:
 - aficiones por los viajes,
 - aficiones por los deportes de aventura.

De entre todas las personas entrevistadas podremos seleccionar para una empresa a la persona cuyo perfil profesional se acerca más a las competencias requeridas para el puesto de trabajo que se necesita cubrir.

La definición operativa del indicador es lo que nos permite recoger información transformable en valores numéricos que nos va a permitir

realizar operaciones de cálculo, estadísticas, para obtener índices que nos van a permitir hacer nuevas operaciones para describir la realidad estudiada, comprenderla, explicar e incluso predecir acontecimientos, siempre en términos probabilísticos, nunca de certeza.

Gráficamente tendríamos:

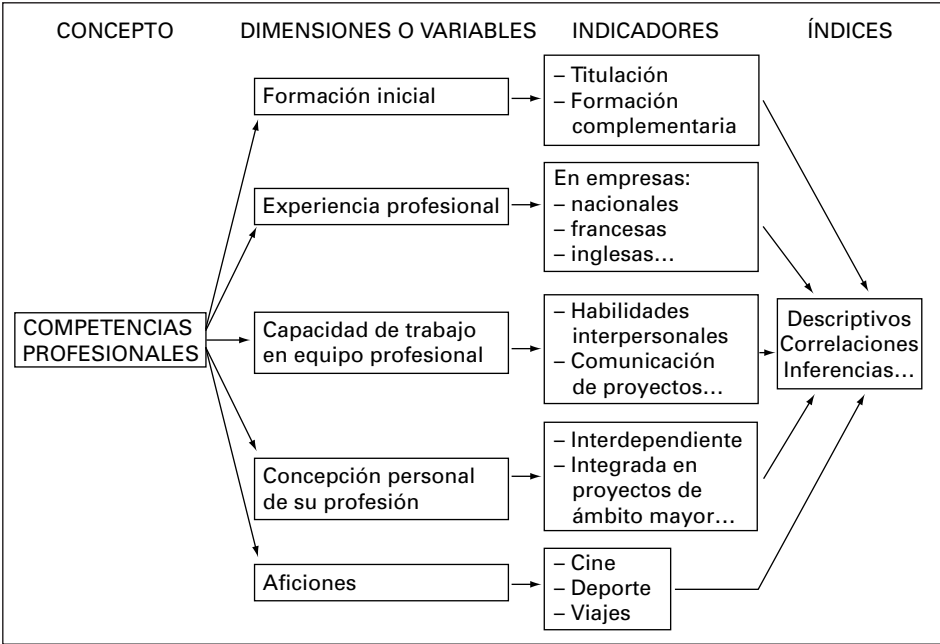


FIGURA 6.1. Del concepto a los índices.

Cuantas más dimensiones, o variables utilicemos para acercarnos al problema, mayor riqueza conseguiremos, pero a su vez, también mayor dificultad tendremos en el análisis de la información. Lo recomendable es llegar a un equilibrio entre la cantidad y la suficiencia y calidad con que el concepto o problema de estudio está representado por las dimensiones y sus indicadores. La contrastación del concepto y las dimensiones con la realidad, y la valoración de expertos y la búsqueda de acuerdos y consenso en la adecuación de las dimensiones a los conceptos y sus indicadores, serán indicios de adecuación de éstos al concepto y fenómeno que tratamos de estudiar.

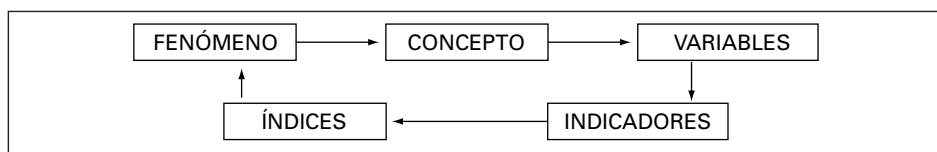


FIGURA 6.2. Proceso circular del fenómeno de los índices.

Los fenómenos o hechos que nos ocupan se delimitan por conceptos. Los conceptos se concretan en dimensiones o variables que se dividen en unidades más concretas y fácilmente observables, identificables y que a su vez se desagregan en indicadores, que deben ser definidos operativamente, y que nos va a permitir operar con ellos para obtener los índices. La información recogida de la realidad, mediante los indicadores nos sirven para describir los hechos que estudiamos, analizarlos, ayudándonos a comprender el fenómeno, permitiéndonos relacionar los obtenidos índices con otros, con el fin de explicar los fenómenos e incluso prever nuevos acontecimientos.

6.3.2. Las variables y sus clases

Denominamos con el término «variable» a las representaciones simbólicas de las dimensiones de los conceptos o de los atributos de las unidades. El término variable también se aplica directamente a los conceptos mismos sin ser descompuestos en dimensiones. Las dimensiones o variables, al representar propiedades de las unidades observadas, nos permiten categorizar o atribuir valores a dichas unidades, dependiendo de que esas dimensiones sean nominales, ordinales o de intervalo. Las propiedades más sencillas que encontramos en las unidades consideradas son los aspectos dicotómicos. Por ejemplo: la dimensión o variable sexo queda dividida en dos categorías: masculino, femenino. La dimensión edad puede clasificar ordinalmente a los sujetos en intervalos de grupos de edades, en tantas categorías como decidamos, o en dos en torno a un valor: mayor de, menor de; o categorías: 1) menos de 20, 2) entre 21 y 35, 3) 36 y más. Y la dimensión calificación obtenida en una prueba objetiva de informática también podemos dividirla en tantas categorías como queramos, al ser variables continuas, frente a la primera que era variable discreta.

Variables activas y variables atributivas. Las variables activas son las que se manipulan y son llamadas también variables experimentales. Las variables que se miden son llamadas variables atributivas.

Variables independiente y variables dependientes. La variable independiente es la supuesta causa de la variable dependiente, que sería su efecto. La variable dependiente es aquella que se predice, mientras que la variable independiente es aquella a partir de la cual se predice. La variable dependiente no se manipula, es observada o medida para comprobar los efectos de la variable independiente. Las variables independientes (V.I.) son aquellas que se introducen en una investigación para tratar de comprobar su posible incidencia en otra que se denomina dependiente (V.D.) y que se mide al final de un tiempo fijado para comprobar la posible influencia de la primera sobre la segunda.

Variables continuas y variables categóricas. Las variables continuas reflejan un conjunto ordenado de valores y son medidas mediante escalas de intervalo; si pueden adoptar infinitos valores fraccionados, son variables continuas. Las variables medidas al nivel de intervalo y de razón son variables continuas. Las variables medidas al nivel ordinal, que adoptan una gradación de menor a mayor, o al revés, son variables igualmente discretas, dado que sus distancias no son equidistantes, y por lo tanto no adaptan valores intermedios, sin embargo muestran una continuidad latente en la variable, y a veces se las considera continuas. Las variables medidas al nivel nominal, mediante categorías, son variables discretas, así, si nos referimos a un grupo de personas y las estudiamos de acuerdo con su nacionalidad, las clasificamos en función de categorías (española, francesa, inglesa), o podemos pretender estudiarlas en función de su nivel de estudios (primarios, secundarios, superior).

Variables nominales, ordinales, de intervalo y de razón. Dependiendo de la escala que formen o nivel de medición, o la manera de asignar códigos numéricos a las características que se observan en las personas u objetos, las variables se clasifican en nominales, ordinales, de intervalo y de razón. El nivel nominal es el más simple en cuanto a la asignación de números a las variables observadas, permitiendo clasificar los individuos y objetos en clases o categorías descriptivas que deben ser exhaustivas en sí mismas y excluyentes. El nivel ordinal además de clasificar ordena en función de que posea en mayor o menor grado la característica que se pretende medir. El nivel de intervalo, clasifica, ordena, y además especifica las distancias existentes entre las distintas categorías. El nivel de razón o de proporción, incluye además, el 0 absoluto, lo que representa ausencia total de la característica que se mide.

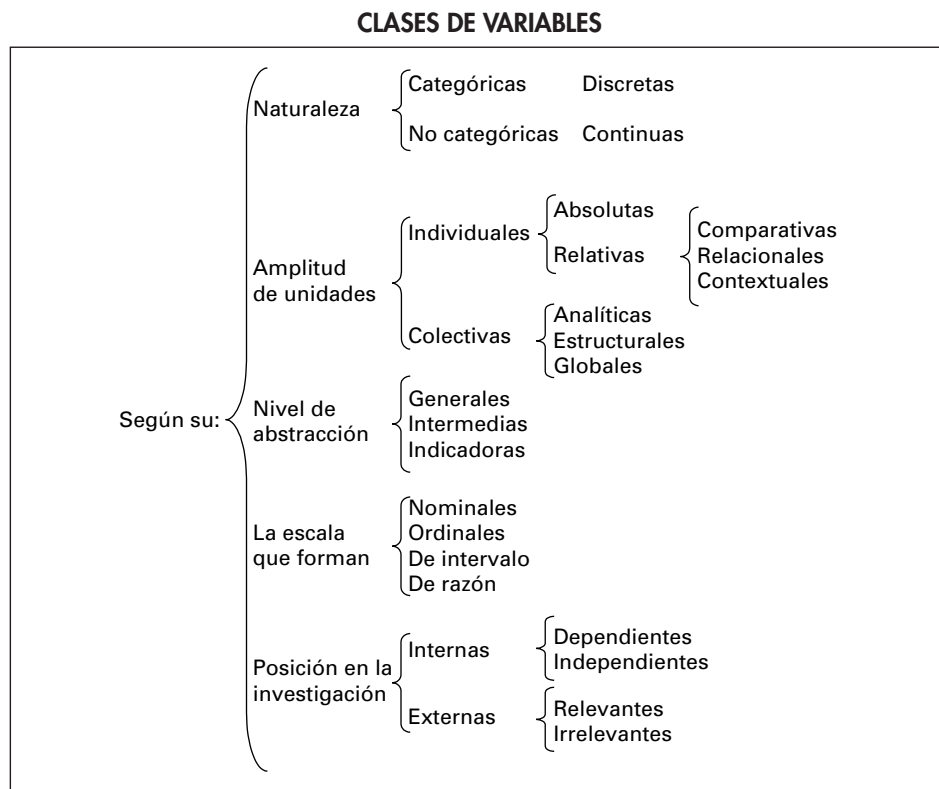


FIGURA 6.3. Clases de variables (tomado de Sierra Bravo, 1995: 106).

6.3.3. Los indicadores

Un indicador es aquello que da señal o cuenta de algo concretándolo. Se entiende por indicador la medida estadística de un concepto o de la dimensión de un concepto o de una parte de aquélla, basado en un análisis teórico previo e integrado en un sistema coherente de medidas semejantes que sirva para describir aquello que estamos estudiando. Los indicadores nos permiten medir conceptos, cuantificar dimensiones ideativas, convertirlas en utilizables a efectos de manipulación estadística, compararlas con otras dimensiones y explicarlas.

Los indicadores son definiciones operacionales, componentes de un modelo, medidas de fenómenos, expresiones numéricas o datos informativos de hechos sociales o concreciones estadísticas. Cuando hablamos de indicadores nos estamos refiriendo a la utilización científica de datos observables con un fin aplicado (González Blasco, 1994: 283).

Los indicadores deben contener dos características esenciales: estar relacionados con el concepto o dimensión que pretenden representar, y ser expresión numérica, cuantitativa, de la dimensión que reflejan. La selección de indicadores para cada dimensión debe ser tal que garantice su relación con la variable estudiada. La adecuación de los indicadores a la variable viene determinada por estudios de relación entre los indicadores y la variable, así como mediante la opinión de expertos que han alcanzado pericia en la elaboración de indicadores y en los temas, dimensiones y variables que estudiamos.

La agrupación de indicadores suele adoptar la forma de inventarios, cuestionarios, hojas de observación, guión de entrevista. La concreción y operativización del concepto hacen posible su medición y manipulación numérica, permitiendo analizarlo relacionarlo y compararlo con otras dimensiones del mismo y de otros conceptos, lo que nos permitirá realizar explicaciones de las realidades sociales que estudiamos.

6.3.4. *Los índices*

Un índice es una medida obtenida por la agrupación de varios indicadores. Los indicadores aportan la información numérica que hace posible el cálculo de los índices. Los índices son los representantes numéricos de una, varias o todas las dimensiones del concepto estudiado operativizado.

La construcción de índices plantea diversas cuestiones a las cuales, siguiendo a González Blasco (1994: 291) se trata de responder del modo siguiente:

- ¿Podemos incluir en un mismo índice uno o varios indicadores que reaccionan de forma distinta a los demás? La solución es considerar de forma diferente cada uno de los indicadores que integran el índice, suponiendo que los indicadores tienen distinta «potencia relativa» y compensando esa diferente potencia por medios de unos «pesos» adjudicados a cada uno de los indicadores.
- Si una variable la dividimos en cuatro dimensiones, o factores, cada uno representado por dos indicadores. Las correlaciones entre las respuestas a cada uno de los indicadores, ocho en total, resultarán altas, y resultarán bajas las relaciones entre estos indicadores y los resultantes de otros factores. Pero esto no siempre sucede al encontrarnos que un indicador es compartido por dos o

más factores. Las técnicas de análisis factorial nos permitirán detectar y reducir los indicadores o ítems compartidos por varios factores.

- Si utilizamos como indicadores de varias dimensiones las respuestas que pueden obtenerse a partir de los ítems, podemos preguntarnos ¿Cuántos ítems debemos utilizar? La respuesta es que al construir índices relativos a conceptos educativos, psicológicos o sociológicos complejos se elige un número de ítems relativamente limitado dentro del conjunto de ítems posibles sugerido por el concepto y su representación literaria en forma de pregunta. Tales índices poseen un rasgo fundamental: su correlación con variables exteriores suele mantenerse sensiblemente estable, cualquiera que sea la muestra de ítems escogida. Este fenómeno, a primera vista sorprendente, recibe el nombre de «intercambiabilidad de los índices».

Las características que debe reunir un índice son:

- Validez o exactitud: grado en que refleja lo que dice medir. Deberá ser único, tener límites definidos de variabilidad, referirse sólo a los indicadores que los forman y reflejar las variaciones que se den éstos, siempre que dichas variaciones sean significativas (estabilidad del índice).
- Potencia: o capacidad para aportar información sobre el concepto que mide, y su grado de normalización que permite su comparabilidad.
- Reproducibilidad: o cualidad de representar el concepto que mide, pudiendo reconstruirse el proceso a partir de la comprensión del propio índice.

Los índices son diversos y relacionados con las características de cada proyecto de investigación y proceso de realización del mismo.

Por ejemplo, para elaborar un índice de satisfacción ante los cambios producidos en la organización en un centro educativo, podemos formular las siguientes preguntas e indicar los distintos tipos de respuesta:

¿Está usted contento con los cambios realizados en la organización del centro?

Sí	No
1 ↓	

¿Cuál es su opinión sobre la introducción de una sesión semanal para el debate sobre el funcionamiento del centro? Indique su grado de acuerdo:

Muy positiva	
Positiva	
Negativa	
Muy negativa	

Para obtener el valor del índice se multiplicaría el peso por el porcentaje y se divide entre 100:

	Pesos	Porcentaje
Muy positiva	4	25
Positiva	3	25
Negativa	2	25
Muy negativa	1	25
Totales		100

$$I_p = \frac{A\% \times 4 + B\% \times 3 + C\% \times 2 + D\% \times 1}{100}$$

6.4. Los niveles de medición: escalas nominal, ordinal, de intervalo y de razón

Con el término datos, queremos referirnos a un conjunto de informaciones expresadas de forma numérica. Normalmente los datos se muestran mediante tablas de modo que una columna suele representar los valores o puntuaciones directas obtenidas por un sujeto en una determinada variable, entendiendo por variable una propiedad o característica que varía de un sujeto a otro. Los valores de la variable pueden medirse, dependiendo de su naturaleza, con escalas de medición diferentes, entre las que se encuentran la escala de medida nominal, la ordinal, la de intervalo y la de razón.

6.4.1. Nivel de medida nominal

Al estudiar los objetos, las personas, o determinadas características, se pueden establecer categorías nominales de dicha variable, y de ese modo situar a cada sujeto en una determinada categoría de la variable entre

varias y contar el número de casos que se asignan, o entran, pudiendo operar estadísticamente con los valores que entran en cada categoría. Lo que nos interesa conocer con este tipo de escala o nivel de medida es la pertenencia o no de lo estudiado a una determinada categoría de las establecidas. Las diferentes categorías deben ser mutuamente excluyentes, y en su conjunto exhaustivas para los objetos considerados. La utilidad de estas categorizaciones nominales de las variables es que permiten realizar relaciones, conexiones, asociaciones, entre distintas categorías.

Son ejemplos de variables categóricas: El género de las personas: 1) masculino 2) femenino; La estatura de jóvenes españoles estableciendo una determinada clasificación como por ejemplo: 1) alta (mayor de 1,80 metros), 2) media (entre 1,65 y 1,80), y 3) baja (menor de 1,65); La referida a clase social; Los ingresos económicos; La nacionalidad de la Comunidad Europea: Británica, Francesa, Española, Italiana, etc.; Las religiones.

A las categorías de las variables podemos asignarles números, pero éstos son irrelevantes para operar, dado que no tiene sentido sumar, restar, multiplicar o dividir la pertenencia a categorías como nacionalidades, nivel de estudios obtenido, aficiones a realizar en el tiempo libre. Si, por ejemplo, queremos conocer la nacionalidad de los turistas que están hospedados en un hotel de la costa en la cuarta semana del mes de Julio, los resultados de dicho estudio podemos mostrarlos mediante una tabla de frecuencias, del siguiente modo:

TABLA 6.1.

Nacionalidad	Frecuencia	Porcentaje
1. Española	278	27,8
2. Británica	96	9,6
3. Francesa	88	8,8
4. Alemana	87	8,7
5. Italiana	73	7,3
6. Países Bajos	68	6,8
7. Escandinava	55	5,5
8. Iraní	34	3,4
9. Saudita	52	5,2
10. Otras	167	16,7
Total	1000	100

Veamos otro ejemplo, acerca de las preferencias literarias de los jubilados, a partir de una encuesta realizada en las bibliotecas de los Centros de la tercera edad de la Comunidad de Castilla y León. Los resultados se muestran en la tabla de frecuencias siguiente:

TABLA 6.2.

Preferencias literarias	Libros leídos	Porcentajes	Proporciones
1. Novela costumbrista	85	28,33	0,28
2. Novela histórica	73	24,33	0,24
3. Novela romántica	42	14	0,14
4. Novela de ciencia-ficción	11	3,66	0,036
5. Teatro	34	11,33	0,113
6. Poesía	26	8,66	0,086
7. Otros	29	9,66	0,096
Total	300	100	1

A partir del recuento de frecuencias podemos calcular porcentajes y proporciones, como queda reflejado en la tabla, para comparar las frecuencias relativas registradas en cada categoría y hacer comparaciones de porcentajes y proporciones entre las distintas categorías en las que hemos clasificado las preferencias literarias de la muestra estudiada.

La reducción de los datos de esta escala nominal se basa en el recuento de frecuencias, que puede dar lugar a porcentajes de presencia en las diferentes categorías, aportando información descriptiva sobre la extensión de aparición de casos en cada categoría y su relación relativa con la totalidad de categorías que se estudia. Los estudios de inferencia hacia la población a partir de los datos obtenidos en esta muestra, con variable medidas a nivel nominal, se apoyan en la utilización de pruebas estadísticas de asociación como el coeficiente de contingencia, basado en la prueba de «ji» cuadrado (χ^2) que pretende probar si las diferencias de frecuencias registradas en cada categoría son o no estadísticamente significativas al nivel de significación fijado por el investigador. La distribución de la variable es discreta.

6.4.2. Nivel de medida ordinal

El nivel de medida ordinal clasifica los objetos, o las personas, de un modo ordenado de acuerdo con una característica o cualidad determi-

nada, de modo que este orden tiene un significado de posición relativa de los elementos con la totalidad. Sitúa a los objetos, o a las personas de un grupo en un orden de acuerdo con un criterio determinado.

Como ejemplo veamos el referido al hecho de un profesor que observa el interés mostrado por sus alumnos en una determinada tarea, mide dicho interés por diversas acciones expresadas mediante 10 indicadores o ítems, valorando cada uno de los ítems por una escala que va de ningún interés a poco, medio, bastante y mucho, y que traducido a valores numéricos equivalen a 1, 2, 3, 4 y 5. Tras la suma de las puntuaciones señaladas en los ítems, se obtienen unos valores que van desde el máximo posible de cada ítem, 5, ($10 \times 5 = 50$), al mínimo 1 ($10 \times 1 = 10$), que expresan el interés observado por el profesor en sus alumnos hacia la tarea objeto de estudio. El conjunto total de la suma de las valoraciones individuales vendría expresado en una columna de datos que ordenaría a los sujetos en función de su mayor o menor interés en la tarea. El nivel o escala de medida ordinal nos permite establecer relaciones entre los individuos del grupo, pudiendo decir que uno está por delante de otro en una determinada variable, sin poder asegurar que nuestra valoración, basada en una apreciación y observación subjetiva, haya establecido una ordenación precisa de los alumnos.

TABLA 6.3.

Interés observado en los alumnos	
Sujetos	Puntuación
1	43
2	39
3	37
4	30
5	29
6	27
7	25
8	22
9	18
10	17
11	17
12	16
Total	320

Las distancias entre un valor, su anterior y su siguiente no son equidistantes por lo que no se puede realizar todo tipo de operaciones con estos valores numéricos, como por ejemplo, hallar la media aritmética. Si se puede hallar la mediana. La variable, aunque discreta, muestra continuidad, por lo que se la considera variable latente continua.

Si quisiéramos buscar el grado de relación entre dos variables que ha sido medidas con una escala que alcanza el nivel ordinal, o una de ellas medida a escala o nivel ordinal y otra en escala de intervalo, tendríamos que utilizar un coeficiente de correlación ordinal, por ejemplo el de Spearman. La información inferencial que puede aportarnos se apoya en la utilización de pruebas estadísticas para la prueba de significación de hipótesis que utilizan el nivel de medida ordinal, pudiendo realizarse comparaciones entre grupos que reciben un tratamiento específico para desarrollar el interés hacia determinada tarea con otros grupos que no reciben dicho tratamiento, y dependiendo de que cumpla otros supuestos exigidos para la utilización de determinadas pruebas estadísticas para el contraste de hipótesis, se utilizará una prueba estadística u otra como veremos en el tema referido a los diseños de investigación experimental.

6.4.3. Niveles de medida de intervalo y de razón

El nivel de medida de intervalo viene medido por aquellas escalas que toman una unidad o valor constante para medir la distancia o posiciones de los distintos objetos medidos con esa unidad. Por ejemplo, el metro, y sus subdivisiones en centímetros. Si medimos la distancia que alcanza un grupo de deportistas en saltos de longitud, la unidad de medida metro —y centímetros— nos permite situar a todos los sujetos a lo largo de un continuo, siendo los intervalos representados por esa unidad equidistantes. Por lo tanto, la escala de intervalo consiste en asignar un número a un sujeto para indicar su posición exacta a lo largo de una escala continua. Si queremos medir la estatura de un grupo de alumnos podremos obtener tanta precisión como nos lo permita el instrumento de medida: metros, centímetros y hasta milímetros, sabiendo que la distancia entre un intervalo y el anterior, y entre ese mismo intervalo y el siguiente, es la misma pudiendo decir que un alumno es más alto que otro con total precisión, indicando la distancia exacta entre ambos.

Los valores numéricos obtenidos por la medición de variables que permiten el nivel de medida de intervalo hacen posible las operaciones aritméticas, permitiendo hallar medias aritméticas y varianzas. De modo que si quisiéramos comprobar el grado de relación existente entre dos

variables medidas con una escala de intervalo, podremos utilizar el coeficiente de relación de Pearson.

La información inferencial que puede aportarnos se apoya en la utilización de pruebas estadísticas para el contraste de hipótesis que utilizan el nivel de medida de intervalo, pudiendo realizarse comparaciones entre grupos que reciben un tratamiento específico para mejorar su tiempo de dedicación a la realización de una determinada tarea, o para mejorar el rendimiento de una asignatura, por ejemplo Matemáticas, medida mediante un instrumento preciso, que asegure en la medida de lo posible, la equidistancia entre una puntuación y otra. Dependiendo de que cumpla otros supuestos exigidos para la utilización de determinadas pruebas estadísticas para el contraste de hipótesis que se basan en la media aritmética y la varianza, podremos utilizar una u otra, tal como veremos en el tema referido a los diseños de investigación experimental.

Si a una escala de intervalo le imponemos la condición de que tengan un punto de origen conocido, estaremos trabajando con una escala de razón. La utilización de un origen y una unidad de medida predeterminados, permiten hallar proporciones y comparaciones en forma de porcentajes.

6.5. Relación entre los indicadores, los niveles de medida, los estadísticos y las pruebas estadísticas

Los valores obtenidos por un grupo de sujetos medidos en cualquiera de estas escalas: nominal o categorial, ordinal, de intervalo y de razón, pueden presentarse en tablas en las que se muestren todas las puntuaciones directas obtenidas, o pueden agruparse por frecuencias cuando un mismo valor ha sido obtenido por varios sujetos; también pueden representarse gráficamente mediante un histograma, diagrama circulares o cualquier otro modo. Estas representaciones gráficas muestran visualmente la distribución de frecuencias de un grupo en una o varias variables. La escala de medición utilizada determinará la elección de las pruebas estadísticas para el análisis de los datos.

De la coherencia en las operaciones realizadas mediante determinados estadísticos o índices y pruebas estadísticas, y la escala de medida que caracteriza a cada indicador, como elemento representante de la variable —concepto o constructo teórico—, que queremos medir, dependerá la calidad de nuestro trabajo y la posibilidad de que podamos concluir y aportar alguna solución sobre aquello que estudiamos.

Ya hemos mencionado que para conocer la media aritmética de una serie de datos requiere que estos alcancen el nivel de medida de intervalo. La mediana exige al menos un nivel ordinal, y la moda un nivel nominal. Por otro lado, el cálculo de la mediana con datos que alcanzan el nivel de intervalo supone una pérdida de precisión, que la aporta la media. Así, a través de un indicador podremos hallar índices con pruebas que exigen un nivel de medida de intervalo si los datos se han recogido mediante escalas que consiguen el nivel de intervalo, o hallaremos índices que exigen un nivel de medida ordinal, si no alcanza el nivel de intervalo. Dependiendo del nivel de medida alcanzado en la variable, y de los supuestos de la muestra, podremos aplicar determinados estadísticos y pruebas para realizar estudios grupales y de inferencia. Estos contenidos serán objeto de su estudio en la Tercera Unidad de esta materia.

6.6. Ejercicios de autoevaluación

6.6.1. *Conceptos fundamentales del tema*

1. Defina los siguientes conceptos: Medida, Variable, Indicador.
2. Notas definitorias de los distintos niveles de medida.

6.6.2. *Cuestiones de síntesis*

3. Elabore una síntesis sobre los contenidos del tema.

6.6.3. Cuestiones sobre los contenidos del tema

4. Clasifique las siguientes variables según su nivel de medición, y si son discretas o continuas:

Variables	Nominal	Ordinal	Intervalo	Discreta	Continua
a) Nacionalidad					
b) Interés en las actividades culturales					
c) Rendimiento en matemáticas					
d) Interés por las materias científicas					
e) Atención en clase					
f) Titulación alcanzada en sus estudios					
g) Edad					
h) Tiempo tardado en realizar a un test					
i) Estado civil					
j) Primer salario licenciados de la UNED					
k) Estatura de los alumnos de una clase					
l) Capacidad de concentración en el estudio					

6.7. Soluciones a los ejercicios de autoevaluación

Consulte para las preguntas de contenido teórico los diferentes epígrafes del tema.

1. Clasifique las siguientes variables según su nivel de medición, y si son discretas o continuas:

Variables	Nominal	Ordinal	Intervalo	Discreta	Continua
a) Nacionalidad	x			x	
b) Interés en las actividades culturales		x		x	
c) Rendimiento en matemáticas			x		x
d) Interés por las materias científicas		x		x	
e) Atención en clase		x		x	
f) Titulación alcanzada en sus estudios		x		x	
g) Edad			x		x
h) Tiempo tardado en realizar a un test			x		x
i) Estado civil	x			x	
j) Primer salario licenciados de la UNED			x		x
k) Estatura de los alumnos de una clase			x		x
l) Capacidad de concentración en el estudio		x		x	

Las variables ordinales las hemos clasificados como discretas, pero ya hemos dicho que mantienen una continuidad de la variable, aunque no admitan operaciones matemáticas como las variables medidas a un nivel de intervalo.

6.8. Referencias bibliográficas

- BUNGE, M. (1976). *La investigación científica*. Barcelona: Ariel.
- GONZÁLEZ BLASCO, P. (1994). «Medir en Ciencias Sociales». En Manuel García Ferrando, Jesús Ibáñez y Francisco Alvira (Comp.). *El análisis de la realidad social. Métodos y técnicas de investigación*. Madrid: Alianza Universidad (2.^a ed.).
- PÉREZ JUSTE, R. (1984). *Pedagogía Experimental I*. Madrid: UNED.
- SIERRA BRAVO, R. (1995). *Técnicas de investigación social. Teoría y ejercicios*. Madrid: Paraninfo (10.^a ed.).
- TORGERSON, W. S. (1958). *Theory and methods of scaling*. New York: John Wiley and Sons.

TEMA 7

Las escalas para la medida de las actitudes. Las escalas de Thurstone

ESQUEMA

- 7.1.** Las escalas para la medida de las actitudes
 - 7.1.1. Las «gradaciones ordinales»
 - 7.1.2. Escalas de rango
- 7.2.** Las escalas de medida de Thurstone
 - 7.2.1. La ley del juicio comparativo
 - 7.2.2. La ley del juicio categórico
- 7.3.** Construcción de una escala de intervalos de Thurstone
 - 7.3.1. Definición del rasgo a medir
 - 7.3.2. Elaboración de los elementos iniciales
 - 7.3.3. Clasificación de las proposiciones. Prueba de los jueces
 - 7.3.4. Criterios para la selección de los elementos.
 - 7.3.4.1. El valor escalar del elemento
 - 7.3.4.2. La ambigüedad del elemento
 - 7.3.4.3. La irrelevancia del elemento
 - 7.3.5. Elaboración de la escala
 - 7.3.6. Calificación de la escala
- 7.4.** Fiabilidad y validez de la escala
- 7.5.** Ejercicios de autoevaluación
- 7.6.** Soluciones a los ejercicios de autoevaluación
- 7.7.** Referencias bibliográficas

OBJETIVOS

Se espera que el alumno, tras el estudio de este tema, sea capaz de comprender y valorar:

- ✓ El concepto de escala de medida
- ✓ Los principales métodos de escalamiento
- ✓ Los fundamentos de los modelos de Thurstone para la elaboración de escalas
- ✓ Los procedimientos para la elaboración de la escala de intervalos aparentemente iguales de Thurstone

7.1. Las escalas para la medida de las actitudes

Una escala de medida es un instrumento cuyo objetivo es realizar mediciones de las variables que pretende estudiar. Un ejemplo sencillo de escala es una regla, o un metro, que permite medir la longitud de los objetos. Los objetos de medición propios de las ciencias sociales son conceptos abstractos, llamados constructos, que se manifiestan a través de las conductas o comportamientos humanos, de naturaleza psicológica, sociológica o educativa, como las actitudes, los intereses, los valores, las opiniones.

El desarrollo de las escalas para la medición de actitudes tiene numerosas aplicaciones en el campo de la pedagogía, la psicología y la sociología, dado que suponen un instrumento fundamental para la medición de las actitudes. Las escalas para la medida de las actitudes nos van a permitir representar las propiedades de dichas variables o construcciones hipotéticas, mediante números, siguiendo unas normas para su elaboración.

Los métodos de escalamiento para la medida de las actitudes psicológicas fueron desarrollados por Thurstone, durante la década de 1920, a partir de los trabajos de Fechner sobre la medición de las respuestas psicofísicas provocadas por estímulos físicos. Thurstone desarrolló un modelo matemático para la medida de las actitudes, intereses, valores u opiniones, conocido con el nombre de *ley del Juicio Comparativo*, a partir del cual dedujo la *ley del Juicio Categórico*.

En la década de los 30, Likert desarrolló un método para elaborar escalas unidimensionales para la medida de las actitudes, más sencillo que el de Thurstone. Gutman, en la década de los 40, desarrolló su método llamado Escalograma, para escalas unidimensionales y multidimensionales. En la década de los 50 Osgood desarrolló el Diferencial Semántico, que en realidad es una ampliación de las escalas de Likert. En la década de los 60 se desarrollaron los métodos de escalamiento multidimensional.

Estos modelos se caracterizan por una serie de reglas que nos permiten elaborar las escalas con el fin de asignar números a las propiedades de los objetos o rasgos a medir. Los supuestos básicos que subyacen en los métodos de escalamiento son:

1. La existencia de un continuo latente a lo largo del cual varía el constructo hipotético objeto de la medición, que no se puede observar directamente sino a través de sus manifestaciones conductuales.

2. La posición, o situación ordenada con la que se pueden representar las manifestaciones de los sujetos sobre el constructo a lo largo de dicho continuo.

En las escalas se presentan al sujeto una serie de estímulos con el fin de que elicite, o emita, una respuesta a partir de dicho estímulo. El estímulo puede ser un objeto, una imagen, unas expresiones verbales tales como adjetivos calificando a un rasgo psicológico, a una actitud, etc. Las respuestas emitidas por cada sujeto a los estímulos nos posibilita diferenciar las opiniones de unos y otros y ordenar sus respuestas, o escalarlas, lo que nos permitirá realizar comparaciones entre los sujetos, y entre grupos de sujetos, e interpretar sus diferencias con respecto al atributo o variable medida.

Dependiendo de las características de las escalas resultantes, tendremos escalas nominales, escalas ordinales, escalas de intervalo y escalas de razón, que permiten obtener mediciones de nivel nominal, ordinal, de intervalo o de razón.

7.1.1. Las «gradaciones ordinales»

Las «gradaciones ordinales» son escalas o instrumentos de medida que se basan en la idea de la clasificación, a partir de las propiedades semánticas de las palabras y las características de los números. Algunas escalas que utilizan este sistema de gradación ordinal son el «Diferencial Semántico», las «Escalas de Rango» y las de «Distancias sociales». Las gradaciones ordinales son medidas de «autoposicionamiento» y su finalidad es ordenar las respuestas dadas por los sujetos para otorgarles datos numéricos. Se suele hacer mediante la indicación de una serie de categorías en las cuales se clasifica la respuesta, o mediante la emisión de la respuesta a la pregunta realizada en una línea graduada en la que se han señalado los extremos opuestos posibles de respuestas.

7.1.1.1. Gradiente en categorías ordinales

Mediante el gradiente en categorías ordinales se enuncia una pregunta para medir un atributo y se acompaña de una serie de categorías graduales de posibles respuestas cerradas. Veamos un ejemplo:

¿Considera que el desarrollo de los contenidos del tema es coherente con los objetivos previamente definidos? Expresa su respuesta entre 5: Muy adecuados, y 1: Muy poco adecuados.

CATEGORIZACIÓN DE TODAS LAS POSIBLES RESPUESTAS	— Muy adecuados	5
	— Adecuados	4
	— Normal	3
	— Poco adecuados	2
	— Muy poco adecuados	1

7.1.1.2. Gradiente en un continuo ordinal limitado

Con esta modalidad se pretende ofrecer a quien responde la pregunta una mayor posibilidad o gama de respuestas. Utiliza un continuo ordinal limitado, representado en una línea, en cuyos extremos se indican las respuestas extremas, señalándose, en la línea, distintas graduaciones posibles de la respuesta, mediante la indicación de intervalos continuos iguales.

Veamos el siguiente ejemplo:

¿El tema desarrollado por el conferenciante ha satisfecho sus expectativas?:

AUTOPOSICIONAMIENTO DE LA RESPUESTA EN EL CONTINUO	1	2	3	4	5	6	7	8	9	10	11
	En absoluto							Totalmente			

7.1.1.3. Escalas de rango

Es una técnica utilizada en los instrumentos de recogida de datos para que los sujetos, a partir de las preguntas que se les hace, ordenen sus respuestas de manera creciente o decreciente, en función de sus preferencias y valoraciones.

Veamos el ejemplo:

¿Cuál es la carrera que quiere cursar? (Ordénalas según sus preferencias):

TABLA 10.1

Estímulos:	Orden de preferencia
Administración de Empresas	
Ciencias Físicas	
Ciencias Matemáticas	
Ciencias Químicas	
Derecho	
Económicas	
Educación	
Filologías	
Filosofía	
Geografía e Historia	
Ingeniería	

A partir de las escalas de rango se pueden obtener:

- *Índices de prestigio*: El mayor porcentaje que obtiene aquello por lo que preguntamos (idioma que preferiría estudiar, profesión que le gustaría ejercer, ciudad en la que le gustaría vivir...).
- *Índices de preferencias*: Los dos porcentajes superiores.
- *Jerarquizaciones*: La lista de los 10 libros más vendidos.

Veamos el siguiente ejemplo:

Después de preguntados un grupo de 239 estudiantes de secundaria sobre la profesión que les gustaría realizar, se calculan los siguientes índices:

TABLA 10.2

Actividad	Índice de prestigio (%)	Jerarquización	Índice de preferencia
Cantante	37	1.º	49
Actor	26	2.º	45
Músico	22	3.º	43
Futbolista	18	4.º	40
Economista	15	5.º	36
Abogado	14	6.º	28
Ingeniero	12	7.º	27
Informático	11	8.º	22
Médico	9	9.º	16
Profesor	8	10.º	17
Total	100		239

El modelo se basa en la percepción personal de los sujetos hacia los estímulos que se les presentan. Estas escalas también reciben el nombre de escalas de juicio dado que los sujetos emiten sus juicios, o valoraciones. Cuando las diferencias entre las diferentes opciones son claras, los sujetos emitirán sus juicios con facilidad. Si los estímulos no ofrecen claras diferencias, la emisión de la respuesta se hará más difícil para el sujeto.

7.2. Las escalas de medida de Thurstone

El concepto de actitud, de acuerdo con Thurstone, se define como el conjunto de inclinaciones y sentimientos, prejuicios o tendencias, temores o convicciones de una persona respecto de un tema determinado. El objetivo de la construcción de una escala para la medida de las actitudes, siguiendo el modelo de Thurstone, es conseguir el escalamiento de los estímulos representativos del tipo de variables que queremos estudiar a lo largo de un continuo. Una vez construida la escala el siguiente paso será aplicarla a los sujetos para diferenciarlos respecto a la variable estudiada, y por lo tanto, servirá para el escalar a los sujetos en función de las respuestas dadas a los estímulos presentados en la prueba o escala.

De entre los diversos métodos desarrollados por Thurstone para la elaboración de sus escalas, el más utilizado es el de los intervalos aparentemente iguales, en el que el continuo de la variable lo divide en 11 categorías. La variable es medida de forma directa, al ser el sujeto consciente de la actitud, o el rasgo que se mide, pudiendo estar sus respuestas influenciadas por el factor de deseabilidad social.

El modelo de Thurstone se basa en la variabilidad de percepción de los sujetos a los que se le presentan los estímulos para que emitan su valoración con respecto a un atributo, asignándole un valor que puede ordenarse a lo largo de un continuo. Cada estímulo originará en el sujeto un proceso discriminante responsable de la valoración que otorga. Estos mismos estímulos, presentados al mismo sujeto en diferentes ocasiones pueden suscitarle valoraciones diferentes, que se distribuirán a lo largo de un continuo.

La elaboración de la escala consiste en un sistema de escalamiento centrado en los estímulos, sobre los cuales los sujetos actúan valorando dichos estímulos, ejerciendo de jueces de dichos estímulos. En este primer paso de construcción de la escala no proyectan su actitud, sino que valoran el grado en que dichos estímulos, concretados en enunciados o proposiciones, reflejan la actitud que se pretende medir. Una vez elaborada la escala definitiva, sí servirá para medir las actitudes en los sujetos y hacer estudios comparativos entre sujetos y grupos en cuanto al grado en que poseen la actitud medida. Si las preferencias podían expresarse de acuerdo con un orden o gradación continua, a ésta podría hacersele corresponder números. Cuando un estímulo es valorado muchas veces su distribución toma la forma de la curva normal. Con esta premisa Thurstone consideró que mediante un sistema de escalas de este tipo, se podría medir cualquier manifestación actitudinal. En lugar de utilizar un solo sujeto para valorar un estímulo o varios estímulos particulares en sucesivas situaciones se pueden tomar muchos sujetos para valorar un estímulo o varios, pudiendo hacerse dicha valoración no sobre los estímulos concretos sino mediante proposiciones representativas de dichos estímulos.

Todas las posibles respuestas emitidas por un sujeto a un estímulo conformarán una distribución de frecuencias a la que Thurstone llamó distribución discriminante, que cuando el número de casos es grande se distribuye siguiendo el modelo de la curva normal. Esta distribución viene definida, igual que la de la curva normal, por dos parámetros, la media μ , que coincide con la mediana y la moda, y la desviación típica σ , que es un índice de heterogeneidad, o de ambigüedad en la respuesta que

el estímulo despierta en el sujeto (Barbero, 1999: 130 y ss). Cuando el estímulo no suscite ninguna ambigüedad en la respuesta dada por el sujeto, la desviación típica será cero. Las respuestas emitidas por un mismo sujeto a varios estímulos producirán diversas distribuciones discriminativas, tantas como estímulos presentados sobre los cuales el sujeto emite sus respuestas. A las medias de cada una de las distribuciones se las denomina valores escalares del estímulo, y a su correspondiente desviación típica se la denominada dispersión discriminante.

Thurstone elaboró dos modelos basados en una serie de ecuaciones matemáticas para la construcción de escalas para la medida de las actitudes. Al primer modelo lo denominó «Ley del Juicio Comparativo», y al segundo modelo, más simplificado que el primero, lo llamó «Ley del Juicio Categórico».

7.2.1. *La Ley del Juicio Comparativo de Thurstone*

La ley del Juicio Comparativo se basa en un sistema de ecuaciones matemáticas que relacionan la proporción de veces que un estímulo K es juzgado mayor que otro estímulo J respecto a un atributo determinado, con los valores escalares (mediana de la distribución de frecuencias de respuesta) y las dispersiones discriminantes (desviaciones típicas de la distribución) de los dos estímulos sobre un continuo psicológico (Torgerson, 1958:159, en Barbero, 1999:133 y ss.). Este modelo se basa en los postulados siguientes:

1. Un estímulo presentado a un sujeto para que emita un juicio sobre algún atributo del mismo, produce en el sujeto un proceso discriminante que le llevará a emitir una valoración subjetiva.
2. Repetido el mismo proceso en ocasiones sucesivas las respuestas se situarán a lo largo de un continuo y su representación adoptará la forma de una distribución normal, denominada por Thurstone distribución discriminante.
3. A la media de dicha distribución la denomina valor escalar del estímulo, a la desviación típica dispersión discriminante.

Si a un mismo sujeto se le presentan dos estímulos en diversas ocasiones cada uno originará una distribución. Si los dos estímulos se presentan a la vez preguntándole cuál es dominante en la dirección del atributo, se producirá una diferencia en la valoración denominada diferencia discriminante. Si se le pregunta esto mismo en situaciones sucesivas producirá una distribución de frecuencias que adopta la forma de la cur-

va normal, cuya media será la media de la diferencia entre los valores escalares de los estímulos.

$$\bar{X}_{1-2} = \bar{X}_1 - \bar{X}_2$$

Y la desviación típica de la distribución de diferencias discriminantes será:

$$s_{1-2} = \sqrt{\left[(X_1 - X_2) - (\bar{X}_1 - \bar{X}_2)\right]^2}$$

$$s_{1-2} = \sqrt{(X_1 - \bar{X}_1)^2 + (X_2 - \bar{X}_2)^2 - 2(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)} = \sqrt{s_1^2 + s_2^2 - 2r_{12}s_1s_2}$$

Siendo los dos primeros términos de la ecuación resultante, las varianzas respectivas de la distribución discriminante de cada uno de los estímulos, y el tercero la covarianza.

Cuando en la mayoría de las ocasiones predomina en la respuesta del sujeto el dominio de un estímulo «1» sobre el otro «2», ambas distribuciones, a lo largo de un mismo continuo, aparecerán una siguiendo a la otra. Si el dominio de un estímulo sobre el otro no se da en la mayoría de las ocasiones, ambas distribuciones compartirán parte de sus áreas. Y si ambos estímulos presentan dificultad para ser discriminados por el sujeto en cuanto a su dominio en el atributo que se juzga referido al estímulo, ambas distribuciones se superpondrán, coincidiendo su valor escalar (su media) y su dispersión discriminante (su desviación típica).

Cuando el número de veces que se han comparado los estímulos es grande, la distribución se ajusta a la curva normal, utilizándose la fórmula de las puntuaciones típicas para hallar la distancia entre los valores escalares.

$$z_{12} = \frac{(\bar{X}_1 - \bar{X}_2)}{s_{1-2}}$$

Siendo z_{12} la distancia entre los valores escalares de los estímulos 1 y 2, tomando como origen la media de la distribución de diferencias y como unidad de medida la desviación típica de la distribución de diferencias discriminantes. Despejando la diferencia de medias y sustituyendo el valor de la desviación típica de las diferencias, tenemos la ecuación de la ley del Juicio Comparativo:

$$(\bar{X}_1 - \bar{X}_2) = z_{12} \sqrt{s_1^2 + s_2^2 - 2r_{12}s_1s_2}$$

Donde:

s_1 : la dispersión discriminativa, o desviación típica del estímulo 1

s_2 : la dispersión discriminativa del estímulo 2

r_{12} : la correlación entre los valores asignados a los estímulos 1 y 2.

7.2.2. *La Ley del Juicio Categórico de Thurstone*

Lo mismo que la ley del Juicio Comparativo, la ley de Juicio Categórico tiene como finalidad la elaboración de las escalas, y para ello se le pide al sujeto, o sujetos que colaboran en la construcción de las mismas, que indique el grado en que un determinado estímulo presenta el atributo que se le muestra, no su actitud personal hacia el mismo, y se basa en un sistema de ecuaciones a partir de determinados supuestos como en la ley vista anteriormente:

1. Las valoraciones dadas por los sujetos a cada uno de los atributos forman un continuo subjetivo.
2. Cada uno de los atributos estudiados da un continuo subjetivo, formado por la valoraciones de los distintos sujetos.
3. Los estímulos presentados a los sujetos en forma de proposiciones producen un proceso discriminante responsable del valor que el sujeto les otorga, que varían de una situación a otra, distribuyéndose sus valores a lo largo de un continuo, llamado distribución discriminante, y cuando el número de veces es grande, adopta la forma de la curva normal.
4. La media de la distribución recibe el nombre de valor escalar del estímulo, y la desviación típica recibe el nombre de dispersión discriminante.
5. La presentación de varios estímulos a un solo sujeto para su valoración, aportará una distribución discriminante para cada estímulo.
6. La nueva ley del Juicio Categórico, añade un nuevo supuesto referido a que el continuo subjetivo de cada sujeto sobre el estímulo que se le presenta puede ser dividido en un número determinado de categorías ordenadas. (En Barbero, 1999: 180 y ss.).

El continuo subjetivo sobre el que cada sujeto puede responder al valorar un atributo, se presenta dividido en un determinado número de categorías ordenadas y, dependiendo de las situaciones y circunstancias, el sujeto señalará una categoría u otra. A partir de las distribuciones de

frecuencias de las valoraciones del sujeto se obtendrán los valores medios y sus desviaciones típicas, que nos aportarán la estimación de los límites de cada una de las categorías a lo largo del continuo.

La ecuación fundamental de la ley del Juicio Categórico es la siguiente:

$$\bar{X}_g - \bar{X}_j = z_{gj} \sqrt{s_g^2 + s_j^2 - 2r_{gj}s_g s_j}$$

Siendo:

j : 1, 2, 3, ...n : número de estímulos

g : 1, 2, 3, ...n : número de categorías

$\bar{X}_j$: el valor escalar del estímulo j

$\bar{X}_g$: la media de la distribución de los valores asignados por el sujeto al límite la categoría g

s_g : la desviación típica de la distribución de los valores asignados al límite la categoría g

s_j : la desviación típica de la distribución de los valores asignados al estímulo j

z_{gj} : la puntuación típica correspondiente a la proporción de veces que el estímulo j ha sido clasificado por debajo del límite de la categoría g

r_{gj} : la correlación entre los valores asignados a los estímulos 1 y 2 en las sucesivas presentaciones.

La ecuación contiene demasiadas incógnitas para hacer factible su resolución, por lo que se asume que la covarianza entre todos los valores se mantiene constante, y que las desviaciones típicas de las distribuciones de los valores asignados a los límites de las categorías son iguales. También se asume que todos los estímulos producen la misma ambigüedad en el sujeto, por lo que sus desviaciones típicas son iguales, y que las correlaciones son nulas, es decir, iguales a 0. La ecuación quedaría así:

$$\bar{X}_g - \bar{X}_j = z_{gj} K$$

Siendo K una constante.

7.3. Construcción de una escala de intervalos de Thurstone

Como venimos diciendo, la técnica de Thurstone para la construcción de escalas, se basa en estudiar y seleccionar una serie de proposiciones

relacionadas con una actitud determinada. Para elaborar la escala se siguen los siguientes pasos:

- 1.º Definición del rasgo, o actitud, que se quiere estudiar
- 2.º Preparación de los elementos iniciales:
 - a) Elaboración de las proposiciones,
 - b) Clasificación de las proposiciones, mediante jueces,
 - c) Selección de las proposiciones
- 3.º Elaboración de la escala piloto
 - d) Aplicación de la escala piloto
 - e) Valoración de las respuestas
- 4.º Escala para la medida de las actitudes lista para ser aplicada.

7.3.1. Definición del rasgo a medir

El primer problema al que se enfrenta el investigador es el de definir con claridad y precisión el rasgo, la variable o constructo que quiere medir. Dicha variable se considera como un continuo sobre el que pueden representar distintas gradaciones en su valoración con respecto a un atributo, desde las de valoración máxima a mínima pasando por neutra o indiferente.

Así, si lo que queremos estudiar es la actitud de los padres frente a la admisión de inmigrantes en el centro educativo de sus hijos, la actitud a medir se referirá al grado de aceptación de extranjeros como compañeros de aula de sus hijos, y las valoraciones que se incluirán irán desde máxima aceptación a máximo rechazo pasando por una actitud indiferente.

7.3.2. Elaboración de los elementos iniciales

Para la elaboración de los elementos que se incluirán en la escala, es conveniente contar con aproximadamente el doble de elementos que se pretende tenga la escala definitiva. Los enunciados deben ser comprensibles para una amplia variedad de personas, de acuerdo con la amplitud de sujetos a los que pretendemos medir y el conjunto de las proposiciones incluidas en la escala. Thurstone (1928, en Barbero, 1999: 228 y ss.) propone los siguientes criterios para la elaboración de los enunciados:

- Ser formulados con claridad para que la emisión de aceptación o rechazo por parte de los sujetos, de acuerdo con su actitud, no ofrezca dificultad.

- Ser representativos de la variable que se quiere medir.
- Ser redactados evitando ambivalencias.
- Ser breves para evitar la fatiga de los encuestados.
- Formular algunos enunciados de modo que los sujetos no los puedan rechazar o aceptar de forma rotunda.
- Formular los enunciados de modo que cualquier persona de la población meta pueda contestarlos.
- Evitar introducir términos como «siempre», «nunca», «todos» ya que puede inducir a respuestas que aportan poca información con respecto a la actitud a medir.

7.3.3. *Clasificación de las proposiciones: prueba de jueces*

Se le pide a un grupo de personas experimentadas en el tema en el que trabajamos que hagan de «jueces» de los elementos iniciales de la escala de medida, que clasifiquen cada una de las proposiciones en función del grado en que represente la actitud, y recojan dicha clasificación en una serie de casillas, que de acuerdo con las recomendaciones de Thurstone, son once, numerando el 1, el 6 y el 11.

Se pretende comprobar la validez de las proposiciones propuestas en relación con la actitud que se mide, por lo tanto, el criterio que deben seguir los jueces para clasificar las proposiciones es considerar si el enunciado de cada proposición es válido en expresar la actitud. Si la proposición expresa bien la actitud debe colocar el enunciado en la casilla 1, si es indiferente para la actitud en la 6 y si es contrario a la actitud en la 11, es decir, el 1 representaría el máximo acuerdo en la expresión de la actitud y el 11 el máximo desacuerdo, y en el resto de las casillas se colocarán en función de que esté mejor definida o representada la actitud entre 2 y 5, o peor definida o representada entre 10 y 7.

Se supone que los «jueces» clasifican las distintas proposiciones o enunciados a partir de una división subjetiva del continuo de la actitud en 11 intervalos aparentemente iguales, y de amplitud, cada intervalo, igual a la unidad. Los límites de los intervalos van desde 0,5-1,5, para señalarlo cuando se considera que dicho elemento mide en muy poco grado la actitud que se quiere medir, hasta el 10,5-11,5, que será señalado cuando un juez considere que dicho elemento representa en gran medida la actitud que se quiere medir. Los límites 5,5-6,5 representan la neutralidad del elemento para medir la actitud pretendida.

Se trata de valorar los enunciados en función del grado en que dicho elemento represente la actitud que se pretende medir, con la máxima objetividad por parte de los jueces, no se trata de una valoración la actitud por parte del juez. Para alcanzar el mayor grado de objetividad en la «prueba de jueces» la muestra de jueces debe ser representativa y suficiente. Si se cuida la representatividad de la muestra de jueces, la objetividad en la valoración y clasificación de los enunciados en torno a la actitud que se pretende que mida la escala, estará garantizada.

7.3.4. Criterios para la selección de los elementos

Los elementos o proposiciones que se seleccionen para construir la escala definitiva deben guardar estrecha relación con la actitud que se pretende medir y representar exhaustivamente las características de dicha actitud que ha debido ser definida y delimitada con anterioridad.

Una vez valorados los diferentes enunciados y clasificados en cada uno de los 11 intervalos, analizaremos cómo quedan clasificados los elementos. Lo primero que tendremos que comprobar es si están cubiertos todos los posibles grados de la actitud, que se corresponden con los intervalos del 1 al 11.

Así, la proposición 1 la podríamos encontrar:

«x» número de veces en el intervalo 1,

«y» número de veces en el intervalo 2,

«z» número de veces en el intervalo 3, etc.

Y del total de las valoraciones otorgadas por los n «jueces», obtendremos una distribución para cada proposición en los diferentes intervalos del 1 al 11. Los criterios a tener en cuenta para seleccionar los elementos que formarán la escala definitiva, de acuerdo con Thurstone, son los siguientes:

7.3.4.1. El valor escalar del elemento

Los elementos que integran la escala conforman un continuo graduado respecto a la variable que se desea medir, por lo que es necesario que haya un número suficiente de elementos por cada uno de los intervalos en que está representada la actitud. Venimos trabajando con 11 intervalos en los que se gradúan los elementos que contribuyen a valorar la actitud que medimos. Y hay que conocer la posición de cada uno de los

elementos en dicho continuo, del 1 al 11, que vendrá determinado por su valor escalar.

Uno de los métodos desarrollados por Thurstone para hallar el valor escalar del elemento (recordemos que el valor escalar venía representado por la media o por la mediana) es el de «los intervalos aparentemente iguales», por lo que se debe calcular el valor de la mediana de la distribución de la clasificación, o punto central de la distribución.

Si la variable, o rasgo, que queremos medir es, por ejemplo «la actitud de los padres en relación con la inclusión de inmigrantes en los centros educativos», y uno de los elementos integrados en la prueba de jueces es: «la presencia de alumnos inmigrantes en el aula producirá un enriquecimiento en el aprendizaje del resto de los alumnos del aula», la valoración otorgada por los N jueces podría dar lugar a la siguiente representación gráfica:

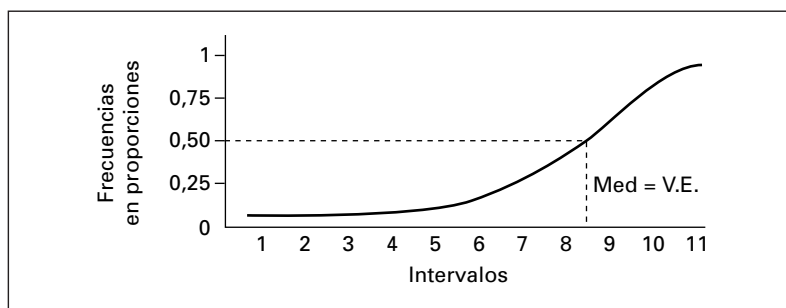


Figura 7.1.

La mayoría de los jueces han valorado el elemento considerando que representa en grado elevado la actitud que se pretende medir, por lo que el valor escalar del elemento se encontrará situado a la derecha de la distribución. Pero para conocer su valor exacto calculamos la mediana de la distribución (dado que el nivel de medida que alcanza la escala es el ordinal y no el de intervalo requerido para calcular la media). La mediana es el valor que deja por debajo de sí el 50% de los valores de la distribución. Aplicando la formula de la mediana:

$$Med_j = Li_1 + \left[\frac{N/2 - (\sum f_c)}{f_d} \right] \times A$$

Siendo:

Med_j : Valor escalar del elemento j

Li_1 : Límite inferior del intervalo asociado a la categoría en la que se encuentra la mediana.

$N/2$: El 50% de los sujetos de la muestra de jueces.

f_c : Número de jueces que clasificaron el elemento j en la categoría en la que se encuentra la de la mediana.

f_d : Número de jueces que clasificaron el elemento k en categorías interiores a la de la mediana.

A : amplitud del intervalo asociado a la categoría donde se encuentra la mediana, que en este caso, al utilizar el método de intervalos aparentemente iguales, es 1.

Calculados los valores escalares de cada uno de los elementos de la prueba de acuerdo con la valoración dada por los jueces (de 1 a 11), debemos comprobar si quedan cubiertos todos los grados de la actitud (los 11), los que la representan en gran medida, los que la representan mínimamente, los neutros y los intermedios. Si no estuvieran cubiertos todos los grados habría que ampliar los enunciados, y repetir el proceso de la prueba de jueces.

7.3.4.2. La ambigüedad del elemento

La ambigüedad hace referencia al grado de acuerdo o desacuerdo entre los jueces en la situación de los elementos en los diferentes grados de la actitud. A partir de la distribución de frecuencias de las posiciones otorgadas a cada elemento, un modo de hallar la ambigüedad es mediante el cálculo de la amplitud intercuartílica, que es la distancia entre el cuartil 1 y el cuartil 3, o lo que es lo mismo, la distancia entre el valor que asignan al elemento el 25% de los jueces y el que le asigna el 75%. La fórmula de la distancia intercuartílica es la siguiente:

$$CA = Q_3 - Q_1$$

Para hallar los cuartiles utilizamos la siguiente fórmula:

$$Q_1 = Li_1 + \left[\frac{m \left(\frac{n}{4} \right) - f_{a(I-1)}}{f_i} \right] a$$

Siendo:

m : el cuartil de que se trate, 1, 2, 3

Li_1 : Límite inferior del intervalo asociado a la categoría en la que se encuentra la frecuencia acumulada en la que se sitúa el cuartil

$m(n/4)$: El 25% de los sujetos de la muestra de jueces

f_i : frecuencia del intervalo en el que se encuentra el cuartil

N : Número de jueces que clasificaron el elemento j en la categoría en la que se encuentra la de la mediana

f_d : frecuencia acumulada hasta Li_1

a : amplitud del intervalo asociado a la categoría donde se encuentra cuartil, que en el caso del método de los intervalos aparentemente iguales, es 1.

Si el valor de ambigüedad obtenido tras aplicar la fórmula es mayor que 2, el elemento es considerado ambiguo y deberá eliminarse. Sólo si se tratara de elementos comprendidos en el intervalo central (5,5,-6,5), al ser estos elementos neutrales se aceptan coeficiente de ambigüedad hasta el valor 3.

Y la suma de las valoraciones otorgadas por los «jueces» la haríamos así:

Tabla 10.3

Proposición 1			
Intervalo	Frecuencias	Porcentajes	Porcentajes acumulados
1	1	1	1
2	3	3	4
3	2	2	6
4	4	4	10
5	8	8	18
6	10	10	28
7	12	12	40
8	14	14	54
9	16	16	70
10	15	15	85
11	15	15	100
Total	100	100	

7.3.4.3. *La irrelevancia del elemento*

Cuando dos o más elementos tienen el mismo valor escalar, lo que supone que representa el mismo grado de la variable que se está midiendo, se puede prescindir de alguno de ellos. Y también podría darse el caso de que dos elementos que tienen el mismo valor escalar contienen algún otro elemento implícito en su redacción que distorsione su valoración, por lo que conviene revisarlos.

Thurstone y Chave (1929, en Barbero, 1999: 234) para dilucidar la irrelevancia de alguno de los elementos, proponen como criterio el siguiente: se les pide a los jueces, o a otra muestra distinta, que marquen con un signo + los elementos que apoyan y con un signo – los que rechazan por algún motivo. Una vez marcados todos los elementos se calcula la media de los valores escalares puntuados con el signo +. Se espera que aquellos elementos que han obtenido el mismo valor escalar, hayan sido calificados con el mismo signo. Y desarrollaron el «índice de semejanza» para comparar los elementos entre sí:

$$IS_{ab} = \frac{n_{ab}}{n_b}$$

Siendo:

IS_{ab} : Índice de semejanza del elemento a con el b

n_{ab} : Número de sujetos que apoyaron tanto el elemento a como el b

n_b : Número de sujetos que apoyaron el elemento b

El índice variará entre los valores 0 y 1. El caso ideal de semejanza se correspondería con el valor 1.

7.3.5. *Elaboración de la escala definitiva*

Una vez seleccionadas las proposiciones que formarán parte de la escala definitiva, siguiendo los criterios de que los elementos cubran todos los niveles de la variable actitud que se desea medir, que su grado de ambigüedad sea aceptable y que sean relevantes, tendremos los elementos de la escala definitiva.

La escala definitiva tiene distinta forma que la prueba de jueces. La forma de respuesta será de dos alternativas con el fin de que los sujetos manifiesten su acuerdo o desacuerdo con la pregunta. Debe incluir instrucciones claras sobre cómo deben responder los sujetos, reflejando de modo sincero su actitud personal ante los enunciados. Los elementos seleccionados se distribuirán aleatoriamente, en un total de 30 a 40.

7.3.6. Calificación de la escala

Una vez terminada la escala definitiva, está lista para ser aplicada a una muestra de sujetos representativa de la población en la cual se quiere medir la actitud. La puntuación obtenida por cada sujeto en la escala vendrá dada por la media de los valores escalares de los elementos contestados favorablemente, es decir, de acuerdo. Recordemos que los valores escalares eran las medias obtenidas por cada elemento en la distribución de frecuencias en la prueba de jueces para la elaboración de la escala.

Supongamos una escala integrada por 30 elementos cuyos valores escalares correspondientes son los siguientes:

Tabla 10.4

Elemento	Valor escalar
1	4,3
2	5,2
3	7,6
4	5,4
5	8,1
6	6,3
7	9,2
8	7,8
9	8,1
...	...
30	8,2

Si el sujeto ha contestado favorablemente a los enunciados o elementos n.º 2, 3, 6, 9, 10, 12, 16, 18, 20, 21, 24, 28 y 30 (14 elementos) y los valores escalares de dichos elementos son los que se indican en la tabla de arriba, la puntuación del sujeto (p) vendrá dada por la fórmula siguiente:

$$p = \frac{\sum VE \text{ de las proposiciones con las que el sujeto está "de acuerdo"}}{n.º \text{ de proposiciones con las que el sujeto está "de acuerdo"}}$$

Que será:

$$p = \frac{5,2 + 7,6 + 6,3 + 8,1 + \dots + 8,2}{14}$$

La puntuación correspondiente a cada sujeto representa el valor numérico que se atribuye a la actitud del sujeto en relación con el tema que se le presenta. Si la puntuación del sujeto fuera 7 significaría que posee dicha actitud en término medio, dado que el intervalo neutral es el central, 6, y el valor de 7 está próximo a dicho intervalo.

La puntuación del grupo se determina mediante la media aritmética de las puntuaciones obtenidas por el grupo. En esta ocasión podemos aplicar la media aritmética dado que la puntuación que aporta la escala puede considerarse de intervalo dado que se apoya en unos elementos obtenidos mediante un escalamiento en intervalos aparentemente iguales, por lo que la distancia entre el intervalo 1 y el 2 se considera casi igual que la distancia entre cualquier otro intervalo consecutivo. Podemos hallar otros índices de tendencia central como la mediana y la moda, y de dispersión como la desviación típica.

Asimismo, se puede hallar la puntuación del grupo en cuanto a los acuerdos, a partir de una tabla de frecuencias, hallando la media aritmética de los valores escalares relativos a las proposiciones que superen un porcentaje del 50% «de acuerdo».

7.4. Fiabilidad y validez de la escala

Las escalas, como cualquier otro instrumento de recogida de información, deben ser fiables y válidas para garantizar su rigor y precisión. Para el cálculo del coeficiente de fiabilidad de las escalas de Thurstone se utiliza el método de las dos mitades, ítems pares e impares, por el procedimiento de Spearman-Brown, en el supuesto de que las varianzas de ambas mitades sean homogéneas, o no existan diferencias significativas entre ellas, cuya fórmula es:

$$R_{xx} = \frac{2r_{xy}}{1 + r_{xy}}$$

Siendo:

R_{xx} : el coeficiente de fiabilidad

r_{xy} : la correlación entre las puntuaciones de los elementos pares e impares.

Para conocer la correlación entre los ítems pares y los ítems impares aplicamos la fórmula de la correlación producto momento de Pearson:

$$r_{xy} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}}$$

Siendo:

r_{xy} : la correlación entre los elementos pares y los impares

X : elementos pares

Y : elementos impares

n : número de sujetos total del grupo.

Teniendo el valor r_{xy} podemos obtener el coeficiente de fiabilidad.

Para conocer la consistencia interna de la escala respecto a la actitud que se quiere medir se utiliza el coeficiente α de Crombach. Su fórmula es:

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum s^2}{s_x^2} \right)$$

Donde:

n : número de elementos de la escala

$\sum s^2$: suma de las varianzas de los elementos pares e impares: $s_p^2 + s_i^2$

s_x^2 : varianzas de las puntuaciones en la escala total.

Para que una escala sea válida los valores escalares otorgados por los jueces deben ser independientes de la muestra de jueces. Así, si la escala es valorada por dos muestras diferentes de jueces, y los valores escalares dados por cada una de las muestras consiguen una correlación alta, su validez es buena. Si los valores escalares otorgados por uno y otro grupo no difieren significativamente, la escala puede considerarse válida.

Para hallar la validez de la escala se puede utilizar el criterio de jueces así como un criterio externo para compararla, pudiendo ser otra escala que mida la misma actitud y haya sido ya validada. En el tema 9 dedicado específicamente a la validez y fiabilidad ampliaremos estos conceptos.

7.5. Ejercicios de autoevaluación

1. Defina el concepto de escala de medida y sus fundamentos.
2. Queremos conocer las opiniones de un grupo de estudiantes acerca de las posibles propuestas para realizar actividades deportivas en el Centro Educativo en el que cursan sus estudios, y para ello, se

les presenta una relación de 6 deportes y se les pide que elijan 3 indicando su orden de preferencia, utilizando la siguiente tabla:

Tabla 10.6

Deportes	Orden de preferencia
Atletismo	
Baloncesto	
Balonmano	
Fútbol	
Judo	
Natación	
Voleibol	

Se le pide que halle los *índices de preferencias, de prestigio y de jerarquización*.

- Defina el concepto de actitud según Thurstone.
- A un grupo de empresarios, con una larga experiencia en empresas de prestigio, se les pide que de entre las diez instituciones universitarias que se le presentan, en las que se ofertan Masters en Administración de Empresas, elijan 4 de ellas y las ordenen de mayor a menor en función de su confianza en la mejor preparación que obtienen sus alumnos.

Tabla 10.7

Instituciones	Elección ordenada de 1 a 4
1. ANCED	
2. CESMA	
3. CFE	
4. EADA	
5. ESIC	
6. IFE	
7. UAM	
8. UAB	
9. UCM	
10. UNED	

7.6. Soluciones a los ejercicios de autoevaluación

1. Defina el concepto de escala de medida y sus fundamentos.

Las escalas para la medida de las actitudes se basan en la formulación de preguntas cuyos enunciados incorporan una cierta graduación de aceptación o rechazo hacia distintas actitudes, intereses, valores u opiniones, en función de un determinado atributo.

3. Defina el concepto de actitud según Thurstone.

El concepto de actitud, de acuerdo con Thurstone, se define como el conjunto de inclinaciones y sentimientos, prejuicios o tendencias, temores o convicciones de una persona respecto de un tema determinado.

4. Si estas instituciones han mostrado sobradamente su calidad y eficacia diferencial en la formación de alumnos en Administración de Empresas, la elección de las cuatro mejores será fácil. Pero si la mayoría han mostrado similar calidad y eficacia, la elección, o la decisión para elegir las cuatro mejores, se hará difícil. El atributo que debe guiar la elección de las instituciones es su calidad y eficacia para la formación de futuros administradores de empresas.

7.7. Referencias bibliográficas

- BARBERO GARCÍA, M. I. (1999). *Psicometría II. Métodos de elaboración de escalas*. Madrid: UNED, Unidades Didácticas. (2.^a ed.)
- GONZÁLEZ BLASCO, P. (1994). «Medir en Ciencias Sociales». En Manuel García Ferrando, Jesús Ibáñez y Francisco Alvira (Comp.): *El análisis de la realidad social. Métodos y técnicas de investigación*. Madrid: alianza Universidad (2.^a ed.).
- LIKERT, R. (1974). «A method of constructing an attitude scale». En G. M. Maranell (ed.): *Scaling: A source book of behavioral scientist*. Chicago: Aldine.
- THURSTONE, L. L. (1928). «Attitude can be measured». *American Journal of Sociology* 33 (529-554).
- THURSTONE, L. L. y CHAVE, E. J. (1929). *The measurement of attitudes*. Chicago: University of Chicago Press.
- TORGERSON, W. S. (1958). *Theory and methods of scaling*. New York: Jonh Wiley and Sons.

TEMA 8

Las escalas para la medida de las actitudes de Likert

ESQUEMA

- 8.1.** Las escalas de medida de Likert
- 8.2.** Construcción de una escala de estimaciones sumadas de Likert
 - 8.2.1. Elaboración de proposiciones
 - 8.2.2. Valoración de las proposiciones
 - 8.2.3. Asignación de puntuaciones a las porposiciones
 - 8.2.4. Análisis de los ítems iniciales
 - 8.2.4.1. Estudio de la unidimensionalidad de la escala
 - 8.2.4.2. Análisis discriminante de los elementos
 - 8.2.5. La escala definitiva. Aplicación e interpretación de las puntuaciones
- 8.3.** Fiabilidad y validez de la escala
- 8.4.** Ejercicios de autoevaluación
- 8.5.** Soluciones a los ejercicios de autoevaluación
- 8.6.** Referencias bibliográficas

OBJETIVOS

Se espera que el alumno, tras el estudio de este tema, sea capaz de comprender y valorar:

- ✓ Los fundamentos del modelo de Likert para la elaboración de escalas
- ✓ La finalidad de las escalas de medida de Likert
- ✓ Los procedimientos para la elaboración de la escala de Likert
- ✓ Los procedimientos seguidos por Likert para conocer la fiabilidad y la validez de sus escalas

8.1. Las escalas de medida de Likert

Likert, en 1929, junto a otros investigadores, se propuso desarrollar técnicas para la medida indirecta de las actitudes vinculadas con problemáticas raciales, económicas, políticas y religiosas. Debido a la complejidad que supone la construcción de las escalas tipo Thurstone, especialmente en lo referente a la prueba de jueces, Likert pretendió elaborar unas escalas más sencillas, aunque garantizando su fiabilidad y validez.

La técnica de Likert tiene como finalidad esencial el escalamiento u ordenación de los sujetos a lo largo de un continuo. Así, las variaciones sistemáticas encontradas en las respuestas dadas por los sujetos son atribuibles a las diferencias individuales entre ellos, pudiendo evaluar dichas diferencias situando las puntuaciones individuales a lo largo de un continuo, o distribución, para realizar estudios comparativos entre cada uno de los sujetos con respecto al grupo, basándose en las medidas de tendencia central, dispersión, apuntamiento y asimetría, pudiendo realizar, asimismo, comparaciones entre grupos aplicando las pruebas estadísticas de los diseños de grupos. Basándose en un sistema de respuesta múltiple para cada elemento de la escala, y comprobando que dichas respuestas tienden a distribuirse siguiendo el modelo de la curva normal, redujo el proceso del cálculo del valor escalar de cada uno de los elementos utilizado por Thurstone. El método de construcción de las escalas seguido por Likert es similar al utilizado en la teoría clásica de los tests psicométricos para la medida de las aptitudes y los tests de rendimiento. (Barbero, 1999: 257 y ss.).

Las escalas de Likert reducen en gran medida el número de proposiciones necesarias en la construcción de las escalas de Thurstone, permitiendo a quien responde optar por un mayor número de posibilidades. Las proposiciones que se proponen para ser valoradas se basan en acciones que se describen y se espera que el sujeto manifieste su actitud valorándolas.

8.2. Construcción de una escala de estimaciones sumadas de Likert

Una vez definida y delimitada con claridad y precisión la actitud que se quiere medir, Likert propuso las siguientes fases para la construcción de sus escalas:

- 1.º Enunciación de proposiciones o elementos iniciales.

- 2.º Valoración de las proposiciones.
- 3.º Asignación de puntuaciones a las proposiciones.
- 4.º Análisis de la unidimensionalidad de los elementos.
- 5.º Aplicación de la prueba.
- 6.º Asignación de puntuaciones a los encuestados.

8.2.1. Elaboración de las proposiciones

Para la elaboración de las proposiciones que integrarán la escala definitiva, primero hay que estudiar las características de las actitudes a medir, revisando la literatura relevante y consultando trabajos previos y los instrumentos existentes en el mercado, junto con la experiencia personal del investigador.

Las proposiciones pretenden recoger las valoraciones de los sujetos sobre su actitud ante una determinada problemática que se le presenta, por lo que se le debe ofrecer al sujeto opciones entre alternativas opuestas, conteniendo elementos que reflejen opciones favorables y desfavorables, al 50 por ciento, hacia la actitud que se quiere medir, mezclados aleatoriamente a lo largo de la de la escala. Por ello, es necesario que los elementos sean valorados por expertos que los clasificaran en positivos, o favorables a la actitud, negativos, o desfavorables a la actitud y neutros en cuanto a la valoración de la actitud, debiéndose eliminar los elementos neutros.

Al formular las proposiciones debe cuidarse que estén orientadas en el sentido de lo deseable, de lo que debería ser, pero su redacción definitiva puede presentarse en estilo directo o contrario para evitar respuestas rutinarias y automáticas y, de este modo, hacer que el sujeto se fije en el significado y sentido de la pregunta siempre que no perjudique más que beneficie en la respuesta. Debe utilizarse la construcción gramatical del presente de indicativo. El número de proposiciones seleccionadas inicialmente debe sobrepasar al menos dos veces la longitud final que se desea tenga la escala definitiva.

8.2.2. Valoración de las proposiciones

La escala presenta para cada proposición, una serie de alternativas de respuesta, que deberán ser tenidas en cuenta por los encuestados a la hora de emitir su respuesta. Frecuentemente se ofrecen cinco posiciones; una central y dos por encima y por debajo, del modo siguiente:

1. Me siento a gusto ayudando a mis compañeros a comprender un tema de estudio:

Muy de acuerdo	Algo de acuerdo	Indiferente	Poco de acuerdo	Nada de acuerdo
☐	☐	☐	☐	☐

El formato de respuesta puede presentarse de este otro modo:

- a) Completamente de acuerdo
- b) De acuerdo
- c) Indiferente
- d) En desacuerdo
- e) Completamente en desacuerdo

También pueden presentarse las alternativas de respuesta del modo siguiente:

Expresa su grado de acuerdo con las siguientes preguntas indicándolo entre 5, máximo acuerdo, a 1, mínimo acuerdo o máximo desacuerdo.

1. El aprendizaje de la asignatura por ordenador me gusta por la autonomía que ofrece.
- | | | | | |
|-----------|---|---|---|-----------|
| 5 | 4 | 3 | 2 | 1 |
| MA | | | | MD |

Se puede optar por una escala de valores del 1 al 6 para evitar un único valor central que suele ser la respuesta más frecuente cuando se contesta de modo rápido.

La opción de alternativa de respuesta en una escala de 1 a 5 fue la elegida por Likert al comprobar que se distribuían de forma parecida a la curva normal.

8.2.3. Asignación de puntuaciones a las proposiciones

Hemos dicho que cada elemento, proposición o ítem se responde mediante la asignación, por parte del sujeto, de entre 1 a 5 valores, teniendo en cuenta su gradación sobre la actitud que se está valorando.

Veamos un ejemplo:

2. Me gusta consultar mis dudas al profesor mediante el correo electrónico.
- | | | | | |
|-----------|---|---|---|-----------|
| 5 | 4 | 3 | 2 | 1 |
| MA | | | | MD |

Una vez construida la escala inicial se aplica a una muestra piloto perteneciente a la misma población para la que se va a elaborar el instrumento definitivo. Cada sujeto responderá marcando en cada uno de los elementos su opción de respuesta. La suma de todos los números marcados en cada una de las proposiciones nos indicará la puntuación total del sujeto que reflejará sus valoraciones sobre la actitud que se quiere medir.

8.2.4. *Análisis de los ítems iniciales*

Supongamos que la escala con la que venimos trabajando para valorar el modo como un profesor realiza la explicación de una lección en una sesión de clase, está formada por 10 ítems y ha sido contestada por los 16 alumnos de la clase, aportando los datos que mostramos en la tabla de datos, tomada del programa informático SPSS, versión 11, junto con la tabla de estadísticos descriptivos totales y por ítems:

TABLA 8.1. Tabla de datos.

Sujetos	it1	it2	it3	it4	it5	it6	it7	it8	it9	it10	Suma its
1	4.00	3.00	4.00	4.00	4.00	4.00	5.00	5.00	3.00	5.00	41.00
2	4.00	3.00	4.00	4.00	4.00	5.00	4.00	5.00	5.00	3.00	41.00
3	5.00	5.00	3.00	3.00	5.00	5.00	4.00	4.00	4.00	5.00	43.00
4	4.00	3.00	5.00	4.00	4.00	4.00	5.00	4.00	4.00	3.00	40.00
5	2.00	5.00	2.00	3.00	3.00	4.00	3.00	5.00	5.00	5.00	37.00
6	3.00	5.00	4.00	3.00	5.00	3.00	4.00	5.00	4.00	4.00	40.00
7	5.00	4.00	3.00	4.00	4.00	4.00	4.00	4.00	5.00	5.00	42.00
8	4.00	3.00	4.00	5.00	5.00	4.00	3.00	5.00	4.00	4.00	41.00
9	3.00	5.00	5.00	4.00	4.00	4.00	4.00	4.00	5.00	5.00	43.00
10	4.00	5.00	4.00	4.00	5.00	3.00	5.00	5.00	4.00	4.00	43.00
11	4.00	3.00	3.00	5.00	3.00	4.00	3.00	5.00	3.00	4.00	37.00
12	5.00	5.00	4.00	4.00	5.00	4.00	5.00	3.00	5.00	2.00	42.00
13	5.00	5.00	4.00	5.00	4.00	4.00	4.00	4.00	4.00	5.00	44.00
14	4.00	3.00	4.00	4.00	4.00	3.00	5.00	5.00	2.00	4.00	38.00
15	3.00	4.00	4.00	3.00	3.00	4.00	4.00	3.00	4.00	5.00	37.00
16	4.00	3.00	4.00	4.00	4.00	4.00	4.00	5.00	4.00	3.00	39.00

TABLA 8.2. Descriptivos.

Estadísticos descriptivos								
	N	Rango	Mínimo	Máximo	Suma	Media	Desv. típ.	Varianza
SUMAS	16	7.00	37.00	44.00	648.00	40.5000	2.33809	5.467

TABLA 8.3. Descriptivos de los ítems.

Estadísticos descriptivos							
	N	Mínimo	Máximo	Suma	Media	Desv. típ.	Varianza
IT1	16	2.00	5.00	63.00	3.9375	.85391	.729
IT2	16	3.00	5.00	64.00	4.0000	.96609	.933
IT3	16	2.00	5.00	61.00	3.8125	.75000	.563
IT4	16	3.00	5.00	63.00	3.9375	.68007	.463
IT5	16	3.00	5.00	66.00	4.1250	.71880	.517
IT6	16	3.00	5.00	63.00	3.9375	.57373	.329
IT7	16	3.00	5.00	66.00	4.1250	.71880	.517
IT8	16	3.00	5.00	71.00	4.4375	.72744	.529
IT9	16	2.00	5.00	65.00	4.0625	.85391	.729
IT10	16	2.00	5.00	66.00	4.1250	.95743	.917
N válido	16						

8.2.4.1. Estudio de la unidimensionalidad de la escala

La aplicación de la escala experimental o piloto a esa primera muestra nos sirve para comprobar si las proposiciones propuestas para medir la actitud están adecuadamente relacionadas con dicha actitud de modo que puedan mantenerse en la escala definitiva, o, por el contrario, eliminarse. Para ello hay que seguir los siguientes pasos:

1. Calcular para cada sujeto la suma de los valores que asignó a cada una de las proposiciones. En la tabla de datos de nuestro ejemplo (Tabla 8.1), figura la suma total obtenida por cada sujeto en la escala, en la última columna.
2. Se calculan las correlaciones existentes entre las puntuaciones obtenidas por cada sujeto en un elemento y las obtenidas en la escala total (correlación ítem-total).

Para hallar las correlaciones se aplica la fórmula de la correlación de Pearson.

$$r_{xy} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}}$$

Siendo:

r_{xy} : correlación entre las puntuaciones obtenidas por los sujetos en el elemento X y en la totalidad de la escala

X : puntuaciones en el elemento que se analiza

Y : puntuaciones en la totalidad de los elementos de la escala

n : número de sujetos de la muestra.

Para calcular las correlaciones siguiendo el procedimiento manual tendríamos que calcular la suma de las puntuaciones obtenidas por todos los sujetos en cada ítem, la suma del producto XY , siendo X la variable ítem (los diez) e Y la suma de los ítems de la escala en cada sujeto. Después calcularíamos la suma de los cuadrados de la variable X (para cada ítem), y la suma de los cuadrados de la variable Y . Después sustituiríamos los valores obtenidos en la fórmula de la correlación de Pearson.

Y utilizando el programa SPSS, mostramos la tabla resumen de los resultados obtenidos a partir de los datos del ejemplo en el que venimos trabajando:

TABLA 8.4. Correlaciones de Pearson

Correlaciones de Pearson			
	N	Sig. (bilateral)	Sumas
IT1	16	.017	.584(*)
IT2	16	.086	.443
IT3	16	.222	.323
IT4	16	.390	.231
IT5	16	.004	.674(**)
IT6	16	.519	.174
IT7	16	.231	.317
IT8	16	.423	-.216
IT9	16	.183	.351
IT10	16	.827	.060

Como se trata de la correlación entre la puntuación obtenida por un elemento con la obtenida en la totalidad de la escala menos la de dicho elemento, hay que aplicar al resultado obtenido en la fórmula anterior una corrección para cada ítem mediante la fórmula siguiente:

$$r_{x(y-x)} = \frac{r_{yx}s_x - s_y}{\sqrt{s_x^2 + s_y^2 - 2r_{yx}s_x s_y}}$$

Esta fórmula nos aporta un índice de homogeneidad del elemento, indicando el grado en que los elementos de la escala miden la misma actitud que pretende medir la escala. Los ítems que no alcanzan un índice de homogeneidad de 0.20 se deben eliminar de la escala definitiva. Este procedimiento resulta muy laborioso para ser realizado sin ayuda de un programa informático para el análisis de datos estadísticos, por lo que se aconseja su utilización.

8.2.4.2. *Análisis discriminante de los elementos*

Otro método utilizado para el análisis de los ítems consiste en ordenar la totalidad de las puntuaciones obtenidas por los sujetos en la escala, de mayor a menor y seleccionar el 25% superior y el 25 % inferior, mediante la siguiente fórmula:

$$n(25\%) = \frac{25N}{100} = \frac{N}{4}$$

que nos daría el número de sujetos correspondiente al 25 %. Ambos grupos tendrán el mismo número de sujetos, es decir, $n_1 = n_2$.

Como la escala de Likert no supera el nivel ordinal aplicamos las pruebas estadísticas no paramétricas, en este caso la prueba U de Mann-Whitney y la prueba de χ^2 , si bien a veces se utilizan las pruebas estadísticas paramétricas t y z.

Para la aplicación de la prueba U de Mann-Whitney se procede del siguiente modo:

1. Se ordenan conjuntamente los sujetos del grupo superior (25% superior) e inferior (25% inferior), asignándoles rangos comenzando por el que obtiene la puntuación más baja, que se le asigna el rango 1. Si dos o más datos tienen la misma puntuación en el ítem, se halla la media de sus rangos, asignándoles el valor medio resultante.
2. Se suman por separado los rangos obtenidos por cada uno de los grupos, y se aplica a ambos totales, ΣR_1 y ΣR_2 , la prueba U de Mann-Whitney.

Veamos un ejemplo. Utilizando los datos de la tabla 8.1, analizamos el ítem número 1. El 25 % de las puntuaciones inferiores y el 25 % de las puntuaciones superiores, ordenadas de menor a mayor, nos daría la siguiente tabla:

TABLA 8.5. Rangos del 25% superior e inferior en el ítem 1.

Sujetos	It1	Rango
5	2.00	1
6	3.00	3
9	3.00	3
15	3.00	3
3	5.00	6.5
7	5.00	6.5
12	5.00	6.5
13	5.00	6.5

Y el valor de U empírico, para muestras muy pequeñas, como es este caso, en el cual consideramos dos grupos, el primero formado por las primeras 4 puntuaciones inferiores, y el segundo formado por las 4 últimas puntuaciones superiores, vendría dado por el número de veces que una puntuación del grupo superior precede a una puntuación del grupo inferior. En este caso el valor de $U = 0$, por que son 0 las veces que sucede el supuesto mencionado.

Para comprobar si podemos rechazar la hipótesis nula, que dice que no se dan diferencias significativas entre los dos grupos, consultamos la tabla de la distribución de U para muestras muy pequeñas y pruebas unilaterales, y comparamos el valor de U empírico, para $n_1 = 4$ y $n_2 = 4$. Vemos que la probabilidad asociada al valor observado (empírico) de $U = 0$ es 0,014, y que es inferior al nivel de significación (alfa) de 0,05, lo que nos permite rechazar la hipótesis nula y confirmar que existen diferencias significativas, por lo que podemos aconsejar mantener el ítem en la escala.

Los cálculos realizados mediante el programa SPSS, nos aportan las siguientes tablas de resultados:

TABLA. 8. 6. Pruebas no paramétricas: Prueba de Mann-Whitney

Rangos				
	Grupo 2	N	Rango promedio	Suma de rangos
Grupo 1	1.00	4	2.50	10.00
	2.00	4	6.50	26.00
	Total	8		

TABLA 8.7. Prueba U de Mann-Whitney

Estadísticos de contraste(b)	
	VAR00001
U de Mann-Whitney	.000
W de Wilcoxon	10.000
Z	-2.428
Sig. asintót. (bilateral)	.015
Sig. exacta [2*(Sig. unilateral)]	.029(a)
a No corregidos para los empates.	
b Variable de agrupación: VAR00002	

Podemos comprobar que los resultados son coincidentes con los obtenidos por el procedimiento manual.

Y si optamos por la aplicación de la prueba de «chí» cuadrado (χ^2), por los mismos supuestos que U, para escalas de medida que no alcanzan el nivel de intervalo, se procede del siguiente modo:

1. Ordenación de las puntuaciones obtenidas, en el elemento que estudiamos, por la totalidad de los sujetos de la muestra, comenzando por la puntuación inferior.
2. Cálculo de la mediana para la totalidad de la muestra (N)

$$Med = Li_1 + \left[\frac{\frac{N}{2} - \sum f_{a(l-1)}}{f_i} \right] \times a$$

Siendo:

Li_1 : límite inferior del intervalo en que se sitúa el Centil 50

$f_{a(l-1)}$: frecuencia acumulada hasta el intervalo en que se sitúa el Centil 50

f_i : frecuencia del intervalo en que se sitúa el Centil 50

a : amplitud del intervalo.

En el caso de una distribución no agrupada en intervalos la mediana vendría dada por el valor central de la distribución, y si la serie estuviera formada por un total de casos pares se hallaría la media aritmética de los dos valores centrales.

3. Escogeríamos el 25 % superior, y el 25 % inferior utilizando la fórmula anterior ($N/4$).
4. Con el valor de la mediana como referencia se forma una tabla de contingencia 2×2 , utilizando las puntuaciones por encima y por debajo de la mediana de ambos grupos:

	Por encima de la Med	Por debajo de la Med
Grupo superior	a	b
Grupo inferior	c	d

Siendo:

- a: número de sujetos del grupo superior que ha obtenido puntuaciones por encima de la mediana en el elemento que se estudia
 - b: número de sujetos del grupo superior que ha obtenido puntuaciones por encima de la mediana en el elemento que se estudia
 - c: número de sujetos del grupo inferior que ha obtenido puntuaciones por debajo de la mediana en el elemento que se estudia
 - d: número de sujetos del grupo inferior que ha obtenido puntuaciones por debajo de la mediana en el elemento que se estudia
5. A continuación se aplicaría la prueba estadística de χ^2 cuya fórmula es:

$$\chi^2 = \frac{N \left(\frac{|ad - cb|}{N} \right)^2}{(a+b)(c+d)(a+c)(b+d)}$$

6. El valor empírico de χ^2 obtenido se compara con el valor teórico en la tablas de χ^2 , con 1 grado de libertad, y al nivel de significación ($\alpha = 0,05$ o $\alpha = 0,01$) fijado. Si el valor empírico es mayor que el nivel teórico que figura en la tabla, las discrepancias son estadísti-

camente significativas, por lo que el elemento debe permanecer para la escala definitiva.

Si trabajamos con muestras grandes, podemos aplicar la prueba de «t» de Student, y si $n > 30$, la prueba z para las diferencias de medias entre los grupos que obtienen el 25% superior y el 25% inferior en los elementos, para comprobar si son discriminativos.

Si utilizamos la prueba «t» de Student para comprobar si los elementos discriminan, utilizamos la fórmula siguiente:

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\left(\frac{n_A s_A^2 + n_B s_B^2}{n_A + n_B - 2} \right) \left(\frac{n_A + n_B}{n_A n_B} \right)}}$$

Para ello se puede proceder del siguiente modo:

1. Se anotan las frecuencias de cada grupo (el 25 % que obtuvo las puntuaciones superiores y el 25% inferior) obtenidas en cada una de las categorías y las frecuencias acumuladas del elemento que se estudia, utilizando el siguiente modelo de tabla:

	Grupo A: 25% sujetos con puntuaciones mínimas						Grupo B: 25% sujetos con puntuaciones máximas					
Categorías	X_i	X_i^2	f	f_a	fX_i	fX_i^2	X_i	X_i^2	f	f_a	fX_i	fX_i^2
Muy de acuerdo	5											
Bastante acuerdo	4											
Normal acuerdo	3											
Poco de acuerdo	2											
Nada de acuerdo	1											
Totales												

Para cada grupo se calcula la media aritmética (a) y la varianza (b):

$$(a) \quad \bar{X}^2 = \frac{\sum fX_i}{n}; \quad (b) \quad s^2 = \frac{\sum fX_i^2}{n} - \bar{X}^2$$

Y se sustituyen los valores obtenidos en la fórmula o ecuación de la prueba «t» de Student, de diferencias de medias, y tras comparar el valor obtenido por «t» con el valor teórico en las Tablas de «t», al nivel de significación «α» fijado, con $(n_1 + n_2 - 2)$ grados de libertad ($gl = 2n - 2$, dado

que $n_1 = n_2$), si el valor empírico es mayor que el valor teórico o crítico que figura en las tablas ($t_e > t_c$), podemos rechazar la hipótesis nula que dice que no existen diferencias significativas entre el grupo superior y el grupo inferior, confirmando la hipótesis alternativa de que sí existen diferencias, y que por lo tanto, el elemento es discriminativo, luego conviene mantenerlo para la escala definitiva.

Si trabajamos con muestras grandes, $n > 30$, o próximas, se utiliza la prueba de z para las diferencias de medias, cuya fórmula es la siguiente:

$$z = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

Debiendo consultar las tablas de z , o de áreas bajo la curva normal. Si $z_e > z_c$, las diferencias son significativas, y se interpreta del mismo modo que hemos indicado en la prueba t .

En la actualidad el análisis de los elementos de las escalas, cuestionarios, pruebas de rendimiento etc. se realiza mediante programas informáticos para el análisis estadístico de los datos, que ahorran esfuerzo y tiempo en la realización de los cálculos. Sin embargo, para su utilización e interpretación es necesario conocer y comprender los fundamentos y procedimientos para la elaboración de las escalas, para su aplicación y para su corrección e interpretación de las puntuaciones obtenidas por los sujetos a quienes se les aplica la escala, y para realizar comparaciones entre sujetos y grupos.

8.2.5. La escala definitiva. Aplicación e interpretación de las puntuaciones

Una vez analizadas todas las proposiciones (enunciados, cuestiones o ítems) de la escala de Likert mediante las pruebas de unidimensionalidad u homogeneidad de los ítems y las pruebas de significación de hipótesis propuestas «U», «t», «z» o « χ^2 », que serán estudiadas en la Tercera Unidad Didáctica, tema 17, seleccionaremos los elementos que muestren mayor índice de homogeneidad y los que discriminen más entre los grupos que alcanza mayor y menor puntuación en los mismos elementos, que deberán ser los incluidos en la escala definitiva.

La forma de puntuar la escala de Likert se hace asignando valores a las alternativas de respuesta de cada uno de los elementos en orden de mayor a menor según que manifieste la actitud en mayor o menor grado.

La interpretación de las puntuaciones obtenidas por los sujetos en la escala, como los tests psicométricos, tiene que ver con las características de la población para la que se ha elaborado la escala, y más concretamente, con la muestra representativa o grupo que se ha utilizado para elaborar la escala. Las puntuaciones obtenidas por los sujetos deben interpretarse en relación con el grupo, considerando su media aritmética como el punto en el origen, y las puntuaciones individuales, desviaciones con relación a la media, por lo que habrá que transformar las puntuaciones directas en puntuaciones diferenciales x :

$$x = X - \bar{X}$$

Y después transformamos las puntuaciones diferenciales en puntuaciones típicas, « z », mediante la fórmula:

$$z = \frac{x}{s_x}$$

Si la muestra es grande ($n > 30$), se considera que su distribución se aproxima a la forma de la curva normal, lo que nos permite, utilizando las tablas de la curva normal, interpretar la posición de un sujeto, en función de la puntuación obtenida en la escala, en relación con el grupo, en porcentajes que deja por debajo de él y por encima dentro de la distribución, de acuerdo con las características de la curva normal, que revisaremos en el tema 13 de la Tercera Unidad Didáctica.

8.3. Fiabilidad y validez de la escala

Likert utilizó la prueba de las dos mitades para el estudio de la fiabilidad de su escala, que como ya hemos visto en el estudio de la fiabilidad de la escala de Thurstone, consiste en correlacionar las puntuaciones obtenidas en los elementos pares con las obtenidas en los elementos impares, mediante la fórmula de la correlación de Pearson:

$$r_{xy} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}}$$

Siendo:

r_{xy} : la correlación entre los elementos pares y los impares,

X : elementos pares,

Y : elementos impares,

n : número de sujetos total del grupo.

Aplicando posteriormente la corrección de Spearman-Brown, cuya fórmula es:

$$R_{xy} = \frac{2r_{xy}}{1 + r_{xy}}$$

Siendo:

R_{xx} : el coeficiente de fiabilidad,

r_{xy} : la correlación entre las puntuaciones de los elementos pares e impares.

El coeficiente consistencia interna « α de Crombach» se aplica para conocer la consistencia de la escala con respecto a la actitud que se pretende medir, siendo su fórmula:

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum s_i^2}{s_x^2} \right)$$

Donde:

n : número de elementos de la escala

$\sum s_i^2$: suma de las varianzas de los elementos $= s_p^2 + s_i^2$.

s_x^2 : varianza de las puntuaciones en la escala total.

En cuanto a la validez de los instrumentos de medida, se dice que la tienen cuando miden lo que pretenden medir. Son diversos los tipos de validez: de contenido, predictiva, convergente, de constructo. Las pruebas de validez de contenido, se basan fundamentalmente en la validación por expertos, y se centran en conocer la adecuación de los ítems para medir el rasgo o constructo que se pretende medir, y, por lo tanto, tiene que ver con la representatividad de los ítems del instrumento del contenido que pretende medir. La validez predictiva, se interesa en comprobar la capacidad de inferencia de la escala en aquella actitud sobre la que fue elaborada. La validez convergente se basa en la relación entre el nuevo instrumento con otro ya validado sobre el mismo constructo, y la validez de constructo pretende comprobar que el método utilizado en la construcción de la escala garantiza la coherencia con los fundamentos teóricos y metodológicos en los que se apoya. Los conceptos de fiabilidad y validez serán objeto de mayor profundización en el tema siguiente.

Veamos un ejemplo para hallar la fiabilidad de una escala por el procedimiento de las dos mitades, de Spearman Brown, y por otros procedimientos.

A un grupo de 10 alumnos de 2.º curso de Bachillerato se les ha aplicado una prueba objetiva de Matemáticas compuesta por 30 ítems de res-

puesta única, calificando cada ítem por 1, acierto, 0, error. Los resultados obtenidos se muestran en la tabla de datos siguiente:

TABLA 8.8. Tabla de datos para el cálculo de la correlación

Sujetos	Puntuaciones totales	Ítems pares (X)	Ítems impares (Y)	XY	X ²	Y ²
1	8	3	5	15	9	25
2	9	4	5	20	16	25
3	11	5	6	30	25	36
4	11	6	5	30	36	25
5	13	7	6	42	49	36
6	15	8	7	56	64	49
7	19	10	9	90	100	81
8	23	12	11	132	144	121
9	23	13	10	130	169	100
10	27	14	13	182	196	169
Totales	159	82	77	727	808	667

Tomamos la fórmula para hallar la fiabilidad por el procedimiento de Spearman-Brown, que es la siguiente:

$$R_{xx} = \frac{2r_{xy}}{1 + r_{xy}}$$

Necesitando hallar previamente la correlación de Pearson. Tal como nos vienen dados los datos en la tabla 8.8, procede utilizar la fórmula de la correlación para puntuaciones directas:

$$r_{xy} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}}$$

Sustituyendo en la fórmula los valores de la tabla, tenemos:

$$\begin{aligned}
 r_{xy} &= \frac{(10 \times 727) - (82 \times 77)}{\sqrt{[(10 \times 808) - 82^2][(10 \times 667) - 77^2]}} = \frac{7270 - 6314}{\sqrt{[8080 - 6724][6670 - 5929]}} = \\
 &= \frac{956}{1002,39} = 0,95
 \end{aligned}$$

Sustituyendo el valor de la correlación r_{xy} en la fórmula de la fiabilidad, tenemos:

$$R_{xx} = \frac{2 \cdot r_{xy}}{1 + r_{xy}} = \frac{2 \cdot 0,95}{1 + 0,95} = \frac{1,80}{1,95} = 0,95$$

Pudiendo comprobar que se trata de una fiabilidad alta, $R_{xx} = 0,95$, próxima a 1.

Veamos ahora los resultados obtenidos por el procedimiento de las dos mitades, mediante el programa informático SPSS:

TABLA 8.9. Análisis de fiabilidad

R E L I A B I L I T Y A N A L Y S I S - S C A L E (S P L I T)	
Reliability Coefficients	
N of Cases = 10.0	N of Items = 2
Correlation between forms = .9537 Equal-length Spearman-Brown = .9763	
Guttman Split-half = .9539 Unequal-length Spearman-Brown = .9763	
1 Items in part 1	1 Items in part 2
Alpha for part 1 = 1.0000	Alpha for part 2 = 1.0000

La tabla nos aporta las correlaciones entre las dos mitades, muy elevada, 0.953, sabiendo que el valor de los coeficientes de fiabilidad van de 0, ninguna, a 1, máxima, los valores obtenidos por los diferentes coeficientes que ha empleado el programa, el de Guttman, para varianzas iguales, el de Spearman- Brown, para mitades con igual y diferente longitud, y el Alfa de Crombach, para la primera mitad, y para la segunda. Todos dan un valor de la fiabilidad de la prueba muy elevado. El de Crombach, perfecto, 1.

8.4. Ejercicios de autoevaluación

1. ¿Finalidad de la técnica de Likert para la medida de las actitudes?
2. ¿Las escalas son técnicas para la medición directa de las actitudes?
3. ¿Cómo se interpreta el índice de semejanza (IS) entre dos elementos?
4. ¿Cómo se halla la puntuación del sujeto en la escala?

5. Elabore una escala para la medida de las actitudes integrada sobre un tema que le interese, formada por 20 ítems, con 5 alternativas de respuesta. Valórela, mediante el procedimiento de los jueces, pidiéndole ayuda a cinco compañeros de estudio. Consulte a su profesor tutor. Aplíquelo a una muestra piloto de 10 personas, compañeros o alumnos, y analice sus datos. Entréguelo como trabajo para que se lo evalúe al profesor de la asignatura.

8.5. Soluciones a los ejercicios de autoevaluación

1. La técnica de Likert para la medida de las actitudes tiene como finalidad el escalamiento u ordenación de los sujetos a lo largo de un continuo, de modo que las variaciones sistemáticas encontradas en las respuestas dadas por los sujetos se deben a las diferencias individuales de los sujetos.
2. ¿Las escalas son técnicas para la medición directa de las actitudes? Sí, puesto que el sujeto sabe que lo que se pretende medir con esa escala es su actitud, y por eso se llama técnicas directas para las medidas de las actitudes.
3. ¿Cómo se interpreta el índice de semejanza (IS) entre dos elementos? Cuando su valor es 1 o próximo a 1 se dice que existe semejanza.
4. ¿Cómo se halla la puntuación del sujeto en la escala? Sumando el total de las valoraciones dadas por el sujeto a cada ítem de la escala.
5. Consulte al profesor de su asignatura.

8.6. Referencias bibliográficas

- BARBERO GARCÍA, M. I. (1999). *Psicometría II. Métodos de elaboración de escalas*. Madrid: UNED, Unidades Didácticas. (2.^a ed.)
- GONZÁLEZ BLASCO, P. (1994). «Medir en Ciencias Sociales». En Manuel García Ferrando, Jesús Ibáñez y Francisco Alvira (comp.): *El análisis de la realidad social. Métodos y técnicas de investigación*. Madrid: Alianza Universidad (2.^a ed.).
- LIKERT, R. (1974). «A method of constructing an attitude scale». En G. M. Maranell (ed.): *Scaling: A source book of behavioural scientist*. Chicago: Aldine.

- THURSTONE, L. L. (1928). «Attitude can be measured». *American Journal of Sociology* 33 (529-554).
- THURSTONE, L. L. y CHAVE, E. J. (1929). *The measurement of attitudes*. Chicago: University of Chicago Press.
- TORGERSON, W. S. (1958). *Theory and methods of scaling*. New York: John Wiley and Sons.

TEMA 9

Características técnicas de los instrumentos de medida desde la teoría clásica de los tests. Fiabilidad, validez

Arturo González Galán

ESQUEMA

- 9.1.** El concepto de fiabilidad
- 9.2.** Tipos de fiabilidad
 - 9.2.1. Fiabilidad como estabilidad
 - 9.2.2. Fiabilidad como equivalencia
 - 9.2.3. Fiabilidad como consistencia interna
- 9.3.** Aspectos a tener en cuenta al valorar un coeficiente de fiabilidad
 - 9.3.1. Variabilidad y tipo de muestra
 - 9.3.2. Tipo de instrumento
 - 9.3.3. Elección del coeficiente de fiabilidad y diferencias entre sus valores
 - 9.3.4. Valores aceptables para el coeficiente «a» de Crombach
 - 9.3.5. Fiabilidad y unidimensionalidad: suma de elementos
- 9.4.** Un ejemplo del cálculo de la fiabilidad de una escala con SPSS
- 9.5.** El concepto de validez
- 9.6.** Tipos de validez
 - 9.2.1. Validez de contenido
 - 9.2.1. Validez de criterio
- 9.7.** Procedimientos para hallar la validez
- 9.8.** Relación entre validez, fiabilidad y longitud de la prueba
- 9.9.** Ejercicios de autoevaluación
- 9.10.** Referencias bibliográficas

OBJETIVOS

Se espera que el alumno, tras el estudio del tema, consiga los siguientes objetivos de aprendizaje:

- ✓ Comprender los conceptos de fiabilidad y validez y su aplicación al campo de la Educación y la Psicopedagogía.
- ✓ Distinguir los tipos de fiabilidad relativa y los coeficientes que implican
- ✓ Ser capaz de interpretar una salida informática relativo al cálculo de distintos coeficientes de fiabilidad
- ✓ Interpretar correctamente un coeficiente de fiabilidad y las causas que hacen variar su valor
- ✓ Conocer y distinguir las distintas formas de abordar la validez de un instrumento de medida
- ✓ Valorar y comprender las diferencias entre fiabilidad y validez y sus mutuas repercusiones

Introducción

En la asignatura de *Métodos de Investigación en Educación Social*, de primer curso, se pudo estudiar una introducción a los conceptos de fiabilidad (tema 15) y validez (tema 16), de los que recomendamos un rápido repaso. En esta ocasión, nos referiremos de nuevo a estos conceptos desde una vertiente más práctica y aplicada. No le quepa duda al estudiante que todo profesional del campo educativo va a tener que relacionarse, de una forma u otra, con distintos instrumentos de medida, ya sean elaborados por él mismo, ya se encuentren disponibles en el mercado. En cualquier caso, el Educador Social debe tener las competencias suficientes para su construcción y análisis, así como para emitir un informe inteligible sobre las variables medidas en función de los destinatarios del mismo.

En este sentido, no basta con saber utilizar un programa estadístico determinado para el análisis de instrumentos de medida, si no se conoce el proceso completo de elaboración y análisis, se es capaz de hacer un plan de análisis y de interpretar adecuadamente los resultados, con sus posibilidades y limitaciones.

Esta forma de análisis de instrumentos de medición se refiere básicamente a aquéllos que recogen datos cuantitativos. Los más utilizados en nuestro campo son los tests, pruebas objetivas de conocimientos y escalas. Además, el estudio de la fiabilidad, la validez y el correspondiente análisis de ítems se enmarca dentro de lo que conocemos como Teoría Clásica de la Medida, si bien existe otra forma de análisis bajo la denominada Teoría de Respuesta al Ítem (TRI).

9.1. El concepto de fiabilidad

El concepto de fiabilidad está relacionado con la precisión de la medida y su estabilidad en el tiempo, es decir, si a un grupo de sujetos le aplicamos el mismo instrumento de medida en dos o más ocasiones relativamente espaciadas (apuntamos relativamente porque en edad escolar lo normal es que los rasgos evolucionen en el tiempo), diremos que el instrumento es fiable si los sujetos tienden a mantener su misma puntuación o, al menos, quedan ordenados de la misma forma en las distintas aplicaciones.

Aunque generalmente se habla de fiabilidad del instrumento de medida, en sentido estricto esta es una expresión incorrecta. La fiabilidad se calcula a partir de las puntuaciones de una muestra de sujetos en un instrumento. Por tanto, la fiabilidad obtenida se refiere a esa muestra de

sujetos y podría variar si lo aplicamos a una muestra distinta y calculamos de nuevo la fiabilidad. Por eso, la mayoría de los tests estandarizados que encontramos en el mercado, están validados con muestras muy grandes de sujetos representativos de la mayoría de los contextos en los que podamos aplicar dicha prueba.

En este sentido, la variabilidad de la muestra (expresada, por ejemplo, por su desviación típica) es un aspecto crucial a la hora de calcular la fiabilidad. Uno de los objetivos principales de los instrumentos de medida en Psicología y Educación es diferenciar a los sujetos en el rasgo que se está midiendo con el fin de realizar diagnósticos y evaluaciones. Supongamos que aplicamos una prueba objetiva de matemáticas a 200 alumnos de un centro escolar. Lo normal es que encontremos diferencias en las puntuaciones de los sujetos, esto es, que exista variabilidad, heterogeneidad (unos tienen más conocimientos, otros menos). Si la prueba está bien construida, podemos obtener una alta fiabilidad en la prueba, fiabilidad calculada, eso sí, a partir de las puntuaciones de esa muestra de 200 sujetos. Supongamos ahora que esa misma prueba se la aplicamos a un grupo de 50 superdotados. Evidentemente, las diferencias entre las puntuaciones de estos sujetos serán más pequeñas que las que existan en una muestra «normal», ya que aquéllos obtendrán puntuaciones muy altas. Si calculáramos ahora la fiabilidad a partir de esta otra muestra, a buen seguro que la fiabilidad obtenida será mucho más baja y, sin embargo, la prueba es la misma.

En otras palabras, el concepto de fiabilidad está también asociado a las diferencias, es decir, a la capacidad del instrumento para apreciar las diferencias entre los sujetos. Un instrumento que no sea capaz de apreciar las diferencias entre los sujetos, no puede decirse que sea fiable. Por ejemplo, si utilizamos un test de matemáticas de Primaria para los alumnos de Bachillerato, casi todos los alumnos alcanzarían la puntuación máxima, no habría variabilidad en las respuestas y, por tanto, no podríamos decir que ese test es fiable (al menos no para esa muestra, aunque sí lo pueda ser para los alumnos de Primaria).

En este punto podemos enlazar con la otra característica con la que se juzgan los instrumentos: la validez. Como es bien sabido, un instrumento es válido si mide lo que dice medir. Así, si utilizamos un test de resolución de problemas, es eso lo que debe medir (la capacidad para resolver problemas) y no, por ejemplo, la capacidad memorística. Apunta Morales (2002), que «la fiabilidad es condición necesaria pero no suficiente para la validez: si se está midiendo muy mal es casi inútil preguntarse qué se está midiendo, pero se puede medir bien (detectar con precisión diferencias

interindividuales, *fiabilidad*) y no aquello precisamente que pretendemos (*validez*)».

Tradicionalmente, al tratar la teoría clásica de la fiabilidad se distinguen dos enfoques de la fiabilidad: fiabilidad relativa y fiabilidad absoluta. Hasta ahora nos estamos refiriendo a la fiabilidad relativa. *Relativa*, es decir, que en distintas aplicaciones del mismo instrumento los sujetos quedan ordenados del mismo modo, sus *posiciones relativas* son iguales en las distintas aplicaciones (aunque la puntuación no sea exactamente la misma). Generalmente, cuando calculamos o nos referimos a la fiabilidad de un instrumento, se trata de la fiabilidad relativa (veremos posteriormente las distintas formas de calcularla). La *fiabilidad absoluta* está más directamente relacionada con lo que conocemos como error típico de medida y, en consecuencia, con la teoría de la inferencia estadística. Su utilidad fundamental es la estimación de la *puntuación verdadera* de un sujeto en un instrumento o, dicho de otra forma, entre qué puntuaciones es más probable que se encuentre su verdadera puntuación, ya que toda medida (especialmente en las Ciencias Sociales) tiene algún margen de error. Otra utilidad de la fiabilidad absoluta es comprobar si la diferencia entre dos puntuaciones en la misma prueba son tan pequeñas que pueden atribuirse a este margen de error o son lo suficientemente grandes como para concluir que dichas diferencias son estadísticamente significativas.

Veamos un ejemplo:

Supongamos que hemos aplicado una prueba objetiva de matemáticas de 100 ítems a un grupo de sujetos. El sujeto X ha obtenido una puntuación de 47. Nos podemos preguntar: ¿si le hubiera aplicado infinitas veces a este sujeto la misma prueba, habría obtenido siempre la misma puntuación, o quizás su nivel de conocimientos corresponde más bien con una puntuación de 52? (probablemente de la respuesta a esta pregunta dependa un aprobado o un suspenso). Pues bien, esto dependerá en buena parte de la fiabilidad de la prueba, de su precisión. A partir de la fiabilidad y de la lógica de la inferencia, se puede llegar a estimar que en el 99 % de los casos de esas hipotéticas infinitas aplicaciones, su puntuación se encontraría entre 44 y 50. Esto, sin duda, puede ser una buena ayuda para tomar la decisión sobre la calificación definitiva.

9.2. Tipos de fiabilidad

Decíamos anteriormente que la fiabilidad relativa se refiere al mantenimiento de las mismas posiciones relativas de un grupo de sujetos en distintas aplicaciones del mismo instrumento. Seguramente, si el alumno piensa qué índice descriptivo indica que en dos variables las puntuaciones

tienden a estar ordenadas del mismo modo, encontrará rápidamente la respuesta: el coeficiente de correlación. Así pues, la fiabilidad, en sus distintos tipos, siempre se va a calcular a partir de coeficientes de correlación (generalmente de Pearson). Recordemos también, como decíamos antes, que la fiabilidad implica variabilidad, capacidad de diferenciación. El coeficiente de correlación expresa la variabilidad conjunta entre dos variables. En consecuencia, si el instrumento no detecta bien las diferencias interindividuales no habrá variabilidad, bajará el coeficiente de correlación y, por tanto, la fiabilidad.

Existen tres formas para calcular la fiabilidad relativa, nos referimos a la fiabilidad como estabilidad, la fiabilidad como equivalencia y la fiabilidad como consistencia interna, siendo esta última la más usada.

9.2.1. *Fiabilidad como estabilidad*

Ha quedado patente que la fiabilidad no es un concepto unívoco, lo que ha dado lugar a estas tres formas para su cálculo. La fiabilidad como estabilidad hace referencia al mantenimiento en el tiempo de las posiciones relativas de las puntuaciones del grupo de sujetos en el mismo instrumento (esto es, los que sacaron las mejores puntuaciones siguen siendo los mejores y los que sacaron las peores conservan los peores puestos). A este método se le suele denominar test-retest, es decir, la aplicación del instrumentos en dos o más ocasiones dejando un intervalo de tiempo adecuado entre las aplicaciones. ¿Cómo se calcula entonces la fiabilidad? Comprobando simplemente si se conservan dichas posiciones relativas entre dos cualesquiera aplicaciones. Por tanto, calculando la correlación entre dichas puntuaciones. Veamos un ejemplo sencillo: puntuaciones de 5 sujetos en una prueba objetiva de 25 ítems (puntuación posible, por tanto, de 0 a 25).

TABLA 9.1.

Sujeto	Test	Retest
1	19	20
2	14	12
3	23	21
4	11	9
5	15	15
Matriz de datos		

TABLA 9.2. Estadísticos descriptivos

	Media	Desviación típica	N
TEST	16.4000	4.66905	5
RETEST	15.4000	5.12835	5

TABLA 9.3. Correlaciones

		TEST	RETEST
TEST	Correlación de Pearson	1	.963**
	Sig. (bilateral)	.	.009
	N	5	5
RETEST	Correlación de Pearson	.963**	1
	Sig. (bilateral)	.009	.
	N	5	5

** La correlación es significativa al nivel 0,01 (bilateral).

Generalmente, cuando calculemos la fiabilidad de un instrumento, lo haremos con muestras más grandes de sujetos. Aquí mostramos la salida del programa estadístico SPSS para calcular la correlación entre dos variables. En primer lugar hemos pedido los estadísticos descriptivos básicos (media y desviación típica) observando que los datos son bastante parecidos. Después se calcula la correlación ($r = 0.963$), resultando de magnitud muy grande, casi perfecta. Como se ve, las puntuaciones no son exactamente las mismas, pero si ordenamos los sujetos de mayor a menor en una variable (aplicaciones), y las puntuaciones obtenidas en la otra variable, veremos que sus posiciones relativas son las mismas. En definitiva, su fiabilidad es muy alta: 0,963.

En la práctica, este procedimiento es muy poco usado, ya que implica tiempo y dinero, obligando a aplicar dos veces el mismo instrumento en momentos distintos.

9.2.2. *Fiabilidad como equivalencia*

Otra forma de calcular la fiabilidad relativa, también en desuso, es la denominada como formas paralelas, procedente del concepto de fiabilidad como *equivalencia*. ¿Por qué es un procedimiento poco utilizado? Simplemente, porque exige la construcción de dos pruebas distintas para medir el mismo rasgo. Esto implica, de nuevo, más tiempo y recursos: no sólo tenemos dos aplicaciones diferentes como antes, sino además

la elaboración de dos pruebas distintas y equivalentes. Si tratáramos, por ejemplo, de un test de conocimiento de palabras inglesas para 2.º de Bachillerato, podríamos hacerlo con 50 palabras. Un test paralelo o equivalente sería uno de otros 50 ítems que respondiera a los mismos contenidos, objetivos, dificultad, etc. que el anterior. Si cogiéramos la tabla 9.1 donde mostrábamos las puntuaciones del test y el retest, bastaría ahora con sustituir los títulos por Forma A y Forma B. La correlación entre las dos formas sería el coeficiente de fiabilidad.

9.2.3. *Fiabilidad como consistencia interna*

La tercera forma de calcular la fiabilidad procede del concepto de fiabilidad como *consistencia interna*. Generalmente, cuando se habla de fiabilidad de un instrumento de medida se hace referencia a esta concepción. La consistencia interna indica en qué medida los distintos ítems que forman parte de un instrumento son consistentes, es decir, tienen que ver unos con otros, miden el mismo rasgo o característica. Se refiere (aunque no se prueba) a la unidimensionalidad del rasgo medido, esto es, que si medimos la actitud hacia los inmigrantes, todos los ítems del instrumento poseen o son parte en cierta medida de ese rasgo.

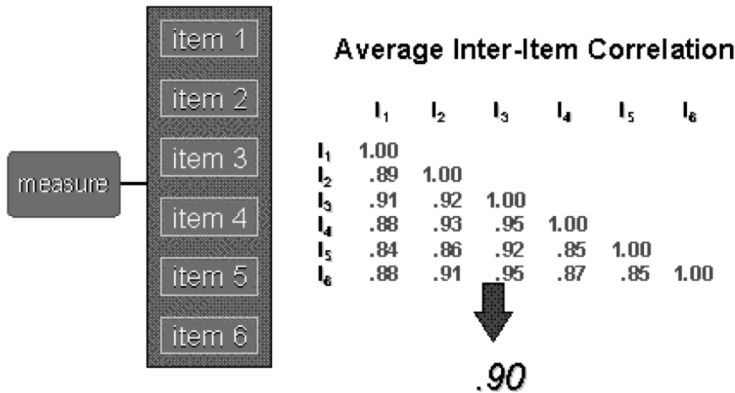
Existen distintos procedimientos para el cálculo de la fiabilidad como consistencia interna, si bien el más utilizado, con diferencia, es el α de Cronbach. Veámoslos:

- **Dos mitades (*split-half*):** Es un procedimiento similar al de las formas paralelas, pero utilizando sólo un instrumento. Dividimos los ítems en dos conjuntos (aleatoriamente, pares-impares...) y calculamos la puntuación en cada una de las mitades. La correlación entre las puntuaciones obtenidas en cada mitad indicaría la fiabilidad del instrumento.

Cuando dividimos un test en dos mitades y calculamos la correlación entre las puntuaciones obtenidas en cada mitad, se podrá comprender que se puede perder cierta variabilidad en las puntuaciones que si hubiéramos utilizado la prueba completa. Por ejemplo, si tenemos una prueba de 50 ítems, el cálculo de la fiabilidad como dos mitades implicaría calcular la correlación entre la puntuación obtenida en cada parte de 25 ítems. Por tanto, la puntuación que podría obtener un sujeto en cada mitad variará entre 0 y 25. Si hubiéramos utilizado el procedimiento de formas paralelas, estaríamos correlacionando dos series de puntuaciones que podrían variar entre 0 y 50. Como se sabe, cuanto mayor es la variabili-

dad, mayor puede ser el coeficiente de correlación. Por esta razón, cuando se utiliza el procedimiento de las dos mitades, puede utilizarse alguna fórmula de corrección al alza de este coeficiente de correlación (o fiabilidad), como las de Spearman-Brown, Rulon y Guttman.

- **Promedio de la correlación inter-ítems:** Bastaría con calcular la correlación entre todos los ítems y calcular la media. Tal y como indica Trochim (2000) gráficamente:



- **Alfa (α) de Cronbach:** La fiabilidad calculada mediante el procedimiento α de Cronbach puede considerarse una extensión del procedimiento de las dos mitades. De hecho, es una estimación de la media de todos los coeficientes de correlación que podrían obtenerse en el caso de dividir el instrumento en todas las mitades posibles. Sería, por tanto, como repetir tantas veces como fuera posible el procedimiento de las dos mitades y luego calcular la media de todos los coeficientes de correlación resultantes. Mediante la fórmula de Cronbach se realiza rápidamente esta estimación. Es importante recordar al lector que la fórmula de Cronbach (1951) y la de Kuder-Richardson 20 (1937) son prácticamente iguales. Los nombres distintos se deben a que los autores difieren en sus modelos teóricos y los desarrollaron en tiempos distintos. La fórmula de Kuder-Richardson 20 se utiliza para ítems dicotómicos y la de Cronbach para ítems continuos. No obstante, programas de análisis estadístico como el SPSS muestran el procedimiento *Alfa* que puede utilizarse tanto para ítems dicotómicos como continuos.

9.3. Aspectos a tener en cuenta al valorar un coeficiente de fiabilidad

La interpretación de resultados es un ejercicio básico para todo profesional en las Ciencias Sociales. Mucho más importante resulta comprender bien que saber calcular, ya que, hoy en día, todas las operaciones las realizamos con ordenadores. Sin embargo, el ordenador es una herramienta al servicio de nuestra inteligencia, por lo que antes de realizar ningún estudio se debe tener un buen diseño de lo que se quiere conseguir.

9.3.1. Variabilidad y tipo de muestra

Como hemos visto a lo largo del capítulo, la fiabilidad es un concepto asociado a la variabilidad de las puntuaciones, ya que está basado en los coeficientes de correlación. En consecuencia, no es posible obtener altos índices de fiabilidad cuando trabajamos con muestras muy homogéneas. Por ejemplo, si aplicamos un test de inteligencia para niños a adultos, no encontraremos casi variabilidad en las puntuaciones obtenidas por los adultos, ya que todos se situarán en torno a la puntuación máxima. Por tanto, un test así utilizado no nos sirve para encontrar diferencias interindividuales, no habría variabilidad, la muestra sería muy homogénea debido al instrumento utilizado, por lo que dicho instrumento no sería fiable para esa muestra. Sin embargo, sí es fiable si lo aplicamos a la edad para el que fue diseñado. Algo similar ocurre cuando la muestra es naturalmente homogénea, por ejemplo, si pasamos el test a un grupo de superdotados.

En el mismo sentido, es bien sabido que el tamaño de la muestra tiene repercusiones en la variabilidad de las muestras, siendo más probable encontrar mayor variabilidad —y por tanto, mayor fiabilidad— en muestras grandes que en muestras pequeñas.

En definitiva, cuando nos encontremos con un coeficiente de fiabilidad inesperadamente bajo, debemos pensar si una causa posible puede ser la homogeneidad (poca variabilidad) de la muestra utilizada, lo cual puede deberse tanto al tipo de muestra como al tipo de instrumento utilizados. Nos bastará calcular la desviación típica para comprobar dicha hipótesis.

9.3.2. Tipo de instrumento

En las Ciencias de la Educación los instrumentos de medida no siempre tienen la misma finalidad que en Psicología. Fíjese el lector que uno

de los objetivos de las pruebas psicológicas es diferenciar a los sujetos en los rasgos conductuales o comportamentales medidos. Sin embargo, en Educación, cuando hablamos, por ejemplo, del rendimiento académico de los estudiantes, al menos idealmente, no buscamos la variabilidad de las puntuaciones. Nos gustaría que todos los estudiantes alcanzaran la máxima puntuación. Evidentemente, los distintos factores que inciden en el rendimiento provocan en la práctica diferencias notables.

No obstante, la evaluación criterial es una práctica pedagógicamente recomendable y de uso creciente. En este sentido, si decidiéramos aplicar una prueba objetiva para realizar una evaluación del rendimiento, podríamos diseñar una parte (por ejemplo, el 50 % de los ítems) para evaluar los objetivos y contenidos mínimos de la materia o simplemente, se podría construir una prueba única para evaluar la adquisición de los contenidos mínimos. Evidentemente, como argumentábamos en el punto anterior, si construimos los ítems a partir de unos criterios mínimos, sería esperable que las puntuaciones obtenidas fueran más altas para todos los alumnos y, por tanto, se homogeneizan las puntuaciones de la muestra. La consecuencia sería una reducción del coeficiente de fiabilidad. ¿Podríamos decir en este caso que la prueba no está bien construida porque tiene una fiabilidad baja? La respuesta es negativa, ya que la misma naturaleza de la prueba no persigue diferenciar a los sujetos y, por tanto, podemos tener una baja fiabilidad y ser una prueba bien construida.

9.3.3. Elección del coeficiente de fiabilidad y diferencias entre sus valores

La elección del coeficiente de fiabilidad entre los distintos coeficientes dependerá, en ocasiones, del tipo de enfoque (estabilidad, equivalencia, consistencia interna) o de las preferencias personales. Actualmente, el coeficiente α de Cronbach es el más utilizado y el que suele reflejarse en los artículos y otras publicaciones científicas. En otras ocasiones se puede presentar más de un coeficiente de fiabilidad. Por ejemplo, si estamos examinando a dos grupos de sujetos sobre la misma asignatura con exámenes distintos (porque se examinan a horas diferentes) podríamos calcular la fiabilidad como la correlación entre las dos formas (formas paralelas) y calcular además el α de Cronbach en cada forma del test.

En cuanto a los valores, generalmente la modalidad test-retest (estabilidad) aporta valores más bajos que las formas paralelas o la consistencia interna.

9.3.4. Valores aceptables para el coeficiente « α » de Cronbach

No existe un criterio ampliamente reconocido para afirmar cuando un coeficiente de fiabilidad es alto o bajo. O, más bien, cuándo es lo suficientemente alto como para considerar la prueba fiable. El coeficiente de fiabilidad varía entre 0 y 1 (aunque podrían encontrarse valores negativos), siendo recomendable que se encuentre lo más cerca de 1 posible. Algunos autores dan valores muy altos —por encima de 0.85— para considerar que el valor es «bueno» o aceptable, independientemente del tipo de prueba. Nosotros coincidimos más con la visión de Morales (2002) en la que distingue los valores aceptables en función de los objetivos de la prueba:

TABLA 9.4.

	Toma de decisiones sobre individuos	Descripción de grupos Feedback a un grupo	Investigación teórica Investigación en general
0.85 o mayor	sí	sí	sí
de 0.60 y 0.85	cuestionable	sí	sí
inferior a 0.60	no	cuestionable	sí, cuestionable

Añadiríamos, además, que se debe ser más exigente con las pruebas tipo test (aptitudes, rendimiento) que con las tipo escala (actitudes, valores).

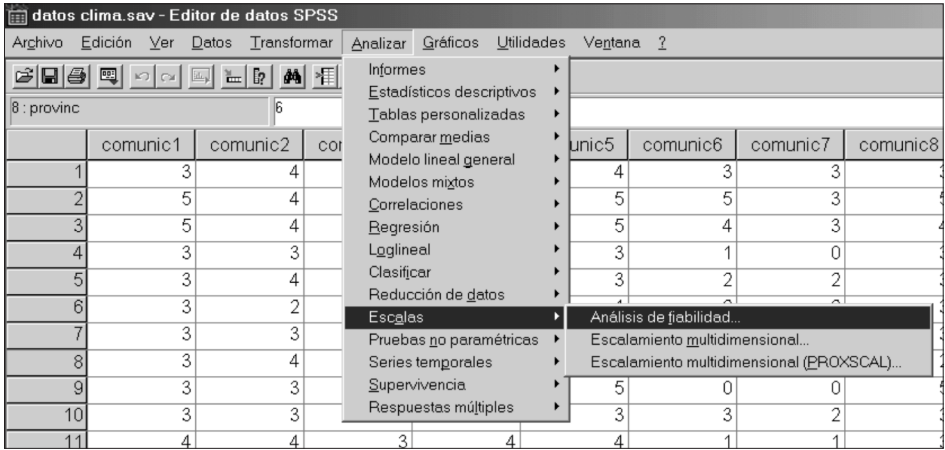
9.3.5. Fiabilidad y unidimensionalidad: suma de elementos

Una de las consecuencias que se suele obtener después de comprobar que un instrumento tiene una alta fiabilidad es considerar que todos los ítems miden el mismo rasgo. Por ejemplo, si estamos midiendo la actitud hacia las matemáticas con una escala de 20 ítems, si obtenemos una alta fiabilidad podríamos interpretar que todos los ítems miden el mismo rasgo, es decir, que el instrumento es unidimensional. Si esto es así, podríamos sumar sin grandes problemas la respuesta dada a cada uno de los ítems para obtener la puntuación total en dicha escala. Sin embargo, esta interpretación, utilizada en sentido estricto, es incorrecta. La fiabilidad no prueba la unidimensionalidad de la prueba, sino que, como apunta acertadamente Morales (2002), la apoya. Para probar dicha unidimensionalidad tenemos otras herramientas más adecuadas, como el análisis factorial (preferentemente confirmatorio, aunque también se utiliza el exploratorio). La unidimensionalidad de la prueba no puede

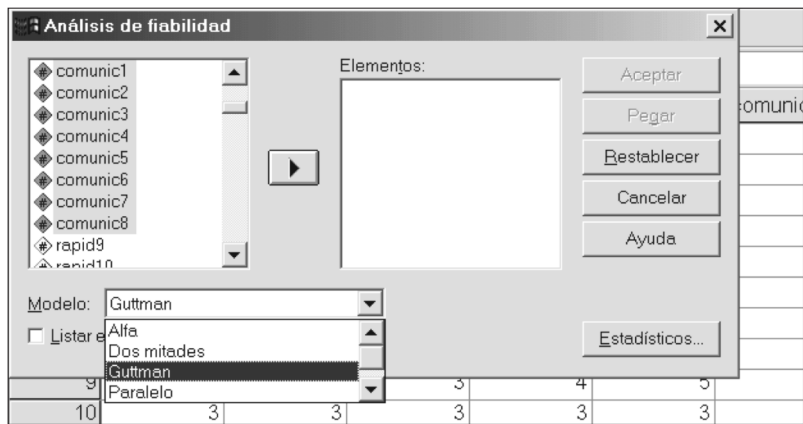
consistir sólo en las intercorrelaciones empíricas entre los ítems, sino que debe estar sustentada teóricamente.

9.4. Un ejemplo del cálculo de la fiabilidad de una escala con SPSS

Veamos un ejemplo del cálculo de la fiabilidad con el programa SPSS (versión 11). Mostramos un ejemplo de una investigación (Martín Bris, González Galán *et al.*) sobre el clima de trabajo en los centros escolares. Se trata de calcular la fiabilidad de una subescala de un instrumento (escala tipo Likert de 6 puntos) para medir el clima de trabajo. Esta subescala se refiere al grado de comunicación existente entre los distintos miembros de la comunidad educativa. Consta de 8 ítems. Lo primero que necesitamos es una matriz de datos donde estén las respuestas de los sujetos a esos 8 ítems:



Para analizar la fiabilidad de cualquier instrumento siempre debemos ir a *analizar* → *escalas* → *análisis de fiabilidad*. Una vez que aceptemos esta opción nos aparece la ventana de la siguiente imagen.



Como se ve, debemos seleccionar los ítems que queremos analizar y elegir el modelo de análisis (Alfa, Dos mitades, Guttman, etc.). Veamos la salida más sencilla que nos proporciona utilizando estos tres modelos.

Esta es la salida utilizando el procedimiento o modelo Alfa de Cronbach:

TABLA 9.5.

R E L I A B I L I T Y A N A L Y S I S - S C A L E (A L P H A)	
Reliability Coefficients	
N of Cases = 928.0	N of Items = 8
Alpha = .9070	

Como vemos en la salida, primero indica que está realizando el análisis de la fiabilidad (*reliability*) por el modelo alfa de Cronbach (*alpha*). Informa sobre el número de casos de la muestra (en este caso eran 928 profesores) y el número de ítems de la escala (ocho). Finalmente proporciona el coeficiente de fiabilidad, $\alpha = .907$, es decir, una muy buena fiabilidad.

Mostramos a continuación la salida del SPSS tras utilizar el procedimiento de las dos mitades:

TABLA 9.6.

R E L I A B I L I T Y A N A L Y S I S - S C A L E (S P L I T)	
Reliability Coefficients	
N of Cases = 928.0	N of Items = 8
Correlation between forms = .7770	Equal-length Spearman-Brown = .8745
Guttman Split-half = .8724	Unequal-length Spearman-Brown = .8745
4 Items in part 1	4 Items in part 2
Alpha for part 1 = .8502	Alpha for part 2 = .8412

Tras la información común, el programa informa de la correlación simple entre las dos mitades ($r = 0.777$), realizando posteriormente las correcciones de Spearman-Brown ($\alpha = 0.8745$) y Guttman ($\alpha = 0.8724$). Como información adicional, añade el α de Cronbach en cada una de las dos mitades (mitad 1, $\alpha = 0.8502$; mitad 2, $\alpha = 8412$).

Finalmente, comentamos la opción que denomina *Guttman*, que aporta el intervalo de confianza para la fiabilidad de la escala, es decir, entre qué valores se encuentra la fiabilidad verdadera de la escala:

TABLA 9.7.

R E L I A B I L I T Y A N A L Y S I S - S C A L E (G U T T M A N)		
N of Cases = 928.0		
Reliability Coefficients		8 items
Lambda 1 = .7936	Lambda 2 = .9097	Lambda 3 = .9070
Lambda 4 = .8724	Lambda 5 = .8832	Lambda 6 = .9142

Como vemos, el intervalo de confianza para la fiabilidad estará entre 0.79 y 0.91.

9.5. El concepto de validez

Como hemos apuntado anteriormente, es discutible afirmar que la fiabilidad sea más importante que la validez o viceversa. Se trata de dos características a las que siempre debemos prestar atención tanto en la construcción como en el uso de instrumentos de medida. Decimos también en el *uso* porque tanto el Educador Social, Pedagogo, Psicopedagogo, Psicólogo... harán uso de instrumentos de medida existentes en el mercado, recibirán «visitas» de los comerciales de las edito-

riales de tests y otras pruebas de las que nos hablarán maravillas. Por tanto, es necesario tener las habilidades suficientes para distinguir lo que es un buen instrumento de lo que no lo es. Para ello, es indispensable revisar el apartado de «características técnicas» y observar atentamente los apartados de fiabilidad, validez y muestra, pues recordemos que tanto la fiabilidad como la validez son características de las muestras de sujetos más que de los propios instrumentos, aunque ordinariamente las consideremos características de los instrumentos.

Nos podrían tratar de vender un peso para medir la altura de nuestros alumnos. Desde luego —afirmaría el comercial— es un instrumento de medida muy fiable (y tendría razón), que permite discriminar muy bien las diferencias entre los sujetos. Sin embargo, no sería válido, no mide lo que dice medir, mide bien, pero mide otra cosa diferente a lo que nos interesa. Esta es la cuestión de la validez, y se nos antoja como un problema de más difícil comprobación que la fiabilidad.

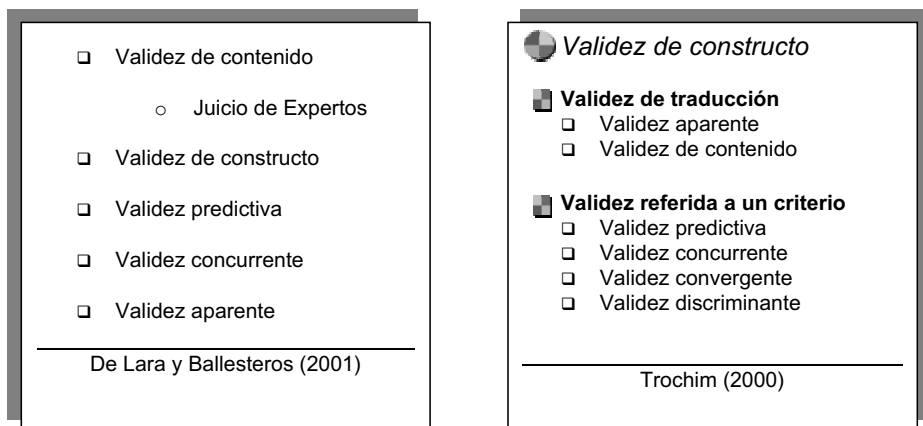
En las Ciencias de la Educación, en Psicopedagogía, la cuestión de la medida es mucho más complicada que en las Ciencias Naturales, donde 1 metro es 1 metro y 1 litro es 1 litro. En nuestra ciencias trabajamos con variables más complejas y abstractas, que venimos a denominar *constructos*. La inteligencia, la motivación, las aptitudes, las actitudes, los conocimientos... ¿qué son y cómo se miden? Todavía no se ha inventado el «casco psicométrico» que al ponerlo sobre la cabeza de un sujeto nos indica en números digitales su cociente intelectual, como si de un termómetro se tratara. En nuestras ciencias, la teoría juega un papel fundamental en la medición, de modo que sólo podemos llegar a una buena medida cuando hemos definido científicamente el significado de los constructos. Así, por ejemplo, la inteligencia es un constructo, no es una propiedad directamente mensurable. En consecuencia, estudiosos e investigadores han ido definiendo las características que conlleva la inteligencia (por ejemplo, razonamiento numérico, verbal, espacial) hasta llegar a las características o variables directamente observables y mensurables. Por eso, en los tests de inteligencia podemos encontrar ítems relativos a operaciones numéricas, listas de palabras o rotación espacial de figuras, como manifestaciones observables de la inteligencia.

Podemos deducir, por tanto, que la validez de un instrumento está muy relacionada con la naturaleza y definición teórica del constructo a medir y con la *operacionalización* del mismo, esto es, con los ítems elegidos para captar sus manifestaciones observables. Por esta razón, la validez de constructo y la validez de contenido se erigen como las más importantes a la hora de comprobar la validez de un instrumento de medida.

9.6. Tipos validez

El alumno pudo estudiar en curso anteriores los distintos tipos de validez. Recordemos rápidamente la clasificación apuntada por De Lara y Ballesteros (2001):

Tipos de validez



Aportamos también la innovadora clasificación de Trochim (2000) y coincidimos con él en que cualquier tipo de comprobación de la validez de un instrumento persigue finalmente la validez de constructo. Trochim apuesta por considerar la validez de constructo para referirse al caso general de traducir un constructo en una operacionalización. De este modo si hablamos del constructo «habilidad matemática», existe una teórica población de ítems que permiten su operacionalización. La validez de constructo nos dirá cuán adecuada ha sido nuestra selección de la muestra de ítems para dar la definición operativa y mensurable del constructo. El resto de formas de entender la validez no serían más —a juicio de Trochim— que distintas formas de mostrar los diversos aspectos de la validez de constructo.

La forma probablemente más correcta de probar la validez de constructo consiste en una técnica estadística compleja denominada análisis factorial confirmatorio. Para realizar este tipo de análisis es fundamental partir de una definición teórica del constructo muy elaborada (así debería ser siempre, aunque lamentablemente no se cumple en la práctica). Después de aplicar el instrumento a una muestra de sujetos suficiente y representativa se prueba en qué medida nuestro modelo de medida (es decir, la agrupación de los ítems del instrumento) se ajusta a los datos. El resultado del análisis es un índice de ajuste que determina la bondad del

ajuste del modelo de medida elaborado (es decir, de la estructura teórica del instrumento de medida) a los datos recogidos en la muestra. Hay distintos programas informáticos que proporcionan este tipo de análisis, como el AMOS (módulo del SPSS) o el EQS¹. No obstante, algunos autores se aproximan a la validez de constructo a través de análisis factoriales exploratorios², pero en nuestra opinión esto sólo supone una aproximación incompleta a este tipo de validez.

En cualquier caso, aunque recomendaríamos la técnica del análisis factorial confirmatorio, más novedosa que el resto de procedimientos para hacernos una idea de la validez del instrumento de medida, es conveniente conocer el resto de técnicas para probar la validez, que pueden utilizarse de modo complementario a la anterior.

9.6.1. *Validez de contenido*

Como hemos visto en la clasificación de Trochim, agrupa la validez aparente y la validez de contenido en lo que denomina originalmente como *validez de traducción*. Con esto se refiere a la forma como se ha traducido o trasladado el constructo teórico a su operacionalización, es decir, si los ítems que definen el constructo (por ejemplo, «actitud hacia la inmigración») son realmente una buena representación de dicho constructo (los ítems indican una buena o mala actitud hacia la inmigración). Para ello distingue dos formas conocidas: La validez aparente y la validez de contenido. La validez aparente es la más básica y sencilla —y pobre, en el sentido de que no garantiza la validez de las medidas—. Se refiere simplemente a la apariencia del instrumento, esto es, si parece que mide lo que dice medir, bajo un juicio subjetivo (lo cual no quiere decir que no se tenga que cumplir, el instrumento debe tener una «buena apariencia»). Por otra parte, la validez de contenido se refiere específicamente al contenido de los ítems que hemos seleccionado para aportar la definición operativa del constructo. Como ya hemos apuntado, un constructo es una variable abstracta que puede ser operacionalizada a través de numerosos ítems y que el instrumento debe ser una buena muestra de esta población de ítems. Para conseguirlo se suele recomendar construir la llamada *tabla de especificaciones* (un cuadro de doble entrada en la que se especifican los contenidos, áreas

¹ Puede verse un ejemplo de su aplicación en González Galán (2000 y 2004).

² Ver, por ejemplo, GARCÍA-LEÓN *et al.*, donde se realiza validez de constructo mediante análisis factorial exploratorio y validez convergente en un cuestionario de agresividad. (http://www.ucm.es/info/Psi/docs/journal/v5_n1_2002/art45.pdf)

y funciones por una parte y objetivos o tareas por otra, en relación con el constructo) y acudir al *juicio de expertos* para la revisión del instrumento y la identificación de lagunas o errores y propuestas de mejora.

9.6.2. *Validez de criterio*

Tenemos, por otra parte, la *validez referida a un criterio*, es decir, cuando existe un criterio de referencia que nos posibilitará saber si nuestro instrumento mide lo que dice medir en teoría o, si por el contrario, una vez recogidos los datos tras aplicar el instrumento, se comprueba que se están midiendo cosas diferentes a las pretendidas (en vez de actitud hacia la inmigración se está midiendo sociabilidad, por ejemplo). Por tanto, en los distintos tipos de validez referida a un criterio, lo que cambia precisamente es el criterio de validez.

En esta línea, seguramente el tipo de validez de mayor utilidad en el diagnóstico en Educación y Psicología sería el de la validez predictiva. Una prueba tiene validez predictiva cuando tiene la capacidad de predecir la respuesta de los sujetos en un rasgo determinado. Por ejemplo, un test de capacidad verbal puede tener validez predictiva sobre los resultados en la calificación de Lengua Española al final de curso. Como es habitual, se dirá que el instrumento es tanto más válido cuanto mayor sea su correlación con el criterio.

La validez concurrente está en la línea de la predictiva, pero el tiempo que pasa entre la aplicación de la prueba y la aplicación del criterio es muy corto o incluso pueden recogerse simultáneamente³.

La validez convergente se refiere a la relación existente entre nuestro instrumento con otro ya validado que mide el mismo constructo. Por tanto, si estamos midiendo la habilidad verbal, nuestro test tendría validez convergente si habiendo aplicado al mismo grupo otro test existente de habilidad verbal, encontramos una alta correlación entre ambos tests. La validez discriminante se centra en el aspecto contrario. Si estamos midiendo habilidad verbal y lo correlacionamos con los resultados en un test de razonamiento lógico, tendríamos que esperar una baja correlación (suponiendo, claro está, que teóricamente no haya relación entre dichas

³ Trochim apunta, sin embargo, que la validez concurrente evalúa la capacidad del instrumento para diferenciar grupos de sujetos que teóricamente tienen que ser distintos. Así, si nuestro instrumento diagnostica a los maniacos-depresivos, debería distinguirlos de los esquizofrénicos paranoides.

variables, ya que una alta correlación no significa necesariamente que se mida lo mismo).

9.7. Procedimientos para hallar la validez

Los procedimientos para hallar la validez de un instrumento son relativamente sencillos. Cuantas más formas utilicemos para probar la validez, más seguridad tendremos de que nuestro instrumento es válido, que mide lo que dice medir.

En primer lugar, si es un instrumento de elaboración propia, recomendamos partir siempre de una buena definición teórica del rasgo o constructo que queremos medir, acudiendo a la revisión de la bibliografía y otras fuentes pertinentes. Si es un instrumento disponible en el mercado, debemos cerciorarnos de que incluye fundamentación teórica seria en su manual sobre las características técnicas del instrumento de medida.

En segundo lugar, debemos referirnos a la validez aparente y de contenido. Para ello seguiremos todas las recomendaciones sugeridas en la construcción de instrumentos de medida y redacción de ítems, cuidando la longitud de la prueba y el nivel de redacción (adaptados a la edad de los sujetos) y tratando también de buscar una maquetación agradable. Si hemos realizado bien el primer paso relativo a la definición teórica, será relativamente fácil construir la matriz de especificaciones para seleccionar los ítems que compondrán el instrumento. Para ambos tipos de validez conviene elegir a algunos expertos (profesores universitarios especialistas en el tema, orientadores que tengan experiencia práctica, etc.) que nos den un juicio de nuestro instrumento.

El resto de procedimientos exigen ya cálculos numéricos que se reducen a coeficientes de correlación o parten de dichos coeficientes. Empecemos por los más sencillos: Para comprobar la validez convergente, se necesita la aplicación simultánea de nuestro instrumento y de otro ya validado que mida lo mismo. En ocasiones, se ha incluido en la prueba a validar una subescala de un test validado como criterio de validez por no alargar excesivamente el tiempo de aplicación (Moraleta, González Galán y García-Gallo, 1999). Tendremos, por tanto, dos series de puntuaciones para el mismo grupo de sujetos (prueba X y prueba Y). El cálculo de la correlación entre esas dos series de puntuaciones proporciona el índice de validez. Pero, ¿cuándo consideramos bueno el índice de validez? Desgraciadamente, no se puede dar una respuesta exacta a esta pregunta. En principio, cuanto más alto, mejor, pero basta con

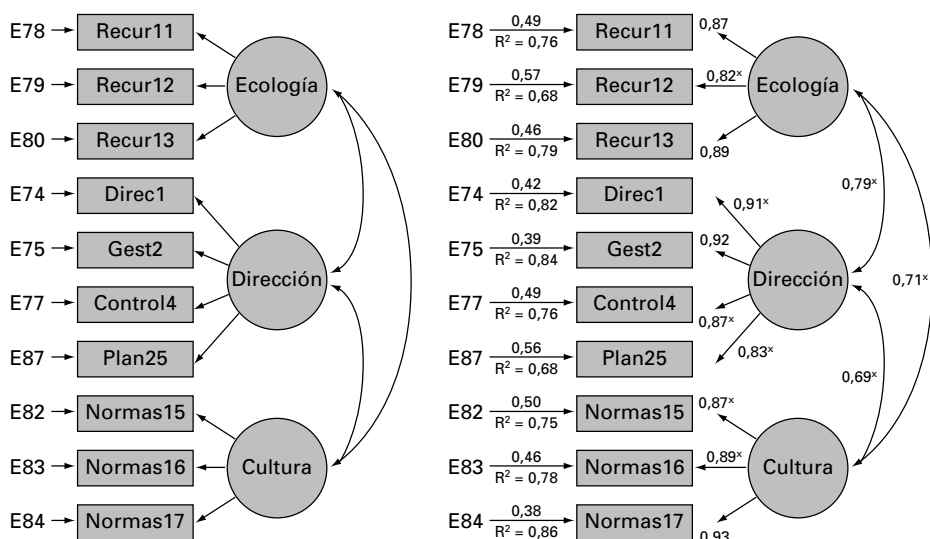
encontrar una relación aceptable que nos indique que no nos alejamos en exceso del constructo que deseamos medir.

Como se podrá deducir, la validez discriminante se calcula también a partir de coeficientes de correlación entre dos pruebas pero que, a diferencia de la validez convergente, miden cosas diferentes. Por tanto, debemos esperar correlaciones bajas (inferiores, por ejemplo, a 0,3). Este tipo de validez suele utilizarse en investigaciones donde se quiere demostrar la especificidad de una prueba para medir un rasgo frente a otras distintas, pero no es necesario incluirla siempre que tratemos de probar la validez de un instrumento.

Tanto en la validez predictiva como en la concurrente, ambas con finalidad diagnóstica, el índice de validez se calcula hallando la correlación entre la prueba construida (predictor, en la terminología de la regresión lineal) y la variable sobre la que se quiere realizar el diagnóstico (criterio). En este caso, claramente buscamos correlaciones cuanto más altas mejor, aunque debemos ser conscientes que en Educación es difícil realizar buenos diagnósticos con un solo predictor, es decir, que no es fácil encontrar correlaciones muy altas entre dos variables (digamos, correlaciones superiores a 0,8, lo que significa un porcentaje de varianza explicada superior a $r^2 = 64\%$). En consecuencia, no podemos hallar este coeficiente de validez hasta que después de pasado un tiempo de la aplicación de nuestro instrumento, hemos recogido los datos del mismo grupo de sujetos en la variable que queríamos predecir. Este es uno de los inconvenientes de estos dos tipos de validez, que no podemos saber su valor hasta que no aplicamos el criterio. ¿Qué sucede, entonces, si encontramos una baja correlación? Sencillamente, que nuestro instrumento carece de la validez predictiva necesaria, ha pasado el tiempo y tenemos que idear un nuevo instrumento para nuestros propósitos.

Finalmente, la forma para calcular la validez de constructo en sentido estricto, es la más compleja y desborda el contenido de esta unidad didáctica. No obstante, vamos a poner un ejemplo sencillo. La técnica se denomina análisis factorial confirmatorio. Supongamos que hemos construido una sencilla escala tipo Likert de 10 ítems para medir el clima de trabajo de un centro educativo. Según hemos derivado después de revisar la literatura de investigación, hay tres factores relacionados (subescalas) que configuran el clima de trabajo: ecología, dirección y cultura, cada uno de los cuales los hemos operacionalizado con tres ítems, excepto dirección que tiene cuatro. Pues bien, a partir de nuestra estructura teórica esperamos una agrupación de los ítems en cada uno de esos tres factores, tal como los hemos definido. La estructura teórica sería la que

aparece en el diagrama de la izquierda y los resultados empíricos en el de la derecha:

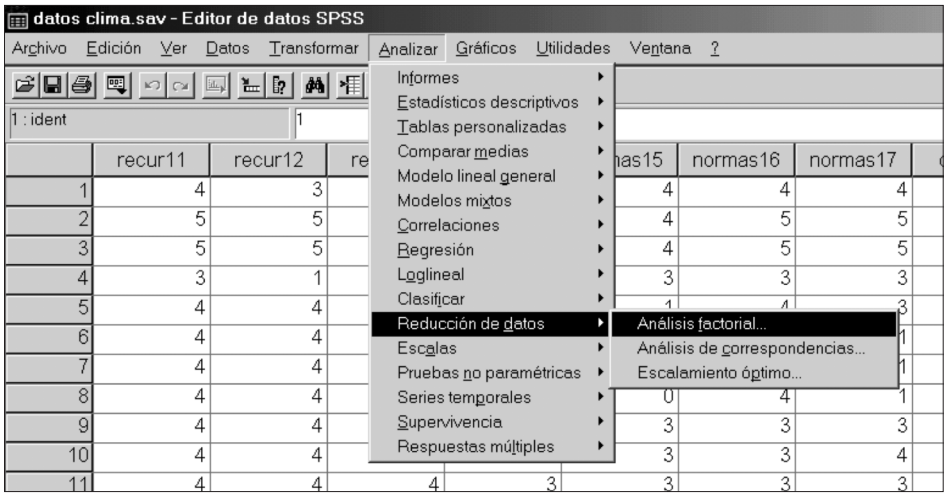


Las circunferencias indican los factores y los rectángulos son las variables medidas, es decir, los ítems del instrumento de medida. Como puede verse en el diagrama de la izquierda, especificamos qué ítems deben definir cada factor (flechas rectas que salen de cada factor a los ítems). Las flechas curvas entre los factores indican que esperamos correlación entre ellos. En el diagrama de la derecha nos dan los valores correspondientes al diagrama hipotético de la izquierda, para lo que se utilizó una muestra de 1.024 profesores (González Galán, 2000). Finalmente, el programa proporciona un índice de ajuste, que es lo que indica si tu modelo de medida está bien construido, es decir, si existe validez de constructo. Este valor debe ser lo más cercano posible a uno, aceptándose valores superiores a 0,9. En este caso, el índice de ajuste RCFI fue de 0,97, indicando una buena validez de constructo, un buen ajuste del modelo teórico a los datos.

Como decíamos en un epígrafe anterior, algunos autores se aproximan a la validez de constructo utilizando análisis factorial exploratorio. Al igual que en el caso anterior se parte de la matriz de correlaciones entre los ítems (o matriz de varianzas-covarianzas). La utilidad básica del análisis factorial consiste en reducir la matriz de correlaciones a un número menor de factores que expliquen la información contenida en dicha matriz de forma más resumida. Cada factor sería un constructo o variable abstracta al que se le da nombre en función de los ítems que han satura-

do en él. La diferencia entre el análisis factorial exploratorio y confirmatorio es que en el primero se parte de la correlación matemática entre los ítems para configurar los factores, mientras que en el segundo es el investigador el que, basado en la teoría, determina qué ítems deben saturar en cada factor, y pone esta hipótesis a prueba.

En cualquier caso, exponemos cómo se haría esta aproximación mediante análisis factorial exploratorio con el programa SPSS. Como siempre, partimos de nuestra matriz de datos y pedimos la opción oportuna:



Cuando se abre la ventana de análisis factorial, se seleccionan los ítems o variables correspondientes a nuestro análisis. No vamos a entrar a explicar las opciones. Sólo diremos que en el ejemplo de la salida que mostramos se ha utilizado el método de extracción de factores de componentes principales con rotación *oblimin*, forzando la solución a tres factores. La salida proporcionada es la siguiente:

TABLA 9.8. Varianza total explicada

Componente	Autovalores iniciales			Suma de las saturaciones Total
	Total	% de la varianza	% acumulado	
1	6.652	66.518	66.518	5.543
2	1.033	10.327	76.845	5.046
3	.772	7.719	84.563	5.224
4	.301	3.015	87.578	
5	.291	2.908	90.486	
6	.251	2.506	92.992	
7	.224	2.238	95.229	
8	.177	1.772	97.001	
9	.163	1.634	98.635	
10	.136	1.365	100.000	

Método de extracción: Análisis de Componentes principales.

- a. Cuando los componentes están correlacionados, las sumas de los cuadrados de las saturaciones no se pueden añadir para obtener una varianza total.

TABLA 9.9. Matriz de configuración.^a

	Componente		
	1	2	3
DIREC1	.991		
GEST2	.920		
CONTROL4	.874		
PLAN25	.585		.248
NORMAS15		.944	
NORMAS17		.925	
NORMAS16		.903	
RECUR13			.943
RECUR12			.919
RECUR11	.205		.752

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Oblimin con Kaiser.

- a. La rotación ha convergido en 6 iteraciones.

TABLA 9.10. Matriz de correlaciones de componentes

Componente	1	2	3
1	1.000	.619	.678
2	.619	1.000	.625
3	.678	.625	1.000

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Oblimin con Kaiser.

En este caso vemos que la agrupación exploratoria responde bastante fielmente a la hipótesis confirmatoria. Sin embargo, esto no suele ser así, lo que no quita para que nuestra agrupación hipotética pueda ser plausible y válida.

Quedan así revisadas las distintas formas de enfrentarse en la práctica a la comprobación y valoración de la fiabilidad y la validez de un instrumento de medida.

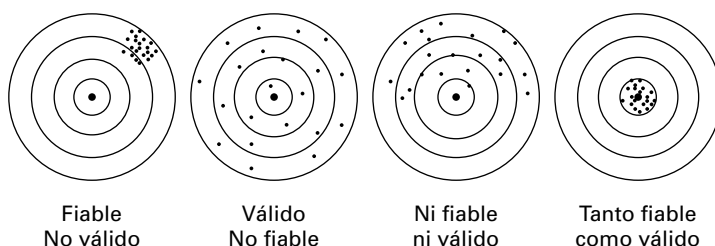
9.8. Relación entre validez, fiabilidad y longitud de la prueba

Nos parece importante recordar y destacar algunos aspectos lógicos que relacionan los conceptos de validez, fiabilidad y longitud de la prueba:

- Tanto la fiabilidad como la validez son índices basados en coeficientes de correlación, es decir, la expresión de la varianza compartida entre dos variables. Por tanto, todo aquello que contribuya a aumentar la variabilidad potencial de las medidas contribuirá a su vez a aumentar tanto la validez como la fiabilidad (por supuesto, todo dentro de la unidad conceptual del constructo o rasgo a medir). Los dos elementos más importantes que pueden contribuir a aumentar dicha variabilidad son:
 - El número de ítems de la prueba: al aumentar el número de ítems aumenta la amplitud de la escala de respuesta (con 5 ítems sería 0-5, mientras que con 10 sería 0-10). Ténganse en cuenta que no se puede aumentar indefinidamente el número de ítems, ya que las pruebas excesivamente largas provocan cansancio y distracción en las respuestas, lo que puede conllevar una disminución de la fiabilidad.
 - El tamaño de la muestra: cuanto más grande sea la muestra, más probable será encontrar variabilidad en las respuestas.

- La heterogeneidad de la muestra: si la muestra es muy homogénea en el rasgo que medimos, todas las respuestas serán muy parecidas, no habrá variabilidad y, por tanto, no podremos esperar altos coeficientes de correlación.
- La validez y la fiabilidad están relacionadas, de manera que las pérdidas o ganancias en una de ellas repercutirá en el mismo sentido en la otra.

A modo de síntesis mostramos el dibujo que representa la fiabilidad y la validez simbolizadas mediante los siguientes tiros sobre las cuatro dianas:



En la primera los tiros muestran consistencia, pero no aciertan con lo que se quiere medir. En la segunda diana los tiros giran en torno a la diana de lo que se pretende medir, pero no consistentemente, no son fiables. En la tercera diana no son ni fiables ni válidos. En la cuarta son fiables, consistente, y lo que es más importante, dan en el centro de lo que se quiere medir.

9.9. Ejercicios de autoevaluación

1. La fiabilidad, en sentido estricto, hace referencia:
 - a) Al instrumento de medida.
 - b) A las puntuaciones de la muestra de sujetos.
 - c) A las puntuaciones de la población.
 - d) A la puntuación en un ítem del instrumento.
2. La fiabilidad, entendida como test-retest, es una forma de fiabilidad como:
 - a) Equivalencia.
 - b) Formas paralelas.
 - c) Consistencia interna.

- d)* Predicción.
3. Si aplicamos una prueba a un grupo de sujetos y obtenemos una fiabilidad de 0.60, podemos concluir:
- a)* Claramente, el instrumento es fiable.
 - b)* Claramente, el instrumento no es fiable.
 - c)* Tendríamos que revisar si la muestra es demasiado homogénea.
 - d)* Podría deberse a que el instrumento es demasiado fácil o difícil para la muestra.
4. La utilización de expertos para garantizar la validez se puede utilizar en la validez...
- a)* Concurrente.
 - b)* Predictiva.
 - c)* De contenido.
 - d)* Discriminante.
5. Para Trochim, todos los tipos de validez se reducen a:
- a)* Validez aparente.
 - b)* Validez referida a un criterio.
 - c)* Validez predictiva.
 - d)* Validez de constructo.
6. La técnica del análisis factorial confirmatorio se utiliza específicamente para probar:
- a)* La validez predictiva.
 - b)* La fiabilidad.
 - c)* La validez de constructo.
 - d)* La fiabilidad y la validez.
7. Un alto coeficiente de fiabilidad es una garantía de la unidimensionalidad del rasgo medido, permitiendo la suma directa de las puntuaciones de los ítems para calcular la puntuación global.
- a)* Verdadero.
 - b)* Falso.

9.10. Referencias bibliográficas

- GARCÍA-LEÓN, REYES, VILA, PÉREZ, ROBLES y RAMOS (2002). *The Aggression Questionnaire: A Validation Study in Student Samples. The Spanish Journal of Psychology*. Vol 5, n.º 1, 45-53.
(http://www.ucm.es/info/Psi/docs/journal/v5_n1_2002/art45.pdf)
- GLEZ. GALÁN, A. (2000). *Calidad, eficacia y clima en centros educativos: modelos de evaluación y relaciones causales*. Tesis Doctoral. Univesidad Complutense de Madrid.
- GLEZ. GALÁN, A. (2004). *La evaluación del clima escolar como factor de calidad y eficacia*. Ed. La Muralla. Madrid.
- LARA, E. DE y BALLESTEROS, B. (2001). *Métodos de investigación en Educación Social*. Madrid: UNED. Unidades Didácticas.
- MARTÍN, M.; GLEZ. GALÁN, A.; TORREGO, J. C. y ARMENGOL, C. (2003). *Clima de Trabajo y Eficacia de Centros Docentes: Percepciones y resultados*. Ed. Universidad de Alcalá.
- MORALEDA, M.; GLEZ. GALÁN, A. y G.^a-GALLO, J. (1998). *AECS: Actitudes y Estrategias Cognitivas Sociales*. Ed. TEA. Madrid.
- MORALES, P. (2002). *Fiabilidad*. Internet WWW page, en URL:
< <http://www.upco.es/personal/peter/estadisticabasica/> >
- TROCHIM, William M. (2000). *The Research Methods Knowledge Base, 2nd Edition*. Internet WWW page, at URL:
<<http://trochim.human.cornell.edu/kb/index.htm>>
(version current as of August 02, 2000).

TEMA 10

Procedimientos para la recogida de información I. La observación sistemática, el cuestionario, la entrevista y las pruebas de rendimiento

ESQUEMA

- 10.1.** Los instrumentos de recogida de información en el proceso general de investigación
- 10.2.** La observación como procedimiento de representación de la realidad
 - 10.2.1. Los sistemas de recogida de información en la observación sistemática
 - 10.2.2. Las unidades de observación
 - 10.2.3. Cuestiones sobre el muestreo en la observación
 - 10.2.4. Técnicas de registro
 - 10.2.5. Fuentes de error en las observaciones
- 10.3.** El cuestionario y sus características
 - 10.3.1. Construcción de los cuestionarios
 - 10.3.2. Tipos de preguntas en los cuestionarios
- 10.4.** La entrevista. Objetivos y características
 - 10.4.1. Modalidades de la entrevista.
- 10.5.** Las pruebas de rendimiento. Evaluación normativa y evaluación criterial
 - 10.5.1. Las pruebas de evaluación normativa
 - 10.5.2. Las pruebas de evaluación criterial
- 10.6.** Ejercicios de autoevaluación
- 10.7.** Referencias bibliográficas

OBJETIVOS

- ✓ Conocer y comprender las características de la observación sistemática.
- ✓ Conocer y comprender algunas de las técnicas utilizadas para el registro de la información en la observación sistemática.
- ✓ Conocer y comprender las características de los cuestionarios.
- ✓ Conocer y comprender las características de la entrevista.
- ✓ Diferenciar los conceptos de evaluación normativa y evaluación criterial.
- ✓ Conocer y valorar los diferentes tipos de pruebas de examen de evaluación normativa.
- ✓ Conocer las características de las pruebas de evaluación criterial.
- ✓ Conocer los procedimientos para la elaboración de pruebas de evaluación criterial.

10.1. Los procedimientos para la recogida de información en el proceso general de investigación

Los instrumentos para la recogida de información, dentro del marco general de investigación, constituyen una parte importante del proceso de investigación, que como ya mencionamos en temas anteriores de estas unidades didácticas, es el siguiente:

- Formulación del problema de investigación.
- Fundamentación teórica: consulta de fuentes y estado de la cuestión sobre el problema y los objetivos de la investigación.
- Enunciación de los objetivos de la investigación.
- Definición de la población y delimitación de la muestra sobre la que se va a realizar el estudio.
- Formulación de las hipótesis: «Sí V.I. entonces V.D.», (y de la hipótesis nula), si procede, y operativización de las variables, o de los objetivos de investigación.
- Concreción del plan de acción, de sus fases y sus objetivos y tiempos.
- Elección de la técnica de recogida de información: Observación, cuestionarios, entrevista.
- Construcción del instrumento si fuera necesario (cuestionario, guión para las entrevistas, hoja de observación).
- Recogida de la información.
- Tratamiento estadístico de datos, (análisis de las variables, análisis descriptivos, análisis relacionales, contraste de hipótesis, ..).
- Discusión de los resultados y conclusiones.
- Significación pedagógica de los resultados.
- Valoración crítica. Análisis del trabajo, de la validez interna y externa.
- Elaboración del informe y difusión del mismo.

Cada uno de los diferentes tipos de instrumentos, por ejemplo, el cuestionario, las pruebas objetivas o la entrevista, será más idóneo que otro dependiendo de las características de la investigación, de sus objetivos y de sus posibilidades. Y a veces, será necesario o recomendable la utilización de varios tipos de instrumentos, por ejemplo, la entrevista, la observación y el cuestionario. Por ello, es necesario que el investigador

conozca las características de los diferentes instrumentos para la recogida de información y desarrolle las habilidades necesarias para seleccionarlos, elaborarlos, aplicarlos e interpretarlos.

La observación es una estrategia fundamental del método científico, dado que toda investigación comienza por la observación de un hecho, de un problema, que se quiere resolver. El propósito de la observación es obtener una descripción o representación de los acontecimientos, procesos y fenómenos así como de los factores que influyen en dichos fenómenos, dentro de un contexto determinado. Utilizada como método de investigación es siempre selectiva, y debe ser sistemática, estructurada y controlada, en un proceso que debe poder explicarse de modo que otras personas puedan comprenderlo y evaluar su adecuación. «Necesita elegir un objeto, definir una tarea, un interés, un punto de vista, un problema. Y su descripción presupone un lenguaje descriptivo..., similitud y clasificación, lo que a su vez presupone interés, puntos de vista y problemas» (Popper, 1963:21, en Evertson y Green, 1997: 309). El observador es el primer instrumento de la observación, influido por sus metas, prejuicios, marco de referencia y habilidades para observar.

Otra estrategia que se utiliza en investigación para recoger información es el cuestionario. El cuestionario es una técnica de recogida de información sobre opiniones, actitudes, habilidades, aplicable a amplias muestras representativas de poblaciones definidas. El término 'cuestionario' sirve para denominar, también, al instrumento concreto de recogida de información de una muestra de la población definida, en torno a las variables que intervienen en la investigación,.

La entrevista es otro de los procedimientos y técnicas utilizadas para la recogida de información. Se diferencia del cuestionario en que suele realizarse de modo individual, y oral, con la presencia del entrevistado y el entrevistador, aunque también grupal, y añade la posibilidad de profundizar en las cuestiones o preguntas que realiza el entrevistador al entrevistado. Las entrevistas estructuradas son encuestas en las que el investigador entra en contacto personal con los encuestados y les hace una serie de preguntas previamente establecidas. En esta modalidad de encuesta, además de cuidar la claridad de las normas o instrucciones y de las cuestiones, los entrevistadores deben tener normas claras sobre como deben dirigirse a los encuestados, qué tipo de aclaraciones pueden hacer y cuáles no por ser objeto del estudio, estrictas, y deben tratar de seguirlas de modo estricto para evitar influir en la dirección de la respuesta. Las instrucciones a los entrevistadores deben estar contenidas en el protocolo de la entrevista.

La encuesta telefónica, en algunos aspectos es similar a la entrevista, al permitir aclaraciones si el investigador supone que el encuestado no ha interpretado bien la pregunta, si bien se diferencia de la entrevista en que en la encuesta telefónica el entrevistador, no puede observar el contexto en el que se desenvuelve el encuestado, pero le permite captar matices relacionados con la expresión, el modo como responde, y el tono de voz. El cuestionario, junto con la entrevista y la encuesta telefónica, es una de las técnicas, más utilizadas en la investigación tipo encuesta.

10.2. La observación como procedimiento de representación de la realidad

La observación es un fenómeno que forma parte de las funciones perceptivas de la persona, esenciales, de una manera tácita, para el funcionamiento cotidiano. *Observar* es ponerse delante de un objeto o poner la mirada en él, y supone el proceso contrario al proceso empático, de sentir con, metiéndose dentro, poniéndose en su lugar. En su origen, la palabra «observar» significaba «ajustarse a lo que está prescrito», acepción en uso.

Es una de las técnicas de investigación más importantes de las ciencias sociales, ya que nada puede reemplazar el contacto directo del investigador con su campo de estudio y ninguna es capaz de sugerir tantas ideas nuevas. Cualquier otra técnica para el estudio y comprensión de la realidad ha de basarse en una primera fase de observación para poder tener una idea real de lo que se va a estudiar.

La observación es *un objetivo* que hay que conseguir, o una aptitud que hay que desarrollar. El acto de la observación *es un proceso* intelectual, que requiere atención —es decir, concentración selectiva—, y está orientado por un objetivo organizador del proceso de observación, que va, por tanto, más allá de la mera percepción, al organizar lo observado en una síntesis integrada en una estructura con significado. La observación supone contemplar de modo sistemático cómo se desarrolla la vida social, sin manipularla, tal cual discurre, y la función del observador es recoger información sobre el objeto que se tome en consideración. Observar a las personas en determinadas situaciones consiste en centrar la atención en sus comportamientos y en las relaciones que establecen entre ellos y su entorno social. La observación recoge información guiada por un objetivo principal: observar para valorar unas necesidades, observar para establecer un diagnóstico, observar para confirmar una hipótesis de investigación, observar para evaluar unos resultados. La observación supone las

operaciones de selección, de estructuración e interpretación de los datos de modo que quede patente una red de significados. También se habla de *técnica de observación* para la recogida de datos, sirviéndose de uno o de varios observadores —observación en sentido estricto—, o de un cuestionario de encuesta —en sentido más amplio—. Por último, se habla de la observación como *resultado* del acto de observar, seguido del acto de interpretar, lo que supone una referencia a un marco teórico. (En Postic y De Ketele, 1988, 17).

Los numerosos sistemas de codificación que existen, podrían agruparse en dos categorías: los sistemas de selección, en los que la información se codifica de un modo sistematizado mediante cuadrículas o parrillas preestablecidas, y los sistemas de producción, en los que el observador, una vez recogida la información mediante notas de campo, elabora el sistema de codificación para su interpretación. Al primer tipo de observación se la denomina sistemática, en la cual las categorías a observar están previamente establecidas. Al segundo tipo de observación se la denomina participativa, y en ella el observador se introduce en el campo sin categorías a observar preestablecidas, las cuales van emergiendo como resultado de la permanencia del observador en el campo. En la observación sistemática la función del observador consiste en contemplar lo que está aconteciendo de un modo distanciado y registrar los hechos mediante anotaciones sobre una hoja de registro de la información, o mediante la utilización de medios técnicos como las videocámaras o las grabadoras. Su característica principal es la ausencia de interacción entre observador y observado. En la observación participante el observador es actor del sistema que estudia, lo que le permite obtener información que de otro modo no podría conseguir, incluida la información procedente de su propia experiencia subjetiva.

La metodología observacional, tanto en su vertiente sistemática como abierta o participante, es una metodología especialmente adecuada en la educación, tanto en los contextos escolares como extraescolares, familiares y sociales. La observación, como estrategia de investigación, para responder a una pregunta, es deliberada y sistemática. Pero también en la vida cotidiana se realizan observaciones sistemáticas, no tanto para responder a preguntas específicas sino para establecer, mantener, controlar, suspender acontecimientos cotidianos y participar en ellos (en Evertson y Green, 1986, 1997: 307). Otro enfoque de investigación observacional es el que se apoya en el enfoque de investigación etnográfica, naturalista o cualitativa y se conoce con el nombre de observación participante. La observación sistemática, será estudiada en este tema. La observación participante será objeto de estudio en el tema siguiente.

La observación es *un modo de representación de la realidad*, que debe acometerse desde varios modos y sistemas de representación, siendo la variedad su principal fuerza, dado que es difícil mantener una única visión de los acontecimientos referidos a una situación determinada, como más exacta que otra. Cada perspectiva para la representación de la realidad social nos va a ayudar a comprenderla mejor, y el observador deberá recoger suficiente y apropiada evidencia para asegurar que la descripción sea lo más exacta posible.

La observación es *un proceso contextualizado* y su consideración puede ayudar al investigador a determinar los factores que limitan y apoyan la observación y a comprender los fenómenos en un contexto dado. Los contextos pueden ser considerados desde distintos aspectos, que pasamos a enumerar:

- El contexto local: comprende el contexto inmediato al acontecimiento a estudiar y los contextos que lo rodean, de mayor amplitud, que coexisten influyéndose mutuamente. Así, podemos considerar el grupo de alumnos, el aula, el Centro, la comunidad etc.
- El contexto histórico de las situaciones: enmarca el contexto local y puede verse a través de las expectativas, tradiciones y estilos de comunicación, y, aunque no es directamente observable sí puede ser explorado a través de anuarios.
- El contexto histórico de los acontecimientos, estudiados a través de informes. Así, por ejemplo, investigaciones realizadas sobre los procesos en el aula han mostrado que los acontecimientos que ocurren en ella tienen su historia, y señalan que las primeras horas, días y semanas de cualquier proyecto con grupos son períodos críticos para el establecimiento de normas, de rutinas esenciales, de expectativas, de estructuras de funcionamiento, que deben estar previstas, teniendo en cuenta que algunas pautas son estables y que otras varían. De esto se deduce la relación entre el muestreo y el tema de estudio, que debe ser cuidadosamente considerado, teniendo en cuenta qué conductas son estables y representativas de lo que estudiamos y cuáles varían y cómo debemos hacer el muestreo en orden a asegurar que nuestras observaciones sean representativas.

El enfoque de investigación observacional adoptado incide en el significado y los niveles de contexto a considerar en el muestreo, los instrumentos y las cuestiones formuladas. Así, podrá adoptar un enfoque de aproximación de exclusividad o de inclusividad. En el extremo de exclu-

sividad todo lo que se observa está previamente determinado, recogido en el instrumento deliberadamente. En el extremo de la inclusividad serán considerados amplios segmentos del contexto, la selectividad estará menos limitada, el proceso de observación será más amplio y flexible que el sistemático del extremo exclusivo y la codificación será hecha de modo retrospectivo. Si bien en el pasado ambos extremos del continuo han tendido a ser excluyentes no hay razones teóricas para que los dos extremos puedan trabajar juntos. Por ejemplo, en una investigación observacional inclusiva el investigador puede identificar modelos de conducta y acontecimientos que pueden originar la construcción de un sistema de categorías o lista de control para centrar la observación sobre aspectos específicos del contexto. Y la información obtenida con un instrumento focalizado puede ser examinada desde un punto de vista más amplio o inclusivo. El punto crucial sería seleccionar el instrumento observacional que permita al investigador recoger los procesos activos, flexibles y adaptativos que ocurran en las situaciones educativas a observar y considerar las claves del contexto que formen las bases de las inferencias (en Evertson y Green, 1986:167). El propósito de la observación guía lo que habrá de hacerse, lo que se podrá obtener y el modo en que se utilizará.

10.2.1. Los sistemas de recogida de información en la observación sistemática

Cuatro son los sistemas de clasificación de los procedimientos de recogida de información mediante la observación: categoriales, descriptivos, narrativos y tecnológicos. Cada uno tiene unas características distintas, que repercuten en las características de los instrumentos para registra los acontecimientos y su análisis posterior y deben seleccionarse dependiendo de cuál sirva mejor para representar el segmento de realidad y fenómeno que quiere observar.

El sistema categorial. Es considerado un sistema de naturaliza cerrado porque contiene un número limitado de categorías o unidades de observación prefijadas. El observador recoge información sólo de los ítems marcados. Utiliza dos tipos de límites en el período de observación: de tiempo y de acontecimientos o conducta. Para el registro de la información utiliza, inventarios, listas de comprobación de la aparición de las conductas seleccionadas, y sus frecuencias. La meta que persigue el investigador que utiliza este sistema es estudiar una amplia serie de situaciones para obtener datos. Los resultados de dichas observaciones, registrados en términos numéricos, forman bases de datos para la reali-

zación de análisis estadísticos y generalización de los resultados a otros casos similares.

El sistema descriptivo. Es un sistema abierto por no contener un número limitado de categorías. Puede trabajar con un rango de categorías fijas para describir los comportamientos o éstas pueden generarse a partir de los modelos observados o identificarse a partir de los datos procesados. Los datos son recogidos dentro de unos márgenes de acontecimientos y contextos, recogiendo información amplia de las conductas que observa, transcribiéndolas con lenguaje verbal. El foco del observador es amplio para captar con la mayor amplitud lo que acontece en la situación de estudio, pudiendo utilizar medios técnicos para realizar grabaciones. La meta del investigador es obtener una descripción detallada del fenómeno para explicar el desarrollo de los procesos e identificar principios generales para explorar procesos específicos y generalizar dentro de los casos a casos y contextos similares y para explorar fenómenos específicos en observaciones posteriores.

El sistema narrativo. Es un sistema abierto por no tener categorías prefijadas. Utiliza el lenguaje oral o escrito para recoger las descripciones y puede ayudarse del uso del magnetófono o de los diarios o anecdotarios del profesor, del educador, o del observador participante. El muestreo de los acontecimientos ocurre dentro de sus límites naturales. Los datos están influidos por la percepción, el marco de trabajo, la experiencia y la riqueza lingüística del observador. La meta del investigador es explorar, describir y explicar el desarrollo de los procesos e identificar principios generales y específicos, como en el sistema descriptivo, y además busca identificar y comprender los factores que influyen en la ocurrencia de dichos acontecimientos. Los hallazgos encontrados son útiles para la comprensión de casos similares.

Los sistemas tecnológicos. Es un sistema abierto, por no partir de categorías prefijadas, el contexto tiene un significado preciso y el muestreo de los acontecimientos ocurre dentro de sus límites naturales, de igual modo que en los dos sistemas anteriores. La utilización de las videocámaras para las grabaciones de los acontecimientos y los procesos en directo, se convierten la materia prima para estudiar los acontecimientos, y seleccionar objetivos de estudio posterior mediante otro tipo de sistema. Pueden ser usados en combinación con otros sistemas en función de las intenciones del investigador y del problema de estudio. El investigador tiene una gran variedad de opciones pudiendo seleccionar la que mejor se acomode al problema y fenómeno que se quiera observar.

10.2.2. *Las unidades de observación*

Las decisiones sobre las unidades de observación son de gran importancia dado que establecen los principios con los que vamos a hacer afirmaciones sobre los acontecimientos que vamos a observar. Pueden ser deductivas o inductivas, seleccionadas a priori o derivadas en el procedimiento posterior de análisis. Las unidades fijadas a priori se basan en la conceptualización desde una perspectiva teórica acerca de lo que se va a observar. La decisión sobre las unidades lleva implícita la delimitación de las relaciones que pueden confirmarse posteriormente dentro de su contexto. Las unidades definen los límites intelectuales de las afirmaciones y de sus relaciones posibles, dentro de su contexto. La determinación de las unidades de observación está realizada con la recogida de datos y su análisis posterior. (En Evertson y Green, 1986: 174).

Las unidades de observación en los sistemas categoriales son derivadas, generalmente, deductivamente o se construyen a partir de un estudio previo con cualquier otro sistema descriptivo, narrativo o con medios técnicos. Estas unidades de observación son unidades discretas y simples reflejando cada categoría una sola conducta, que puede considerarse como una variable independiente. Los instrumentos para recoger la información pueden ser escalas, por ejemplo, tipo Likert, o el diferencial semántico de Osgood.

Las unidades de observación que utilizan los sistemas descriptivos son variadas: inductivas, deductivas, naturales, discretas, inferidas, de conducta, moleculares, molares, funcionales y situacionales. Esta variedad de unidades se debe a las características propias de los sistemas descriptivos que tienden a basarse en el análisis de los acontecimientos recogidos aunque también son esbozadas desde la investigación teórica. La relación que se establece entre las unidades tiene una base conceptual. Las microunidades son discretas y originan de un modo inductivo unidades molares. Las unidades de observación desde una perspectiva naturalista, descriptiva, al realizarse sobre un continuo de conducta personal y relacional en un contexto natural, son tan amplias como se han mencionado en el párrafo anterior, y es que la observación, desde este enfoque, no se realiza sobre conductas o atributos pre-especificados, realizándose las observaciones tanto sobre aspectos verbales, relacionales, situacionales, e interpretados por el observador.

Las unidades de observación en los sistemas narrativos son similares a los sistemas descriptivos, pero sobre información textual. Los sistemas tecnológicos recogen información de modo no selectivo y su función

principal es ayudar a los otros sistemas. Las unidades en los sistemas tecnológicos son más amplias que las de los anteriores sistemas, limitadas por la situación y tiempo que dure la grabación. Las grabaciones proporcionan las bases para:

- Un análisis con profundidad del desarrollo de las conductas.
- La identificación de los modelos de conducta dentro de las secuencias seguidas.
- La identificación y construcción de nuevas unidades de observación.
- La comprobación de los modelos identificados a través del tiempo y las situaciones.
- El análisis de los datos desde una variedad de perspectivas complementarias.

Las grabaciones facilitan al investigador y a los participantes la estimulación del recuerdo para obtener las perspectivas de éstos sobre los acontecimientos y validar los modelos observados en una forma de triangulación entre observadores. La meta del investigador que utiliza medios tecnológicos es obtener un registro permanente de un acontecimiento o fenómeno para poder estudiarlo en profundidad en un momento posterior. Los medios tecnológicos pueden utilizarse para complementar los otros sistemas, categorial descriptivo y narrativo.

10.2.3. Cuestiones sobre el muestreo en la observación

Seleccionar el momento, lugar, el tiempo, y los centros de observación asociados a la cuestión de estudio son aspectos de fundamental importancia relacionados con las cuestiones del muestreo. Una selección errónea puede invalidar los hallazgos. (Evertson y Green, 1986: 180). Son cinco las áreas de decisiones sobre el muestreo:

- el número de sujetos,
- la longitud total del tiempo de observación,
- la distribución del tiempo de observación,
- la probabilidad de que la observación sea representativa del fenómeno objeto de estudio (qué se observa, cuándo y dónde) y
- las alternativas u opciones.

Todas estas decisiones tienen que estar informadas por el marco teórico, las investigaciones anteriores y por la realización de una fase piloto

‘de orientación de la situación’ que contribuye a garantizar la representatividad de la muestra. El momento de la decisión sobre el muestreo está en función del sistema de observación adoptado. En el sistema categorial las decisiones se adoptan de un modo apriorístico. En los sistemas descriptivo, narrativo y tecnológico, las decisiones sobre el muestreo se adoptan a lo largo del proceso de investigación. No obstante, en todos los sistemas el observador debe estar familiarizado con las situaciones de observación para asegurar que el muestreo es representativo del fenómeno que se quiere observar.

El muestreo del tiempo es usado, al menos, de tres modos: para especificar los límites generales del período de observación, para especificar los intervalos de grabación o para designar segmentos pequeños de tiempo dentro de un período general de la observación. Para el muestreo de los acontecimientos el investigador determinará el asunto específico, que será el objeto de la observación, y después determinará las situaciones posibles en las que el fenómeno puede ocurrir. Después decidirá sobre el muestreo de los acontecimientos y los centros de observación donde pueden manifestarse (centro escolar, aula, centro de trabajo, actividad del profesor, actividad del alumno...) de un modo representativo de la realidad, para garantizar la fiabilidad y validez de la observación.

10.2.4. Técnicas de registro

La sistematización de los instrumentos de observación depende del marco de referencia teórico del observador. El grado de estructuración de los mismos puede depender también del momento o fase en la que se efectúa la observación, así, en una primera aproximación al estudio de problema, puede estar poco estructurada para estar abierto a las situaciones contextuales y relacionales de la problemática objeto de estudio. En una segunda aproximación, los instrumentos de observación estarán más estructurados, también en diferente grado dependiendo de que se trate de un sistema de observación categorial, o un sistema de observación descriptivo naturalista.

Los instrumentos de registro de observaciones se clasifican dentro de un continuo de menor a mayor estructuración, como siguen:

- Registros narrativos abiertos, en contextos naturales, en los que el observador toma notas sobre lo que acontece en la situación que observa, que posteriormente completará con sus apreciaciones.
- Escalas de valoración, inventarios de conducta, en los cuales las conductas seleccionadas son valoradas mediante una escala de

grados, por ejemplo, de 1 a 5 de menor a mayor acuerdo con la conducta especificada.

- Registros de conducta, mediante tablas de doble entrada, en las cuales en las columnas se especifican las conductas y en las filas se indica la ocurrencia o no de esa conducta. Esta forma de registro se utiliza para indicar la presencia o ausencia de una determinada conducta durante el período temporal de observación fijado. Es una forma predeterminada, o prefijada, de registro dado que las conductas se especifican antes de que sucedan. Las conductas se registran tras una breve lectura de las categorías que figuran en la escala indicando la intensidad con la que se produce la conducta. Puede ser utilizada para registrar la aparición de determinadas conductas en un sujeto o para estudiar una conducta en el grupo, pudiendo servir para otras opciones.

Ejemplo de registro individual:

REGISTRO INDIVIDUAL

Nombre:			
Conductas	Grado de aparición		
	Frecuente	Rara vez	Ausente
Atiende las intervenciones de los demás			
Colabora en las actividades			
Sugiere nuevas actividades			
Disfruta en las actividades grupales			
Se mantiene aislado			
...			

- Matrices de interacción, mediante las cuales se busca constatar las interacciones entre las conductas y el ambiente y el momento en el que se producen, e incluso, con qué intensidad se da la interacción. Por ejemplo, una conducta referida a un alumno: presta atención, cruzada por: en la mañana, tarde, noche, y matizada en: clase, actividad, comedor, etc.

10.2.5. Las fuentes de error en las observaciones

Las fuentes de error se consideran desde aspectos de fiabilidad, validez y rigor en las observaciones y se ocupa de los límites sobre la certeza

de los datos obtenidos. Los errores están relacionados con el observador, el sistema de representación elegido, el marco conceptual del fenómeno estudiado y los procedimientos de recogida de datos. Los errores van unidos al sistema perceptivo lo que lleva a los investigadores a adoptar variedad de medidas para reducir la posibilidad de error, asegurar el rigor en la observación y obtener una representación adecuada de la realidad.

La validez se refiere al grado en que la información obtenida por un instrumento realmente describe lo que el investigador se había propuesto describir y la fiabilidad se refiere a la exactitud y consistencia con que lo hace; al considerar la fiabilidad en las investigaciones observacionales hay que hacer dos distinciones conceptuales entre los índices de fiabilidad procedentes de los coeficientes de fiabilidad, utilizados para controlarla mediante, por ejemplo, registros paralelos, o el procedimiento de recogida de información sucesiva, a cuyos resultados numéricos se les aplica pruebas estadísticas para hallar coeficientes de fiabilidad, como los utilizados en las pruebas de fiabilidad de las escalas de Likert, o explicados en el tema referido a la fiabilidad y validez de los instrumentos de medida. También se pueden hallar indicadores de fiabilidad mediante la utilización de coeficientes de acuerdo entre varios observadores sobre los acontecimientos que observan y describen, y cuando el registro de las observaciones se realiza mediante medios técnicos, el visionado repetido de las grabaciones, y pos varios observadores, garantizará la fiabilidad de las interpretaciones y de los resultados de la observación. (en Evertson y Green, 1986:184).

El tipo de registro elegido afecta tanto a la fiabilidad como a la validez de los datos. Estas exigencias de los instrumentos de medida se verán favorecidas si se cuida:

- La claridad en la formulación de las definiciones conductuales a observar.
- El número de conductas o categorías a observar reducido.
- Los códigos estandarizados con las suficientes garantías científicas, en los estudios piloto.
- El muestreo de tiempo suficiente en la observación para garantizar el registro emitido.

Otro tipo de fiabilidad para los estudios descriptivos se basa en la necesidad de que el lenguaje de la descripción a un nivel primario de la recogida de los datos esté en consonancia con las teorías de la acción que

estén usando los participantes, aconsejándose dos modos para obtener la fiabilidad para los datos descriptivos: 1) el uso grabaciones audiovisuales como garantía de los datos, la revisión inmediata de la grabación audiovisual para asegurar la fiabilidad de la descripción en una forma de triangulación de los datos, y 2) la colaboración de otros investigadores o participantes como garantía de evidencia primaria.

Las cuestiones como el muestreo de sujetos, la variabilidad y estabilidad de la conducta o variables que se estudian, o el medio en que la observación tiene lugar, inciden en la validez de las investigaciones.

10.3. El cuestionario y sus características

Como dijimos al comienzo de este tema, las técnicas y los instrumentos para la recogida de información, constituyen una parte esencial del proceso de investigación. El cuestionario es una técnica de recogida de información sobre opiniones, actitudes, habilidades, aplicable a amplias muestras representativas de poblaciones definidas. Su finalidad es recoger información de modo sistemático, de acuerdo con los objetivos del investigador, de las variables que intervienen en la investigación, de una muestra de la población definida. El término «cuestionario» también designa al instrumento concreto de recogida de información, y, junto con la entrevista y la encuesta telefónica, es una de las técnicas, más utilizadas en la investigación tipo encuesta.

Los cuestionarios son instrumentos que se presentan de forma escrita y que deben ser completados por los propios encuestados. Los cuestionarios se caracterizan por presentar una serie de preguntas, elementos o estímulos predeterminados que, a diferencia de la entrevista, no pueden modificarse durante la recogida de información. Los cuestionarios deben incluir instrucciones escritas claras sobre cómo responderlos, definiciones de algunos términos difíciles para el encuestado, e indicaciones de cómo y cuando devolver el cuestionario una vez contestado.

Algunas de las características que lo diferencian de otras técnicas de recogida de información, como son la entrevista, la encuesta telefónica o la observación, quedan recogidas a continuación.

- Permiten recoger una gran cantidad de información referida a hechos diversos.
- Las preguntas están de antemano delimitadas por el investigador en torno a las variables que investiga, garantizando de este modo cierta estabilidad en las respuestas.

- Las informaciones no se recogen directamente sino que se reflejan mediante expresiones verbales, generalmente escritas, a partir de unas cuestiones concretas formuladas en el instrumento.
- La información que se recoge se centra en aspectos tanto objetivos como subjetivos o de opinión.
- La persona que responde lo hace desde su propia interpretación de la pregunta.

El cuestionario supone una ventaja económica sobre las otras técnicas de recogida de información como son la observación y la entrevista, al poder recoger gran cantidad de información de una gran muestra poblacional. Sin embargo, hay que indicar que no siempre es seguro que sea el sujeto seleccionado para responder el que lo haga. Este problema se resuelve cuando el cuestionario es contestado en presencia del encuestador, mediante la entrevista en persona o en la encuesta telefónica.

10.3.1 Construcción de los cuestionarios

El propósito de la investigación, sus objetivos, es el criterio fundamental que guiará la construcción del cuestionario. Pero también debe tenerse en cuenta la población a la que se destina y los recursos con los que se cuenta. El propósito tiene que ver con los objetivos de la investigación, con la cuestión a la cual la investigación trata de responder, con la formulación operativa del problema basada en una revisión del estado de la cuestión que ayudar a delimitar las variables esenciales que guiarán la recogida de información, en torno a las cuales se elaborarán los indicadores, los ítems, del cuestionario. La población a la que se dirige determinará el modo de elaboración del cuestionario, su estructura, el tipo de preguntas, su redacción sencilla o compleja. Y los recursos condicionarán cuestiones como la aplicación a una muestra amplia o reducida, en función de que contemos con posibilidades de pagar a aplicadores y correctores de los cuestionarios para su análisis, gastos de correo, teléfono, etcétera.

La construcción de los cuestionarios debe cumplir las exigencias de la investigación educativa y social de los instrumentos de medida como son su fiabilidad y su validez, de acuerdo con las recomendaciones que se hacen en el tema referido a las características técnicas de los instrumentos de medida. Toda investigación, sabemos, comienza con el planteamiento del problema y continua con la formulación de las hipótesis como tentativa de solución de solución del problema. La formulación de

las hipótesis va a determinar las características de los instrumentos de recogida de información que debemos utilizar. En concreto, para la elaboración del cuestionario hay que tener en cuenta:

- El rasgo a medir, concretado en las variables que se van a estudiar.
- Los objetivos a conseguir.
- La selección de los elementos.
- La formulación de los elementos.
- La longitud de la prueba.
- El tipo de respuestas.

Después de elaborado el cuestionario, la primera prueba que debe pasar es la del juicio de expertos, que valorará la adecuación de las preguntas al rasgo que pretende medir, a los objetivos a conseguir, su adecuación a la muestra a la cual se dirige, la adecuación en la formulación de los elementos, y el grado de inteligibilidad. Y una vez validados por el juicio de expertos, sería aplicados a una muestra de sujetos para observar las posibles dificultades que encuentran en la comprensión de los ítems. La corrección de los cuestionarios servirá para observar el modo de respuesta de los mismos, y para hallar los índices de fiabilidad, validez y homogeneidad de los ítems del cuestionario. Si supera de modo positivo los pasos mencionados, el cuestionario estará listo para ser aplicado en la investigación, y sus datos serán adecuados para continuar el proceso de la investigación.

En el siguiente diagrama puede verse la fase de selección o elaboración de los instrumentos de recogida de datos, del cuestionario en este caso, en el proceso de la investigación.

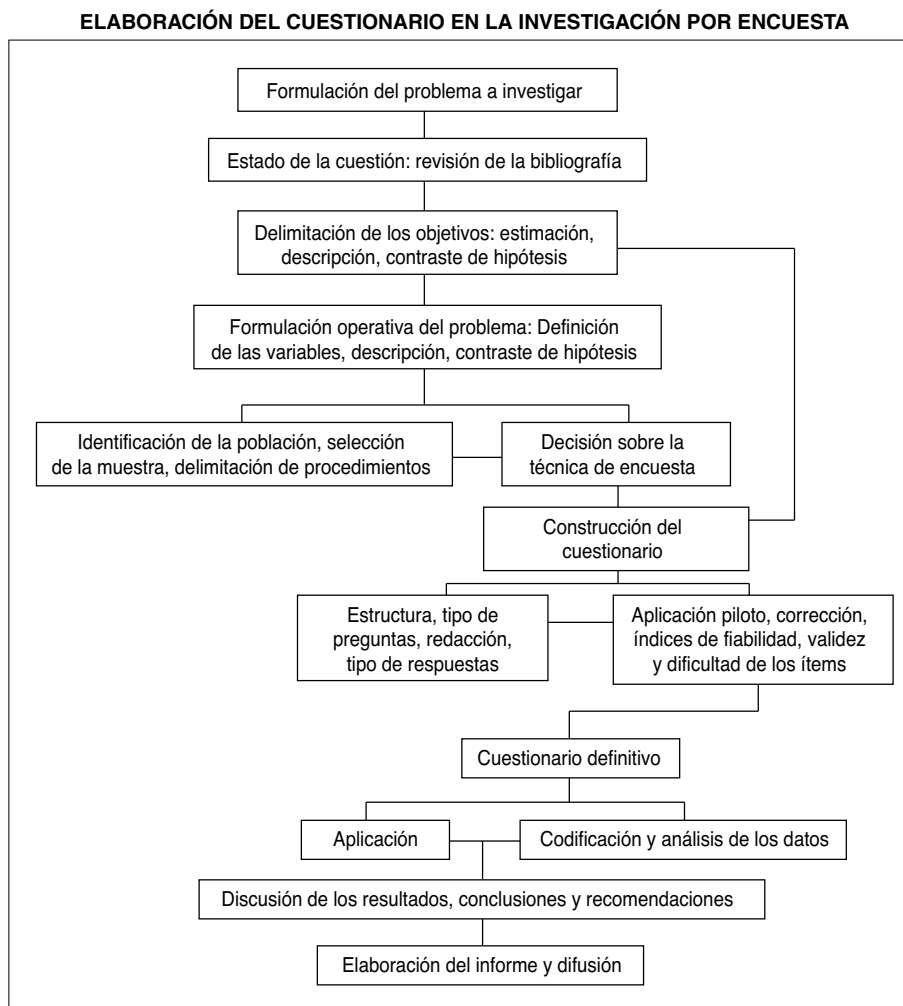


FIGURA 10.1.

Los pasos comentados para la elaboración de escalas, en los temas referidos a las escalas de Thurstone y Likert, son aplicables a la elaboración de los cuestionarios. Así mismo el tema dedicado a la fiabilidad y validez, es pertinente para hallar los índices de fiabilidad y validez del cuestionario.

10.3.2. Tipos de preguntas en los cuestionarios

Al formular las preguntas de los cuestionarios debemos tener en cuenta:

- Evitar formular preguntas generales si lo que se pretende es recoger información específica.
- A veces es necesario formular diversas preguntas sobre una misma cuestión para alcanzar mayor precisión en la respuesta.
- De información.
- De intención.

Las preguntas que integran un cuestionario necesariamente tendrán que ver con los objetivos de la investigación, es decir, con la naturaleza de la información que se quiera recoger. Otros elementos a tener en cuenta son la dimensión temporal, la estructura formal o la extensión y profundidad informativa que se pretenda conseguir. Los tipos más frecuentes de preguntas que se pueden introducir en un cuestionario se clasifican en abiertas y cerradas. Las preguntas abiertas, también llamadas de «respuesta libre» se caracterizan por que al sujeto no se le imponen ningunas categorías determinadas de respuesta, ni delimitación específica de espacio para responder. Se suele utilizar este tipo de preguntas en investigaciones exploratorias en las que el investigador necesita conocer de un modo amplio la información que determinado grupo posee sobre determinadas cuestiones y, posteriormente, buscar las categorías esenciales en torno al tema que investigamos, y a partir de esta categorización realizar una análisis de contenido de la información aportada.

Veamos alguno de los tipos de preguntas que suelen ir en los cuestionarios.

Las preguntas cerradas pretenden recoger el grado de acuerdo o desacuerdo del encuestado con la pregunta, o afirmación, formulada, pudiendo ser su respuesta de:

- SÍ
- NO

O de expresión de su grado de acuerdo entre cinco, o más o menos, gradaciones de respuesta, como por ejemplo:

- Totalmente en desacuerdo.
- Muy en desacuerdo.
- Parcialmente de acuerdo.
- Muy de acuerdo.
- Totalmente de acuerdo.

En otras ocasiones, las preguntas de respuesta cerrada pueden presentar un abanico posible de respuestas a elegir por el encuestado, como en el siguiente ejemplo

Cuando Vd. llega a una conferencia que ya ha comenzado, se decide por:

- Entrar y pasar a las primera filas.
- No entrar.

Otro tipo de preguntas es aquella que presenta varias opciones pero se le pide al encuestado que elija de entre todas las posible respuestas las 3 que prefiere y que las numere en orden de preferencia. Por ejemplo: Elija los tres deportes, indicando el orden de preferencia sobre el que más le gusta practicar al aire libre:

- Atletismo.
- Baloncesto.
- Balonmano.
- Equitación.
- Esquí.
- Fútbol.
- Golf.
- Natación.
- Tenis.
- Voleibol.

En este tipo de preguntas el encuestador debe ser lo más exhaustivo posible en cuanto a las opciones de respuesta que ofrece. Con el fin de subsanar el hecho de ser preguntas cerradas; para subsanar esta dificultad, a veces, los cuestionarios se caracterizan por presentar una mayoría preguntas cerradas con alguna pregunta abierta. O al revés. Del modo siguiente:

Elija los tres deportes, indicando el orden de preferencia 1: el que más, 2, y 3, que más le gusta practicar al aire libre:

- Baloncesto.
- Balonmano.
- Fútbol.
- Natación.
- Tenis.
- Voleibol.

Otros (especifíquelos y numérelos):

Otro tipo de tipologías seguidas en la clasificación de las preguntas se relaciona con la naturaleza de la información a recoger:

- Sobre hechos.
- Sobre acciones.
- De información.
- De intención.
- De opinión.
- ...

O del modo en que presentamos el siguiente cuestionario en formato de escala:

Cuestionario sobre la sesión de clase

Este cuestionario tiene como objetivo conocer sus impresiones sobre la clase que acaba de finalizar. Es anónimo. Los resultados se utilizarán con fines de mejora. Contestar al cuestionario le llevará alrededor de 10 minutos.

Muchas gracias.

Exprese su grado de acuerdo sobre las proposiciones que se le muestran a continuación, de modo que refleje sus percepciones sobre la sesión de clase usando la siguiente escala:

- 5 = Muy de acuerdo.
- 4 = De acuerdo.
- 3 = Ni de acuerdo ni en desacuerdo.
- 2 = Muy en desacuerdo.
- 1 = Muy en desacuerdo.

1. El profesor presenta al inicio de la clase los objetivos de la lección.

5	4	3	2	1
MA				MD

2. El profesor, a medida que explica los contenidos realiza ejemplos.

5	4	3	2	1
MA				MD

3. El profesor después de explicar una pequeña cantidad de la lección y realizar el ejemplo, pregunta a algún alumno para comprobar su comprensión de la explicación.

5	4	3	2	1
MA				MD

10.4. La entrevista. Objetivos y características

La técnica de la entrevista es especialmente útil para obtener una información más personalizada y completa, que la aportada por otras técnicas como pueda ser el cuestionario. A veces, como técnica complementaria para completar la información obtenida mediante otras técnicas como puede ser la observación, el cuestionario o los tests psicométricos, e incluso como técnica indispensable como son las entrevistas para selección de personal en las empresas, de promoción interna, o la entrevista psiquiátrica.

La característica esencial de la entrevista, y su aportación principal como técnica de recogida de datos, es que permite profundizar más allá de la formulación de la pregunta, e incluso añadir nuevas preguntas para obtener una información más completa y precisa, y esto es lo que hace de la entrevista una técnica compleja requiriendo de una gran profesionalidad y experiencia del entrevistador, para que sea provechosa. Este tipo de entrevistas en profundidad suele utilizarse para completar la información recogida con otros instrumentos, como pueda ser la observación o el cuestionario, y se dirige a informantes claves para la investigación, como pueda ser el responsable de un programa, el que lo aplica y algún miembro de los que se beneficia del programa.

Las entrevistas suelen realizarse siguiendo un guión estructurado en torno a la información que pretendemos obtener de acuerdo con diversos y variados objetivos. Así, por ejemplo, si se trata de una entrevista profesional, el guión se atendería al perfil y competencias requeridas para el puesto de trabajo que se desea cubrir, y las preguntas guardarían estrecha correspondencia con el perfil del puesto de trabajo. Otros ámbitos de aplicación son la educación, la psicología o la psiquiatría. En algunos casos puede caracterizarse por no seguir un guión muy estructurado, utilizándose como guía de la misma el propio desarrollo de las respuestas dadas por el entrevistado. Si se tratara de una entrevista de investigación, la estructura y estilo de las preguntas se realizarían de acuerdo con los objetivos de la propia investigación. La entrevista como método de obtención de información no estandarizado, dificulta la comparación de la información recogida con otros grupos, por lo que se debe tener especial cuidado en interpretarla con códigos estandarizados.

Siguiendo a Fernández Ballesteros, (1983: 360), las características de la entrevista como técnica de recogida de información son las siguientes:

- Existencia de un objetivo o propósito definido, lo cual la diferencia de la mera conversación.

- Relación directa, presencial, entre dos o más personas.
- Existencia de dos canales de comunicación, uno en el ámbito verbal, centrado en el contenido que se expresa, y otro corporal, relacionado con la forma como se expresa dicho contenido.
- Asignación de roles específicos cada uno de los participantes, lo que hace que la entrevista se convierta en una relación asimétrica. El rol del entrevistador sería definido y desempeñado por un experto en ciencias humanas, con el cometido de recabar información de acuerdo con los objetivos de la entrevista, y el rol de entrevistado estaría vinculado, fundamentalmente, en aportar información atendiendo a las cuestiones que se le plantean, de la forma más exacta y válida posibles.

Con la técnica de la entrevista se consigue una representatividad muestral de los entrevistados más fiable que la obtenida por la técnica del cuestionario, dado que con la primera está asegurado que quien responde es la persona seleccionada de acuerdo con la técnica de muestreo elegida, y no así en el cuestionario; por el contrario, un mayor tamaño de la muestra conseguible por la técnica del cuestionario, puede compensar la anterior debilidad mencionada para conseguir la representatividad de la muestra.

10.4.1. *Modalidades de la entrevista*

La modalidad de la entrevista depende del enfoque de investigación, de sus objetivos y del tipo de información a recoger.

Dimensiones	Modalidades de la entrevista
Finalidad	Clínica o psiquiátrica / De orientación / De selección / De investigación
Estructuración	Estructurada / No estructurada
Directividad	Directiva / No directiva
Número de participantes	Individual / Grupal / Varios grupos

La entrevista psiquiátrica y psicológica. En uno de los primeros ámbitos en los que se ha utilizado la técnica de la entrevista como método de recogida de información, ha sido en psiquiatría, como técnicas de «diagnóstico conversacional». A partir de las respuestas aportadas por el entrevistado, el psiquiatra formula hipótesis que trata de confirmar mediante las siguientes cuestiones que le pueden llevar a confirmar o negar sus

hipótesis y a plantear otras nuevas que guiarán la formulación de otras preguntas. El entrevistador se apoyará en la información aportada para tratar de responder a sus objetivos que tienen que ver con la solución de los problemas que han llevado a la consulta clínica.

Dependiendo del modelo psicológico que fundamenta la formación del entrevistador y que le caracterizan, nos encontramos con entrevistas que siguen el modelo psicoanalítico de Freud, el de Rogers, o el de cualquier otro enfoque o modelo psicológico. La conducción de la entrevista, y la interpretación de la información, guardan estrecha dependencia con las teorías que sustentan los modelos en las que se apoyan. También tiene una larga tradición la entrevista con fines a la realización de un informe de orientación en el ámbito de la Educación, y en la selección de personal, en el ámbito de la empresa y del desarrollo de recursos humanos.

La entrevista de investigación, el ámbito de las Ciencias Sociales, es donde más se ha desarrollado la entrevista como técnica de recogida de información y datos. La entrevista forma parte de las técnicas utilizadas en la «investigación tipo encuesta», cuyo objetivo es recoger información de una muestra representativa de una población definida, en torno a determinadas actitudes, opiniones o motivaciones en las que está interesado en profundizar el investigador. Dentro de la investigación, la entrevista suele utilizarse como técnica complementaria del cuestionario, con el fin de ampliar y precisar la información que de otro modo, por la rigidez del cuestionario, no podría conseguirse.

La entrevista, cuando es *estructurada y sistemática*, también se guía por unas cuestiones previamente delimitadas por el investigador, pero ofrece la oportunidad al investigador de aclarar la pregunta, e incluso de añadir o profundizar en la misma, en función de la respuesta dada por el entrevistado. Además el entrevistador puede recoger información del contexto de la entrevista y del comportamiento del entrevistado, fruto de su observación.

Sea en la modalidad de entrevista, o de cuestionario, el investigador debe cuidar de que las preguntas sean interpretadas y comprendidas de un único modo. Lo cierto es que este objetivo es casi imposible de conseguir, pero el cuidado en la construcción gramatical de las preguntas, la claridad en las instrucciones, la univocidad de las cuestiones, la cuidada definición de algunos términos, reducirá la ambigüedad en la comprensión de las mismas.

10.5. Las pruebas de rendimiento. Evaluación normativa y evaluación criterial

Por su utilización en el ámbito de la educación, pasamos a describir las pruebas de rendimiento, comenzando por la delimitación de dos conceptos evaluativos, como son la evaluación normativa y la evaluación criterial. Después describiremos el tipo de pruebas de rendimiento utilizados dentro de cada una de estas concepciones evaluativas.

Fue Glaser, en 1963, quien utilizó las expresiones evaluación basada en pautas de normalidad y evaluación basada en criterios; con el concepto de evaluación basada en normas estadísticas describía los tests de rendimiento en los que se destaca la posición relativa de un sujeto dentro de un grupo, llamado grupo normativo, que ha realizado el mismo test. El rendimiento del sujeto se interpreta en relación con el rendimiento del grupo normativo. La evaluación referida a normas utiliza tests basados en pautas de normalidad destinados a determinar la posición de un sujeto examinado en relación con el rendimiento de un grupo de sujetos.

El concepto de evaluación basada en criterios lo refería a la evaluación realizada mediante pruebas o tests que identificaban el dominio, o su ausencia mediante «conductas específicas» por parte del sujeto sobre las preguntas del examen. La evaluación criterial determina el rendimiento de un sujeto mediante pruebas de rendimiento elaboradas con referencia a un dominio de conocimientos, habilidades, destrezas... que deben ser dominados por el alumno.

La necesidad de la evaluación basada en criterios coincidió, en la década de los 60, con el desarrollo de la enseñanza por objetivos conductuales, definido como un propósito instructivo que describe lo que los educandos podrán hacer, o deberían ser capaces de hacer, después de la instrucción y que antes no podían hacer.

10.5.1. Las pruebas de examen en la evaluación normativa

En la década de los 70, como consecuencia de las exigencias que imponían la formulación de los objetivos educativos en términos de conducta; una vez formulado el objetivo a alcanzar y realizada la instrucción para lograrlo, no queda otra cosa que comprobar el rendimiento conseguido en coherencia con lo pretendido, y se necesitaban tests de rendimiento adecuados a los objetivos tan claramente formulados. Si el examen no es congruente con el objetivo la evaluación no sirve. Y fue

Popham (1978, 1983:36 y ss.), principalmente, quien desarrolló este concepto evaluativo, y en el que nos apoyamos para describir estas pruebas de examen, que pueden incluirse en las categorías de pruebas de selección de respuestas y pruebas de elaboración de respuestas.

La pruebas de selección de respuestas ponen al examinando en situación de elegir entre las opciones que figuran en el mismo, siendo las más comunes de verdadero / falso y de elección múltiple. La pruebas de elaboración de respuesta requieren del examinando la creación de una respuesta, generalmente escrita, de un modo breve o extenso según se especifique.

Cada una de estas pruebas, y sus diferentes modalidades, está indicada para evaluar un tipo diferente de aprendizaje. Si lo que se pretende evaluar es la información o conocimiento de hechos que tiene un sujeto resulta muy adecuado un test de selección de respuesta, pero, sin embargo, las estas pruebas no pueden valorar la capacidad de un sujeto para expresarse, para sintetizar ideas o su capacidad crítica y creativa ante un problema, un acontecimiento o un hecho histórico sobre el que se le pide un comentario personal. Las pruebas de selección de respuesta también se utilizan para medir aptitudes intelectuales complejas, actitudes, intereses, valores, discriminaciones sutiles, asociaciones, etc. de manera similar a como hemos visto en el tema referido a la construcción de escalas para la medida de las actitudes.

El tiempo requerido para su corrección es mucho menor que el requerido para una prueba de elaboración; sin embargo el proceso de elaboración de las pruebas de selección supone mayor tiempo por la precisión exigida para la formulación de las preguntas, y se alcanza mayor objetividad en la calificación que en los de elaboración, siempre que las preguntas sean precisas, bien formuladas y representativas de los contenidos que se pretenden evaluar.

La influencia que una y otra prueba puede tener en el aprendizaje de los alumnos tiene más que ver con el hecho del uso generalizado como única técnica de evaluación que por la prueba en sí. Si un profesor emplea mayoritariamente pruebas de selección de respuesta sus alumnos se centrarán más en la información de hechos puntuales de una manera comprensiva y desarrollaran las habilidades intelectuales en las que más inciden los ítems, o preguntas, que aparecen en el test. Y si el profesor utiliza fundamentalmente pruebas de elaboración de respuesta, llevará a que sus alumnos se preocupen por estudiar temas amplios y a desarrollar su capacidad para organizar y presentar ideas de modo escrito.

Sea cual fuere la prueba para la evaluación de los aprendizajes elegida, conviene cuidar una serie de aspectos a la hora de elaborar las pruebas de examen, como son las siguientes:

- Formular las preguntas de un modo preciso.
- Evitar términos ambiguos.
- Evitar el uso de pistas involuntarias.
- Cuidar la sintaxis, de modo que no sea innecesariamente complicada.
- Utilizar un vocabulario adaptado al nivel de aprendizaje de que se trate.
- Incluir las indicaciones necesarias en el examen para que el examinado pueda conocer lo que se pregunta, y adecuar su respuesta.

10.5.1.1. Las pruebas de selección de elección binaria

Las pruebas de elección binaria están integradas por preguntas sobre las cuales se le pide al sujeto con conteste si es verdadero / falso, También se les puede pedir que conteste sí / no, correcto / incorrecto. Se utilizan para comprobar la habilidad para identificar la corrección de enunciados sobre hechos o definiciones. Entre sus ventajas están las de poder abarcar una gama muy amplia de contenidos de manera breve. Entre sus desventajas se encuentra la posibilidad de acertar la respuesta por pura casualidad el 50 por 100 de las veces, además del gran número de pistas inadvertidas que suelen mostrar estos ítems, lo que eleva más las posibilidades de aciertos por casualidad. El uso frecuente de este tipo de pruebas fomenta el aprendizaje memorístico de los contenidos enunciados de un texto o de las explicaciones del profesor. Un problema añadido a este tipo de pruebas se debe al hecho de que no siempre los enunciados son totalmente verdaderos o totalmente falsos. Su modo de corrección es muy sencillo pero sus puntos débiles no siempre la hacen la prueba de evaluación más adecuada.

Para su redacción, se aconseja evitar formulaciones ambiguas, la inclusión de pistas inadvertidas, las frases extensas y complejas, la inclusión de más de un concepto en un único enunciado, el empleo de enunciados negativos. En cuanto al número de ítems a incluir, se aconseja utilizar un número similar de ítems verdaderos y falsos.

10.5.1.2. Las pruebas de asociación o confrontación

Una prueba de confrontación consiste en la presentación de dos listas paralelas de palabras, siendo la columna de la izquierda la de las premisas y la de la derecha la de las respuestas. Las columnas van precedidas de las indicaciones necesarias para comprender lo que se espera del sujeto para realizar el test. Entre las dos columnas el alumno debe formar pares asociados de acuerdo con la base lógica introductoria. Cada una de las listas debe ser homogénea, referida a un único tema o característica. Cuando la relación de la derecha contiene más elementos que la de la izquierda se llama confrontación imperfecta, y aumenta la eficacia de la prueba. Suelen utilizarse estas pruebas para evaluar asociaciones sencillas como definiciones y términos, hechos históricos y fechas. Entre sus ventajas se señalan la obtención de gran cantidad de información en breve espacio y tiempo; su elaboración es relativamente fácil así como su corrección y objetividad en la misma. Entre sus desventajas destacamos que un uso excesivo de este tipo de pruebas favorecería un aprendizaje memorístico en detrimento de un aprendizaje de mayor profundidad.

10.5.1.3. Las pruebas de elección múltiple

Es el tipo de prueba de elección más utilizada, y se utiliza para detectar la capacidad del sujeto para reconocer conocimientos de hechos, para detectar habilidades intelectuales complejas, o para conocer actitudes, valores o intereses en relación con un tema.

Un ítem de elección múltiple está formado por una pregunta o un enunciado incompleto, al que se le denomina tronco y es el estímulo al cual el sujeto debe responder mediante la selección de la respuesta entre las alternativas posibles que se le presentan. Entre las alternativas hay varias respuestas incorrectas, denominadas factores de distracción, y al menos una correcta, aunque pueden presentarse varias alternativas correctas, con distinciones muy sutiles, difíciles de discernir, y el examinando deberá escoger la mejor, lo que le obligará a tener un conocimiento completo del tema que se evalúa.

Entre sus ventajas cabe señalar su versatilidad para utilizarse en la evaluación de aprendizajes muy diferentes, y su calificación objetiva. Al ofrecer más alternativas de respuesta que el de elección binaria aumenta la dificultad de ser contestado por casualidad, aumentando la fiabilidad de cada ítem. También permite incluir entre las alternativas de respuestas errores con fines diagnósticos. Entre sus limitaciones está la posibilidad

que ofrece al estudiante de reconocer una respuesta correcta, que por sí mismo no hubiera sido capaz de emitir o de elaborar. El examinando tampoco cuenta con la posibilidad de realizar sus propias síntesis o de responder de un modo creativo, crítico o con matices personales.

En cuanto a las reglas a tener en cuenta para la elaboración de pruebas de respuesta múltiple, caben señalar las siguientes:

- El tronco debe presentar una pregunta o un problema completo, y estar redactado con claridad. El tronco debe recoger la mayor parte del contenido del ítem de modo que permita seleccionar rápidamente la respuesta de entre las alternativas posibles. Cuando el tronco deba contener mucha información, ésta se puede equilibrar con alternativas más largas.
- Conviene evitar los enunciados negativos, dado que pueden introducir confusiones al no advertirse el no, a no ser que la esencia de la pregunta se centre en cosas que no deberían hacerse.
- Cada una de las alternativas debe ser congruente gramaticalmente con el tronco del ítem, de lo contrario supondrían pistas que invalidarían su eficacia.
- Se debe procurar que haya una sola alternativa que represente la respuesta correcta o la más correcta.
- Las alternativas deben ser respuestas plausibles, evitando que las alternativas distractoras resulten claramente incorrectas.
- El número de alternativas aconsejable es de cuatro o cinco. Si se presentan sólo tres el examinando tienen una probabilidad del 33 por 100 de acertar por casualidad.
- La posición de las respuestas correctas en los diferentes ítems se deben alternar al azar, utilizando, por ejemplo, una tabla de número aleatorios.
- Se deben evitar las alternativas del tipo «ninguno de los anteriores» o «todos los anteriores».

Las pruebas de elección múltiple ofrecen muchas ventajas pero debe evitarse su uso en exclusividad olvidando las pruebas de elaboración de respuestas.

10.5.1.4. Las pruebas de respuesta breve

Este tipo de pruebas se caracterizan por la brevedad de la respuesta que se requiere del sujeto. Ante una pregunta directa se le pide que ela-

bore una frase ante, o mediante un enunciado incompleto que escriba la palabra o palabras que faltan. Son adecuados para evaluar aprendizajes sencillos, conocimientos de hechos, cuestiones complejas, definiciones, conceptos, diferencias, características, permitiendo al examinando reflejar un conocimiento profundo acerca de la cuestión que se le pregunta.

Entre sus ventajas están su fácil construcción y exigir de los alumnos la elaboración de la respuesta correcta. Entre sus desventajas se encuentran una menor objetividad en su calificación que la de las pruebas de selección de respuesta, dado que el profesor puede encontrarse con gran variedad de respuestas, y el mayor tiempo requerido para su corrección.

Pasamos a comentar alguna de las reglas a tener en cuenta en la elaboración de las preguntas de respuesta breve:

- El ítem de respuesta breve debe ser construido de forma que invite a una respuesta concisa y clara, debiendo proporcionarse instrucciones claras al comienzo de la prueba.
- Cuando se trata de enunciados con espacios para completar, éstos deben tener la misma extensión para no introducir pistas indeseadas.

10.5.1.5. Las pruebas de ensayo

Es la modalidad más común de pruebas de elaboración de respuesta, indicadas para evaluar aprendizajes complejos que exijan la realización de síntesis, comparaciones, contrastes, explicaciones, valoraciones. Los ítems pueden estar formulados de modo que demanden una respuesta muy breve o extensa. Una manera de restringir la respuesta es indicar un espacio limitado, aunque esto discrimina en función del tamaño de letra.

Entre sus ventajas está la posibilidad de evaluar aprendizajes complejos que permiten al estudiante expresarse de un modo original exponiendo, explicando, caracterizando, diferenciando, relacionando, contrastando, sintetizando y/o valorando todos los elementos relevantes a la cuestión formulada. Las desventajas que tienen son el excesivo tiempo que lleva su corrección y la dificultad de calificar con objetividad lo que reduce su fiabilidad y su validez si no se cuidan estos inconvenientes. Si debido al excesivo tiempo que exige su corrección ésta se hace de un modo rápido y superficial, reduce aún más la validez de la evaluación. Otra de sus desventajas es la oportunidad que tienen los estudiantes con escasos conocimientos para responder a lo que se les pregunta, escribiendo de una manera aparentemente convincente apoyándose en un

vocabulario cuidado y en una buena capacidad intelectual. La utilización exclusiva de las pruebas de ensayo para evaluar los aprendizajes de un alumno, dejaría fuera de la evaluación la mayor parte del contenido de la materia.

Las reglas a tener en cuenta para la redacción y calificación de las preguntas de ensayo son las siguientes:

- La pregunta debe expresar claramente lo que se espera que conteste el sujeto y debe incluir suficientes guías para que el sujeto sepa qué se espera que responda o qué orientación debe dar a la respuesta. Una pregunta mal formulada, llevaría a interpretaciones diferentes por parte de los examinandos.
- Se debe indicar el tiempo aproximado para cada pregunta así como el peso de cada pregunta en la calificación.
- La inclusión de preguntas opcionales supone la realización de exámenes diferentes para un mismo grupo de alumnos.
- Los criterios claves de calificación deben prepararse antes de iniciar la corrección de los exámenes, incluyendo la incidencia que tendrán las faltas de ortografía, de puntuación, redacción o presentación, en la calificación final.
- La corrección de una pregunta en todos los ejercicios, antes de pasar a la siguiente, favorece la objetividad en la corrección, así como su evaluación de manera anónima.

10.5.2. Las pruebas de evaluación criteriales

La evaluación referida a criterio (ERC) es una modalidad de evaluación educativa cuyo objetivo es recoger información mediante un instrumento estandarizado con la finalidad de conocer el conjunto de conocimientos y habilidades adquiridas por un sujeto sobre un dominio de referencia. Los elementos fundamentales de la ERC son dos: la definición exhaustiva del dominio instruccional y el establecimiento de un criterio decisorio sobre el dominio que debe lograr el estudiante. El cometido de la evaluación basada en criterios es incrementar la capacidad de los tests para ofrecer descripciones claras del rendimiento del examinando. Un test basado en criterios se emplea para determinar la posición de un individuo con respecto a un dominio de la conducta perfectamente definido. Seguimos en la exposición de este apartado los trabajos de Jornet y Suárez, 1989, 1994).

Las principales fases para la elaboración de tests criteriales de clase son las siguientes:

- Especificación del dominio educativo.
- Elaboración de un plan descriptivo del test.
- Creación de ítems.
- Análisis de los ítems.
- Determinación de estándares y puntos de corte.
- Fiabilidad y validez de los ítems.

10.5.2.1. Especificación del dominio educativo

Un campo o dominio de conducta consiste en un conjunto de habilidades que los examinandos ponen en juego en una situación de examen. El rendimiento vinculado al dominio se refiere a una forma de competencia o de habilidades para la realización de una determinada tarea.

El atributo más importante de un test basado en criterios consiste en brindar una descripción clara de lo que realmente significa el rendimiento medido. Otra cuestión que preocupa a la hora de elaborar un test basado en criterios es la determinación del grado de generalidad del campo de conducta que se mide.

La definición del dominio tiene que ver con lo que se denomina universo de medida. El dominio puede definirse desde una concepción genérica los elementos que componen un determinado programa de aprendizaje o desde una concepción métrica el conjunto de todas las unidades que lo componen, que pueden ser objetivos o ítems. La definición del dominio por objetivos es más ambigua que por ítems.

De la calidad de la definición del dominio depende la calidad de la interpretación de las puntuaciones obtenidas en el TRC; por ello es conveniente trabajar con el dominio de una lección o una unidad didáctica para poder elaborar los ítems en congruencia con el dominio.

10.5.2.2. Plan descriptivo de un test basado en criterios

Un test basado en criterios requiere de un plan descriptivo de las especificaciones que debe realizar el test, y sirve a dos funciones: 1) Facilitar la redacción de los ítems en congruencia con el dominio a evaluar, indicando si debe predominar un enfoque abstracto, genérico o concreto, aplicativo, relacional o sintético y concretar los aspectos a eva-

luar con relación al dominio. 2) Comunicar a quienes han de responder, lo que pretende medir dicho test, para su interpretación precisa.

10.5.2.3. Creación de ítems

Cuando delimitamos los tipos de ítems que pueden emplearse para evaluar la posición de un sujeto con respecto a las unidades del dominio definidas operativamente referidas a competencias intelectuales, habilidades o actitudes estamos definiendo operativamente el propio dominio de conducta. Cuanto más y mejor delimitadas estén las especificaciones del test más fácil será la elaboración de los ítems.

La elaboración de todos los ítems posibles relativos a un dominio debe regirse por reglas de generación de ítems, para facilitar su escritura, lo que permite plantear estructuras genéricas, diferenciando partes fijas y partes variables, para facilitar la redacción de los ítems. «Mediante el análisis de conceptos se identifican los atributos críticos o características universales, y las variables o características particulares, y mediante la combinación entre atributos y variables se generan ejemplos y contraejemplos a utilizar en la formulación de ítems y en la enseñanza. Una vez realizado el análisis y especificado el dominio y la aplicación de las reglas de generación de ítems, se puede formar un banco de ítems que permitirá obtener pruebas paralelas aplicables a diferentes grupos y momentos». (Jornet y Suárez, 1994:433).

En la creación de ítems hay que cuidar tanto la fidelidad al dominio como su redacción.

- La relación de la raíz, o tronco del ítem, con las alternativas de respuesta posible, debe guardar coherencia gramatical, sin sugerir la respuesta mediante una asociación verbal particular.
- La respuesta correcta debe poder situarse en todas las posiciones posibles, y presentar una extensión similar, de modo que no ofrezca pistas involuntarias para su elección. Las alternativas distractoras deben serlo realmente, debiendo ser creíbles, evitando que se eliminen por sí mismas.
- Los contenidos deben centrarse en aspectos relevantes, no en cuestiones de detalle o triviales; tampoco en cuestiones de opinión o discutibles, a no ser que sea éste el objetivo perseguido.
- Para salvaguardar la coherencia entre ítems e instrucción, debe evitarse el vocabulario rebuscado, las pistas involuntarias, o incluir elementos absolutistas como siempre o nunca.

- Cada ítem debe centrarse en sólo un aspecto, de un modo claro, preciso, no ambiguo y no excesivamente largo.

Las principales cuestiones a constatar en la generación de ítems y en el análisis de los mismos son dos: la representatividad, con relación al universo de medida (congruencia ítemobjetivo) y su eficacia para apreciar el cambio producido por la intervención pedagógica.

10.5.2.4. Análisis de ítems

De acuerdo con Berk (1978, en Jornet y Suárez, 1994), el análisis de ítems presenta dos características: la revisión lógica y la revisión empírica.

- La revisión lógica busca lograr la validez de contenido del ítem, y se centra en tres aspectos: la congruencia ítemobjetivo, la calidad técnica y los sesgos. En relación con la congruencia ítemobjetivo, una vez que se ha especificado el dominio, se ha estructurado, jerarquizado y operativizado se generan ítems buscando la congruencia con el objetivo siguiendo un patrón de generación específico para cada objetivo. Su análisis posterior se realiza recurriendo a jueces de expertos. En cuanto a la calidad técnica los ítems deben apreciar con fidelidad los aprendizajes logrados tras los procesos de enseñanza. En relación con los sesgos de tipo cultural es conveniente evitarlos, pudiendo hacerse recurriendo al análisis de jueces de expertos.
- La revisión empírica suele centrarse en el análisis de índices de dificultad, discriminación y homogeneidad. Los índices de dificultad comparando sus resultados entre grupos pre y postest, o que superan o no el criterio seleccionado para apreciar la capacidad de cada ítem para detectar los cambios producidos por la instrucción. Los índices discriminación, también con grupos pre y postest. Los índices de homogeneidad, basados en la congruencia ítemobjetivo, en la validez de los ítems, referida a la validez de contenido, la calidad técnica y los sesgos.

10.5.2.5. Determinación de «estándares» y «puntos de corte»

En los TRC es clave, junto con la definición del dominio, la determinación de estándares y puntos de corte. El término «estándar» se utiliza para referirse a puntuaciones verdaderas o teóricas, libres de error. La mayoría de los procedimientos para la determinación de estándares se

basan en el juicio de expertos, siendo conveniente que éstos sean los mismos que han definido el dominio y que han realizado el análisis de los ítems y la finalidad de la prueba.

La expresión «punto de corte», referido a las puntuaciones observadas en los alumnos, consiste en establecer una puntuación que diferencie entre el sujeto que ha alcanzado un nivel mínimo de competencia que le permite promocionar, aprobar, del que no lo ha alcanzado. Para la determinación del «punto de corte» se suele utilizar el método de validación de grupos de criterio, utilizando como criterio de contraste dos grupos de sujetos: los que han superado con éxito el aprendizaje de una unidad y los que aún no han iniciado su aprendizaje.

10.5.2.6. Fiabilidad y validez de los tests basados en criterios

La elaboración de ítems debe hacerse de modo que cumplan las exigencia de fiabilidad y validez para que sean reconocidos como correctos por ellos mismos y por otros expertos en evaluación.

La determinación de la fiabilidad en este tipo de test se refiere a la consistencia en la asignación de los sujetos a las categorías de apto / no apto, ya sea mediante la aplicación de dos formas paralelas de la misma prueba a un mismo grupo de sujetos o por la aplicación del mismo test en dos ocasiones consecutivas. El problema de la validez suele referirse al de la validez de contenido en clara referencia a la congruencia entre el ítem y el objetivo. Para lograrla se requiere: 1) Definir, con la mayor riqueza y precisión, el dominio que será objeto de valoración. 2) Especificar el tipo de información que debe ofrecer el test. 3) Redactar un número adecuado de ítems representativos. 4) Comprobar la bondad de tales especificaciones e ítems.

Las operaciones fundamentales de las pruebas criterioles como son la especificación del dominio, el análisis de los ítems, la determinación de los estándares y la fiabilidad y validez deben ser continuamente revisadas hasta lograr un ajuste satisfactorio, siendo la más importante la especificación del dominio (Jornet y Suárez, 1994: 443). En conclusión, la elaboración de pruebas basadas en criterios requiere definir el dominio y especificar con precisión la información que debe recoger el test, determinar el número de ítems a partir del cual obtendremos una información fiable sobre el rendimiento alcanzado por el alumno en ese dominio, decidir con la puntuación de corte para determinar si es suficiente el dominio alcanzado por un alumno en la resolución de la tarea para promocionar o no y aprovechar la información recogida para la toma de decisiones de

mejora del programa y de su enseñanza, del profesor y del contexto organizativo.

10.6. Ejercicios de autoevaluación

1. Realice una síntesis sobre las características de la observación sistemática.
2. Elabore una hoja de registro de conductas individuales sobre un tema en el que esté interesado en el ámbito de la intervención educativa-social.
3. Elabore una escala para la autovaloración de las conductas individuales por parte de los miembros de un grupo de jóvenes sobre los que vd. tenga relación.
4. Elabore un ejemplo de cada una de las pruebas utilizadas en la evaluación del rendimiento.

10.7. Referencias bibliográficas

- BERK, R. A. (1978). *The applications of structural facet theory to achievement tests construction*. Educational Research Quarterly, 3, 6272.
- EVERTSON, C. M. y GREEN, J. L. (1986). «Observation as inquiry and method», en WITTRICK: *Handbook of research on teaching*. N. Y. London: MacMillan. (162213). Traducido al español: WITTRICK, M.: *La investigación de la enseñanza I y II*. Ed. Paidós-MEC. Barcelona, 1989.
- GLASER, R. (1963): «Instructional technology and the measurement of learning outcomes: Some questions», en *American Psychologist*, 18 (51921).
- JORNET, J. M. y SUAREZ, J. M. (1989): *Conceptualización del dominio educativo*, Bordón.
- JORNET, J. M. y SUAREZ, J. M. (1994): «Evaluación referida al criterio» en *Tratado de Educación Personalizada*, Madrid, Rialp.
- POPHAM, W. J. (1983): *Evaluación basada en criterios*. Madrid: Magisterio Español.
- POSTIC, M. y DE KETELE, J. M. (1992). *Observar las situaciones educativas*. Madrid, Narcea.

TEMA 11

Procedimientos para la recogida de información II. La observación participante, el análisis de contenido, la técnica Delphi y los grupos de discusión

ESQUEMA

- 11.1.** La observación participante. Orígenes y características
 - 11.1.1. La función del observador en la observación participante
 - 11.1.2. El diseño para la realización de la observación participante
 - 11.1.3. El proceso de recogida de datos en la observación participante
 - 11.1.4. El análisis de la información en la observación participante
 - 11.1.5. La elaboración de informe
- 11.2.** El análisis de contenido. Definición y características
 - 11.2.1. Las fases del análisis de contenido
 - 11.2.2. Las unidades de análisis y su contexto
 - 11.2.3. La inferencia como objetivo intelectual básico del análisis de contenido y la validez como criterio de éxito
- 11.3.** El análisis y la evaluación de necesidades
 - 11.3.1. Las fases seguidas en la valoración de necesidades
- 11.4.** La técnica Delphi.
 - 11.4.1. Características y posibilidades
 - 11.4.2. El proceso de recogida de datos mediante el Delphi
- 11.5.** Los grupos de discusión
 - 11.5.1. Características y posibilidades
 - 11.5.2. El proceso de recogida de datos mediante los grupos de discusión
- 11.6.** Ejercicios de autoevaluación
- 11.7.** Referencias bibliográficas

OBJETIVOS

- ✓ Conocer y comprender las características de la metodología de observación participante.
- ✓ Conocer y comprender algunas de las técnicas utilizadas para el registro de la información en la observación participante.
- ✓ Conocer y comprender las características del análisis de contenido.
- ✓ Conocer y valorar las características del análisis y valoración de necesidades.
- ✓ Conocer y comprender las características de la técnica Delphi.
- ✓ Conocer y comprender las características de los grupos de discusión.
- ✓ Valorar las ventajas e inconvenientes de uno y otro procedimiento y enfoque para la recogida de información y las situaciones y ámbitos más adecuados para la aplicación de cada uno de ellos.

11.1. La observación participante. Orígenes y características

La observación participante es una estrategia de campo amplia que combina el uso de varias técnicas como son el análisis de documentos, la entrevista a grupos y personas, considerados como informantes claves, la participación directa del observador en el campo, y la introspección.

La observación participante se puede definir como una técnica de investigación que se orienta al estudio de la vida cotidiana de los grupos humanos con el objeto de explorar, describir y comprender sus formas de vida y su significado. La observación participante es una estrategia metodológica utilizada en el trabajo de campo, la etnografía y la antropología. El trabajo de campo tiene dos acepciones, la primera se refiere a un referente geográfico en el sentido que da idea que el objeto de estudio no se encuentra en el espacio cotidiano del investigador; la segunda acepción señala el conjunto de técnicas necesarias para obtener la información empírica deseada, entre ellas la observación participante.

La etnografía es la descripción de los grupos humanos, que se consigue tras una prolongada permanencia en el campo con el grupo que se va a estudiar, durante la cual y mediante la observación participante y el empleo de informantes clave se obtienen la información y los datos que se analizan. La etnografía utiliza la metodología de la observación participante para realizar sus investigaciones, centrada en el estudio detallado de un pequeño número de casos, o único, con el fin de explorar fenómenos particulares en su propio contexto, a partir de la recogida de datos inestructurados y su análisis, con el objetivo de realizar interpretaciones explícitas del significado de las acciones humanas que acontecen, resultando en descripciones y explicaciones textuales, sobre las cuales se pueden realizar análisis estadísticos como una función complementaria.

En la investigación sociológica y en la Antropología clásica, a principios del siglo xx, la observación directa, con diferentes grados de participación por parte del investigador, fue utilizada en los trabajos de Park y Burgess, (1921), de la escuela sociológica de Chicago, y en los del antropólogo cultural Malinowski, sobre los isleños de Tobriand, del Pacífico occidental (1922), que con el énfasis especial de la recogida de información de primera mano y su análisis, de modo rigurosos, contribuyeron al desarrollo de la moderna antropología cultural y la etnografía. La etnografía también se ha enriquecido a por la influencia del historicismo, y de la hermenéutica, utilizadas para el estudio de la comprensión de textos

históricos, lo que suponía el estudio de personas que vivieron en otros períodos de tiempo con culturas diferentes a las de nuestros días. Por eso se vio adecuada su utilización para el estudio de culturas diferentes de las occidentales, en la sociedad contemporánea, como son las culturas del Pacífico occidental, y aquellas culturas que aún viviendo entre nosotros son reconocidas como diferentes.

La observación participante se orienta hacia el conocimiento de grupos específicos, comunidades, grupos étnicos, grupos de minorías y organizaciones, implicando una inmersión en la cultura del grupo seleccionado. Es especialmente indicada para entender la realidad social desde el punto de vista de sus actores y organizaciones. La utilización de los métodos de la etnografía se han ampliado y se ha mostrado eficaz en los ámbitos educativos y sociales, por su carácter descriptivo, exploratorio e interpretativo, de acuerdo con la idea de que en la interacción con los demás existe siempre el objetivo de obtener información sobre las personas en las situaciones en las que viven y se desarrollan.

11.1.1. La función del observador en la observación participante

La observación participante es el método en el que el observador participa en la vida diaria de los grupos que estudia, ya sea abiertamente en el papel de investigador o encubiertamente en algún otro papel, observando las cosas que ocurren, escuchando lo que se dice y preguntando a la gente, durante un período de tiempo largo para hacerse familiar al grupo, dejar de ser un extraño, para que el grupo se manifieste de modo natural. En la observación participante el investigador se integra, tanto como pueda, en las situaciones que estudia. El propósito de tal participación es desarrollar un punto de vista interno, del propio contexto, acerca de lo que está ocurriendo. Sin embargo, el profesional de la observación tiene que combinar el lado del observador y el de participante, de modo que sea capaz de comprender la situación desde un punto de vista interno, a la vez que aportar la perspectiva del profesional externo. Y es que, aunque la observación siempre es subjetiva, es necesario que el observador mantenga una cierta distancia profesional con el fin de favorecer la objetividad de la observación, adoptando una doble perspectiva que le permita percibir la realidad «desde dentro» y «desde fuera». Entre el observador y el grupo observado, se establece una comunicación, normalmente de carácter no verbal, en la que el observador está alerta a las claves de la realidad que observa, y, a través de ellas interpreta lo que ocurre, obteniendo un conocimiento profundo de la situación que observa.

Desde la observación participante se defiende el análisis desde la perspectiva de los propios actores del grupo estudiado, tratando de captar sus puntos de vista. Sin embargo, el observador, aunque está implicado en las situaciones, debido a su esfuerzo consciente de objetividad para observar los acontecimientos y los cambios que allí se produzcan, alcanza, en cierta medida, una postura marginal, para captar dichos cambios.

En relación con la función del observador la observación varía a lo largo de un continuo en los aspectos siguientes:

- De la observación totalmente participativa a la pasiva.
- De la comunicación del papel del observador: va desde la observación abierta a la observación encubierta, donde el personal y los participantes no son conscientes de la presencia del observador.
- De la comunicación del propósito de observación: desde su comunicación a todos los miembros, pasando por una explicación parcial, por observaciones encubiertas donde no se da ninguna explicación ni al personal del programa ni a los participantes hasta una explicación falsa sobre el propósito de la investigación al personal y participantes del programa.
- De la duración de las observaciones: varía desde una única observación en un tiempo limitado a múltiples y extensas observaciones.
- Del objetivo de las observaciones: va desde el extremo en que el objetivo es concreto, centrado en un aspecto o componente singular de la situación y cuestiones a observar, a un enfoque amplio, con una visión holística, global, de la situación. (En Patton, 1990: 217).

En el caso de la estrategia abierta, los principios de la observación participante que debe seguir el observador, siguiendo a Caplow (1972: 177), son:

- La explicación del papel del investigador debe ser breve y sencilla.
- La explicación deber ser lo suficientemente general para cubrir todas las categorías de los trabajos que se desean llevar a cabo. Las futuras actividades parecerán razonables sin que sea necesario recurrir a nuevas explicaciones.
- Hay que buscar ante todo el apoyo de personas clave: los demás moldearán su actitud conforme a la de aquellas.
- Hay que conseguir la colaboración de aquellas personas estudiadas que sean observadores hábiles y que estén bien situadas para

observar, dado que serán una ayuda inestimable en la investigación, tanto para la recogida de información como para la interpretación de las observaciones. Es lo que el etnólogo llama el informador privilegiado.

- No hay que buscar una «asimilación» total. No es posible ni deseable.
- No es necesario actuar exactamente como los otros; basta demostrar un interés amistoso hacia ellos y hacia sus actividades. Cualquier muestra de desaprobación moral o de condescendencia puede amenazar el éxito de la investigación.
- Hay que mantener contactos frecuentes. La observación no descubrirá grandes cosas hasta que no se haya vivido con la gente un cierto tiempo. Hay que evitar tomar partido en un conflicto que divida al grupo, a menos que el estudio se limite al partido que se sostenga.

Algunos de estos principios son también aplicables a las estrategias encubiertas. En la estrategia abierta hay una serie de fases en función de la relación de reciprocidad del investigador con los observados y de la duración de su estancia en el grupo observado, cada una con sus propias reglas de comportamiento, que ofrecen distintas oportunidades para el desarrollo de la investigación. Estas fases son: de recién llegado, de miembro provisional, de miembro categórico, de observador como persona y de migrante inminente.

Otro modo de acceder al escenario es hacerlo de manera encubierta, realizando el investigador su papel sin informar a los observados del proceso de investigación. Esta posición, especialmente recomendada cuando se quiere ocultar a los ojos del público en general que se va a estudiar la realidad social en la que viven, plantea problemas éticos a algunos investigadores, aunque otros investigadores piensan que al informar a los observados muchos aspectos de la conducta humana quedan oscurecidos como consecuencia de la presencia de un extraño que lleva a las personas a modificar sus comportamientos, tratando de presentar su mejor imagen, de acuerdo con las expectativas sociales.

Las formas de actuar el investigador en ambos casos son distintas: en la estrategia abierta el investigador debe manejar la tensión que su presencia produce en los observados (Ruiz Olabuénaga, 1999: 159). La colaboración de los observados es posible que termine cuando se evidencian las disonancias entre lo que se dice con lo que se hace. En la investigación encubierta no se plantean estos problemas, pero existen otros. El inves-

tigador maneja la información que transmite para conseguir que su identidad social permanezca oculta; en cualquier caso los observados se forman una imagen del observador, que implica un marco de referencia del observador y una imagen de lo que puede preguntarles y, por tanto, contestarle. El observador pasa por una serie de fases y sentimientos durante el trabajo de campo: desde la euforia, a la ansiedad, la desilusión, la emoción, la monotonía de las anotaciones, las entrevistas con los informantes y la observación de fenómenos repetitivos, de los sentimientos de amistad a la observación desapasionada e imparcial.

11.1.2. El diseño para la realización de la observación participante

El proceso de investigación mediante la observación participante conlleva las siguientes fases:

- Delimitación de los objetivos y escenarios.
- Identificación de las fuentes de datos, los procedimientos de acceso a los escenarios y las unidades de observación.
- Determinación de los procedimientos, instrumentos de recogida de información y los tiempos. Selección de informantes clave.
- Acceso al campo, a los escenarios. Recogida de información. Sistemas de registro.
- Análisis de la información y datos. Sistemas de codificación y de representación de los datos.
- Sistemas para la validación de la información. La triangulación.
- Interpretación de los resultados, realización de inferencias, verificación.
- Elaboración del informe.

Dado que los diseños de investigación cualitativa se caracterizan por no ser absolutamente lineales y apriorísticos, el orden no tiene por que ser fijo; así, los objetivos pueden verse ampliados y modificados, los modos de recogida de información pueden igualmente ampliarse, los procedimientos de análisis de datos se dan de modo simultáneo a su recogida, y la interpretación de los resultados y los sistemas de validación pueden aconsejar una ampliación de la recogida de información. De este modo, es posible que la investigación difiera del diseño original, influida por la propia realidad, por el punto de vista de los actores sociales que

pueden modificar la perspectiva del observador, especialmente en las primeras etapas de la permanencia en el campo, lo que le permitirá definir el problema y los nuevos elementos para continuar el estudio. Y es que el contexto confiere un significado específico a los problemas que se estudian, dado que los hechos son significativos desde el contexto en que ocurren, desde las personas que integran los grupos, desde sus relaciones, y desde los códigos culturales de la propia comunidad para interpretarlos. La observación suele comenzar con un necesario conocimiento amplio, descriptivo, de lo que acontece, para poco a poco ir centrándose en los aspectos claves de la problemática a estudiar, y así se manifiesta en los procesos de recogida y análisis de la información.

En este tipo de investigación, la observación participante, el propio observador es el principal instrumento, sus ojos y sus oídos son instrumentos esenciales, así como el diario o cuaderno de campo, donde se debe indicar el día, hora, lugar y persona de los hechos, debiendo escribir las notas de campo inmediatamente ocurrido un suceso, en presencia de los observados, siendo importante anotar aquellos comportamientos no verbales que sirvan para confirmar lo que se ha dicho verbalmente. Otros instrumentos a utilizar son los documentos escritos, cuestionarios, cámaras fotográficas o de video, así como las entrevistas a informadores clave, historias profesionales y encuestas. Otros documentos a utilizar son las biografías de personas, planos, relatos escritos, censos, etc.

11.1.3. El proceso de recogida de datos en la observación participante

La observación participante es una estrategia de campo amplia, que comienza antes de introducirse en el campo, con la situación de lo que se pretende estudiar, la delimitación de los objetivos de estudio y del plan de lo que se va a hacer, el medio físico, las personas y sus interacciones, la concreción de los aspectos prácticos de inmersión en el campo, el contacto con los grupos. La observación combina, simultáneamente, el uso de varias técnicas de recogida y análisis de la información, como son el análisis de documentos, la entrevista a personas en la situación e informantes específicos, la participación directa, la observación y la introspección. El conocimiento de la historia previa a los acontecimientos que van a ser observados es fundamental para comprender el medio social.

Observar lo que ocurre en una situación de estudio dada, las actividades, lo que dice la gente, cómo interactúa, la naturaleza del medio físico, etc. es importante en un enfoque comprensivo de trabajo de cam-

po, pero también es importante observar lo que no ocurre y sería esperable y deseable que ocurriera; si las metas de un programa, su diseño de implementación y las propuestas de ciertas cosas que deben ocurrir o se espera que ocurran, y no es así, debe ser anotado por el observador. Ello estará influido por la preparación, sentido común y experiencia del observador.

Una de las funciones del observador consiste realizar, mediante notas de campo, descripciones de lo que está siendo vivido y observado, y son importantes para determinar la calidad de los análisis e interpretaciones posteriores. Las notas de campo son la base fundamental de los estudios de caso en la investigación cualitativa, y deben recoger citas literales de la gente observada, de sus conversaciones, de los sentimientos y reacciones del observador sobre lo que está observando y de síntesis o los «insight» que genera la observación y su interpretación. El observador debe seguir ciertas estrategias para hacer las anotaciones, las grabaciones, de modo que sea capaz de conseguir que su trabajo no afecte a su participación; una norma básica a tener en cuenta es escribir rápida y brevemente y completar las notas de campo tan pronto como sea posible fuera de la situación de campo. La extensión o la amplitud depende de las decisiones del observador. Las clases de registro que suelen utilizarse para organizar la información recogida son: el registro básico de notas tomadas durante las observaciones, entrevistas o grabaciones, que deben conservarse sin ninguna alteración; el registro temático, basado en las descripciones y reflexiones del investigador en torno a diversos campos temáticos; el registro logístico, reflejando las dificultades encontradas, los acontecimientos significativos, las reflexiones personales; registros complementarios. (En López-Barajas, 1998: 27).

Goetz y LeCompte (1988: 128), sugieren las siguientes áreas en torno a las cuales se puede organizar la observación:

- ¿Quiénes están en el grupo o en la escena?
- ¿Qué está sucediendo? ¿Qué comportamientos son repetitivos y cuáles anómalos? ¿Qué recursos se emplean en dichas actividades y cómo son asignados? ¿Cuál es la naturaleza de la participación y de la interacción? ¿Qué estatus y roles aparecen en la interacción? ¿Cuál es el contenido de sus conversaciones? ¿Qué temas son comunes y cuáles poco frecuentes? ¿Qué relatos, anécdotas y admoniciones intercambian? ¿Qué lenguajes verbales y no verbales utilizan para comunicarse? ¿Qué estructuras tienen sus conversaciones?

- ¿Dónde está situado el grupo o la escena? ¿Con qué recursos cuentan? ¿Cómo asigna y emplea el espacio y los objetos físicos? ¿Qué sensaciones y sentimientos se detectan en los contextos del grupo?
- ¿Cuándo se reúne e interactúa el grupo? ¿Con qué frecuencia y cuánto tiempo se prolongan las reuniones? ¿Cómo emplea y distribuye el tiempo? ¿Cómo perciben los participantes su pasado y su futuro?
- ¿Cómo se interrelacionan los elementos identificados, tanto desde el punto de vista de los participantes como desde la perspectiva del investigador? ¿Cómo se mantiene la estabilidad? ¿Cómo surge y es orientado el cambio? ¿Cómo se organizan todos estos elementos? ¿Qué reglas, normas o costumbres rigen en la organización social? ¿Cómo se relaciona este grupo con otros grupos, organizaciones o instituciones?
- ¿Por qué funciona el grupo como lo hace? ¿Qué significados atribuyen los participantes a su conducta? ¿Cuál es la historia del grupo? ¿Qué símbolos, tradiciones, valores y concepciones del mundo se pueden descubrir en él?

Veamos un ejemplo de hoja de registro:

Lugar de la observación:

Fecha:

- ¿Quiénes están en la escena?
- ¿Qué está sucediendo?
- ¿Qué interacciones se producen?
- ¿Qué estatus y roles aparecen en la interacción?
- ¿Cuál es el contenido de sus conversaciones?
- ¿Qué lenguajes verbales y no verbales utilizan para comunicarse?
- ¿Dónde está situado el grupo o la escena?
- ¿Con qué recursos cuentan?
- ¿Qué sensaciones y sentimientos se detectan en los contextos del grupo?
- ¿Qué significados atribuyen los participantes a su conducta?
- Otras informaciones (¿Cuál es la historia del grupo?):
- Comentarios del observador.

En esta fase se pueden cometer errores irremediables si no se cuida recoger bien la información, comprobar los datos recogidos, evitando posibles errores de registro, de enumeración, de transcripción, de identificación, de tipografía o de clasificación, de pérdida de material, errores cometidos por los encuestadores o por los observadores, respuestas falsas o poco reflexivas. El observador debe, además, tener en cuenta que los indicadores pueden estar deformados por el hecho de que la situación está siendo observada, y que las personas pueden comportarse de modo diferente a como lo harían si la observación no tuviera lugar. Una observación no contaminada únicamente se logra cuando las personas no saben que están siendo observadas, o el observador permanece suficientemente tiempo en la situación hasta que pueda ser considerado como uno más del grupo.

La observación participante combina la observación y la entrevista informal. El observador debe ser capaz de conocer el lenguaje específico de los ambientes que observa así como la comunicación no verbal. Durante las actividades informales no planificadas suele ser difícil organizar las observaciones; es imposible anticipar qué tipo de cosas pueden emerger, pero es importante recordar que todo lo que ocurre en torno a las cuestiones objeto de la observación puede ser de interés. Las entrevistas completan la información recogida mediante la observación, pero la información que aportan los entrevistados puede ser erróneas o ser autojustificativas, porque los participantes informan de sus percepciones y desde su punto de vista. Por eso las observaciones proporcionan una comprobación de lo que se recoge en las entrevistas; éstas permiten al observador ir más allá del comportamiento externo para explorar los estados internos de las personas que han sido observadas.

Otra fuente de información son los documentos y registros sobre los temas de estudio, siendo importante que al comienzo del trabajo el investigador negocie su acceso a dichos documentos y archivos, dado que proporcionan información valiosa a la vez que estímulos para generar cuestiones a comprobar a través de la observación directa y de las entrevistas. Los análisis de documentos proporcionan otra visión de las escenas, y su información puede ser utilizada para guiar y comprobar la información recogida por otras fuentes. Cada fuente de datos tiene sus ventajas y desventajas pudiendo un enfoque compensar la debilidad del otro. El trabajo de campo no es un método o técnica simple. El observador está en el sitio en el que se desarrolla el programa observando y participando con la gente, entrevistándola y consultando los documentos básicos del programa. Y debe permanecer atento a todas las posibles fuentes de error, verificando la información recogida. La combinación de

estas tres técnicas: observación, entrevista y análisis de documentos, ayudará al trabajador de campo a aclarar los constructos desarrollados y a corregir los sesgos que aparecen cuando el fenómeno se estudia por un solo procedimiento. Todo junto resulta un enfoque multimétodo, que contrastado mediante la triangulación incrementa la validez y la fiabilidad de los datos de la investigación.

11.1.4. El análisis de la información en la observación participante

Cuando el observador ha acumulado suficiente información para realizar el análisis, los procedimientos básicos son la codificación, la búsqueda de categorías de significado, la relación entre categorías, la interpretación de los datos y la exposición de los hallazgos utilizando tablas, diagramas, análisis estadísticos si fueran necesarios, intercalados en la redacción escrita que serán incluidas en el informe. Mediante la codificación de las informaciones recogidas se va reduciendo la información y haciéndola manejable para la búsqueda de categorías de significado, necesarias para su interpretación. El análisis de los datos puede seguir los siguientes pasos:

- 1.º Lectura de la información textual.
- 2.º Asignación de códigos a las unidades de significado, indicándolos al margen del texto.
- 3.º Desarrollo de categorías para agrupar los códigos que representan segmentos e ideas con significado similar.
- 4.º Agrupación de categorías formando categorías de orden superior.
- 5.º Revisión de las categorías, creación de nuevas categorías y reagrupaciones y reasignaciones si fuera necesario.
- 6.º Relación entre las categorías, explicación según sus procedencia original y su relación con otras teniendo en cuenta las situaciones y sus momentos.
- 7.º Ampliación de la información si fuera necesario para una comprobación de los datos y verificación de la información.

La interpretación de los resultados puede recorrer también diversos procesos según la perspectiva formal del análisis: de categorías, temático, taxonómico, componencial, hermenéutico, semiótico, temático, simbólico, de valores, fenomenológico, filosófico e histórico. Los datos pueden ser analizados utilizando los programas informáticos de análisis textual

que existen en la actualidad, como puede ser el AQUAD, el NUDIST, el SPAD.T, o el SPSS, entre otros.

11.1.5. La elaboración del informe de la observación participante

El observador participante debe ser descriptivo en las notas de campo, reunir una variedad de información desde diferentes perspectivas, seleccionar los informantes clave para aprovechar sus conocimientos, sus habilidades, su posición, que pueden resultar de gran ayuda, usar citas representativas de los participantes del programa en sus propios términos, captando el punto de vista de los participantes, de sus experiencias, con sus propias palabras.

El observador, para la elaboración del informe, debe tener en cuenta los diferentes estadios del trabajo de campo, y en consecuencia deberá reflejar:

- Los motivos que le llevaron a interesarse por el tema y a iniciar el estudio.
- La revisión de la información disponible y literatura relevante al tema.
- La formulación del problema a investigar y de los objetivos del estudio.
- Las negociaciones para estudiar la situación en el contexto.
- El plan de acceso al campo, el modo de introducirse en las situaciones y relacionarse con las personas y los grupos.
- Las técnicas de recogida de información utilizadas y los análisis de información y ampliación de la permanencia en el campo, y la búsqueda de información complementaria.
- La validación de los hallazgos y de sus interpretaciones, y devolución de las mismas al grupo.
- La presentación de los resultados, elaboración de las conclusiones y recomendaciones.
- La presentación del informe final.

Patton, (1990:273) considera que el observador participante tenga en cuenta las siguientes recomendaciones:

- Comunicar la verdad e informar de un modo sincero en el estadio inicial al entrar a los sitios del trabajo de campo, teniendo en

cuenta que el investigador-observador es también observado y evaluado.

- Ser disciplinado y consciente en la toma de notas detalladas en todos los estadios del trabajo de campo y permanecer alerta durante la mayoría las situaciones rutinarias de la fase intermedia del trabajo de campo.
- Estar tan implicado como sea posible en experimentar el programa de un modo global, mientras se mantiene una perspectiva analítica en el trabajo de campo.
- Centrarse en unificar la información, dentro de unas líneas determinadas, en una síntesis útil del trabajo de campo, separando, nítidamente, las descripciones de las interpretaciones y de los juicios.
- Proporcionar un feedback formativo como parte del proceso de verificación del trabajo de campo.
- Incluir en las notas de campo y en los informes las experiencias propias del observador, sus pensamientos y sus sentimientos; Éstos son también datos de campo.
- Cuidar la entrega de la información a tiempo y observar su impacto.

11.2. El análisis de contenido. Definición y características

El análisis de contenido es una técnica de investigación destinada a formular, a partir de ciertos datos, normalmente recogidos en un soporte físico como es un documento escrito, una grabación, u otro medio, inferencias fiables y válidas aplicables a su contexto. El análisis de contenido ha llegado a ser un método científico capaz de ofrecer inferencias a partir de datos esencialmente verbales, simbólicos o comunicativos. Para Ander-Egg el análisis de contenido es una técnica de recopilación de información que permite estudiar el contenido manifiesto de una comunicación, clasificando sus diferentes partes, conforme a las categorías establecidas por el investigador, con el fin de identificar de manera sistemática y objetiva dichas categorías dentro del mensaje (1987:330).

Intuitivamente, el análisis de contenido podría caracterizarse como un método de investigación del significado simbólico de los mensajes, pero hay que decir que 1) los mensajes no siempre tienen un único significado, pudiendo contemplarse desde múltiples perspectivas; en suma, un men-

saje puede transmitir multiplicidad de contenidos incluso a un único receptor, y 2) no es necesario que exista coincidencia acerca de los significados, si bien el consenso o el acuerdo intersubjetivo sobre lo que significa un mensaje simplifica el análisis de su contenido, pero difícilmente la coincidencia puede servir como presupuesto para el análisis de contenido, puesto que esta coincidencia depende de que si se comparten la misma perspectiva cultural o sociopolítica. El rasgo más característico de los mensajes es que informan a una persona de forma vicaria, proporcionando al receptor un conocimiento acerca de hechos o sucesos que se producen en un lugar lejano, acerca de objetos que quizás existieron en el pasado o acerca de las ideas de otras personas. La naturaleza vicaria de las comunicaciones simbólicas es lo que obliga al receptor a formular inferencias específicas, a partir de los datos que le proporcionan sus sentidos, en relación con su medio empírico llamado contexto de los datos.

Todas las teorías del significado y de los fenómenos simbólicos, se asemejan entre sí por la importancia que conceden a la relación entre los datos y su contexto. Los datos son estímulos físicos, pero el interés por su significado parte de un nivel superior de abstracción, como un documento escrito, una película cinematográfica, una fotografía, una grabación verbal, siendo el contexto su propio medio por el que se manifiesta. La definición del análisis de contenido establece el objeto de investigación y sitúa al investigador en una posición concreta frente a la realidad. El marco referencial dentro del cual puede representarse el papel que desempeña el investigador, recurre a los siguientes conceptos básicos:

- El objetivo del análisis de contenido.
- Los datos y su contexto.
- Las unidades de análisis.
- Las fases del análisis de contenido.
- La inferencia como tarea intelectual básica del análisis de contenido.
- La validez como criterio de éxito.

Este marco de referencia tiene tres finalidades: prescriptiva, analítica y metodológica. Es prescriptivo, en el sentido de que debe guiar la conceptualización y el diseño de los análisis de contenido práctico en cualquier circunstancia. Es analítico en el sentido de que debe facilitar el examen crítico de los resultados del análisis de contenido efectuado por otros; y es metodológico en el sentido de que debe orientar el desarrollo y perfeccionamiento sistemático de los métodos de análisis de contenido.

11.2.1. Las fases del análisis de contenido

La finalidad del análisis de contenido es proporcionar conocimientos, nuevas comprensiones y guía práctica para la acción en relación con el tema que se estudia. Como todo proceso de investigación, debe iniciarse con la enunciación clara de la finalidad y objetivos de lo que el analista quiere conocer y conseguir con el análisis de contenido, y dado que el análisis de contenido suministra un conocimiento vicario, es decir, una información acerca de algo que no puede observarse directamente, el objetivo debe quedar situado dentro de su contexto y de los datos disponibles (Krippendorff, 1990).

Para la realización del análisis de contenido pueden establecerse las siguientes fases:

1. Establecer la unidad de análisis: un documento, un artículo, las narraciones sobre un tema específico aportadas por un grupo, las sesiones de clase, las anotaciones de la discusión en grupo, las anotaciones de la observación participante, una grabación en vídeo, etc.
2. Seleccionar la muestra del material de análisis: las anotaciones de tres sesiones de clase, los cinco primeros minutos, intermedios y finales de una grabación en vídeo, etc.
3. Determinar las categorías de análisis: según el objetivo del análisis de contenido, de la temática, serán necesarias diversas categorías: de significado, de valor, de personas. Las categorías vendrán determinadas por el propio contenido, por sus actores, por la frecuencia de aparición de determinados significados.
4. Realizar inferencias a partir de las categorías y su relación entre ellas, elaborando teorías explicativas.
5. Elaborar el informe, explicando los objetivos pretendidos, los materiales, el contexto, el proceso, los resultados, las conclusiones, y las fuentes documentales para verificar las conclusiones. Si el objetivo del análisis de contenido está enunciado de manera clara, inequívoca, podrá juzgarse si se ha podido realizar, adjuntando las pruebas, las fuentes de datos, que permitirán validar los resultados.

La codificación de unidades textuales mediante categorías conceptuales se realiza a partir de inferencias, siendo estas inferencias la principal preocupación del análisis de contenido. El análisis de contenido es una técnica de investigación para formular inferencias identificando de

manera sistemática y objetiva ciertas características especificadas dentro de un texto. Todo análisis de contenido debe realizarse en relación con el contexto de la información y los datos, y justificarse en función de éste.

11.2.2. Las unidades de análisis y su contexto

El elemento esencial de la comunicación, es la palabra. La unidad de análisis en el análisis de contenido es el elemento concreto de comunicación que se va a analizar. Se pueden considerar dos tipos de unidades básicas: las de tipo gramatical y las no gramaticales. En cuanto a las primeras, se suelen analizar los vocablos, especialmente teniendo en cuenta su significado dentro de las frases, su frecuencia de aparición y la relación de significado de unos con otros, así como el modo en que viene expresado. Las segundas hacen referencia a otros modos de manifestar la comunicación como es a través de fotografías, gráficos, o de documentos audiovisuales.

En todo análisis de contenido debe quedar claro qué datos se analizan, de que manera se definen y de qué población se extraen. Los datos son los elementos básicos, primitivos, del análisis de contenido, y constituyen la superficie que el analista debe tratar de comprender. Los datos pueden venir dados mediante documentos escritos, pudiendo ser documentos de prensa, comunicados, documentos oficiales, documentos personales, documentos gráficos documentos audiovisuales, obras literarias, cinematográficas, sobre los cuales el analista realiza un análisis a posteriori. Pero también puede realizar el análisis de la comunicación en un contexto real, en vivo, sobre la cual el investigador observa y toma notas, o registra mediante instrumentos técnicos, combinando la observación y el análisis textual, completándose el análisis con la revisión de las notas y la audición y visionado de las grabaciones. En la realización a esta segunda situación del análisis de contenido hay que tener en cuenta los siguientes elementos:

- El emisor, persona que comunica o los productores del contenido y las causas del de la comunicación, identificando sus intenciones, y determinando las características de las personas y de los grupos que emiten el mensaje.
- El código, o lenguaje empleado para establecer la comunicación, el mensaje, o contenido de la comunicación, lo que se quiere decir y el canal, o medio por el cual se emite la comunicación. Las características del contenido, como son describir tendencias en el contenido de la comunicación, comparar los distintos niveles de

comunicación, el estudio de su significado, el estudio de las formas de estilo, etc.

- El efecto del contenido, en relación con el receptor o persona o público al que se dirige el contenido de la comunicación, o que la recibe, reflejando sus intereses, actitudes y valores, los aspectos que más le llaman su atención del contenido y las actitudes y respuestas emitidas ante las comunicaciones recibidas.
- El contexto en el que se produce la comunicación debe hacerse explícito y descrito por el analista con el fin de incluir todas las circunstancias circundantes, antecedentes o consecuentes al acto de la comunicación y a lo que se comunica.

Cualquier trabajo de investigación debe definir los límites más allá de los cuales no podrá extenderse el análisis.

11.2.3. La inferencia como objetivo intelectual básico del análisis de contenido y la validez como criterio de éxito

En todo análisis de contenido el objetivo último es la formulación justificada de inferencias a partir de los datos originales, teniendo en cuenta el contexto. Para realizar y justificar esas inferencias, el analista de contenido debe contar con relaciones relativamente estables entre los datos y el contexto (o construir una teoría operacional de esas relaciones), incluida la aportación de los factores mediadores. Se realizará una construcción analítica teórica de esas relaciones, formuladas de manera que los datos aparezcan como sus variables independientes y el objetivo forme parte de sus variables dependientes. Una construcción analítica ofrece reglas de inferencia sirviendo como puente lógico entre los datos disponibles y la finalidad incierta situada en su contexto, debiendo conocer la forma específica de esas relaciones.

Más allá de su continuo compromiso con cuestiones educativas, psicológicas, sociológicas y políticas, el interés por el uso de la técnica del análisis de contenido, viene unido a la importancia de establecer criterios adecuados de validez. En todo análisis de contenido hay que especificar con claridad el tipo de pruebas necesarias para validar sus resultados, las inferencias sobre los fenómenos que se estudian, con el fin de que otras personas puedan comprobar si las inferencias son exactas. La metodología del análisis de contenido sólo puede progresar si se efectúan esfuerzos sistemáticos para validar sus resultados. Los análisis de contenido que no ofrecen pruebas adecuadas de su validez, resultan empíricamente inútiles,

y contribuyen poco al desarrollo de la metodología del análisis de contenido.

11.3. El análisis y la evaluación de necesidades

Los estudios de evaluación de necesidades consisten en la reunión de información y sus procesos de análisis correspondientes, con la finalidad de identificar las necesidades de los individuos, los grupos, las instituciones, las comunidades y la sociedad con el fin de identificar las áreas en las que existen deficiencias, con el objetivo último de la utilización de los resultados para la realización de acciones, en muchos casos concretadas en programas de intervención, tendentes a subsanar y mejorar la situación detectada, planificar su realización, y valorar sus resultados para comprobar si se ha logrado satisfacer las necesidades y si es necesario iniciar nuevas mejoras.

La definición más utilizada de «necesidad» en el contexto de valoración de necesidades es la que se refiere a la discrepancia entre el nivel deseable y el conseguido, y se dice que existe una necesidad en un determinado campo cuando el nivel actual es inferior al deseado. Una segunda acepción de la definición de necesidad hace referencia al deseo de una preferencia. Y una tercera acepción de necesidad se basa en las deficiencia. Entonces, existe una necesidad cuando: 1) Lo conseguido es inferior a lo que se pretendía, 2) Se desea algo que no se tiene y 3) Se da una deficiencia que hay que subsanar. (Kaufman, 1972)

El motivo más frecuente que llevan a la realización de una valoración de necesidades es la de proporcionar información para planificar acciones tendentes a solucionar las necesidades detectadas. Puede resultar en la identificación de necesidades que deberán ser resueltas con nuevos proyectos, la conveniencia de introducir cambios en proyectos en curso, o la especificación de dónde concentrar esfuerzos y recursos para mejorar proyectos bien ideados que se están realizando. La valoración de necesidades se suele utilizar para desarrollar planes de acción inmediata o a medio y largo plazo.

La valoración de necesidades es considerada un componente de la evaluación, siendo una de las actividades de todo proceso de investigación evaluativa, circunscrita a la fase inicial, y de la fase última conclusiva del mismo proceso evaluativo que cierra y abre un nuevo ciclo con la enumeración áreas de necesidad prioritaria a tener en cuenta en los planes de mejora, dentro de los procesos de mejora continua de las instituciones.

11.3.1. Las fases seguidas en la valoración de necesidades

Los estadios seguidos en los estudios de valoración de necesidades, son los siguientes:

1. *Los propósitos de una valoración de necesidades:* Como otras formas de investigación, planificar un estudio de valoración de necesidades requiere clarificar los propósitos de valoración y delimitar las áreas en las cuales los estudios de valoración de necesidades se van a concentrar.
2. *Las normas para la identificación de necesidades:* Cuando se planifica un estudio de valoración de necesidades en educación, las metas educativas, los resultados a lograr, son, frecuentemente, las normas frente a las cuales se compara el nivel obtenido para determinar las discrepancias. Otro modo de identificar las necesidades es la valoración subjetiva por parte de los informantes.
3. *Los diseños de valoración de necesidades:* Los diseños para la valoración de necesidades comienzan con una especificación del objetivo del estudio, lo cual supone una delimitación de los propósitos específicos y los ámbitos en los que se van a valorar las necesidades. Para ello se requerirá elaborar instrumentos con los indicadores representativos de los propósitos y ámbitos específicos, recoger la información, analizarla y realizar informes. Los instrumentos para recoger opiniones suelen ser los propios de la de investigación tipo encuesta: cuestionarios, inventarios, hojas de registro, entrevistas, etc.
4. *La asignación de prioridades de las necesidades:* El resultado de los estudios de valoración de necesidades es la identificación de muchas necesidades. Para maximizar la utilidad del estudio las necesidades identificadas deben ser ordenadas de acuerdo con el grado de consenso alcanzado, la prioridad según la importancia otorgada, la posibilidad que tiene de ser resueltas con éxito, y el impacto que va a tener en la satisfacción de la necesidad.
5. *El uso de los resultados:* Una de las características de los estudios de valoración de necesidades radica en su finalidad que es la utilización de los resultados en aquello por lo que se ha realizado el estudio. Supone el último estadio del proceso: ser activo en el uso de los hallazgos.

Dado que los análisis para la valoración de necesidades se realizan en diferentes situaciones, las estrategias utilizadas también varían. Suelen

utilizarse técnicas como la técnica delphi, los grupos de discusión, las entrevistas en profundidad, la observación, e instrumentos como el cuestionario, los inventarios, las hojas de registro.

11.4. La técnica Delphi. Características y posibilidades

Es una técnica basada en la consulta a un grupo de expertos, cuya finalidad es la detección y análisis de necesidades, y la búsqueda de soluciones, sobre una determinada cuestión que se aborda desde un enfoque prospectivo, anticipatorio, y que afecta a grupos poblaciones amplios, con el fin de elaborar programas de intervención que atajen los problemas antes de que se manifiesten, a partir del mayor grado de consenso logrado entre los expertos. Es una técnica originalmente utilizada para recoger información de las aportaciones personales de un grupo de expertos, sobre los diferentes aspectos que definen un determinado problema y el modo posible de solucionarlo, haciendo partícipes de las opiniones subjetivas, de una manera anónima, a los otros miembros, con el fin de obtener el mayor grado de consenso para solucionar dicho problema, estableciendo prioridades. (Hernández López, 1996, en Fernández-Ballesteros). Se utiliza para el análisis de problemas diversos relacionados con la planificación y toma de decisiones, muy utilizada en el mundo empresarial para analizar tendencias y anticiparse a problemas futuros, ofreciendo soluciones y planificando el modo de conseguirlas.

Los resultados aportados sobre las necesidades en torno a una determinada cuestión, y la soluciones posibles aportaciones, guiarán la elaboración de planes y programas específicos dirigidos a dar respuesta a dichas necesidades. Así, por ejemplo, si se trata de recoger las opiniones y valoraciones de expertos para tratar de abordar y solucionar un problema como puede ser la integración escolar de alumnos procedentes de diferentes países, las valoraciones sobre un cuestionario guía, y las opiniones sobre cuestiones abiertas y otras aportaciones dadas por un grupo suficientemente amplio y procedente de diversos campos, como puede ser el formado por: maestros, pedagogos, psicólogos, sociólogos, políticos, educadores sociales, médicos, padres, etc. Como puede deducirse por el ejemplo, los expertos lo son en el sentido de que deben dar su opinión sobre la cuestión a estudio, bien por su formación sobre el tema, bien por su relación con él, pudiendo ser considerados como informadores claves.

Desde la perspectiva de la búsqueda del consenso, la técnica delphi, o Delfos, es considerada como una técnica de recogida de información cualitativa, aunque utiliza el instrumento de recogida de información

del cuestionario con preguntas formuladas de modo abierto, pero también cerrado, similar al utilizado en las escalas tipo Likert; para el análisis de las cuestiones abiertas utiliza el análisis de contenido y para las cerradas, los análisis estadísticos. Por ello puede ser considerada más apropiadamente una técnica mixta, desde el punto de vista del enfoque de investigación.

11.4.1. El proceso de recogida de información y datos mediante el Delphi

La técnica delphi, enmarcada en un proceso de investigación, debe comenzar, como siempre, con la formulación precisa del problema a estudiar, y una cuidada planificación de los pasos a realizar para resolverlo. El muestreo de expertos es así mismo, esencial, y los criterios para su selección deben ser especificados y argumentados. El modo de dirigirnos a ellos, el tipo de instrumento y los procedimientos para el análisis de la información y devolución de la información, son los pasos siguientes, y también deben estar previstos en la fase de planificación del trabajo. (Padilla Carmona, Moreno Sánchez y Vélez González, 2002).

Una vez concretado el tema de estudio, y profundizado en él a través de la revisión del estado de la cuestión, lo que conviene hacer a continuación es pensar en los expertos que por su posible formación, experiencia e implicación en el tema de estudio, pueden aportar su opinión, para estudiar con la mayor profundidad el problema, y los posibles modos de abordar la búsqueda de soluciones. El grupo de expertos debe ser, en un principio, lo suficiente amplio, para garantizar la representatividad de sus opiniones, y que sea número suficiente, teniendo en cuenta la pérdida de informantes que a lo largo del proceso de consulta pueda producirse.

El grupo de expertos, lo es, desde diversas perspectivas. Si se pretende que el grupo de expertos tenga cierta homogeneidad en su especialización, puede ser suficiente entre 10 y 15 miembros. Y si lo que se requiere es diversidad de expertos, por que así lo aconseja las características del problema a estudiar, el número de expertos a consultar será mayor, entre 30 y 50, procurando que de cada especialidad haya tres o más, para enriquecer las aportaciones, y prever, asimismo, la pérdida de informantes durante el proceso.

Después deberemos dirigirnos a cada miembro seleccionado, indicándoles los motivos para pedirle su colaboración, y explicándoles con claridad y precisión nuestro proyecto de investigación, las características

de la técnica delphi, y lo que esperamos de él, garantizándole la discrecionalidad en el uso de sus aportaciones, opiniones y valoraciones, así como el tiempo que se prevé que dure el proceso de consulta. El modo de dirigirnos a los informantes puede ser a través de carta, teléfono si les conocemos, correo electrónico, entrevista, o cualquier otro. Durante todo el proceso debemos cuidar la presentación de nuestros objetivos de la manera más motivadora posible para conseguir y mantener la colaboración de los informantes o expertos.

El instrumento para la recogida de información y datos que utiliza es el cuestionario individual, normalmente enviado por correo, y que tras el análisis de las respuestas emitidas por el grupo de expertos, se les reenvían los resultados grupales, junto con nuevas cuestiones, con el objetivo de alcanzar una mayor profundidad en las tentativas de solución del problema, a la vez que el mayor grado de consenso posible. El reenvío de las opiniones y soluciones aportadas por los expertos, puede hacerse hasta dos o más veces. Los expertos consultados no tienen por qué conocerse entre sí. De hecho el anonimato, o mejor la confidencialidad, en la consulta debe garantizarse. Sin embargo, aunque no hay interacción personal entre los miembros del grupo, si es posible la influencia mutua dado que se comparte la información de las aportaciones del grupo.

Una vez aceptada su colaboración se le enviará el primer cuestionario, con la información de los plazos para su devolución. El primer cuestionario se recomienda que sea en un formato de preguntas lo suficientemente abierto para que se pueda expresar el mayor número de ideas posibles, desde su propia perspectiva, el experto. Una vez devueltos los cuestionarios contestados se realizará el análisis de la información, buscando el consenso en las aportaciones así como las ideas únicas y originales. El análisis de las preguntas abiertas puede utilizarse siguiendo el análisis cualitativo de datos textuales, codificando la información y buscando las categorías de significado común a las aportaciones, y las ideas originales. Basándose en las categorías Con ello se elaborará un segundo cuestionario, con un mayor número de preguntas cerradas, formadas a partir de los ítems sugeridos por las categorías del análisis del primer cuestionario, que les será enviado a los expertos para recoger su valoración.

El segundo cuestionario estará formado por ítems redactados de un modo preciso, para que el experto emita sus respuestas mostrando su grado de acuerdo o desacuerdo, mediante preguntas tipo Likert, u otras en las que ordene de acuerdo a su importancia en la incidencia de la cuestión de estudio, o de cualquier otro modo, dejando también posibilidades para que amplíe las cuestiones y aporte sugerencias.

El análisis de los datos del segundo cuestionario se realizará utilizando las técnicas adecuadas al tipo de preguntas, y el posible nivel de medida que alcance. Así podrán utilizarse análisis estadísticos basados en el nivel de medida ordinal, o de intervalo o cuasi intervalo para ofrecer puntuaciones de tendencia central, de dispersión, o incluso contrastes de grupos, junto con representaciones gráficas, en la búsqueda de acuerdos. Con los resultados, en los cuales se muestren los acuerdos logrados en las propuestas, junto con las propuestas divergentes con la petición de que se argumente dicha divergencia, y se enviará como tercer cuestionario.

Si se ha conseguido consenso total en las propuestas entre los expertos, probablemente el investigador considere que no se necesiten más consultas, pero si el investigador, o equipo de investigadores, lo considerara necesario, podría realizar un cuarto cuestionario mostrando las cuestiones que han alcanzado mayor grado de acuerdo tras el análisis de la información del cuestionario tercero, así como las ideas distintas aportadas por alguno de los expertos, para someterlas a una cuarta consideración.

Wilkin (1984), argumentando que con la técnica delphi se pretende la búsqueda de consenso, considera que la medida más adecuada indicadora de consenso sería la moda, como la expresión de la opinión que más se repite, más que la media aritmética. También aconseja la utilización de tablas de porcentajes. Se informará a los expertos de las conclusiones del estudio.

Los resultados aportados por los expertos, finalmente, serán publicados, y se utilizarán para la elaboración de las medidas de mejora, basadas en muchos casos, en planes concretados en programas de intervención, y seguimiento de los mismos, para tratar de solucionar el problema que originó el estudio, y las necesidades detectadas.

Entre las ventajas de la técnica delphi, podrían destacarse la posibilidad de consultar a diversos expertos dispersos a lo largo y ancho de la geografía sin necesidad de reunirlos, la posibilidad de que cada uno de ellos manifiesten sus puntos de vista sin influencia de los otros, y la oportunidad de tras la primera consulta abierta y libre, tener la información de otros expertos para considerarla, contrastarla con la suya propia, y hacer síntesis enriquecedoras que favorezcan la amplitud y profundidad para atisbar los problemas, anticiparse a las necesidades y promover planes y programas de intervención para tratar de resolverlos.

11.5. Los grupos de discusión

La técnica de los grupos de discusión se utiliza para la recogida de información de grupos de sujetos, creado de modo artificial, con la única finalidad de recoger información sobre un tema específico. Los miembros del grupo son reunidos durante un periodo de tiempo limitado, con el fin de debatir sobre un tema específico propuesto por el investigador. Los miembros del grupo deben ser homogéneos con el fin de posibilitar una discusión libre entre ellos, recomendándose, en algunos casos, que los miembros del grupo no se conozcan, para evitar prejuicios en su participación. Los datos de análisis serán los aportados en la discusión, por lo tanto, fundamentalmente textuales. Una vez profundizado, a lo largo de las reuniones que fueran necesarias, en el tema en cuestión, y obtenidos acuerdos y conclusiones, el grupo ha terminado con la misión para la que fue formado, y se disolverá,

Como la técnica delphi, los grupos de discusión suelen utilizarse para profundizar en un tema, para detectar las posibles necesidades sobre la cuestión de modo que se aporten posibles soluciones para elaborar planes, programas y acciones concretas de intervención. También puede utilizarse para evaluar programas ya en acción, y valorar sus resultados, efectos, y necesidades de mejora.

Las técnicas que venimos comentando en este tema, son perfectamente complementarias entre sí. De modo que cuando se está investigando una cuestión, puede utilizarse la observación participante, la consulta a un grupo de expertos mediante la técnica delphi, la entrevista en profundidad a informantes clave, conocedores del tema, o la formación de un grupo de discusión formado por especialistas en el tema o cuestión sobre la que se investiga.

Las primeras fases del trabajo en el grupo de discusión suelen ser fundamentales para explorar de un modo amplio la problemática que se estudia, y a partir de ello ir focalizando, concretando, consensuando, los aspectos claves del tema, que guiarán la formulación de posibles hipótesis a comprobar, la realización de instrumentos de recogida de información como el cuestionario, guiones de entrevista, pautas de observación, etc. También se utilizan los grupos de discusión para comentar los resultados obtenidos de una aplicación piloto de un programa, para valorar los resultados obtenidos tras la aplicaciones de instrumentos de recogida de información, para comentar el progreso en alguna investigación.

Los grupos de discusión se ha constituido como una técnica esencial a utilizar en los grupos de investigación, en muchos casos multidiscipli-

nares, e incluso procedentes de distintos contextos geográficos. Con ello se trata de enriquecer las perspectivas de grupos formados por investigadores procedentes de la misma disciplina, y de los mismos centros de trabajo. Los encuentros se programan a lo largo del periodo que se estima dure la investigación, y los temas de discusión, y las tareas se fijan de antemano, revisándose al final de cada reunión, de acuerdo con los progresos de la investigación.

11.5.1. Aspectos a considerar para la formación y funcionamiento de los grupos de discusión

Conviene tener en cuenta alguna de estas indicaciones para la formación de los grupos de discusión:

- El tamaño del grupo y su composición.
- El número de grupos.
- Las características de los sujetos que deben formar parte de los grupos, contactar con ellos y conseguir su compromiso de colaboración y participación.
- Definir el papel de cada sujeto en su colaboración en la investigación y conseguir su compromiso.
- La duración del proceso de investigación.
- El número de reuniones a mantener, el lugar y su duración.

Dado que se trata de recoger información amplia y especializada de los diferentes miembros que formen el grupo, y a la vez se busca homogeneidad, acuerdos, habrá que buscar tantos grupos como aspectos se consideren de interés para abordar el tema en profundidad y diversidad. Las reuniones en los grupos homogéneos se mantendrán hasta que se llegue al consenso, después algún representante de los grupos homogéneos se reunirán en el grupo multidisciplinar para avanzar desde cada perspectiva que integra la investigación, en la misma. El tipo de muestreo es, por tanto, deliberado, o intencional, según el propósito de la investigación. El tamaño de los grupos suele aconsejarse entre 6 y 12. Las características personales y profesionales de los miembros de los grupos son de gran importancia, por que de ello depende el éxito en la investigación. Deben ser personas preparadas en el tema de estudio, abiertas al diálogo, responsables con sus compromisos. Estos compromisos son más fáciles de alcanzar, dado que la colaboración comporta un esfuerzo en tiempo y trabajo, si los miembros reciben algún tipo de incentivos, que cuando se

trata de investigación subvencionadas, suelen ser incentivos económicos y de autoría en los hallazgos.

La duración de las reuniones suele ser de una hora a dos horas, como máximo, y debe estar fijada tanto la hora de inicio como de finalización. El responsable de la investigación suele ser el moderador de la reunión, si bien se puede asignar un coordinador para cada uno de los diferentes grupos, que puede hacer a la vez el papel de relator, que va tomando nota de los acuerdos, o estas notas pueden ir realizándose por otro miembro del grupo, o dependiendo de las características del mismo, ir consensuando y concretando los acuerdos y las propuestas de trabajo pendientes a realizar para ir avanzando en la investigación. También se puede realizar grabación de las discusiones por medios técnicos como son cámaras de video o magnetófono. Es conveniente mantener contactos durante los procesos intermedios a las reuniones de trabajo, pudiendo realizarse mediante acuerdos de colaboración y comunicación a través del correo electrónico o cualquier otro medio.

Al final, los representantes de cada uno de los grupos serán los responsables de elaborar los informes de progresos y avance, y contribuir a la redacción del informe final de la investigación. El informe deberá reflejar todo el proceso de la investigación, el punto de partida, los objetivos, las hipótesis, el trabajo de campo realizado en los períodos intermedios cuando se trata de un grupo de investigación, los instrumentos para la recogida de datos, el análisis de la información la y su incorporación en el informe general del trabajo de investigación.

Como puede verse se ha expuesto la técnica de los grupos de discusión para su utilización con diferentes fines. Para profundizar en un determinado tema, o como estrategia para la puesta en común procesual entre los miembros de un grupo de investigación.

11.6. Ejercicios de autoevaluación

1. Defina la observación sistemática recogiendo sus notas características.
2. Realice una observación participante en el contexto en el que Vd. trabaje, utilizando, y completando, el modelo de registro de información que viene en el apartado 11.1.3, referido al proceso de recogida de datos en la observación participante.
3. Características y finalidad del análisis de necesidades.

4. Realice una síntesis sobre las características de la técnica delphi, y reflexione sobre las posibilidades de su aplicación en los ámbitos de trabajo del educador social.
5. Realice una síntesis sobre las características de los grupos de discusión, y valore su utilidad en los ámbitos de trabajo del educador social.

11.7. Referencias bibliográficas

- ANDER-EGG, E. (1987). *Técnicas de investigación social*. México: El Ateneo.
- BALLESTER, I. y OLIVER, J. L. (2000). «Paneles delphi: metodología de investigación aplicada con educadores». En AIDIPE (Comp.) *Nuevas realidades educativas, nuevas necesidades metodológicas*. Málaga: CEDMA, 374-380.
- CANALES, M. y PEINADO, A. (1999). «Grupos de discusión». En J. M. Delgado, y J. Gutierrez, (Eds.) *Métodos y técnicas cualitativas de investigación en ciencias sociales*. Madrid: Síntesis.
- CAPLOW, T. (1972). *La investigación sociológica*. Barcelona: Laia, Ediciones de Bolsillo.
- COOK, T. D. y REICHARDT, CH. S. (1986). *Métodos cualitativos y cuantitativos en investigación evaluativa*. Ediciones Morata, S. A. Madrid,
- ETXEBERRÍA, J., GARCÍA, E., GIL, J. y RODRÍGUEZ, G. (1995). *Análisis de datos y textos*. Madrid: RA-MA.
- FILSTEAS, W. J. (1970). *Qualitative Methodology. Firsthand Involvement with the Social World*. Chicago: Rand McNally.
- GIL FLOREZ, J. (1994). *Análisis de datos cualitativos. Aplicaciones a la investigación educativa*. Barcelona. PPU.
- GOETZ, J. P. y M. D. LeCOMPTE (1988). *Etnografía y diseño cualitativo en investigación educativa*. Madrid: Morata, S. A.
- KAUFMAN, R. A. (1972). *Educational system planning*. Englewood Cliff, N. J.: Prentice Hall.
- HERNÁNDEZ LÓPEZ, J. M. (1996). «Procedimientos de recogida de información en evaluación de programas». En R. Fernández-Ballesteros (Ed.). *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*. Madrid: Síntesis Psicológica.
- KRIPPENDORFF, K. (1990). *Metodología de análisis de contenido*. Barcelona, Buenos Aires, Paidós.
- LÓPEZ-BARAJAS ZAYAS, E. (Coor.) (1998): *La observación participante*. Madrid: UNED. Actas y Congresos.

- MALINOWSKI, B. (1922). *Argonauts of the western Pacific*. (Los Argonautas del Pacífico occidental). London: Routledge & Kegan Paul.
- PADILLA CARMONA, M. T., MORENO SÁNCHEZ, E. y VELÉZ GONZÁLEZ, E. (2002). «La técnica delphi en la evaluación de necesidades: una aplicación al tratamiento del género en los centros escolares». *Bordón. Revista de Pedagogía*. 54, 83-94.
- PARK, R. y BURGESS, E. (Eds.) (1921). *Introduction to the science of sociology*. Chicago: University of Chicago Press.
- PATTON, M. Q. (1990). *Qualitative Evaluation and Research Methods*. Newbury Park, CA.; Sage.
- PÉREZ-CAMPANERO, M.^a P. (1991). *Cómo detectar las necesidades de intervención socioeducativa*. Madrid: Narcea.
- RUIZ OLABUÉNAGA, J. I. (1999). *Metodología de la investigación cualitativa*. Bilbao: Universidad de Deusto.
- TAYLOR, S. J. y BOGDAN, R. (1998). *Introducción a los métodos cualitativos de investigación*. Barcelona: Paidós.
- WITKIN, B. R. y ALTSCHULD, J. W. (1995). *Planning and conducting need assessments. A practical guide*. Thousand Oaks, CA: Sage.

Tercera Unidad Didáctica

ESTADÍSTICA DESCRIPTIVA E INFERENCIAL
PARA EL ANÁLISIS DE LA INFORMACIÓN.
LOS DISEÑOS DE INVESTIGACIÓN

TEMA 12

Distribuciones de variables y representaciones gráficas

ESQUEMA

12.1. Distribuciones de variables

12.1.1. Distribuciones numéricas y reducción de datos

12.1.2. La distribución binomial

12.1.3. La distribución continua

12.1.4. La distribución normal de probabilidad o curva normal

12.2. Representaciones gráficas

12.2.1. Diagrama de barras

12.2.2. Histograma

12.2.3. Diagrama circular de sectores

12.2.4. Gráficas de frecuencias

12.3. Ejercicios de autoevaluación

12.4. Solución a los ejercicios de autoevaluación

12.5. Referencias bibliográficas

OBJETIVOS

- ✓ Conocer y comprender el concepto de distribución y sus tipos.
- ✓ Comprender el modo de reducción de los datos como primer acercamiento a su análisis.
- ✓ Conocer y comprender las diferentes representaciones gráficas de las distribuciones.
- ✓ Comprender y realizar las diferentes representaciones gráficas de las distribuciones.

12.1. Distribuciones de variables

12.1.1. Distribuciones numéricas y reducción de datos

Las puntuaciones obtenidas mediante los instrumentos de medida para la recogida de información en una determinada variable se denominan «puntuaciones directas» y para un primer análisis de dichas puntuaciones se suelen presentar en forma de distribuciones mediante tablas.

Veamos el siguiente ejemplo:

Se trata de representar las puntuaciones obtenidas por un grupo de 17 alumnos en una prueba de Lengua Española. La variable es los conocimientos de un grupo en Lengua Española, medida mediante una prueba objetiva precisa de rendimiento, que podemos considerar que alcanza el nivel de medida de intervalo. Las puntuaciones obtenidas por los sujetos en la prueba, en una escala de 0 a 10 puntos, se muestran a continuación:

3, 3, 6, 4, 4, 6, 6, 5, 5, 5, 6, 8, 8, 9, 7, 7

Lo primero que hay que hacer es situar las puntuaciones ordenadas de menor a mayor en una tabla, como sigue:

TABLA 12.1.

Calificaciones	
Sujetos	Puntuaciones
1	3
2	3
3	4
4	4
5	5
6	5
7	5
8	6
9	6
10	6
11	6
12	7
13	7
14	7
15	8
16	8
17	9

El análisis de datos estadísticos a partir de la distribución de puntuaciones directas, comienza con la reducción de los datos, pudiendo hacerse agrupando los datos por frecuencias, por porcentajes, por proporciones, que nos daría los datos de la siguiente tabla:

TABLA 12.2.

Calificaciones en Lengua Española en frecuencias, proporciones y porcentajes						
X_i	f_i	f_a	p_i	p_a	P_i	P_a
3	2	2	0,117	0,117	11,7	11,7
4	2	4	0,117	0,235	11,7	23,5
5	3	7	0,176	0,411	17,6	41,1
6	4	11	0,235	0,647	23	64,7
7	3	14	0,176	0,823	17	82,3
8	2	16	0,117	0,941	11,7	94,1
9	1	17	0,058	1	5,8	100
	$n = 17$		1,00		100	

Siendo:

X_i : las puntuaciones en la variable

f_i : frecuencias, o número de veces que se repite cada uno de los valores de la variable

n : número total de sujetos

f_a : frecuencias acumuladas en valores absolutos: las frecuencias anteriores se suman a la siguiente

p_i : frecuencia relativa en proporciones; es el cociente entre la frecuencia absoluta de cada clase y el número total de sujetos

p_a : frecuencia relativa acumulada en proporciones

P_i : frecuencia relativa en porcentajes: es el cociente entre la frecuencia absoluta de cada clase y el número total de casos multiplicado por 100

P_a : frecuencia relativa en porcentajes acumulados.

Las frecuencias (f_i) representan el número de veces que se repite cada uno de los valores de la variable. Las frecuencias acumuladas en valores absolutos se representan por f_a , y suponen la suma acumulativa de las frecuencias anteriores a la siguiente.

La frecuencia relativa en proporciones, representada por p_i , es el cociente entre la frecuencia absoluta de cada clase y el número total de casos, en este caso sujetos:

$$p_i = \frac{f_i}{n}$$

La frecuencia relativa acumulada en proporciones (p_a) es el resultado de la suma acumulativa de la proporción anterior a la siguiente. La frecuencia relativa en porcentajes (P_i) viene dada por el cociente entre la frecuencia absoluta de cada clase y el número total de casos multiplicado por 100:

$$P_i = \frac{f_i}{n} \times 100$$

La frecuencia relativa en porcentajes acumulados (P_a) es el resultado de la suma acumulativa de la proporción anterior con la siguiente.

Si los datos los analizáramos mediante un programa informático, por ejemplo el SPSS, una vez introducidas las calificaciones en el editor de datos, en vista de variables, pedimos analizar descriptivos, seleccionando frecuencias, y una vez ejecutada la petición nos aporta la siguiente tabla que coincide con los valores hallados con calculadora, si bien con el programa SPSS ahorramos tiempo y esfuerzo y ganamos en exactitud:

TABLA 12.3.

Calificaciones en Lengua Española en frecuencias y porcentajes					
Calificaciones	Frecuencia	Frecuencia acumulada	Porcentaje	Porcentaje válido	Porcentaje acumulado
3.00	2	2	11.8	11.8	11.8
4.00	2	4	11.8	11.8	23.5
5.00	3	7	111.6	111.6	41.2
6.00	4	11	23.5	23.5	64.7
11.00	3	14	111.6	111.6	82.4
8.00	2	16	11.8	11.8	94.1
9.00	1	17	5.9	5.9	100.0
Total	17		100.0	100.0	

El paso siguiente en el análisis de los datos estadísticos, para su mejor manejo y comprensión, consiste en representarlos mediante gráficas.

12.1.2. La distribución binomial

El ejemplo clásico de variable discreta binomial es la distribución formada por la cantidad de veces que sale cara cuando se lanza una moneda n veces. Los n intentos son independientes, resultando cada uno cara o cruz, éxito o fracaso, con probabilidades π y $(n - \pi)$ poblacionales y p y $(n - p)$ muestrales. El número de éxitos, en este caso, se llama variable aleatoria binomial y se representa por la letra s .

$$p(s) = \binom{n}{s} \pi^s (1 - \pi)^{n-s}$$

Veamos algunos ejemplos de variables binomiales:

TABLA 12.4. Ejemplos de variables binomiales. En Wonnacott y Wonnacott 1979: 84

Intento	Éxito	Fracaso	π	n	s
Lanzamiento de una moneda	Cara	Cruz	1/2	n lanzamientos	Número de veces que sale cara
Nacimiento terneras	Tenera	Tenero	Casi 1/2	Número de terneros/as en el rebaño	Número de terneras en el rebaño
Lanzamiento de dos dados	7 puntos	Cualquier otra cantidad	6/36	n lanzamientos	Número de siete
Extracción de un votante en una encuesta	Demócrata	Republicano	Proporción de demócratas en la población	Tamaño de la muestra	Número de demócratas en la muestra

Las representaciones gráficas de las variables discretas se hacen mediante el diagrama de barras.

12.1.3. La distribución continua

Cuando las variables aleatorias que medimos son continuas, o básicamente continuas, su representación gráfica puede hacerse mediante histogramas, y cuando hay una gran densidad de valores de la variable medida tienen una forma de curva bien delineada. Cuando la distribución es muy grande la curva adopta la forma de campana, llamada curva normal o distribución normal, y también distribución de Gauss, o campana de Gauss, en honor a quien la definió.

12.1.4. La distribución normal de probabilidad o curva normal

La distribución normal es la distribución de probabilidad más útil e importante en estadística. Constituye la base de un gran grupo de pruebas estadísticas para la inferencia. La distribución normal estándar se denomina z . La distribución normal es simétrica y tiene forma de campana. El área total debajo de la curva representa una probabilidad de 1,0. La escala horizontal debajo de la distribución representa los valores medidos en una variable. El punto medio de esta escala es la media, o el promedio de todos los valores y su valor en puntuaciones z es igual a 0, siendo su desviación típica igual a 1.

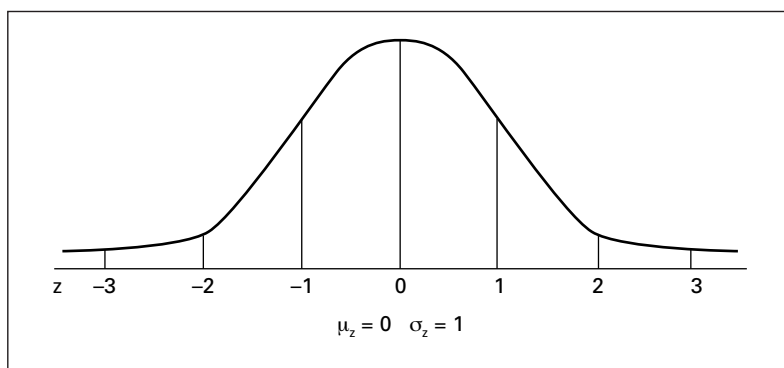


FIGURA 12.1.

La probabilidad equivalente al área comprendida entre la curva normal y el eje de abscisas por debajo de ella es igual a 1. Y la parte de área comprendida entre unos determinados valores de z y la curva que la cierra al trazar unas líneas perpendiculares desde los valores a la curva, corresponde a una determinada probabilidad, bien calculada y que se puede consultar en las tablas de probabilidad bajo la curva normal de la distribución de la variable z . Cualquier variable aleatoria puede transformarse a la forma estándar de la distribución de la curva normal. A esta transformación de la variable aleatoria se le denomina tipificación de la variable, y la nueva transformación de la variable sufre una transformación lineal de la inicial mediante la fórmula:

$$z = \frac{X_i - \mu}{\sigma}$$

Y tiene como propiedades fundamentales que el valor de su media es 0 y su desviación típica es 1.

12.2. Representaciones gráficas

Las representaciones gráficas es una técnica diferente a la de las tablas de distribución de datos para reducir los datos, presentándolos de modo que nos facilita su análisis e interpretación visual. Las representaciones gráficas aportan información valiosa que en muchas ocasiones no afloran en los índices estadísticos, recomendándose que sean previas al estudio de los índices estadísticos para realizar una exploración inicial de la información aportada por los datos.

El tipo de gráfica que elijamos para la representación de los datos va a depender de las características de la variable objeto de estudio. La selección correcta del gráfico deberá tener en cuenta si la variable es continua o discreta, si ha venido medida por instrumentos que alcanzan el nivel de medida de intervalo, ordinal o nominal, y si las puntuaciones vienen dadas de modo directo o acumuladas. Los gráficos deberán ir acompañados de una explicación que haga mención al tipo de datos que utiliza y la fuente que permitirá consultar la información que representa.

Los gráficos estadísticos se fundamentan en el sistema de coordenadas cartesiano. Está formado por dos ejes perpendiculares que se cruzan en un punto en el que se sitúa el valor cero (0). El eje horizontal es llamado eje de abscisas (X) y en él se sitúan los valores de la variable. El eje vertical es llamado eje de ordenadas (Y) y en él se sitúan las frecuencias, o número de casos que se repite en un determinado valor de la variable, normalmente variables continuas.

12.2.1. Diagrama de barras

Son gráficas que representan cada valor mediante una barra proporcional a la frecuencia obtenida en la variable. Se utilizan para representar valores de variables discretas, no continuas, por lo que las barras deben estar separadas. La base de la barra adopta un ancho igual para toda las barras, elegido por el investigador, y la altura viene determinada por el valor o la frecuencia obtenida.

Utilizando los datos de las calificaciones obtenidas por los alumnos de 1.º de ESO en Lengua Española, con el programa SPSS, si pedimos un diagrama de barras, nos muestra la gráfica siguiente en la hoja de resultados:

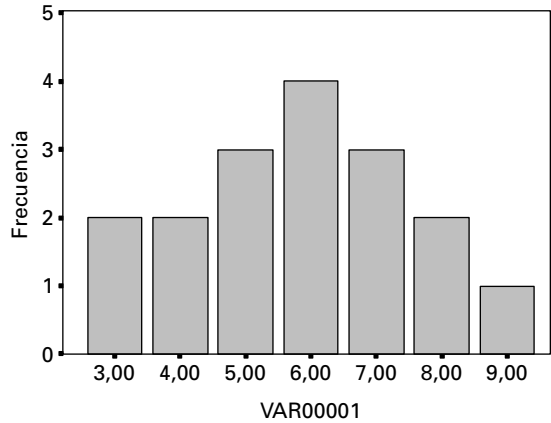


FIGURA 12.2. Diagrama de barras de las calificaciones obtenidas en Lengua Española por un grupo de alumnos de 1.º de ESO.

En otro ejemplo podemos ver representadas las nacionalidades de un grupo de niños hijos de inmigrantes asistentes a un campamento de verano organizado por la Comunidad de Madrid. En el eje de ordenadas aparecen las cuatro nacionalidades: búlgara, rumana, polaca y yugoslava, y en el eje de abscisas aparecen las frecuencias en porcentajes.

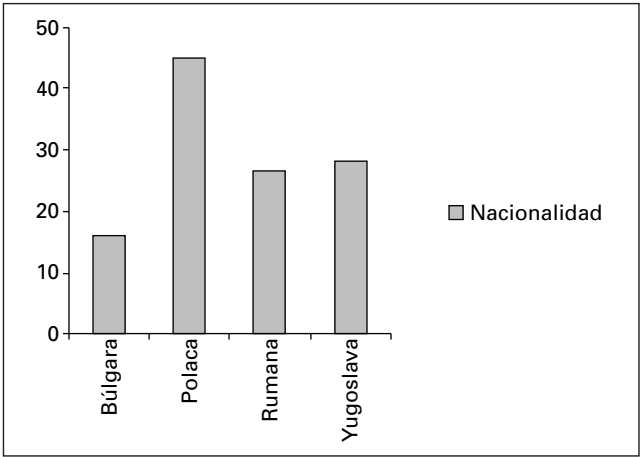


FIGURA 12.3. Nacionalidades de los hijos de inmigrantes asistentes a un campamento de verano en la Comunidad de Madrid.

12.2.2. Histograma

Se utiliza para representar variables continuas agrupadas en intervalos. Las barras se presentan unidas, dado que la variable es continua. La altura de la barra indica la frecuencia correspondiente al intervalo o valor de la variable X , y en la base se indica la marca de clase o valor medio del intervalo.

Para construir un histograma necesitamos, primero agrupar los datos en intervalos, para lo cual, como dijimos anteriormente es necesario:

1.º Conocer las características de la distribución, el número de sujetos y las puntuaciones obtenidas por cada uno de ellos.

2.º Determinar el rango de la distribución:

$$R = \text{Mayor valor} - \text{Menor valor}$$

3.º Determinar el número de intervalos y su amplitud siguiendo la siguiente pauta:

Número de datos de la distribución	Número de intervalos
Menos de 30	Ninguno o de 2 a 4
Entre 30 y 50	De 5 a 7
Entre 51 y 100	De 6 a 10
Entre 101 y 250	De 7 a 12
Más de 250	De 10 a 20

TABLA 12.5.

El número de intervalos resultante determina el número de barras del histograma. Y la amplitud del intervalo se calcula dividiendo el rango de la distribución entre el número de intervalos:

$$A = \frac{R}{K}$$

Siendo:

A: la amplitud del intervalo

R: el rango

K: el número de intervalos.

4.º Construir los intervalos: El valor más bajo de la distribución estaría en el primer intervalo, comenzando éste por el número entero

inferior más próximo al valor más bajo de la distribución, convirtiéndose en el primer límite inferior más bajo. El segundo límite inferior vendría determinado por la suma al primer límite inferior de la amplitud del intervalo, y sería el siguiente valor, sin incluirlo, del límite superior del primer intervalo, con el fin de que cada dato de la serie sea incluido en sólo un intervalo o clase.

- 5.º Construir una tabla de frecuencias basada en los valores de la distribución. En la primera columna se incluirían las clases, y en la segunda las frecuencias de los datos de la distribución para cada clase.

La tabla de frecuencias nos da una aproximación a una representación gráfica de los datos similar a un histograma y al polígono de frecuencias, permitiéndonos hacer una interpretación y valoración de las características de la distribución de los datos. Por ejemplo, si miramos la siguiente serie de datos referida a las edades de las personas que visitaron una exposición sobre los hallazgos prehistóricos encontrados en la localidad «X» el primer día de su apertura:

11- 12- 12- 14 -14- 15- 18- 17- 15- 18- 17-16-17-16-17 - 21- 22- 21- 24- 23- 21-22- 21- 24-23-21 - 22 - 21- 24 - 23 - 25- 25- 26- 27- 28-26- 25- 26- 29-28-26-28-25- 27-28-29- 28-27-26- 31- 32- 32- 34- 31- 31- 32- 32- 34-31-33-32- 32-34-31-32-32-32 -34-31-34-32 -32-32-34-31- 34 -35- 35- 36- 37- 38- 36- 35-37- 36-39-38 -36- 38 - 35 - 37-38-39- 38- 37- 36 -41- 42- 42- 44- 41-42- 44-43-41-42-42-43-44-41-43-42-42- 43-44- 43-41- 47- 48- 45- 46- 45- 46- 49-48- 46-48- 45-47- 48-49-48- 47- 52- 53- 52- 54- 51-54-52- 52-54-51-53- 55- 56-55-56- 59-58

Tenemos:

- Número total de personas o casos: 150
- Número de intervalos aconsejados: de 7 a 12. Nos decidimos por 10
- Rango o amplitud: $59 - 11 = 48$
- Amplitud del intervalo: $48 : 10 = 4,8$; redondeamos a 5
- Límite inferior del primer intervalo: 10

La tabla de frecuencias por intervalos nos da una aproximación de las características del grupo por edades:

TABLA 12.6.

Intervalos	Frecuencias	Total
10-14	/////	5
15-19	///// /////	10
20-24	///// ///// /////	15
25-29	///// ///// ///// /////	19
30-34	///// ///// ///// ///// ///// //	27
35-39	///// ///// ///// /////	20
40-44	///// ///// ///// ///// //	21
45-48	///// ///// ///// //	16
50-54	///// ///// //	11
55-59	///// //	6
Total		150

6.º Dibujar el histograma: Para ello hay que trazar dos ejes perpendiculares. En el vertical se marcarían las frecuencias en orden ascendente, y en el horizontal se marcarían las clases. Dibujar las barras de cada clase: la altura de cada barra viene determinada por el número de frecuencias para ese intervalo o clase.

Si introducimos los datos en el editor de datos del programa SPSS, y pedimos un histograma con la curva normal, nos proporciona la siguiente gráfica que trasladamos desde el SPSS al fichero de word con el que estamos trabajando este texto:

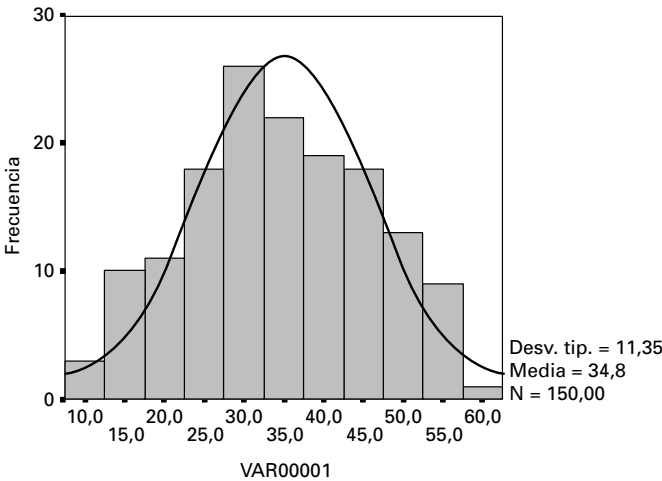


FIGURA 12.4. Gráfico por edades con curva normal.

Como podemos observar, en el margen derecho de la gráfica aparecen los estadísticos descriptivos más representativos como son la media de edad, 34,8 años, y su desviación típica, 11,35.

- 7.º Titular la gráfica.
- 8.º Identificar el patrón de variación de la distribución. Señalar los casos extremos y extraños. Elaborar una explicación lógica. Determinar la necesidad o no de ampliar el estudio, buscar nuevos datos para elaborar posibles conclusiones.

12.2.3. Diagrama circular de sectores

El diagrama de sectores se utiliza para representar variables discretas, aunque también puede utilizarse con variables continuas. La amplitud de los sectores del diagrama circular es proporcional a su frecuencia. El ángulo del círculo, 360°, representa la totalidad de los valores u observaciones con los que se trabaja, y el ángulo de cada sector indica la frecuencia relativa, en porcentajes, de cada una de las categorías integradas en el estudio. Si pedimos al programa que aparezcan los porcentajes en la gráfica, queda más completa la información que aporta. Con los datos de las calificaciones de Lengua Española, tenemos el siguiente gráfico de sectores:

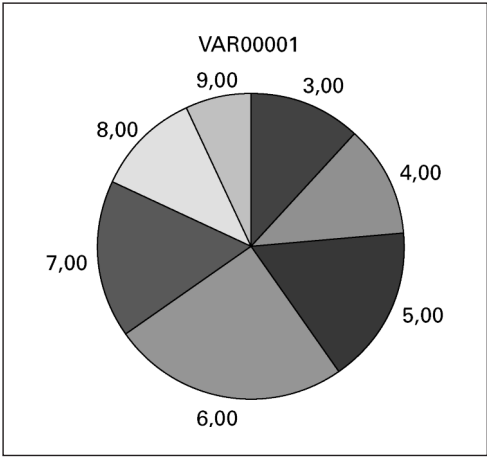


FIGURA 12.6. Gráfico de las calificaciones obtenidas por un grupo de alumnos de 2.º de ESO en Lengua Española.

12.2.4. Gráficas de frecuencias

Los diagramas de barras de variables continuas y los histogramas también pueden representarse mediante gráficas de frecuencias. En el eje horizontal se sitúan los valores de la variable y en el vertical las frecuencias que obtienen dichos valores.

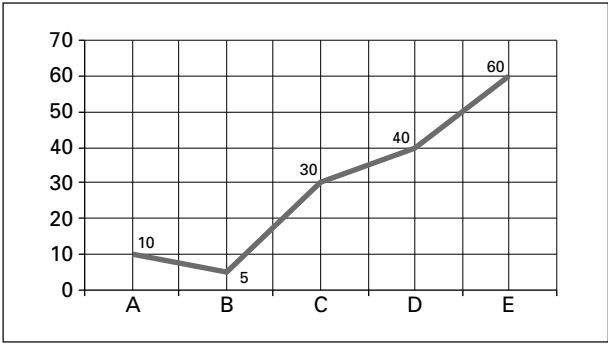


FIGURA 12.7.

A partir de un diagrama de barras se puede realizar un polígono de frecuencias, uniendo mediante líneas rectas el punto medio de cada intervalo. Cuando el diagrama de barras representa datos medidos en escala nominal carece de sentido representarlos mediante un polígono de frecuencias.

Veamos el siguiente ejemplo:

Se trata de representar, mediante un gráfico de sectores, las edades de las personas que han participado en unas jornadas de Educación Medioambiental a partir de la tabla de datos siguiente:

TABLA 12.6. Participantes en las Jornadas de Educación Medioambiental

Edades	Frecuencias	Porcentajes
15-24	10	6,66
25-34	30	20,00
35-44	50	33,33
45-54	40	26,66
55-64	20	13,33
Total	150	100,00

La amplitud de cada sector se halla mediante una regla de tres, lo que nos aporta los porcentajes de cada intervalo de edad. El diagrama de sectores proporciona una visión clara del conjunto de cada valor con respecto a la totalidad. Los sectores aparecen ordenados de menor a mayor porcentaje.

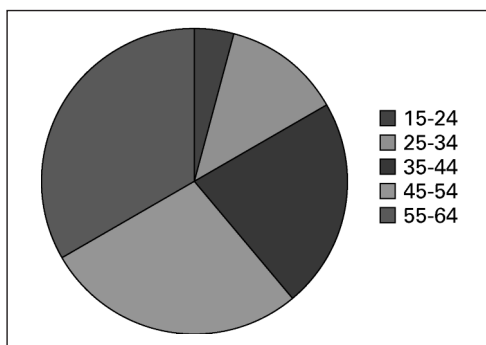


FIGURA 12.8. Gráfico de sectores por edades.

12.3. Ejercicios de autoevaluación

Problema 1

Las puntuaciones obtenidas por un grupo de 15 alumnos en un cuestionario de 10 ítems valorados de 1 a 5, sobre la actitud ante el uso de ordenadores para el aprendizaje, han sido las siguientes:

$$X_i = (46, 45, 38, 42, 39, 36, 44, 42, 43, 44, 47, 45, 44, 38, 44)$$

Represente las puntuaciones mediante un histograma.

Problema 2

Represente las puntuaciones del problema anterior mediante un diagrama por sectores de porcentajes.

12.4. Solución a los ejercicios de autoevaluación

Problema 1

Las puntuaciones obtenidas por un grupo de 15 alumnos en un cuestionario de 10 ítems valorados de 1 a 5, sobre la actitud ante el uso de ordenadores para el aprendizaje, han sido las siguientes:

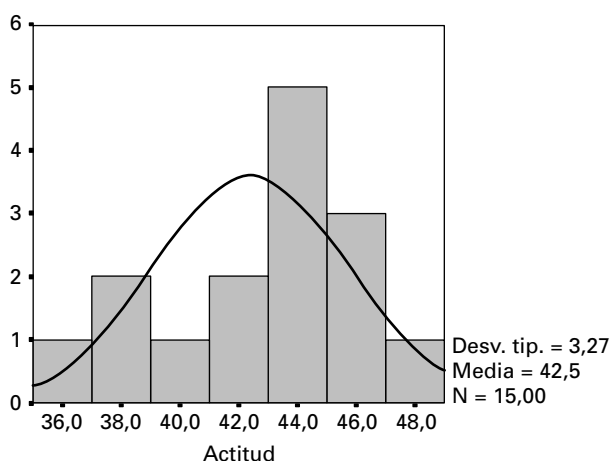
$$X_i = (46, 45, 38, 42, 39, 36, 44, 42, 43, 44, 47, 45, 44, 38, 44)$$

Represente las puntuaciones mediante un histograma y mediante un diagrama circular por sectores.

Primero ordenaremos los datos situándolos en una tabla por frecuencias.

X_i	f_i
36	1
38	2
39	1
42	2
43	1
44	4
45	2
46	1
47	1

Después trazaremos un eje de coordenadas situando en el eje de abscisas los valores obtenidos en la variable actitud, y en el de ordenadas las frecuencias, resultando el siguiente diagrama que hemos realizado mediante el programa SPSS, y que nos aporta los principales estadísticos descriptivos:



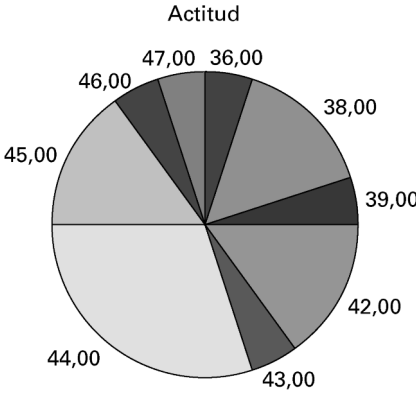
Problema 2

Represente las puntuaciones del problema anterior mediante un diagrama por sectores de porcentajes.

Para realizar el diagrama circular por sectores necesitamos hallar los porcentajes proporcionales a las frecuencias:

ACTITUD			
Puntuaciones	Frecuencia	Porcentaje	Porcentaje acumulado
36.00	1	6.7	6.7
38.00	2	13.3	20.0
39.00	1	6.7	26.7
42.00	2	13.3	40.0
43.00	1	6.7	46.7
44.00	4	26.7	73.3
45.00	2	13.3	86.7
46.00	1	6.7	93.3
411.00	1	6.7	100.0
Total	15	100.0	

Quedando la representación del diagrama de este modo:



12.5. Referencias bibliográficas

SPSS (2002). *Programa informático para el análisis de datos estadísticos*. Versión 11.

WONNACOTT, T. H. y WONNACOTT, R. J. (1997). *Introducción a la estadística*. México: Limusa.

TEMA 13

Características de las distribuciones de frecuencias: medidas de tendencia central, de posición y de dispersión. Puntuaciones individuales

ESQUEMA

- 13.1.** Medidas de tendencia central. La moda. La mediana. La media
- 13.2.** Medidas de posición. Los cuantiles
 - 13.2.1. Percentiles. Cuartiles. Deciles. La amplitud intercuartílica
- 13.3.** Medidas de variabilidad y de dispersión
 - 13.3.1. La amplitud
 - 13.3.2. La desviación media
 - 13.3.3. La varianza, la desviación típica y el coeficiente de variación
 - 13.3.4. Significado de la varianza y la desviación típica
- 13.4.** Puntuaciones individuales
 - 13.4.1. Las puntuaciones diferenciales
 - 13.4.2. Las puntuaciones típicas
 - 13.4.3. La curva normal
 - 13.4.4. Normalización de las distribuciones
 - 13.4.5. Asimetría y curtosis
- 13.5.** Ejercicios de autoevaluación
- 13.6.** Solución a los ejercicios de autoevaluación
- 13.7.** Referencias bibliográficas

OBJETIVOS

- ✓ Conocer y comprender los estadísticos de tendencia central.
- ✓ Conocer y comprender las medidas de posición.
- ✓ Conocer y comprender las medidas de variabilidad y dispersión.
- ✓ Conocer y comprender el significado de la varianza y la desviación típica.
- ✓ Conocer y comprender el significado de las puntuaciones individuales.

13.1. Medidas de tendencia central: la moda, la mediana y la media

La tendencia central se refiere al valor que se encuentra hacia el centro de la distribución, y suele venir representado por la moda, la mediana y la media de la distribución. Las medidas de tendencia central dan idea de los valores más representativos del conjunto de una serie de datos.

La moda. La moda (Mo) de una serie de datos se define como el valor más frecuente. Es el valor que más se repite en una serie de datos, el que más predomina. Y si los datos están agrupados en intervalos se hablará de clase o intervalo modal.

La mediana. De una serie de datos ordenados de menor a mayor, el valor del dato que ocupa el lugar intermedio se llama mediana, y deja por debajo y por encima de sí el 50% de los casos. La mediana (Me) es el valor medio. Si el número de datos es par la mediana sería la media aritmética de los dos valores intermedios. La mediana coincide con la media en una distribución simétrica.

El cálculo de la mediana requiere operar con datos medidos al menos en escala ordinal (nivel de medición ordinal). Para datos agrupados en intervalos, su formula es:

$$Me = L_1 + \left[\frac{\frac{n}{2} - \sum f_{a(t-1)}}{f_a} \right] \times a$$

Siendo:

L_1 = Límite exacto inferior del intervalo crítico,

n = número total de casos, o de observaciones,

$f_{a(t-1)}$ = frecuencia absoluta acumulada por debajo del intervalo crítico,

f_a = frecuencia absoluta del intervalo crítico,

a = amplitud del intervalo.

La media aritmética o promedio. La media aritmética, media o promedio es la medida de tendencia central más usual, y se calcula sumando todos los valores obtenidos por un grupo en una variable y dividiéndolos por el número total de valores. O la suma de todas las puntuaciones de una distribución, dividida por el total de puntuaciones. La utilización de la media exige operar con datos medidos en nivel de medición de intervalo, o de razón. Su fórmula es:

$$\bar{X} = \frac{\sum X_i}{n}$$

Siendo:

$\sum X_i$: la suma de las puntuaciones directas,

n : el número de casos (sujetos si se trata de un grupo de personas).

La media aritmética para datos agrupados por frecuencias adopta esta ecuación:

$$\bar{X} = \frac{\sum X_i f_i}{n}$$

Siendo:

$\sum X_i f_i$: la suma de las puntuaciones directas por su frecuencia,

n : el número de sujetos.

Veamos el siguiente ejemplo.

Las puntuaciones obtenidas por 12 alumnos en las materias de Lengua Española, Matemáticas e Inglés han sido las siguientes:

TABLA 13.1. Puntuaciones en las materias

Sujetos	Lengua	Matemáticas	Inglés
1	3	4	3
2	3	5	6
3	4	6	5
4	4	4	6
5	5	6	6
6	6	5	7
7	6	4	6
8	7	6	7
9	7	7	8
10	7	9	8
11	8	7	9
12	9	8	8

Para hallar la puntuación media que ha obtenido el grupo en éstas asignaturas utilizamos la fórmula de puntuaciones directas, necesitando obtener previamente los sumatorios en cada una de las asignaturas y dividirlos después entre el número de casos.

En Lengua, el sumatorio de las puntuaciones obtenidas por el total de alumnos es: $\sum X_i = 69$. Y su media aritmética es:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{69}{12} = 5,75$$

En Matemáticas: $\sum X_i = 71$

Y su media aritmética es = 5,916

En Inglés: $\sum X_i = 79$

Y su media aritmética es = 6,583

Si hacemos un cuadro resumen de las tres medias aritméticas, tenemos los siguientes datos:

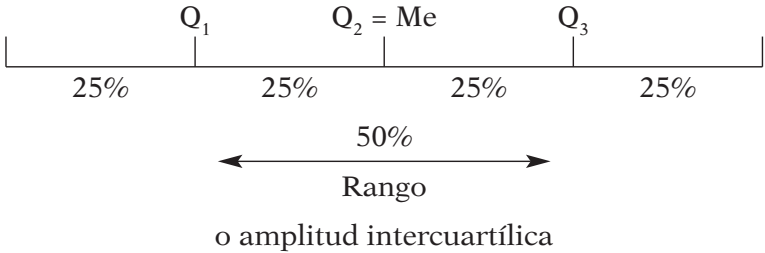
TABLA 13.2.

	Lengua Española	Matemáticas	Inglés
Medias	5,75	5,92	6,58

13.2. Medidas de posición. Cuantiles

Al estudiar las medidas de tendencia central hemos visto que la mediana representa el valor central de la serie, dejando por debajo de dicho valor el 50% de los casos de la serie de datos u observaciones de la distribución. La mediana, por lo tanto, se manifiesta como la posición central de la serie. A partir de esta valoración de la mediana, se pueden establecer distintas posiciones en la serie, con el fin de conocer la posición que ocupa una determinada puntuación dentro de la serie de datos.

Si la serie la dividimos entre cuatro a cada una de las partes la denominamos cuartil, si es entre diez, a cada parte la denominamos decil. El término genérico es el de cuantil.



13.2.1. Percentiles, cuartiles, deciles y amplitud intercuartílica

Los percentiles. Si una distribución la dividimos entre 100, cada una de los valores que toma se denomina centil o percentil. En una distribución de datos agrupados en intervalos, el percentil 'm' que ocuparía un determinado valor de la distribución se halla mediante la fórmula siguiente:

$$P_m = L_1 + \left[\frac{m\left(\frac{n}{100}\right) - f_{a(l-1)}}{f_a} \right] \times a$$

Los cuartiles. Para hallar los cuartiles se debe dividir el conjunto de valores de una variable en cuatro partes, cada una de las cuales debe contener el mismo número de valores. Primero se divide en dos partes, buscando la mediana para clasificar los valores por encima y por debajo de ella.

$$Q_m = L_1 + \left[\frac{m\left(\frac{n}{4}\right) - f_{a(l-1)}}{f_a} \right] \times a$$

Los deciles. Si la serie la dividimos entre diez, a cada parte la denominamos decil.

$$D_m = L_1 + \left[\frac{m\left(\frac{n}{10}\right) - f_{a(l-1)}}{f_a} \right] \times a$$

Siendo:

m = el cuantil que deseamos conocer; en este caso, el percentil,

L_1 = límite inferior del intervalo en el que se encuentra el cuantil,

n = número total de casos,

$f_{a(l-1)}$ = frecuencia acumulada del intervalo anterior al que se encuentra m ,

a = amplitud del intervalo.

La amplitud intercuartílica. La amplitud intercuartílica es la diferencia entre el cuartil inferior y el cuartil superior. Por definición, el recorrido intercuartílico no toma en consideración las puntuaciones por encima del tercer cuartil, ni las puntuaciones por debajo del primer cuartil. Sólo toma el 50 % de las puntuaciones centrales de la serie, no considerando los valores extremos. El recorrido intercuartílico, y la

amplitud semi-intercuartílica, que es la mitad, se hallan mediante las fórmulas siguientes:

$$\text{Amplitud intercuartílica} = (Q_3 - Q_1)$$

$$\text{Amplitud semi-intercuartílica} = \frac{(Q_3 - Q_1)}{2}$$

Veamos un ejemplo.

En un centro educativo se ha aplicado a los alumnos de los tres grupos de 6.º curso de Primaria una prueba objetiva de Matemáticas compuesta por 50 ítems. Los datos obtenidos se muestran en la tabla de datos. Se le pide:

1. Calcular el valor del cuartil 3, el percentil 30.
2. Calcular el valor del percentil 30.
3. Representar los datos de la tabla mediante una curva o polígono de frecuencias
4. Señalar la posición del cuartil 3 y el centil 30.

TABLA 13.3.

Calificaciones en Matemáticas		
Intervalos	f_i	f_a
5-9	2	2
10-14	5	7
15-19	11	18
20-24	14	32
25-29	25	57
30-34	16	73
35-39	12	85
40-44	3	88
45-49	1	89
total	89	

Resolución del problema:

1. Cálculo del valor del cuartil 3,

$$Q_m = L_i + \left[\frac{m(\frac{n}{4}) - f_a(I-1)_1}{f_i} \right] \times C; \quad \frac{m \times n}{4} = \frac{3 \times 89}{4} = 66,75;$$

$$Q_3 = 29,5 + \left[\frac{66,75 - 57}{16} \right] \times 5 = 29,5 + \frac{9,75}{16} \times 5 = 29,5 + 3,046 = 32,546$$

El valor del cuartil 3 es 32,55.

2. Cálculo del valor del percentil 30.

$$P_m = L_i + \left[\frac{m(\frac{n}{100}) - f_a(I-1)_1}{f_i} \right] \times C; \quad \frac{m \times n}{100} = \frac{30 \times 89}{100} = 26,67;$$

$$P_{30} = 19,5 + \left[\frac{26,67 - 18}{14} \right] \times 5 = 19,5 + 0,619 \times 5 = 19,5 + 3,096 = 22,596$$

3. Representación de los datos de la tabla anterior mediante un polígono de frecuencias.
4. Situación en el gráfico del cuartil 3 y del percentil 30.

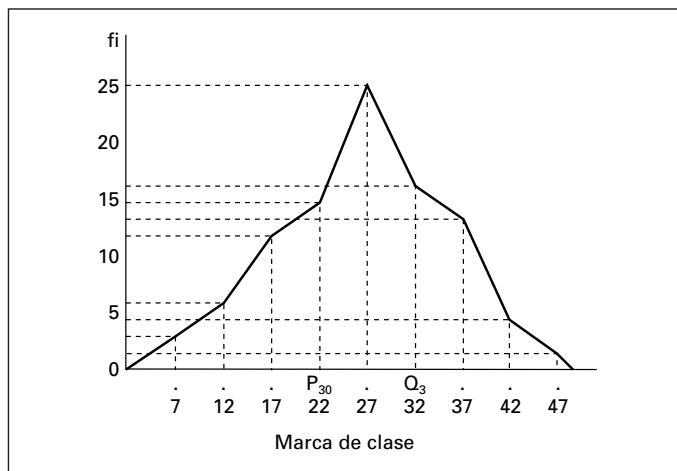


FIGURA 13.1.

13.3. Medidas de variabilidad o de dispersión

Las medidas o índices de variabilidad o de dispersión dan idea del alejamiento de los datos grupales respecto de las medidas de centralización. Estos índices aportan una información esencial sobre el grupo en relación con las medidas de centralización. Los índices de variabilidad informan sobre el grado de concentración o dispersión de los valores de la distribución en torno al valor central. Si el índice está cercano al promedio se dice que la distribución es homogénea. Si está lejano se dice que es heterogénea.

13.3.1. La amplitud

La amplitud, o amplitud total de un conjunto de datos, es la diferencia entre el valor más pequeño y el más grande de los datos. Es la medida de dispersión más simple. La amplitud total representa el recorrido de la variable, y se halla restando a la puntuación final, o mayor, la puntuación inicial, o menor, y sumándole una unidad.

$$A_T = X_f - X_i + 1$$

Siendo:

X_f = la puntuación máxima o final de la serie,

X_i = la puntuación mínima o inicial de la serie.

13.3.2. La desviación media

Se basa en la desviación de cada uno de los valores que toma la variable con relación a la media, pudiéndose tomar en valores absolutos. La desviación media es la media de la suma de las desviaciones de cada valor con respecto a la media, en valores absolutos.

$$DM = \frac{|X_1 - \bar{X}| + |X_2 - \bar{X}| + \dots + |X_n - \bar{X}|}{n} = \frac{\sum |X_i - \bar{X}|}{n}$$

La desviación media con respecto a la media es igual a la suma de cada una de las puntuaciones directas menos la media, dividida entre el número total de sujetos o puntuaciones de la serie.

13.3.3. La varianza, la desviación típica y el coeficiente de variación

En la desviación media, un elemento que está próximo a la media contribuye poco a la desviación media. La contribución de cada elemento del conjunto a la desviación media es directamente proporcional a su valor. Los valores alejados de la media tienen una influencia desproporcionada en la medida de dispersión. Con la desviación media queda desdibujada esa mayor dispersión de algunos valores. Para resaltar esa desviación de las desviaciones importantes sobre las pequeñas se elevan al cuadrado todas las desviaciones.

13.3.3.1. La varianza

Sumando el cuadrado de las diferencias de cada puntuación con respecto a la media, o puntuaciones diferenciales, y obteniendo su media, se obtiene la varianza

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

Para su cálculo es más eficiente esta otra fórmula:

$$s^2 = \frac{\sum X_i^2}{n} - \bar{X}^2$$

Con puntuaciones agrupadas en intervalos la fórmula que se utiliza es:

$$s^2 = \frac{\sum x_i^2 n_i}{n}$$

Siendo $x_i : (X_i - \bar{X})$ la puntuación diferencial con respecto a la media.

La varianza es una medida estadística muy utilizada para conocer las diferencias existentes entre dos o más muestras. Al tratarse de un valor obtenido en función de las distancias cuadráticas de cada puntuación de la distribución con respecto a la media aritmética, adopta siempre valores positivos.

Retomando el problema planteado en el tema anterior que decía: Las puntuaciones obtenidas por un grupo de 15 alumnos en un cuestionario

de 10 ítems valorados de 1 a 5, sobre la actitud ante el uso de ordenadores para el aprendizaje, han sido las siguientes:

$$X_i = (46, 45, 38, 42, 39, 36, 44, 42, 43, 44, 47, 45, 44, 38, 44)$$

Calcule la varianza de la distribución. Aunque se trata de un cuestionario de actitud, cuyas medidas no son de intervalo, por tratarse de la valoración de una variable subjetiva como es la actitud, vamos a trabajar con los datos como si fueran de intervalo).

Tomamos la fórmula de la varianza:

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

Viendo las incógnitas que hay que resolver, calculamos la media:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{637}{15} = 42.46$$

Ordenamos los datos situándolos en una tabla, calculamos la media, sus puntuaciones diferenciales y el cuadrado de éstas y las sumamos:

TABLA 13.4.

X_i	x_i	x_i^2
36	-6.47	41.82
38	-4.47	19.95
38	-4.47	19.95
39	-3.47	12.02
42	-.47	.22
42	-.47	.22
43	.53	.28
44	1.53	2.35
44	1.53	2.35
44	1.53	2.35
44	1.53	2.35
45	2.53	6.42
45	2.53	6.42
46	3.53	12.48
47	4.53	20.55
$\Sigma = 637$	$\Sigma = 0$	$\Sigma = 149,73$

Y sustituyendo los datos en la fórmula de la varianza, tenemos:

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n} = \frac{149,73}{15} = 9,982$$

13.3.3.2. La desviación típica

Como medida descriptiva se utiliza más la desviación típica que la varianza, dado que el valor resultante de hallar la varianza, con respecto a la media, es muy distante de ésta por estar sus unidades elevadas al cuadrado. La desviación típica se puede obtener a partir de la varianza por ser la raíz cuadra de la varianza.

$$s = \sqrt{s^2}$$

Con muestras de tamaño pequeño, $n < 30$, al denominador se le resta una unidad, quedando del siguiente modo:

$$s = \sqrt{\frac{\sum x_i^2 n_i}{n - 1}}$$

13.3.3.3. Significado de la desviación típica y la varianza

Al hablar de las medidas de tendencia central, o valores promedio de una distribución, mencionamos que los índices que nos aportan más información sobre la calificación típica o más representativa del grupo son la media y la mediana.

Sin embargo, cuando queremos tener información acerca de las puntuaciones obtenidas en una determinada variable por un grupo formado por un elevado número de sujetos, no nos basta con su media aritmética como medida de tendencia central, sino que necesitamos una medida de variabilidad, ya sea la desviación típica o la varianza. Así, cuando obtenemos el valor de la media aritmética de un grupo y la desviación típica, ésta nos informa de la homogeneidad o heterogeneidad de las puntuaciones del grupo. Nos indica cuantos sujetos se sitúan alrededor de la media, cuántos por debajo y cuántos por encima de esa media y qué tan próximos a ella. Cuánto menor sea el valor de la desviación típica, más sujetos están próximos a la media. Cuánto mayor sea la desviación típica, más dispersas están las calificaciones a ambos lados del valor de la media.

La desviación típica es la medida de variabilidad más utilizada, por ser la más fiable, en el sentido de que es la que menos varía de una muestra a otra. La principal ventaja de la desviación típica sobre la varianza es que la desviación típica viene dada en las mismas unidades que la media.

13.3.3.4. *El coeficiente de variación*

El coeficiente de variación se define como el cociente entre la desviación típica de una distribución y su media aritmética multiplicado por cien. Este índice nos va a permitir realizar comparaciones entre variables de distinta naturaleza. Su fórmula es:

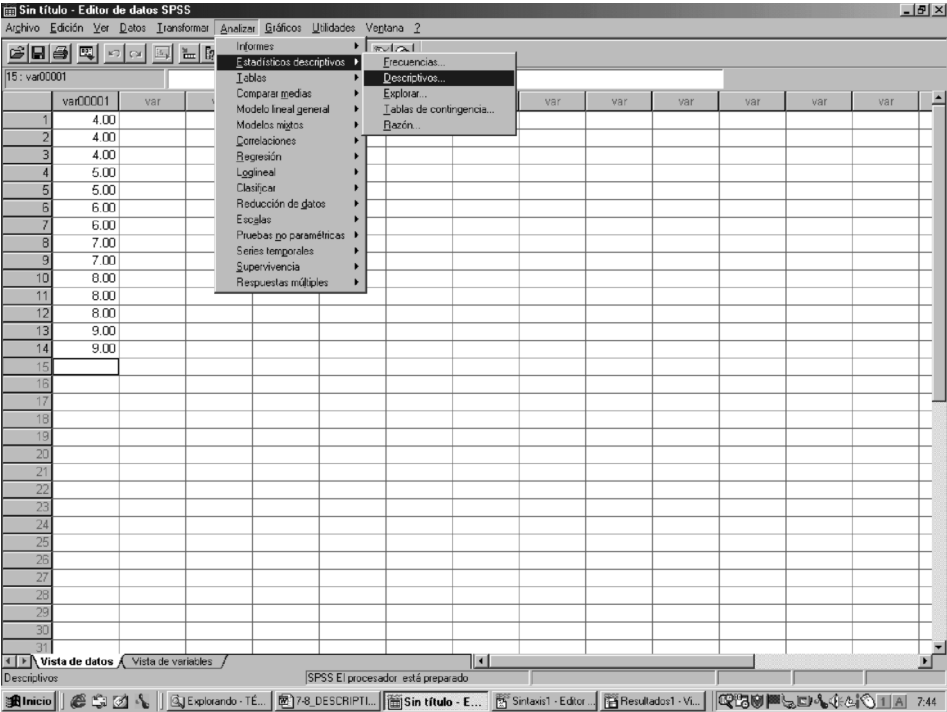
$$CV = \frac{S_x}{\bar{X}} \times 100$$

Veamos un ejemplo.

Queremos calcular la media, desviación típica y la varianza de las calificaciones obtenidas por un grupo de 14 alumnos en una prueba de Matemáticas, cuyos resultados han sido los siguientes: 4, 6, 7, 5, 4, 6, 8, 9, 7, 8, 9, 5, 4, 8.

Si lo resolvemos utilizando el programa informático SPSS, lo primero que hacemos es introducir los datos en un fichero de datos, en la primera columna de la pantalla, que se nombra como Variable1. Si en la barra de herramientas del fichero de datos seleccionamos: Datos>ordenar datos>descendente, en nuestra columna de la variable 1 aparecen los datos ordenados de mayor a menor (este paso no es imprescindible para nuestro propósito).

Una vez introducidos las puntuaciones directas, o datos observados en el fichero de datos, en la barra de herramientas marcamos: Analizar > Estadísticos descriptivos > descriptivos



Y en el cuadro de diálogo > opciones > varianza > pegar, lo que nos da un archivo de sintaxis, seleccionamos y marcamos el contenido de la sintaxis, y en la barra de herramientas de este nuevo cuadro de diálogo, pulsamos ejecutar, que viene representado por una punta de flecha, y nos amplía el anterior archivo de resultados, en el que aparece una tabla con los estadísticos descriptivos que le hemos pedido, del modo como queda reflejado a continuación:

TABLA 13.5. Estadísticos descriptivos

Estadísticos descriptivos						
	N	Mínimo	Máximo	Media	Desv. típ.	Varianza
VAR00001	14	4.00	9.00	6.4286	1.82775	3.341
N válido (según lista)	14					

Ahora le pedimos a Vd. que calcule el coeficiente de variación y que interprete los resultados en términos de valoración del rendimiento obtenido por el grupo de alumnos en Matemáticas.

13.4. Las puntuaciones individuales

Hasta ahora hemos mencionado las puntuaciones directas X_i y las puntuaciones grupales. A partir de estas puntuaciones se puede estudiar la relación de unas puntuaciones directas de un sujeto con las del grupo.

13.4.1. Las puntuaciones diferenciales

La diferencia entre una puntuación directa individual X_i y la media aritmética del grupo se denomina puntuación diferencial e indica la desviación de una puntuación directa de un sujeto con respecto a la media del grupo.

$$x_i = X_i - \bar{X}$$

Las propiedades de estas puntuaciones diferenciales son las siguientes:

- La media de las puntuaciones diferenciales es igual a cero:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{\sum (X_i - \bar{X})}{n} = \frac{\sum X_i - \sum \bar{X}}{n} = \frac{\sum X_i}{n} - \frac{n\bar{X}}{n} = \bar{X} - \bar{X} = 0$$

- La varianza de las puntuaciones diferenciales es igual a la varianza de las puntuaciones directas:

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum x_i^2}{n} = \frac{\sum (X_i - \bar{X})^2}{n}$$

Las puntuaciones diferenciales de dos series de datos diferentes podrían coincidir en su valor pero no significar lo mismo, no permitiendo la comparación. Las puntuaciones típicas aportan una solución a esta dificultad.

13.4.2. Las puntuaciones típicas

Una puntuación típica se define como la diferencia entre una puntuación directa en una variable y su media (puntuación diferencial) dividida entre la desviación típica de esa misma variable.

$$z_i = \frac{x_i}{s_x} = \frac{X_i - \bar{X}}{s_x}$$

La puntuación típica indica el número de desviaciones típicas que una determinada puntuación directa se aparta de la media del grupo.

Mediante la fórmula de las puntuaciones típicas obtenemos una nueva variable que es una transformación lineal de la inicial; hemos transformado la variable aleatoria X_i en una variable tipificada. Ya mencionamos esto en el capítulo anterior cuando hablamos de la distribución normal. Las propiedades de la distribución de las puntuaciones típicas z , son las siguientes:

- 1.^a La media de la distribución de las puntuaciones típicas z es igual a cero:

$$\bar{z}_x = \frac{\sum z_x}{n} = \frac{\sum \left(\frac{x_i}{s_x} \right)}{n} = \frac{\frac{1}{s_x} \sum x_i}{n} = \frac{\sum x_i}{ns_x} = \frac{0}{ns_x} = 0$$

- 2.^a La varianza y la desviación típica de la distribución de las puntuaciones típicas z es igual a uno, 1:

$$s_{z_x}^2 = \frac{\sum (z_x - \bar{z}_x)^2}{n} = \frac{\sum (z_x)^2}{n} = \frac{\sum \left(\frac{x}{s_x} \right)^2}{n} = \frac{\frac{1}{s_x^2} \sum x^2}{n} = \frac{1}{s_x^2} \frac{\sum x^2}{n} = \frac{1}{s_x^2} s_x^2 = 1$$

Las puntuaciones típicas reflejan las relaciones entre las puntuaciones directas y las grupales de una serie de datos a otra independientemente de la unidad de medida utilizada. Así, cuando trabajamos con las puntuaciones típicas de dos series, y son iguales, se dice que son equivalentes.

13.4.3. La curva normal

La curva normal, o distribución normal, es la representación gráfica mediante un polígono de frecuencias, o un histograma, de la distribución de una serie teóricamente infinita de datos, que adopta la forma de campana y que se la conoce con el nombre de campana de Gauss, en honor al astrónomo y matemático alemán, que junto con Laplace, astrónomo y matemático francés, hicieron importantes contribuciones al conocimiento de esta distribución.

La distribución normal presenta las siguientes propiedades:

- 1.^a Es una distribución simétrica en torno a su media aritmética, dejando por debajo y por encima de ella el 50 % de la totalidad de las puntuaciones o casos.

- 2.^a Por ser simétrica, media, mediana y moda coinciden.
- 3.^a Es asintótica con respecto al eje de abscisas (no lo toca); tiene una amplitud infinita, por lo que la curva se extiende desde $-\infty$ hasta $+\infty$, sin tocar el eje de abscisas.
- 4.^a Tiene dos puntos de inflexión situados a una desviación típica con respecto a la media, a ambos lados de la misma, pasando la curva de ser cóncava a ser convexa.

Existe una familia de distribuciones normales que difieren entre sí en cuanto a su media y a su variabilidad, simbolizadas genéricamente por la expresión:

$$N(\bar{X}, s_x)$$

Si transformamos las puntuaciones directas de una distribución en puntuaciones típicas, se obtiene la distribución normal, o distribución estándar o tipificada, que tiene de media = 0 y varianza y desviación típica = 1, y se simboliza así: $N(0,1)$.

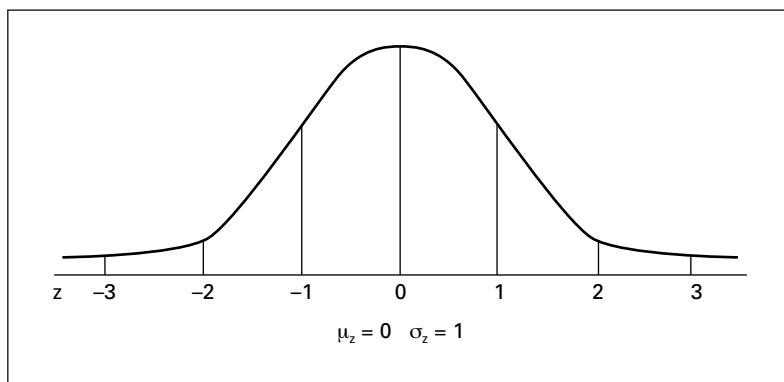


FIGURA 13.2. Curva normal.

En cualquier conjunto de datos, son válidas las siguientes relaciones:

1. Por lo menos el 56% de los individuos, o los casos, tienen valores situados dentro de $1,5 \sigma$ (o s , o desviaciones típicas).
2. Por lo menos el 75% de los individuos tienen valores situados entre 2σ a ambos lados de la media.
3. Por los menos el 89% está situado entre la media y 3σ a ambos lados de la media.

4. Por lo menos el 94 % de los sujetos está situado entre la media y 4σ a ambos lados.
5. Todo ello independientemente de la forma que tenga la distribución de frecuencias de los datos.

En las distribuciones que toman la forma de curva normal, o campana de Gauss, la aproximación mencionada en los cinco puntos anteriores es mayor.

Para los valores de z siguientes, el área que deja por debajo de sí la curva normal es la siguiente:

$z = \pm 1 = 68,26$ % del área bajo la curva normal.

$z = \pm 1,96 = 95,00$ % del área bajo la curva normal.

$z = \pm 2,58 = 99,00$ % del área bajo la curva normal.

Las áreas bajo la curva normal se interpretan como la probabilidad de que un valor elegido al azar esté comprendido entre los z valores típicos dados, o como el porcentaje de casos que hay entre dos valores de z dados con respecto al total. Esto tiene una gran importancia para la inferencia estadística.

13.4.4. Normalización de las distribuciones

Cuando se trabaja con datos procedentes de muestras tomadas de la realidad, la distribución de los datos tomados de una determinada variable no siempre se ajusta a la normal, pero se sabe que esta variable en la población sí adopta la forma de curva normal. Por eso conviene normalizar la distribución adaptándola a la curva normal, manteniendo su media y su desviación típica. A este proceso se le conoce con el nombre de normalización de la distribución, y consiste en la realización de los siguientes pasos:

- 1.º Se calculan las puntuaciones típicas correspondientes a cada uno de los límites exactos de cada intervalo, excepto el de los límites extremos de la distribución.
- 2.º Se determinan las proporciones correspondientes a cada intervalo.
- 3.º Se transforman las proporciones anteriores en su correspondiente frecuencia.

Veamos un ejemplo.

A un grupo compuesto por 600 alumnos se les ha aplicado una prueba en la que han obtenido unas calificaciones cuya media y desviación típica son:

$$\bar{X} = 6,25$$

$$s = 2,5$$

Supuesta la condición de distribución normal de los datos, ¿Cuántos alumnos quedan por encima y por debajo de aquél cuya puntuación en la prueba ha sido $X_i = 8$?

1.º Debemos realizar el cálculo de z

$$z_i = \frac{X_i - \bar{X}}{s}; z = \frac{8 - 6,25}{2,5} = 0,70$$

2.º Consultamos en la tabla de áreas bajo la curva normal la probabilidad que deja por debajo de sí y hasta la media, la puntuación obtenida y nos da una probabilidad por encima de la media igual a 0,2580. Como cada mitad del área de la curva normal a ambos lados de la media equivale al 0,5 de probabilidades, entonces:

$$0,5 + 0,2580 = 0,7580$$

3.º Representamos los valores en la curva normal:

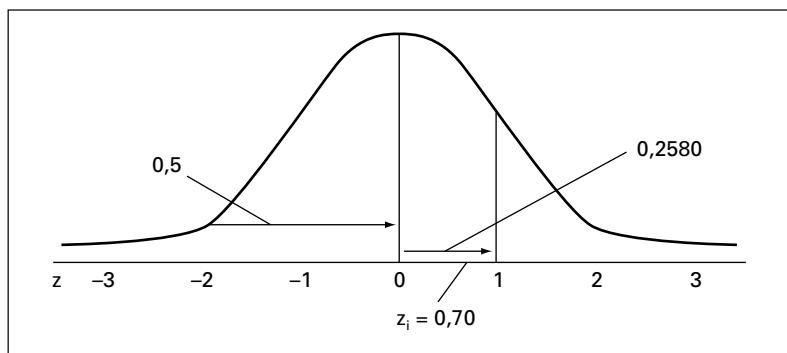


FIGURA 13.3.

4.º Comprobamos que el sujeto cuya puntuación z es de 0,70, deja por debajo de sí un área equivalente a:

$$0,5 + 0,2580 = 0,7580$$

5.º Mediante una regla de tres, hallamos el número de sujetos que corresponde a esta área: Si la totalidad de sujetos, 600, están dis-

tribuidos bajo la curva normal, cuya área total equivale a la probabilidad = 1, el sujeto que obtuvo la puntuación equivalente en típicas de 0,70, deja por debajo de sí un área equivalente a la probabilidad = 0,7580, luego:

$$\frac{600}{1} = \frac{x}{0,7580}; \Rightarrow x = 600 \times 0,7580 = 454,8$$

6.º Concluimos: el sujeto cuya puntuación obtenida en la prueba ha sido 8, deja por debajo de sí un total de 454 sujetos, y por encima de sí, 155, ocupando él la posición aproximada de 455.

Veamos otro ejemplo:

A un grupo de 400 alumnos se le ha aplicado una prueba objetiva de Matemáticas, obteniendo una Media = 20 y desviación típica = 5, supuesta la condición de que dichas puntuaciones se distribuyen de acuerdo con la curva normal. ¿Cuántos sujetos se sitúan entre el A y el B, cuyas puntuaciones son: $X_A = 23$ y $X_B = 29$?

1.º Calculamos las puntuaciones típicas:

$$z_i = \frac{X_i - \bar{X}}{s}; z_A = \frac{23 - 20}{5} = 0,60; z_B = \frac{29 - 20}{5} = 1,80$$

2.º Consultamos las tablas de probabilidades bajo la curva normal, para ver la probabilidad equivalente al área que dejan por debajo de sí las dos puntuaciones típicas halladas. Ambas puntuaciones son superiores a la media de z cuyo valor sabemos es 0. La puntuación $z_A = 0,60$, deja entre la media y ella un área con una probabilidad asociada de 0,2257 y la puntuación $z_B = 1,80$, deja entre la media y ella un área con una probabilidad asociada de 0,4641. Entre ambas puntuaciones z hay una probabilidad asociada al área bajo la curva normal:

$$0,4641 - 0,2257 = 0,2384$$

3.º Dibujamos la curva normal y representamos los valores de z y las áreas correspondientes a dichos valores:

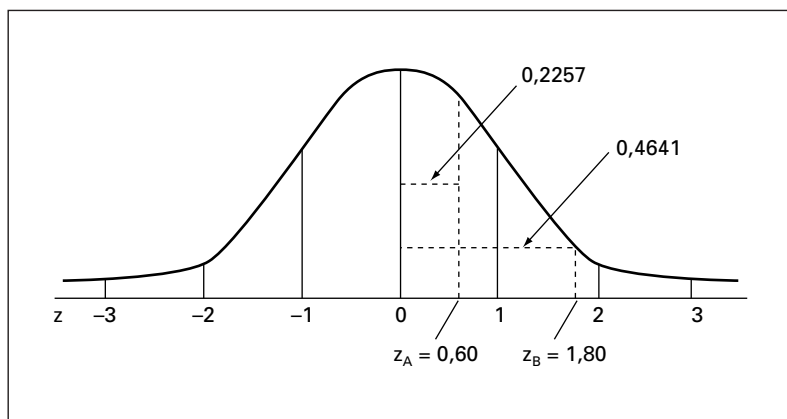


FIGURA 13.4.

- 4.º Como en el ejercicio anterior, mediante una regla de tres, o una proporción, hallamos los sujetos equivalentes al área:

$$\frac{x}{400} = \frac{0,2384}{1}; \quad x = \frac{0,2384 \times 400}{1} = 95,36$$

- 5.º Conclusión: Son 95 alumnos los que se sitúan entre las dos puntuaciones z halladas.

13.4.5. Asimetría y curtosis

La asimetría o sesgo de una distribución de frecuencias hace referencia a la proporción de observaciones por encima y por debajo del promedio. Si están equilibradas se dice que la distribución es simétrica. Si hay mayor número de valores inferiores a la media que superiores se dice que la distribución es asimétrica positiva. Y si hay mayor número de valores superiores a la media se dice que es asimétrica negativa.

La simetría de una distribución de frecuencias de una variable respecto a la línea trazada desde el eje de abscisas a partir de un valor central, como puede ser la media, queda reflejada en la gráfica de la distribución. La distribución de frecuencias de una variable es simétrica cuando sus estadísticos de tendencia central media, moda y mediana coinciden, como es el caso de la distribución de frecuencias de la curva normal.

El coeficiente de sesgo o asimetría se halla dividiendo la diferencia de la media y la moda por el valor de la desviación típica:

$$\text{Coeficiente de asimetría} = \frac{\bar{X} - Mo}{s}$$

Si el coeficiente es igual a cero, la distribución es simétrica.

Si el coeficiente es mayor de 0, la distribución es sesgada a la derecha.

Si el coeficiente es menor de 0, la distribución es sesgada a la izquierda.

La curtosis o apuntamiento se refiere a la forma que adopta la distribución de frecuencias de una variable en relación con la acumulación de puntuaciones en torno al promedio. Si existe mucha acumulación en torno al promedio, resulta una curva apuntada, elevada, y se denomina leptocúrtica. Si se da poca concentración es platicúrtica, y si se da en un grado intermedio es mesocúrtica.

El coeficiente de apuntamiento se halla dividiendo el valor de la media de la distribución entre el valor de su desviación típica elevado a la cuarta potencia, y disminuido el resultado de la división en tres unidades:

$$\text{Coeficiente de apuntamiento} = \frac{\bar{X}}{s^4} - 3$$

Si el coeficiente es igual a cero, o casi cero, la distribución es similar a la curva normal y se denomina mesocúrtica.

Si el coeficiente es mayor de 0, la distribución es más estrecha y elevada que la curva normal y se denomina leptocúrtica.

Si el coeficiente es menor de 0, la distribución es más ancha y baja que la curva normal y se denomina platicúrtica.

13.5. Ejercicios de autoevaluación

Problema 1

Tomando la tabla de datos vista en el ejemplo sobre la media, referido a las puntuaciones obtenidas por 12 alumnos en las materias de Lengua Española, Matemáticas e Inglés:

TABLA 13.6. Puntuaciones en las materias.

Sujetos	Lengua	Matemáticas	Inglés
1	3	4	3
2	3	5	6
3	4	6	5
4	4	4	6
5	5	6	6
6	6	5	7
7	6	4	6
8	7	6	7
9	7	7	8
10	7	9	8
11	8	7	9
12	9	8	8

Le pedimos que halle las medias, desviaciones típicas, varianzas y coeficientes de variación en las tres variables y valore los resultados en términos educativos.

Problema 2

Tenemos dos series de datos que se presentan en torno a dos columnas; X_i representa las puntuaciones que han obtenidos los alumnos de dos Centros Educativos en una prueba de levantamiento de pesos, y f_i es el número de alumnos que coinciden en levantar un determinado peso, y significa la frecuencia.

TABLA 13.7. Series de datos de dos Centros.

Serie I		Serie II			
X_i	f_i	X_i	f_i	X_i	f_i
1	1	1	1	12	150
2	22	2	1	13	90
3	80	3	5	14	80
4	200	4	10	15	70
5	400	5	30	16	50
6	200	6	50	17	30
7	80	7	70	18	10
8	20	8	80	19	5
9	1	9	90	20	1
		10	150	21	1
		11	200		

1. Represente gráficamente dichos datos mediante tres formas diferentes, siendo una de ella un polígono de frecuencias que tendrá la forma de una distribución normal.
2. Halle la desviación típica para cada serie de datos y su varianza.
3. Compruebe las propiedades de la desviación típica, en relación con la curva normal, en estas dos series de datos, y diga el número de sujetos y su porcentaje, que se encuentran entre 1 desviación típica a ambos lados de la media.
4. Compare los resultados que obtenga, a partir de las dos series de datos, y valórelos.
5. Halle el coeficiente de variación de ambas series y compárelo

Problema 3

Un grupo de 25 alumnos de un Centro Educativo de Primaria de 6 curso ha obtenido en la asignatura de Matemáticas las calificaciones siguientes mediante una prueba de rendimiento precisa.

Calificaciones: 3, 3, 5, 5, 4, 7, 6, 9, 8, 6, 7, 5, 6, 4, 6, 5, 8, 5, 7, 6, 6, 5, 4, 7, 6

- Elabore una tabla por intervalos representando las frecuencias absolutas, acumuladas y relativas, en proporciones y porcentajes.
- Represente gráficamente los datos mediante un polígono de frecuencias, y comente las características de la distribución
- Halle los estadísticos siguientes: Media, mediana, moda, distribución típica, varianza.

Problema 4

Retomando el problema del apartado referido a la varianza, que decía: Las puntuaciones obtenidas por un grupo de 15 alumnos en un cuestionario de 10 ítems valorados de 1 a 5, sobre la actitud ante el uso de ordenadores para el aprendizaje, han sido las siguientes:

$$X = (46, 45, 38, 42, 39, 36, 44, 42, 43, 44, 47, 45, 44, 38, 44)$$

Ahora le pedimos que calcule la varianza de la distribución por la fórmula de frecuencias, la desviación típica y el coeficiente de asimetría y curtosis (La media era: 42,47, y la varianza: 9,98).

Problema 5

Los resultados obtenidos en una prueba de razonamiento por una muestra de 500 sujetos, que se distribuyen de acuerdo con la curva normal, y cuya Media = 40 y $s = 5$. ¿Cuántos sujetos se sitúan entre aquellos cuyas puntuaciones han sido $X_A = 38$ y $X_B = 32$? Indíquelo numérica y gráficamente.

Problema 6

La media del salario mensual del total de la población activa en España, cifrada en 5 millones, se sitúa en 350.000 pesetas, siendo su desviación típica de 5,75. Cuántas personas se sitúan entre aquellos que ganan un salario de 175.000 y 250.000 pesetas. Dibuje la curva normal y el área que deja entre sí estos dos sujetos.

Problema 7

En un grupo de 400 alumnos de Secundaria se ha aplicado una prueba objetiva en Química cuyos resultados cumple el requisito de normalidad. ¿Cuántos alumnos se sitúan entre el alumno A cuya puntuación es 20, y el B, cuya puntuación es 28, siendo la media 25 y la desviación típica 4? Represéntelo gráficamente.

13.6. Soluciones a los ejercicios de autoevaluación

Presentamos las soluciones de algunos de los problemas propuestos.

Problema 1

TABLA 13.9. Descriptivos

	Lengua Española	Matemáticas	Inglés
Medias	5,75	5,916	6,583
Desviación típica	1,959	1,621	1,621

Como podemos ver existen diferencias mínimas entre las puntuaciones obtenidas en las tres asignaturas, y asimismo sus desviaciones típicas están próximas.

Problema 2

La desviación típica de la primera serie es: $s = 1,411$. La de la segunda serie es $s = 1,82$.

Problema 4

Las puntuaciones obtenidas por un grupo de 15 alumnos en un cuestionario de 10 ítems valorados de 1 a 5, sobre la actitud ante el uso de ordenadores para el aprendizaje, han sido las siguientes:

$$X = (46, 45, 38, 42, 39, 36, 44, 42, 43, 44, 47, 45, 44, 38, 44)$$

(La media era: 42,47, y la varianza: 9,98).

a) Para calcular la varianza tomamos la fórmula de puntuaciones agrupadas y muestras pequeñas:

$$s^2 = \frac{\sum x_i^2 n_i}{n - 1}$$

Ordenaremos los datos situándolos en una tabla por frecuencias y calculamos las puntuaciones diferenciales, sus cuadrados y el producto con su respectiva frecuencia:

TABLA 13.10. Tabla para calcular la varianza

X_i	f_i	x	x_i^2	$(x_i^2 f_i)$
36	1	-6.47	41.86	41.86
38	2	-4.47	19.98	39.96
39	1	-3.47	12.04	12.04
42	2	-0.47	0,22	0.44
43	1	0.53	0.28	0.28
44	4	1.53	2.34	9.36
45	2	2.53	6.40	12.80
46	1	3.53	12.46	12.46
47	1	4.53	20.52	20.52
	15			$\Sigma = 146.72$

Sustituyendo los datos en la fórmula, tenemos:

$$s^2 = \frac{\sum x_i^2 n_i}{n-1} = \frac{146.72}{14} = 10,48$$

El resultado de la varianza es algo diferente al hallado anteriormente (9,98) debido a que aquí hemos aplicado el factor de corrección $(n - 1)$ para muestras pequeñas, como es el caso.

- b) La desviación típica. Sabemos que la desviación típica se puede obtener a partir de la varianza por ser la raíz cuadra de la varianza. Sustituyendo los datos en la fórmula, tenemos:

$$s = \sqrt{s^2} = \sqrt{10,48} = 3.237$$

- c) El coeficiente de asimetría La fórmula del coeficiente de sesgo o asimetría se halla dividiendo la diferencia de la media y la moda por el valor de la desviación típica:

$$\text{Coeficiente de asimetría} = \frac{\bar{X} - Mo}{s}$$

El valor de la moda es 44, que es el que más se repite, entonces sustituyendo los datos en la fórmula tenemos:

$$\text{Coeficiente de asimetría} = \frac{42,77 - 44}{3,24} = \frac{-1,53}{3,24} = -0,47$$

Como el coeficiente es menor de 0, la distribución tiene un sesgo hacia la izquierda.

- d) El coeficiente de curtosis o de apuntamiento se halla dividiendo el valor de la media de la distribución entre el valor de su desviación típica elevado a la cuarta potencia, y disminuido el resultado de la división en tres unidades:

$$\begin{aligned}\text{Coeficiente de apuntamiento} &= \frac{\bar{X}}{s^4} - 3 = \frac{42,47}{3,24^4} - 3 = \frac{42,47}{110,19} - 3 = \\ &= 0,385 - 3 = -2,61\end{aligned}$$

Como el coeficiente es menor de 0, la distribución es más ancha y baja que la curva normal y se denomina platicúrtica.

Problema 7

En un grupo de 400 alumnos de Secundaria se ha aplicado una prueba objetiva en Química cuyos resultados cumple el requisito de normalidad. ¿Cuántos alumnos se sitúan entre el A cuya puntuación es 20, y el B, cuya puntuación es 28, siendo la media 25 y la desviación típica 4? Representélo gráficamente.

$$z = \frac{X_i - X}{s}; z_A = \frac{20 - 25}{4} = \frac{-5}{4} = -1,25; z_B = \frac{28 - 25}{4} = \frac{3}{4} = 0,75$$

$z_A = -1,25$ deja por debajo de sí una probabilidad $p = 0,1056$

$$z_B = 0,75 \quad p = 0,5 + 0,2734 = 0,7734;$$

$$p_B - p_A = 0,7734 - 0,1056 = 0,6678.$$

Veámoslo gráficamente:

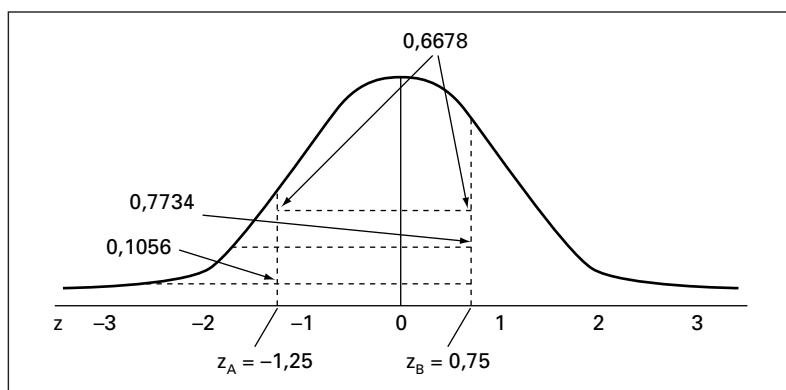


FIGURA 13.5.

Mediante una proporción, o una regla de tres, calculamos el número de sujetos correspondientes al área de probabilidad entre ambas puntuaciones típicas.

$$\frac{400}{x} = \frac{1}{0,6678}; x = \frac{0,6678 \times 400}{1} = 267,121 \cong 267$$

Y la solución es que son 267 los sujetos que se sitúan entre el A y el B.

13.7. Referencias bibliográficas

- KERLINGER, F. N. (1985). *Investigación del comportamiento*. México, Interamericana.
- SPSS (2002). *Programa informático para el análisis de datos estadísticos*. Versión 11.

TEMA 14

Correlaciones bivariadas

ESQUEMA

- 14.1.** Las distribuciones bidimensionales y la correlación
- 14.2.** Las tablas de correlación y las tablas de contingencia
- 14.3.** La covarianza
- 14.4.** El diagrama de dispersión
- 14.5.** Las medidas correlacionales bivariadas y su interpretación
 - 14.5.1. La correlación producto momento de Pearson
 - 14.5.2. Interpretación de la correlación producto momento de Pearson
 - 14.5.3. Cálculo de la correlación producto momento de Pearson mediante el programa informático SPSS
 - 14.5.4. La correlación ordinal o de rangos de Spearman
 - 14.5.5. Interpretación de la correlación ordinal o de rangos de Spearman
 - 14.5.6. Cálculo de la correlación ordinal de Spearman mediante el programa informático SPSS
 - 14.5.7. El coeficiente de contingencia o de asociación C de Pearson
 - 14.5.8. Interpretación del coeficiente de contingencia C de Pearson
- 14.6.** Ejercicios de autoevaluación
- 14.7.** Solución a los ejercicios de autoevaluación
- 14.8.** Referencias bibliográficas

OBJETIVOS

- ✓ Conocer y comprender el concepto de covarianza.
- ✓ Conocer y comprender el significado de la relación.
- ✓ Conocer y comprender los diferentes coeficientes de correlación.
- ✓ Calcular e interpretar el coeficiente de correlación producto momento de Pearson.
- ✓ Calcular e interpretar el coeficiente de correlación ordinal de Spearman.
- ✓ Calcular e interpretar el coeficiente de asociación C de Pearson.

14.1. Las distribuciones bidimensionales y las relaciones

La ciencia no es el conocimiento de simples datos particulares, sino el conocimiento del modo en que se relacionan entre sí dichos datos.

Las distribuciones bidimensionales son el resultado de las observaciones realizadas sobre una determinada muestra, o una población, de dos características de dicha muestra o población, de manera simultánea. A cada uno de los casos, o sujetos, observados, le corresponden dos observaciones que pueden ser dos valores numéricos, si se trata de características cuantificables, o dos categorías o denominaciones, si se trata de variables categóricas. La distribución bidimensional de las variables observadas X e Y viene definida del siguiente modo (x_i, y_j, n_{ij}) , siendo el primer elemento el valor de la característica observada en la variable X , el segundo elemento el valor de la característica observada en la variable Y , y el tercero la frecuencia conjunta.

Una relación queda establecida por un conjunto de pares ordenados, estando constituidos cada par por dos objetos o elementos que aparecen en un orden fijo.

Las relaciones se determinan mediante un producto cartesiano. Por ejemplo, si tenemos el conjunto $A = (a_1, a_2, a_3, a_4)$ y el conjunto $B = (b_1, b_2, b_3, b_4)$, el conjunto $A \times B$ sería el resultado del producto cartesiano entre cada uno de los elementos de ambos conjuntos:

$$A \times B = \{(a_1, b_1), (a_1, b_2), (a_1, b_3), (a_1, b_4), (a_2, b_1), (a_2, b_2), (a_2, b_3), (a_2, b_4), (a_3, b_1), (a_3, b_2), (a_3, b_3), (a_3, b_4), (a_4, b_1), (a_4, b_2), (a_4, b_3), (a_4, b_4)\}$$

El conjunto del primer miembro recibe el nombre de Inicial (I), y el del segundo miembro recibe el nombre de Final o imagen (F).

Y gráficamente se representa así:

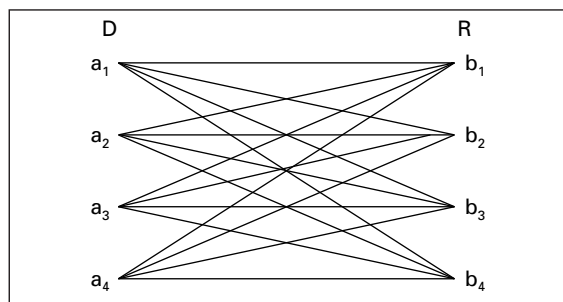


FIGURA 14.1. Relaciones entre D y R.

Pudiendo representarse también, mediante los ejes cartesianos. En el eje de abscisas se registrarían los valores de la variable X o dominio, y en el eje de ordenadas se registran los valores de la variable Y o extensión. Los puntos de cruce entre ambos valores son los pares ordenados de los valores de ambas variables.

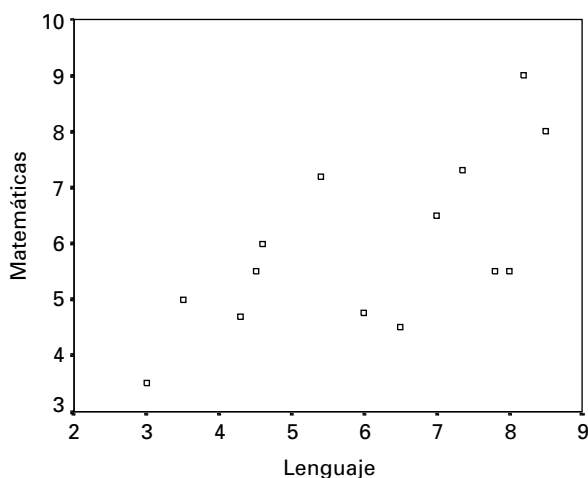


FIGURA 14.2. Relaciones entre X e Y .

14.2. Las tablas de correlación y las tablas de contingencia

Los valores observados en las variables se representan mediante una tabla de doble entrada. Dependiendo del nivel de medida de las variables observadas, las tablas reciben el nombre de tablas de correlación, para medidas de intervalo y ordinales, y tablas de contingencia, para variables categóricas o nominales.

Veamos los siguientes ejemplos:

Tenemos la siguiente información sobre una familia integrada por los padres y sus cinco hijos:

Andrés, padre, 48, 174 cm de estatura, 75 kg.

Rosa, madre, 41, 165 cm de estatura, 55 kg.

José, hijo, 14 años, 156 cm de estatura, 56 kg.

Ana, hija, 12 años, 147 cm de estatura, 40 kg.

Isabel, hija, 10 años, 135 cm de estatura, 38 kg.

Cristina, hija, 8 años, 120 cm de estatura, 28 kg.

Antonio, hijo, 6 años, 112 cm de estatura, 30 kg.

Si queremos ver la relación entre los nombres y el género, primero formamos la siguiente tabla, (el dominio es el nombre y el rango es el género):

TABLA 14.1. Nombres y género

	Inicial	Final
Sujetos	Nombre (X)	Género (Y)
1	Antonio	Hombre
2	Cristina	Mujer
3	Isabel	Mujer
4	Ana	Mujer
5	José	Hombre
6	Rosa	Mujer
7	Andrés	Hombre

Si, por ejemplo, queremos ver la relación entre edad y estatura, ordenamos los pares de datos en una tabla de doble entrada, formamos una tabla de correlación:

TABLA 14.2. Edad y estatura

Sujetos	Edad (X)	Estatura (Y)
1	6	112
2	8	120
3	10	135
4	12	147
5	14	156
6	35	165
7	42	174

Podemos observar que la información está ordenada por el criterio de edad, de menor a mayor. En esta tabla, en la que ambas variables, edad y estatura, vienen medidas a un nivel de intervalo, se llama tabla de correlación.

Y ahora queremos relacionar la edad de los miembros de la familia con el género, a partir de la siguiente tabla:

TABLA 14.3. Edad y género

	Inicial	Final
Sujetos	Edad (X)	Género (Y)
1	6	Hombre
2	8	Mujer
3	10	Mujer
4	12	Mujer
5	14	Hombre
6	35	Mujer
7	42	Hombre

A partir de la información recogida en la tabla referida a la edad y género, en la que una de las variables viene medida a nivel nominal, no podemos sino elaborar una tabla de contingencia. Los criterios van a ser, para las filas, la edad, en dos categorías: menores de 30 e iguales o mayores de 30, y en las columnas la categoría género: hombre, mujer. Veámoslo:

TABLA 14.4. Edad y género

Tabla de contingencia			
Género	Hombre	Mujer	Totales
Edad			
A(< 30)	// (2)	/// (3)	5
B (30 ó >)	/ (1)	/ (1)	2
Totales	3	4	7

Podemos observar que en las cuadrículas se sitúan las frecuencias observadas que cumplen la regla de correspondencia definida: hombres menores de 30 años y de 30 años y más, y mujeres menores de 30 años y de 30 años y más.

La regla de correspondencia entre los valores de ambas variables se llama función y suele designarse por las letras $f, g, \dots$ la cual asigna a un miembro, en este caso un valor, particular del conjunto inicial otro miembro del segundo conjunto, otro valor final o imagen. Todas las funciones son relaciones de conjuntos de pares ordenados, pero no significan que una cosa dependa de la otra, o sea una cosa causada por la otra. A cada miembro del conjunto inicial se le asigna uno y sólo un miembro del con-

junto final. Se ha establecido, de este modo una función mediante una regla de correspondencia:

$f = ((x,y); x \text{ es el valor del conjunto inicial; } y \text{ el valor del conjunto final})$

Las funciones se expresan del modo siguiente:

$$y = f(x)$$

Se lee « y es función de x ». « x » representa a los objetos del dominio y « $f(x)$ » designa a los objetos que se asignan a la x del dominio, es decir, los objetos del rango y . Todas las funciones son relaciones pero no todas las relaciones son funciones. Las funciones matemáticas pueden utilizarse para hacer deducciones y predecir variables.

14.3. La covarianza

Cuando analizamos los estadísticos descriptivos de una variable estudiamos la media, la desviación típica y la varianza, entre otros. Pero si estudiamos dos variables sobre un mismo grupo de sujetos, como podría ser la actitud hacia el aprendizaje con ordenador y el rendimiento en Matemáticas, además de conocer los estadísticos de cada una de las variables interesa analizar la distribución conjunta, siendo el estadístico más adecuado el de el coeficiente de la covarianza.

Recordemos el procedimiento utilizado para hallar la varianza, a partir de las puntuaciones obtenidas por un grupo de 15 alumnos en un cuestionario de actitud hacia el aprendizaje de contenidos curriculares por ordenador, mediante una escala de 10 ítems sobre los cuales el alumno podría expresar su grado de acuerdo entre 5 máximo y 1 mínimo:

TABLA 14.5. Tabla para el cálculo de la varianza

Alumnos	X	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
1	36	-6.47	41.82
2	38	-4.47	19.95
3	38	-4.47	19.95
4	39	-3.47	12.02
5	42	-.47	0.22
6	42	-.47	0.22
7	43	.53	0.28
8	44	1.53	2.35
9	44	1.53	2.35
10	44	1.53	2.35
11	44	1.53	2.35
12	45	2.53	6.42
13	45	2.53	6.42
14	46	3.53	12.48
15	47	4.53	20.55
$N = 15$	$\bar{X} = 42,47$	$\Sigma = 0$	$\Sigma = 149,73$

Y sustituyendo los datos en la fórmula de la varianza, tenemos:

$$s_x^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

Y para muestras pequeñas:

$$s_{n-1}^2 = \frac{\sum x^2}{n-1} = \frac{149,73}{15-1} = 10,48$$

Su desviación típica es: $s_x = 3,237$, es decir, la raíz cuadrada de 10,48.

Si además de las puntuaciones en el cuestionario de actitud tenemos las puntuaciones obtenidas por estos mismos 15 alumnos en una prueba de Matemáticas, de las cuales hallamos su varianza:

TABLA 14.6. Tabla para el cálculo de la varianza

Alumnos	Y	$(Y_i - \bar{Y})$	$(X_i - \bar{Y})^2$
1	6.00	-0.07	0.00
2	5.00	-1.07	1.14
3	11.00	0.93	0.87
4	4.00	-2.07	4.27
5	11.00	0.93	0.87
6	6.00	-0.07	0.00
7	11.00	0.93	0.87
8	5.00	-1.07	1.14
9	5.00	-1.07	1.14
10	6.00	-0.07	0.00
11	4.00	-2.07	4.27
12	6.00	-0.07	0.00
13	11.00	0.93	0.87
14	8.00	1.93	3.74
15	8.00	1.93	3.74
N = 15	$\bar{Y} = 6,067$	$\Sigma = 0$	$\Sigma = 22,93$

Sustituyendo los datos en la fórmula de la varianza para muestras pequeñas:

$$s_y^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} = \frac{\sum x^2}{n-1} = \frac{22,93}{14} = 1,638$$

Siendo su desviación típica, $s_y = 1,28$ (la raíz cuadrado de 1,638).

Hay que indicar que se trata de los mismos alumnos, y en el mismo orden (su identificación es el número: 1, 2, 3 ...15. Y esto es muy importante, puesto que se trata de conocer la relación entre pares ordenados de valores de dos variables un mismo sujeto. Ahora vamos a estudiar ambas variables simultáneamente, la distribución en su conjunto, siendo la covarianza el coeficiente más adecuado para el análisis estadístico. Su fórmula, para puntuaciones agrupadas, es la siguiente:

$$s_{xy} = \frac{1}{N} \sum_{i=1}^h \sum_{j=1}^k (X_i - \bar{X})(Y_j - \bar{Y}) n_{ij}$$

Y la fórmula para puntuaciones sin agrupar es:

$$s_{xy} = \frac{\sum_{i=1}^h \sum_{j=1}^k (X_i - \bar{X})(Y_j - \bar{Y})}{n}$$

Si ambas distribuciones varían en torno a sus medias aritméticas en el mismo sentido, es decir, los mismos alumnos obtiene las puntuaciones en el mismo orden aproximado, de menor a mayor, o viceversa, en ambas variables, el signo del coeficiente de covarianza será positivo, y si varían en sentido contrario, es decir, en una variable obtienen las puntuaciones en orden creciente, y en la otra decreciente, la covarianza tendrá signo negativo.

Vamos a considerar ahora las dos distribuciones referidas a las puntuaciones obtenidas por un grupo de 15 alumnos en actitud y en Matemáticas en su conjunto. Calculadas las medias y las puntuaciones diferenciales de cada distribución, y el producto de las puntuaciones diferenciales de una y otra variable, tenemos la siguiente tabla de datos:

TABLA 14.7. Tabla para el cálculo de la covarianza

Puntuaciones obtenidas en un cuestionario de Actitud hacia el aprendizaje por ordenador y calificaciones obtenidas en Matemáticas					
Alumnos	X	Y	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$
1	36	6.00	-6.47	-.07	0.43
2	38	5.00	-4.47	-1.07	4.76
3	38	7.00	-4.47	.93	-4.17
4	39	4.00	-3.47	-2.07	7.16
5	42	7.00	-.47	.93	-0.44
6	42	6.00	-.47	-.07	0.03
7	43	7.00	.53	.93	0.50
8	44	5.00	1.53	-1.07	-1.64
9	44	5.00	1.53	-1.07	-1.64
10	44	6.00	1.53	-.07	-0.10
11	44	4.00	1.53	-2.07	-3.17
12	45	6.00	2.53	-.07	-0.17
13	45	7.00	2.53	.93	2.36
14	46	8.00	3.53	1.93	6.83
15	47	8.00	4.53	1.93	8.76
N = 15	$\bar{X} = 42,47$	$\bar{Y} = 6,067$	$\Sigma = 0$	$\Sigma = 0$	$\Sigma = 19,53$

Y sustituyendo las incógnitas en la fórmula de la covarianza, para muestras pequeñas, por sus correspondientes valores numéricos, tenemos:

$$s_{xy} = \frac{\sum_{i=1}^h \sum_{j=1}^k (X_i - \bar{X})(Y_j - \bar{Y})}{n-1} = \frac{19,53}{14} = 1.385$$

Si no hubiéramos aplicado el valor de corrección para muestras pequeñas, es decir, situando en el denominador $N = 15$, el valor del coeficiente de covarianza que obtendríamos sería $s_{xy} = 1,3$.

El valor obtenido en el coeficiente de covariación de 1,38, nos indica que existe covariación entre ambas variables, las calificaciones obtenidas por el grupo en Lenguaje y en Matemáticas, y al ser su signo positivo nos indica que covarían en la misma dirección, existe, por lo tanto, una tendencia de los alumnos a obtener notas con incremento en la misma dirección.

Vemos que el procedimiento seguido para calcular la covarianza es muy similar al seguido para el cálculo de la varianza. Las diferencias entre ambas las observamos cuando se trata de realizar el producto de las puntuaciones diferenciales, por sí mismas, intra, en la varianza, y entre sí, en la covarianza. La variación del conjunto de pares ordenados recibe el nombre de covarianza, y es un índice de la relación entre ambos conjuntos de puntuaciones.

La covarianza es una forma de la varianza para conocer en qué modo varían dos variables relacionadas. La covarianza estudia la relación entre dos variables pero no nos indica si la relación es fuerte o débil. Esto nos lo indica la correlación. El valor de la covarianza nos permite conocer la covariación en un grupo de casos, o sujetos determinados, pero no nos permite realizar comparaciones entre casos distintos. Este problema se soluciona buscando una unidad comparable de relación mediante la transformación de la covarianza en unidades de relación al dividirla entre el producto de la raíz cuadrada de la varianza de ambas variables, obteniendo de este modo un coeficiente de correlación.

$$\text{Correlación}_{xy} = \frac{s_{xy}}{s_x s_y}$$

Los valores que se obtienen están comprendidos entre el 0 y el ± 1 , y se interpreta como la parte de la variación de las variables X e Y que lo hacen de modo conjunto. Ahora sustituimos los valores de $s_{xy} = 1,385$, $s_x = 3,237$ y $s_y = 1,28$:

$$\text{Correlación}_{xy} = \frac{1,385}{3,237 \times 1,28} = \frac{1,385}{4,143} = 0,33$$

El resultado de la correlación nos indica que se trata de una correlación baja. Los datos recogidos en las tablas ya nos indicaban este posible resultado. Volvamos sobre ellos dos páginas hacia atrás; los datos están

ordenados por la variable actitud, y sus correspondientes valores en la variable matemáticas no siguen un orden uniforme creciente.

14.4. El diagrama de dispersión

Si queremos representar gráficamente los valores de una tabla de correlaciones de dos series de datos de pares ordenados mediante los ejes de coordenadas, situaremos en el eje de abscisas una variable, la X , y en el de ordenadas la otra variable, la Y , lo que nos permite ver la posición que muestran los pares de valores, situándose en torno a una nube de puntos más o menos dispersos, los cuales reflejan el grado de relación entre ellos.

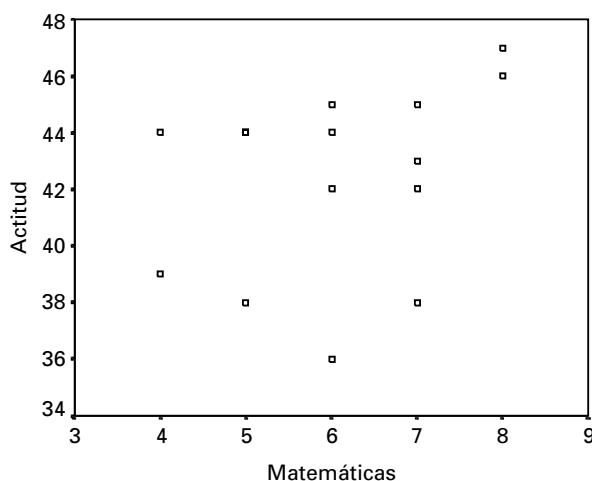


FIGURA 14.3. Diagrama de dispersión.

En este diagrama podemos observar que la relación de las variables es baja, dado que los puntos no se concentran en torno a una línea, y también podemos decir que es positiva, puesto que se observa una cierta dirección ascendente en las dos variables.

Veamos el diagrama obtenido a partir de la tabla de datos que presentamos a continuación referida a las calificaciones obtenidas por un grupo de alumnos en sendas pruebas de rendimiento de Lenguaje (variable X) y Matemáticas (variable Y):

TABLA 14.8. Calificación de los alumnos

Alumno	Lenguaje (X)	Matemáticas (Y)
1	3.00	3.50
2	3.50	5.00
3	4.30	4.70
4	4.50	5.50
5	4.60	6.00
6	5.40	7.20
7	6.00	4.75
8	6.50	4.50
9	7.00	6.50
10	7.35	7.30
11	7.80	5.50
12	8.00	5.50
13	8.20	9.00
14	8.50	8.00

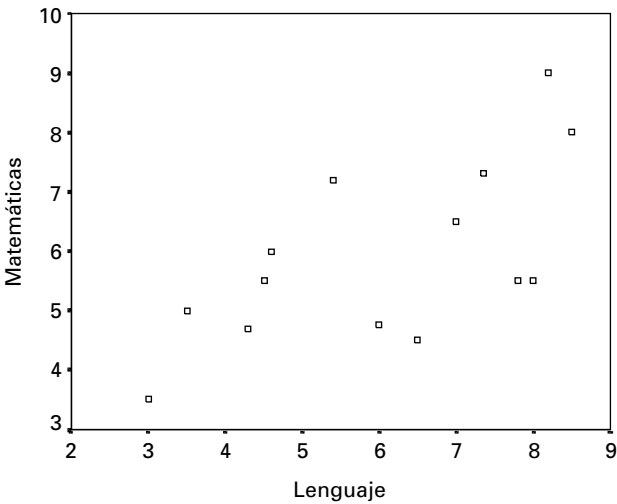


FIGURA 14.4. Diagrama de dispersión.

Se trata de un diagrama de dispersión obtenido con el programa informático de análisis estadístico de datos SPSS. Los puntos en el diagrama representan los pares de observaciones en ambas variables. Podemos observar que se confirma el valor aportado por el coeficiente de covariación, que nos indicaba que la relación era directa, ambas variables varían en la misma dirección, y relativamente intenso el grado de relación dado que se observa una cierta concentración en torno a una línea imaginable llamada recta de regresión.

14.5. Las medidas correlacionales bivariadas y su interpretación

La correlación es una medida del grado de relación entre dos variables. Las relaciones entre dos variables, como hemos observado en las tablas de correlación, son conjuntos de pares ordenados. Los coeficientes de correlación aportan información sobre la intensidad y la dirección con que dos variables están relacionadas y permiten establecer comparaciones entre diversos casos.

Los coeficientes de correlación se basan en la variación concomitante de los miembros del conjunto de pares ordenados. Si covarían, es decir, varían juntos y en forma correspondiente, por ejemplo valores altos con valores altos, valores bajos con valores bajos se dice que hay correlación positiva. Si covarían valores altos con valores bajos, entonces se dice que hay correlación negativa. La correlación bivariada, por tanto, expresa el grado y sentido de la variación concomitante de dos series de datos.

Veamos el siguiente ejemplo, adaptado a partir del de Kerlinger (1987: 45):

TABLA 14.9. Relaciones

Serie I		Serie II		Serie III		Serie IV		Serie V	
X	Y	X	Y	X	Y	X	Y	X	Y
1	1	1	2	1	9	1	7	1	1
2	2	2	4	2	8	2	8	2	2
3	3	3	6	3	7	3	4	3	3
4	4	4	8	4	6	4	5	4	4
5	5	5	10	5	5	5	6	5	5
6	6	6	12	6	4	6	9	6	9
7	7	7	14	7	3	7	1	7	8
8	8	8	16	8	2	8	3	8	7
9	9	9	18	9	1	9	2	9	6
$r = 1,00$ Relación directa perfecta positiva		$r = 1,00$ Relación directa perfecta positiva		$r = -1$ Relación inversa perfecta negativa		Relación imperfecta		Relación imperfecta	

En I y II es posible predecir con toda exactitud de X a Y , pero en III no es posible calcular los valores de Y , por que entre ellas se da una función o regla de correspondencia definida. En este ejemplo sencillo, fácilmente

podemos predecir que al siguiente valor de la variable X , el 10, en la serie I sería el 10, en la serie II sería el 20, en la serie III sería el 0, y en la serie IV, de modo intuitivo no se puede indicar. La serie V puede observarse que la relación entre las dos variables va en una dirección hasta que llegado un punto se invierte.

Dependiendo del valor que alcance la correlación, tenemos que existe una máxima correlación cuando alcanza el valor de 1. No existe ninguna correlación cuando el valor obtenido es $= 0$. Si el valor es $+1$ es una correlación perfecta, directa o positiva, si es positiva, dado que la variación de las series lo hacen en el mismo sentido, o perfecta negativa si su valor es -1 , dado que la variación de las series lo hace en sentido inverso, a medida que una aumenta disminuye la otra. Si el valor está entre el 0 y el ± 1 , la correlación es imperfecta, positiva o negativa. En síntesis podemos verlo del siguiente modo:

Correlación perfecta: $r = \pm 1$ (positiva o negativa)

Correlación imperfecta: $0 < r < 1$ positiva, y $-1 < r < 0$ negativa

Correlación nula: $r = 0$, o ausencia de correlación.

Veamos los valores de las correlaciones existentes entre las distintas series y sus representaciones gráficas mediante diagramas de dispersión:

Series I y II:

TABLA 14.10. Correlación Series I y II

Correlaciones Serie I y Serie II			
		VAR0001	VAR0002
VAR0001	Correlación de Pearson	1	1.000(**)
	N	9	9
VAR0002	Correlación de Pearson	1.000(**)	1
	N	9	9
** La correlación es significativa al nivel 0,01 (bilateral).			

La correlación arroja un valor $r = 1,00$, lo que supone una relación directa perfecta positiva. Las correlaciones de las dos primeras series son iguales, y su diagrama de dispersión también. Es significativa al nivel de significación del 0,01 (hipótesis bilateral).

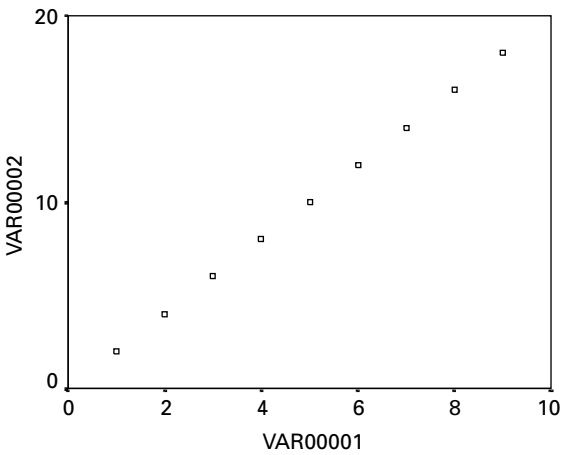


FIGURA 14.5. Diagrama de dispersión. Series I y II.

El diagrama muestra la intensidad de la relación entre las dos variables, en torno a una línea, o mejor en este caso, sobre la línea, y la dirección ascendente, positiva.

Serie III:

TABLA 14.11. Correlación Serie III

Correlación Serie III			
		VAR00001	VAR00003
VAR00001	Correlación de Pearson	1	-1.000(**)
	N	9	9
VAR00003	Correlación de Pearson	-1.000(**)	1
	N	9	9
** La correlación es significativa al nivel 0,01 (bilateral).			

La correlación arroja un valor $r = -1,00$, lo que supone una relación perfecta inversa o negativa. Es significativa al nivel de significación del 0,01 (hipótesis bilateral).

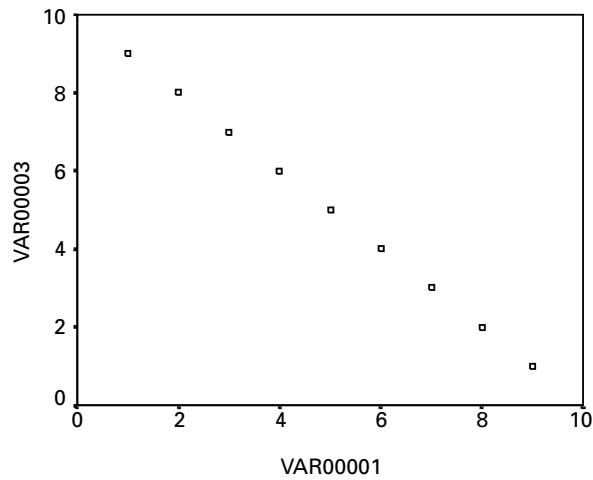


FIGURA 14.6. Diagrama de Dispersión. Serie III.

El diagrama muestra la misma intensidad de relación entre las dos variables que en el caso anterior, pero la dirección de los valores de ambas variables es contraria, cuando una aumenta la otra decrece.

Serie IV:

TABLA 14.12. Correlación Serie IV

Correlación Serie IV			
		VAR00001	VAR00004
VAR00001	Correlación de Pearson	1	-.617
	Sig. (bilateral)	.	.077
	N	9	9
VAR00004	Correlación de Pearson	-.617	1
	Sig. (bilateral)	.077	.
	N	9	9

La correlación arroja un valor $r = -0,617$, lo que supone una relación imperfecta y negativa, o inversa. No es significativa $p = 0,077 > 0,05$.

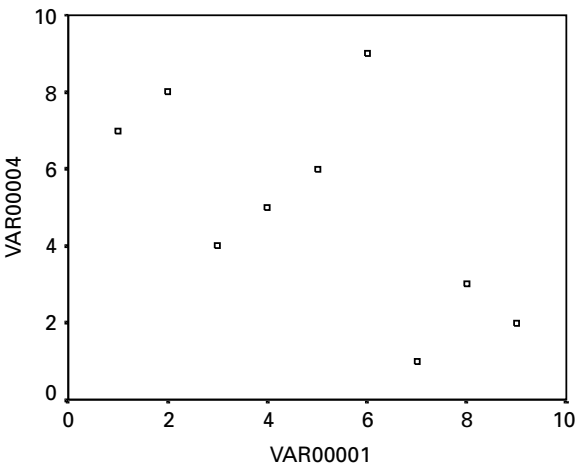


FIGURA 14.7. Diagrama de dispersión. Serie IV.

El diagrama muestra una relación inversa entre las dos variables. Los puntos tienden a concentrarse en torno a una línea; muestran la dirección inversa, confirmado por el valor de r obtenido, que es negativo.

Serie V:

TABLA 14.13. Correlación Serie V

Correlación Serie V			
		VAR00001	VAR00005
VAR00001	Correlación de Pearson	1	-.833(**)
	Sig. (bilateral)	.	.005
	N	9	9
VAR00005	Correlación de Pearson	-.833(**)	1
	Sig. (bilateral)	.005	.
	N	9	9

La correlación obtiene un valor elevado negativo, lo que indica una correlación inversa. El diagrama de dispersión muestra el punto de inflexión que adelantaban los datos numéricos. Es significativa ($0,005 < 0,05$).

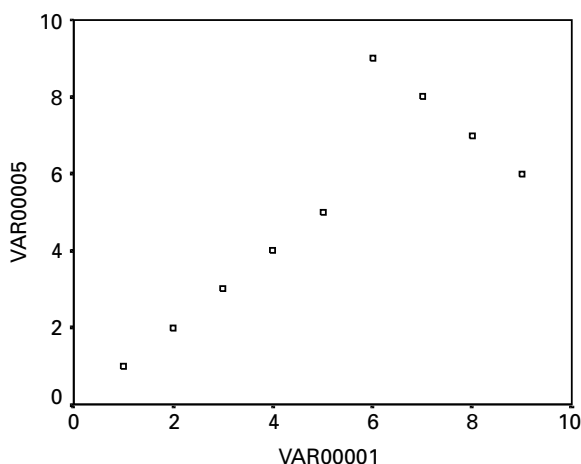


FIGURA 14.8. Diagrama de dispersión. Serie V.

El diagrama muestra una pauta de relación entre las dos variables diferente a las vistas hasta ahora. Los puntos dibujan dos líneas.

Podría darse el caso de que la relación pareciera baja si se calcula mediante un coeficiente de correlación lineal de Pearson, y sin embargo se tratara de una correlación curvilínea. Cuando se observa que la relación puede ser curvilínea, se recomienda la utilización del coeficiente «eta» o razón de correlaciones, la cual refleja la varianza estimada mediante la línea de ajuste óptima, ya sea curva o recta.

Dependiendo de que las observaciones hayan sido medidas en escalas de intervalos, ordinales o categoriales, el coeficiente utilizado para calcular el grado de relación existente entre ambas variables será distinto. Así utilizaremos el coeficiente de correlación de Pearson para las variables que alcanzan nivel de medida de intervalo, el coeficiente de Spearman para el nivel de medida ordinal, y el de «chi cuadrado» y contingencia para el nivel de medida categorial.

14.5.1. El coeficiente de correlación producto momento de Pearson

La correlación de Pearson, o correlación producto-momento, entre dos variables se define como el cociente entre la covarianza entre las variables X e Y , dividida entre el producto de las desviaciones típicas de cada una de las variables:

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

Se lee correlación de la variable X con la variable Y . Esta correlación se utiliza con datos medidos en escala de medida de intervalo. La ecuación de la correlación de Pearson, r_{xy} , para puntuaciones directas, es la siguiente:

$$r_{xy} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2] \times [n \sum Y^2 - (\sum Y)^2]}};$$

Este nuevo índice de correlación sí nos permite realizar comparaciones entre el grado de relación existente entre distintos casos y/o grupos de sujetos, y conocer la intensidad de la relación.

Veamos un ejemplo sencillo: ¿Cuál será el grado de relación entre las dos series de datos siguiente, referida al aumento de volumen que alcanza un globo a medida que se aumente la temperatura? (Los datos son inventados)

TABLA 14.14. Tabla de correlación

Serie A	Serie B
0	12
1	19
2	23
3	34
4	56
5	62

La serie A representa la variable X_i o valores de las temperaturas, y la serie B representa la variable Y_i , o serie de datos del volumen. El objetivo del ejemplo pretende mostrar la relación entre dos variables medidas a un nivel de intervalo, preciso, por lo que el coeficiente de correlación a utilizar es el producto momento de Pearson. Dado que los datos vienen dados en puntuaciones directas, utilizaremos la ecuación de la r de Pearson para este tipo de puntuaciones.

$$r_{xy} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2] \times [n \sum Y^2 - (\sum Y)^2]}};$$

Y comprobamos que debemos calcular las sumas de cada una de las variables, la suma del producto de ambas variables, y la suma de los cuadrados de cada una de las variables. Con una calculadora manual procedemos a hallar los datos que necesitamos y que mostramos en la siguiente tabla:

TABLA 14.15. Tabla para el cálculo de la Correlación de Pearson

X_i	Y_i	$X_i Y_i$	X_i^2	Y_i^2
0	12	0	0	144
1	19	19	1	361
2	23	46	4	529
3	34	102	9	1156
4	56	224	16	3136
5	62	310	25	3844
$\Sigma = 15$	$\Sigma = 206$	$\Sigma = 701$	$\Sigma = 55$	$\Sigma = 9170$

Y en la fórmula, sustituimos los valores calculados:

$$\begin{aligned}
 r_{xy} &= \frac{6 \times 701 - (15 \times 206)}{\sqrt{[6 \times 55 - (15)^2] \times [6 \times 9170 - (206)^2]}} = \frac{4206 - 3090}{\sqrt{(330 - 225) \times (55020 - 42436)}} = \\
 &= \frac{1116}{\sqrt{105 \times 12584}} = \frac{1116}{\sqrt{132320}} = \frac{1116}{1149,4868} = 0,97
 \end{aligned}$$

El valor obtenido, $r_{xy} = 0,97$, muy próximo a 1, nos indica que existe una correlación muy elevada entre ambas series, casi perfecta.

14.5.2. Interpretación de la correlación producto momento de Pearson

Para interpretar con precisión el valor del coeficiente del correlación obtenido, se comprueba si es significativo a un nivel de confianza aceptable científicamente, por ejemplo, al nivel de confianza del 95% o al más exigente del 99%, (o sus correspondientes niveles de significación $\alpha = 0,05$, y $\alpha = 0,01$, respectivamente). Estos conceptos se verán en los temas 15 y siguientes, dedicados a la inferencia estadística. Se trata de comprobar si el valor empíricamente obtenido ha podido deberse al azar o es fruto de la existencia de relación significativa al nivel de confianza fijado y para el tamaño de nuestra muestra. Para ello se consulta la tabla de probabilidades de la distribución correspondiente.

La hipótesis nula afirma que el valor del parámetro es 0, o lo que es lo mismo:

$$H_0: \rho = 0$$

Debemos utilizar la fórmula del estadístico t , en la significatividad de r :

$$t = \frac{r_{xy}}{\sqrt{\frac{1-r_{xy}^2}{n-2}}} = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$$

Y sustituyendo los valores de nuestro ejemplo en la fórmula

$$t = \frac{0,97\sqrt{15-2}}{\sqrt{1-0,97^2}} = \frac{0,97\sqrt{13}}{\sqrt{0,591}} = \frac{3,497}{0,243} = 14,39$$

El estadístico t sigue la distribución de la t de Student con $n - 2$ grados de libertad. El valor crítico de t , para una hipótesis bilateral, al nivel de significación de $\alpha = 0,05$ es $t_c = 2,160$, y g.l. = $15 - 2 = 13$. Y al nivel de significación de $\alpha = 0,01$ es $t_c = 3,012$. (Véase en el Apéndice de Tablas Estadísticas.)

$$t_e = 14,39 > t_c = 2,160$$

Como el valor empírico de t es mayor que el valor crítico, podemos rechazar la hipótesis de nulidad, pudiendo afirmar que entre ambas variables se da una relación significativa al nivel de significación de 0,05. También lo es al nivel más exigente del 0,01. Existe una tabla de significación de r , especialmente útil para muestras pequeñas.

14.5.3. Cálculo de la correlación producto momento de Pearson mediante el programa informático SPSS

Ahora mostramos el procedimiento seguido para realizar el análisis de los datos del ejemplo mediante el programa informático SPSS. Para ello hemos introducido los datos de la primera variable en el fichero de datos, en su correspondiente columna denominada, por defecto, variable 1, y los datos de la segunda variable en la segunda columna, denominada variable 2. Después hemos marcado en la barra de herramientas «analizar», y en el menú desplegado, correlaciones, de Pearson, y opcional, estadísticos descriptivos y covarianzas, después marcamos pegar, y aceptamos. De este modo obtenemos el fichero de sintaxis, y marcando la información de la que queremos ver sus resultados, marcamos en la barra de herramientas del fichero de sintaxis «ejecutar», que viene representado por una punta de flecha, y nos da un nuevo fichero con los resultados. Si en la

barra de herramientas del menú resultados marcamos el icono de importar, nos guarda las tablas de resultados en un fichero html, en la carpeta que decidamos, que podremos insertar y manipular en nuestro archivo del informe en el procesador de texto de word.

A continuación mostramos las tablas de los estadísticos descriptivos de medias, desviación típica, y las tablas de correlaciones, con las sumas de cuadrados, covarianzas y correlaciones, importado del fichero html creado a partir del fichero de resultados obtenido mediante el programa SPSS, que podemos manipular en nuestro fichero de word.

TABLA 14.16. Descriptivos

Estadísticos descriptivos			
	Media	Desviación típica	N
VAR X	2.5000	1.87083	6
VAR Y	34.3333	20.48089	6

TABLA 14.17. Correlaciones y otros

Correlaciones, Suma de cuadrados y productos cruzados y Covarianzas			
		VAR X	VAR Y
VAR X	Correlación de Pearson	1	.971(**)
	Sig. (bilateral)	.	.001
	Suma de cuadrados y productos cruzados	17.500	186.000
	Covarianza	3.500	37.200
	N	6	6
VAR Y	Correlación de Pearson	.971(**)	1
	Sig. (bilateral)	.001	.
	Suma de cuadrados y productos cruzados	186.000	2097.333
	Covarianza	37.200	419.467
	N	6	6
** La correlación es significativa al nivel 0,01 (bilateral).			

Pudiendo comprobar que se obtiene el mismo valor de la correlación mediante ambos procedimientos. En la tabla de resultados nos indica que entre ambas series de datos existe una correlación muy elevada, $r = 0,971$, significativa al nivel 0,01, el mismo nivel de significación que comprobamos al consultar las tablas de la distribución de t siguiendo la fórmula del estadístico t para la significación de la correlación.

14.5.4. El coeficiente de correlación ordinal de Spearman

Cuando se trabaja con variables medidas con un nivel ordinal, significa que las puntuaciones obtenidas mediante un cuestionario, o una escala de observación, conforman una sucesión de valores ordenados de mayor a menor.

Su fórmula es la siguiente:

$$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)};$$

Siendo:

D : la diferencia de rango entre la variable X y la variable Y ,

n : número de casos.

Veamos un ejemplo:

Se trata de conocer si existe alguna relación entre las variables «claridad expositiva» de un profesor de Matemáticas, que la nombramos como X_i , y su «capacidad para motivar a los alumnos», que la nombramos como Y_i , valoradas por 12 alumnos, mediante sendos cuestionarios de 12 ítems valorados entre 5, máxima valoración y 1, mínima valoración. La puntuación máxima posible es $12 \times 5 = 60$. Los resultados en ambas variables se reflejan, ordenados por la variable X , de mayor a menor puntuación, en la siguiente tabla:

TABLA 14.18. Tabla de correlación

Alumno	X_i	Y_i
1	50	43
2	47	50
3	45	46
4	40	40
5	40	40
6	35	32
7	32	37
8	27	35
9	25	28
10	25	30
11	22	22
12	19	26

Por tratarse de una escala que alcanza un nivel de medida ordinal, utilizaremos el coeficiente de correlación de Spearman, debiendo ordenar

primero las puntuaciones en rangos para luego hallar su diferencia D y elevarla al cuadrado.

$$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)};$$

Procedemos haciendo los cálculos para sustituir las incógnitas que demanda la fórmula:

TABLA 14.19. Tabla para el cálculo de la red de Spearman

Alumno	X_i	Y_i	Rango X_i	Rango Y_i	D	D^2
1	50	43	1	3	-2	4
2	47	50	2	1	1	1
3	45	46	3	2	1	1
4	40	40	4,5	4,5	0	0
5	40	40	4,5	4,5	0	0
6	35	32	6	8	-2	4
7	32	37	7	6	1	1
8	27	35	8	7	+1	1
9	25	28	9.5	10	-0,5	0,25
10	25	30	9.5	9	0,5	0,25
11	22	22	11	12	-1	1
12	19	26	12	11	+1	1
						$\Sigma D^2 = 14,5$

Sustituyendo los datos calculados en la fórmula, tenemos:

$$r_s = 1 - \frac{6 \times 14,5}{12(12^2 - 1)} = 0,95$$

El valor del coeficiente de correlación obtenido $r_s = 0,95$ es elevado, próximo a 1, lo que indica la existencia de relación muy alta entre las variables claridad expositiva y capacidad motivacional del profesor.

14.5.5. Interpretación del coeficiente de correlación ordinal de Spearman

Como decíamos al interpretar el valor obtenido de r en la correlación de Pearson, científicamente se puede asegurar que se da una correlación significativa, si al nivel de significación aceptable de 0,05 ó 0,01, se puede rechazar la hipótesis de nulidad.

Para ello, en el caso de muestras pequeñas podemos consultar las tablas de la correlación de Spearman (ver apéndice), teniendo en cuenta el tamaño de la muestra y el nivel de significación fijado. Con muestras mayores de 8, podemos utilizar el estadístico de t , dado que a partir de este tamaño la distribución de r_s se distribuye como la distribución de t . Tomando la fórmula de t para interpretar la significación de la correlación:

$$t = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}}$$

Si tras aplicar la fórmula obtenemos un valor empírico de t igual o mayor que el valor crítico de t , que vemos en sus tablas, con $gl = n - 2$, y $\alpha = 0,05$, o el nivel más exigente de $\alpha = 0,01$, podemos rechazar la hipótesis nula y afirmar que la correlación entre las dos variables es significativa al nivel de significación fijado. Interpretemos, por tanto, el valor de $r_s = 0,95$, obtenido, con $g.l. = 12 - 2 = 10$ y $\alpha = 0,05$ y $\alpha = 0,01$. Sustituyendo los valores de nuestro ejemplo en la fórmula del estadístico tenemos:

$$t = \frac{0,95\sqrt{12-2}}{\sqrt{1-0,95^2}} = \frac{0,95\sqrt{10}}{\sqrt{0,0975}} = \frac{3,004}{0,3122} = 9,62$$

Consultando las tablas de la distribución de t , con 10 grados de libertad y un nivel de significación de $\alpha = 0,05$, hipótesis bilateral, el valor crítico de t es 2,228. Como el valor empírico obtenido es $9,62 > 2,228$, podemos rechazar la hipótesis nula y afirmar que se dan relaciones entre las dos variables, a un nivel de significación del 0,05. Podemos comprobar si es también significativo al nivel más exigente de $\alpha = 0,01$, vemos que el valor de t crítico, con $g.l. = 10$ es 3,169. Como nuestro valor de t empírico $9,62 > 3,169$, podemos rechazar la hipótesis de nulidad a este nivel de significación.

14.5.6. Cálculo de la correlación ordinal de Spearman mediante el programa informático SPSS

Mediante el programa de análisis estadístico de datos SPSS, obtenemos una «rho» de Spearman con valor prácticamente igual al hallado con calculadora, y nos indica que es un valor significativo al nivel de confianza del 0,01, muy elevado:

TABLA 14.20.

Correlaciones				
			VAR X	VAR Y
Rho de Spearman	VAR X	Coeficiente de correlación	1.000	.949(**)
		Sig. (bilateral)	.	.000
		N	12	12
	VAR Y	Coeficiente de correlación	.949(**)	1.000
		Sig. (bilateral)	.000	.
		N	12	12
** La correlación es significativa al nivel 0,01 (bilateral).				

El valor del coeficiente obtenido es coincidente con el obtenido mediante la fórmula de la correlación ordinal de Spearman. Y lo mismo podemos decir de la interpretación científica de su significación. Es coincidente con el nivel máximo del 0,01 obtenido mediante la fórmula del estadístico *t*.

14.5.7. *El coeficiente de contingencia o de asociación C de Pearson*

El coeficiente de contingencia o asociación *C* de Pearson se utiliza para ver el grado de asociación existente entre variables medidas a nivel nominal o categorial. El coeficiente *C* viene definido por la siguiente fórmula:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

El coeficiente alcanza valores comprendidos entre 0 y 1, y se valora como el resto de coeficientes, el 0 representa ausencia de relación entre las dos variables, y el 1 indica total asociación o relación de las variables, aunque el valor 1 sólo se alcanza en el caso ideal de infinitas modalidades. En el caso de una tabla de contingencia cuadrada, siendo el número de filas igual al número de columnas, el valor máximo de *C* viene dado por la fórmula

$$S = \sqrt{\frac{h-1}{h}}$$

Siendo h el número de filas, y en este caso igual al número de columnas. Para resolver el coeficiente de contingencia necesitamos conocer el valor de «ji», χ^2 , cuya fórmula es:

$$\chi^2 = \sum_{i=1}^c \frac{(f_o - f_e)^2}{f_e}$$

Siendo:

f_o = frecuencias observadas

f_e = frecuencias esperadas

Necesitando hallar el valor de las frecuencias esperadas

$$f_e = \frac{f_f \cdot f_c}{f_t}$$

Siendo:

f_f = frecuencias de las filas

f_c = frecuencias de las columnas

f_t = frecuencias totales

Veamos un ejemplo: Queremos determinar el grado de asociación, o relación que existe entre el nivel de estudios (Primaria y Secundaria) y el nivel de integración alcanzado por los alumnos en el Centro Educativo (Baja, Media, Alta), de acuerdo con los datos siguientes:

TABLA 14.21. Tabla de contingencias

Integración (categorías) Nivel de estudios (variable)			
	Baja	Media	Alta
Primaria	30	40	50
Secundaria	20	40	60

La fórmula del coeficiente de asociación C es la siguiente:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

Lo primero que debemos hacer para resolver el problema es calcular los totales de las filas y de las columnas, y el total. El total de cada columna multiplicado por el total de cada fila, dividido entre la suma

total de observaciones, nos da el valor de las frecuencias esperadas (teóricas posibles) de cada cruce (de filas y columnas).

Las frecuencias esperadas (f_e), de los alumnos con nivel de estudios de primaria y baja integración se calculan del siguiente modo:

$$f_{e(a,c)} = \frac{f_f \cdot f_c}{f_t} = \frac{120 \times 50}{240} = 25;$$

Las frecuencias esperadas teóricas (f_e), que quedan reflejadas en la tabla siguiente entre paréntesis, (las letras a,b,c,d,e, han sido incluidas en este caso para orientar la situación de las distintas frecuencias teóricas esperadas):

TABLA 14.22. Tabla para el cálculo de «chi» cuadrado

Integración				
Nivel de estudios	Baja (c)	Media (d)	Alta (e)	Total
Primaria (a)	30 (25) (a,c)	40 (40) (a,d)	50 (55) (a,e)	120
Secundaria (b)	20 (25) (b,c)	40 (40) (b,d)	60 (55) (b,e)	120
Total	50	80	110	240

Con los valores de frecuencias esperadas hallamos el valor de χ^2 :

$$\begin{aligned} \chi^2 = \sum_{i=1}^c \frac{(f_o - f_e)^2}{f_e} &= \frac{(30-25)^2}{25} + \frac{(40-40)^2}{40} + \frac{(50-55)^2}{55} + \frac{(20-25)^2}{25} + \\ &+ \frac{(40-40)^2}{40} + \frac{(60-55)^2}{55} = 1 + 0 + 0,45 + 1 + 0 + 0,45 = 2,90 \end{aligned}$$

Y sustituimos su valor en la fórmula del coeficiente de contingencia:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} = \sqrt{\frac{2,9}{240 + 2,9}} = \sqrt{\frac{2,9}{242,9}} = \sqrt{0,011939} = 0,109$$

El valor máximo de C no se puede hallar por no ser una tabla de contingencia cuadrada, es decir, no tener el mismo número de filas y columnas.

14.5.8. Interpretación del coeficiente de asociación C de Pearson

Los coeficientes de correlación obtenidos a partir de tablas de contingencia, por no alcanzar el nivel de medida de intervalo ni ordinal, y la

distribución de sus datos no sigue el modelo de la curva normal, especialmente cuando el n es pequeño, son interpretados de modo diferente a los que hemos visto hasta ahora, que eran medidos a nivel de intervalo, la correlación de Pearson, y ordinal, la correlación de Spearman.

Así, el coeficiente de Contingencia, para un nivel de medida nominal o categorial, que utiliza para su cálculo el estadístico de χ^2 , se apoya en la distribución de este estadístico para interpretar la significación del valor del coeficiente obtenido. Entonces, podemos decir que la significación del coeficiente de Contingencia C se halla probando la significación del valor obtenido en el cálculo previo de χ^2 .

La hipótesis de nulidad dice que no existen diferencias significativas al nivel de significación fijado, aceptable científicamente, como es el de 0,05, o el más exigente del 0,01, entre el valor de χ^2 obtenido empíricamente y el valor crítico de χ^2 que podemos encontrar en la tabla de la distribución de χ^2 , con los grados de libertad $= (f - 1)(c - 1)$ y el nivel de significación fijado.

Si el valor empírico de χ^2 es igual o mayor que el valor crítico, entonces podemos rechazar la H_0 y afirmar que el valor obtenido mediante el coeficiente de asociación C , es significativo al nivel fijado, y que por lo tanto se da asociación entre ambas variables. Pero si el valor crítico de χ^2 fuera mayor que el valor empírico calculado, entonces se confirma la hipótesis de nulidad.

En el ejemplo que hemos puesto en el epígrafe anterior, obteníamos un valor de $\chi^2 = 2,90$, con grados de libertad:

$$\text{g.l.} = (f - 1)(c - 1) = (2 - 1)(3 - 1) = 2.$$

Consultando la Tabla χ^2 , para g.l. = 2, y un $\alpha = 0,05$, nos da un valor de $\chi^2 = 5,991$. Como $2,90 < 5,991$, se acepta la hipótesis de nulidad. El grado de asociación aportado por el coeficiente de contingencia es muy bajo como para poder considerarlo significativo al nivel de confianza del 95 %.

14.6. Ejercicios de autoevaluación

Problema 1

Se desea saber el grado de correlación existente entre la edad, en meses, (variable X) y la estatura, en centímetros, (variable Y) de 18 niños de un centro de educación infantil. Los datos son los siguientes:

X	18	20	20	21	22	23	24	25	27	27	28	30	32	34	36	38	39
Y	75	76	80	87	90	78	93	90	85	94	95	96	99	100	105	109	108

- Elabore la tabla de correlaciones.
- Aplice el coeficiente de correlación adecuado y valore su resultado.
- Represente el diagrama de dispersión correspondiente.

Problema 2

Un grupo de alumnos que ha sido calificado mediante sendas pruebas de rendimiento en las asignaturas de Lenguaje (variable X) y Matemáticas (variable Y), y han obtenido las calificaciones siguientes:

Tabla 14.23. Tabla de correlación

Alumno	X	Y
1	3.00	3.50
2	3.50	5.00
3	4.30	4.70
4	4.50	5.50
5	4.60	6.00
6	5.40	7.20
7	6.00	4.75
8	6.50	4.50
9	7.00	6.50
10	7.35	7.30
11	7.80	5.50
12	8.00	5.50
13	8.20	9.00
14	8.50	8.00

Se le pide que:

- Calcule el coeficiente de covarianza entre las dos series de datos.
- Represente la correlación mediante un diagrama de dispersión. Sitúe sobre el eje de coordenadas, los valores de la variable Y en las ordenadas o eje vertical, y los valores de la variable X en las abscisas o eje horizontal. Señale los puntos de cruce de ambas variables.
- Comente el diagrama.

Problema 3

¿Cuál es la correlación existente entre las dos series de puntuaciones obtenidas por un grupo de 12 alumnos, en sendas pruebas elaboradas por el profesor en Composición escrita (X) y en comprensión lectora (Y). Justifique el coeficiente de correlación elegido.

Tabla 14.24. Tabla de correlación

Alumno	X	Y
1	47	43
2	45	42
3	32	34
4	43	45
5	28	33
6	40	38
7	36	30
8	41	37
9	39	35
10	25	28
11	27	25
12	33	43

Problema 4

A partir de la tabla de datos del problema 2, referida a los alumnos que han aprobado y suspendido en Lenguaje y en Matemáticas, elabore una tabla de contingencia y calcule e interprete el grado de asociación existente entre la variable asignaturas y las categorías suspende y aprueba (calificación) obtenidas por los alumnos.

Problema 5

Las calificaciones obtenidas por un grupo de 16 alumnos de 3.º de Secundaria, en las pruebas de Matemáticas y Física (valoración de 10 a 1), han sido los siguientes:

Calificaciones en Matemáticas: 4, 3, 6, 8, 7, 5, 4, 6, 5, 9, 7, 8, 4, 5, 4, 8.

Calificaciones en Física: 4, 6, 8, 5, 4, 8, 7, 5, 7, 6, 8, 9, 7, 8, 9, 6.

a) Elabore la tabla de correlaciones.

b) Calcule la media y la desviación típica.

- c) Calcule el coeficiente de variación.
- d) Calcule el coeficiente de correlación. Justifique la elección del coeficiente.
- e) Realice el diagrama de dispersión.

14.7. Soluciones a los ejercicios de autoevaluación

Problema 1

Observe que para confeccionar la tabla de correlaciones lo primero que hay que hacer es ordenar las puntuaciones de los sujetos en la variable, en este caso la X , referida a edad, situando en la columna de la variable Y , referida a estatura, el valor correspondiente a cada sujeto en esta segunda variable. Si los pares de puntuaciones de cada sujeto no están bien ordenados, los resultados lo único que hacen es confundir.

El coeficiente elegido ha debido ser el de Pearson, por venir los datos medidos con un nivel de intervalo.

Problema 2

Un grupo de alumnos que ha sido calificado mediante sendas pruebas de rendimiento en las asignaturas de Lenguaje (variable X) y Matemáticas (variable Y), han obtenido los valores que se observan en la siguiente tabla:

Tabla 14.25. Tabla de correlación

Alumno	Lenguaje (X)	Matemáticas (Y)
1	3.00	3.50
2	3.50	5.00
3	4.30	4.70
4	4.50	5.50
5	4.60	6.00
6	5.40	7.20
7	6.00	4.75
8	6.50	4.50
9	7.00	6.50
10	7.35	7.30
11	7.80	5.50
12	8.00	5.50
13	8.20	9.00
14	8.50	8.00

Utilizando la tabla de datos de las calificaciones obtenidas por los alumnos, vamos a calcular la covarianza entre ambas variables. Para ello, primero tendremos que calcular sus medias aritméticas:

$$\bar{X} = \frac{\sum X_i}{n}; \quad \bar{X} = \frac{84.65}{14} = 6.046; \quad \bar{Y} = \frac{82.95}{14} = 5.925$$

Teniendo el valor de la media aritmética calculamos las puntuaciones diferenciales con relación a sus medias y después el producto de cada par de puntuaciones diferenciales, lo que queda reflejado en la siguiente tabla:

Tabla 14.26. Tabla para el cálculo de la covarianza

Alumnos	X	Y	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$
1	3.00	3.50	-3.05	-2.43	11.40
2	3.50	5.00	-2.55	-.93	2.37
3	4.30	4.70	-1.75	-1.23	2.15
4	4.50	5.50	-1.55	-.43	.67
5	4.60	6.00	-1.45	.07	-.10
6	5.40	11.20	-.65	1.27	-.83
7	6.00	4.75	-.05	-1.18	.06
8	6.50	4.50	.45	-1.43	-.64
9	11.00	6.50	.95	.57	.54
10	11.35	11.30	1.30	1.37	1.78
11	11.80	5.50	1.75	-.43	-.75
12	8.00	5.50	1.95	-.43	-.84
13	8.20	9.00	2.15	3.07	6.60
14	8.50	8.00	2.45	2.07	5.07
	Media = 6.05	Media = 5.93	-0.05	-0.07	23.49

Ya tenemos todos los datos para sustituir en las incógnitas en la fórmula de la covarianza, del siguiente modo:

$$s_{xy} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n} = \frac{23.49}{14} = 1.68$$

El valor obtenido en el coeficiente de covariación de 1,68, nos indica que existe covariación entre ambas variables, las calificaciones obtenidas por un grupo de sujetos en Lenguaje y en Matemáticas, y al ser su signo positivo nos indica que covarían en la misma dirección. Existe, por tanto, una tendencia de los alumnos a obtener notas con incremento en la misma dirección. Pero para conocer la correlación existente entre ambos pares de variables de este grupo de alumnos, necesitamos calcular el índice de correlación.

Problema 3

¿Cuál es la correlación existente entre las dos series de puntuaciones obtenidas por un grupo de 12 alumnos, en sendas pruebas elaboradas por el profesor en Composición escrita (X) y en comprensión lectora (Y)? Justifique el coeficiente de correlación elegido. Consideramos que por tratarse de pruebas elaboradas por el profesor sobre composición y expresión escrita, no alcanzan el nivel de medición de intervalo, eligiendo el coeficiente de correlación de rangos de Spearman para datos ordinales.

Lo primero que debemos hacer es ordenar las puntuaciones por rangos, para lo cual se ordenan las puntuaciones en orden descendente, de mayor a menor, se hacen los rangos para cada variable, se hallan las diferencias de rango y se elevan al cuadrado:

Tabla 14.27. Tabla para el cálculo del coeficiente r de Spearman

Alumnos	X	Rango X	Y	Rango Y	D	D ²
1	47	1	43	2	-1	1
2	45	2	42	3	-1	1
3	32	9	34	7	2	4
4	43	3	45	1	2	4
5	28	10	33	8	2	4
6	40	5	38	4	1	1
7	36	7	30	9.5	-2.5	6.25
8	41	4	37	5	-1	1
9	39	6	35	6	0	0
10	25	12	28	11	1	1
11	27	11	25	12	-1	1
12	33	8	30	9.5	-1.5	2.25
						Σ = 26,5

Y ahora sustituimos los datos en la fórmula:

$$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 26,5}{12(12^2 - 1)} = 1 - \frac{159}{1716} = 1 - 0,092 = 0,908$$

El valor obtenido de $r = 0,908$, casi 0,91, está próximo a 1, lo que significa una correlación elevada que nos indica que el interés por la lectura está asociado a un alto nivel en comprensión lectora, o viceversa.

Problema 4

Estudie la información de la tabla 14,25 del problema 2, referida a los alumnos que han aprobado y suspendido en Lenguaje y en Matemáticas. Elabore una tabla de contingencia y calcule e interprete el grado de asociación existente entre las asignaturas y la calificación obtenida por los alumnos.

En este caso tendríamos que hacer una tabla de contingencia en la que se reflejarían las frecuencias observadas en las dos variables, en este caso ambas categóricas: aprobado, suspenso, y Lenguaje, Matemáticas. La tabla queda del modo siguiente:

Tabla 14.28. Tabla de Contingencia

	Lenguaje	Matemáticas	Totales
Suspensos	///// (5)	//// (4)	9
Aprobados	////////// (9)	////////// (10)	19
Totales	14	14	28

Si nos fijamos en las filas, podemos observar que son 5 los sujetos que suspenden Lenguaje, 4 los que suspenden en Matemáticas, lo cual da un total de 9 suspensos; por otro lado, son 9 los que aprueban en Lenguaje y 10 en Matemáticas, dando un total de 19 sujetos los que aprueban en ambas asignaturas. Y si nos fijamos en las columnas, vemos que en Lenguaje suspenden 5 y aprueban 9, lo que nos da el total de presentados, 14. Y en Matemáticas, suspenden 4 y aprueban 10, dándonos el mismo total puesto que se trata del mismo grupo de alumnos, 14. El coeficiente aplicado a estos datos nos daría el grado de asociación entre los suspensos y aprobados en Lenguaje y Matemáticas.

Problema 5

Las calificaciones obtenidas por un grupo de 16 alumnos de 3.º de Secundaria, en las pruebas de Matemáticas y Física (valoración de 10 a 1), han sido los siguientes:

Calificaciones en Matemáticas: 4, 3, 6, 8, 7, 5, 4, 6, 5, 9, 7, 8, 4, 5, 4, 8.

Calificaciones en Física: 4, 6, 8, 5, 7, 5, 7, 6, 8, 9, 7, 8, 9, 6, 4, 8.

a) Elaboración de la tabla de correlaciones:

Ordenados los datos de menor a mayor en las calificaciones obtenidas por los alumnos en Matemáticas, y en la columna siguiente ponemos la calificación obtenida por el mismo sujeto en Física. Mediante el programa informático SPSS, en el ordenador, seleccionamos en la barra de herramientas datos -> ordenar casos -> Variable 1 y 2 -> ordenar ascendente, y lo hace respetando el orden de los pares de puntuaciones para cada caso o sujeto (el sujeto 1, en Matemáticas 3 y en Física 6, ... el 16 , 9 y 9).

Tabla 14.29. Tabla de correlaciones

Sujetos	Matemáticas X	Física Y
1	3.00	6.00
2	4.00	4.00
3	4.00	4.00
4	4.00	7.00
5	5.00	5.00
6	5.00	6.00
7	5.00	8.00
8	6.00	6.00
9	6.00	8.00
10	7.00	7.00
11	7.00	7.00
12	8.00	5.00
13	8.00	8.00
14	8.00	8.00
15	9.00	9.00
16	9.00	9.00

En la barra de herramientas seleccionamos: Analizar > correlaciones > bivariadas, y en el cuadro de diálogo marcamos Pearson, y opciones, media, desviación típica y productos cruzados, continuar, pegar, lo que nos da el archivo de sintaxis, y seleccionamos el texto, y marcamos ejecutar (la punta de flecha) lo que nos da un archivo de resultados, marcamos el icono de Exportar, y le indicamos a donde. Luego, desde

nuestro archivo de trabajo de word, lo insertamos, teniendo en el texto las tablas de resultados siguientes:

Tabla 14.30. Descriptivos

Estadísticos descriptivos			
	Media	Desviación típica	N
VAR X	6.1250	1.92787	16
VAR Y	6.6875	1.62147	16

Tabla 14.31. Correlaciones

Correlaciones			
		VAR X	VAR Y
VAR X	Correlación de Pearson	1	.632(**)
	Sig. (bilateral)	.	.009
	Suma de cuadrados y productos cruzados	55.750	29.625
	Covarianza	3.717	1.975
	N	16	16
VAR Y	Correlación de Pearson	.632(**)	1
	Sig. (bilateral)	.009	.
	Suma de cuadrados y productos cruzados	29.625	39.437
	Covarianza	1.975	2.629
	N	16	16
** La correlación es significativa al nivel 0,01 (bilateral).			

Vemos los valores obtenidos en los estadísticos descriptivos, el valor de la correlación de Pearson y su significatividad (0,01).

Aplique la fórmula del coeficiente de correlación de Pearson, y compare los resultados que obtiene, con los aportados en las tablas.

Realización del diagrama de dispersión: El diagrama de dispersión, mediante el programa SPSS, se realiza marcando en la barra de herramientas de la hoja de datos → gráficos → diagrama de dispersión → seleccionando las dos variables → pegar → y en el cuadro de sintaxis seleccionamos el texto, y ejecutamos marcando el icono de punta de flecha, lo que nos da el archivo de resultados, y seleccionando el diagrama, lo copiamos y pegamos aquí. Este es el diagrama realizado:

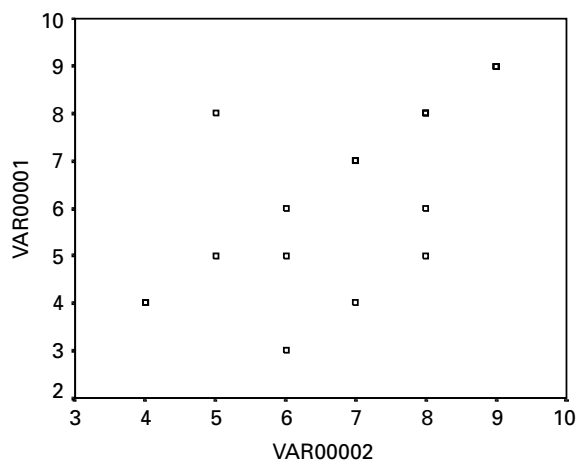


Figura 14.9. Diagrama de dispersión.

Que, como podemos ver, muestra una correlación directa positiva.

14.8. Referencias bibliográficas

- JIMÉNEZ, C., LÓPEZ-BARAJAZ ZAYAS, E. y PÉREZ JUSTE, R. (1987). *Pedagogía Experimental II*. Unidades Didácticas de la UNED. Madrid: UNED.
- JIMÉNEZ, C., y PÉREZ JUSTE, R. (1991). *Formulario y Tablas de Pedagogía Experimental*. Madrid: UNED.
- KERLINGER, F. N. (1987). *Investigación del comportamiento*. México. Interamericana.
- PÉREZ, C. (2001). *Técnicas Estadísticas con SPSS*. Madrid, Pearson Educación, S. A. Prentice Hall.

TEMA 15

Probabilidad y muestreo. Distribución muestral y estimación de parámetros

ESQUEMA

- 15.1.** La probabilidad
- 15.2.** El muestreo
 - 15.2.1. Población y muestra
 - 15.2.2. Aleatoriedad y aleatorización
 - 15.2.3. El muestreo y sus tipos
 - 15.2.3.1. Muestreo aleatorio.
 - 15.2.3.2. Muestreo determinístico
 - 15.2.4. Muestra, tamaño de la misma y representatividad
- 15.3.** La distribución muestral
 - 15.3.1. La suma muestral, la media muestral y la varianza muestral
 - 15.3.2. El error típico
 - 15.3.3. El teorema del límite central
- 15.4.** La estimación de parámetros
 - 15.4.1. Estimación puntual y estimación por intervalos
 - 15.4.2. El intervalo de confianza y los límites de confianza para la estimación de parámetros. Estimación de la media poblacional
 - 15.4.3. Propiedades de los estimadores
- 15.5.** Ejercicios de autoevaluación
- 15.6.** Solución a los ejercicios de autoevaluación
- 15.7.** Referencias bibliográficas

OBJETIVOS

- ✓ Comprender y valorar la finalidad de la estadística inferencial.
- ✓ Comprender y valorar los supuestos con los que opera la estadística inferencial.
- ✓ Conocer y comprender los conceptos de población y muestra
- ✓ Comprender y valorar los procedimientos para la selección de las muestras.
- ✓ Conocer y comprender los conceptos de estadístico, parámetro, significación estadística e intervalo de confianza.
- ✓ Conocer y comprender el concepto de distribución muestral y conceptos asociados.
- ✓ Comprender el concepto de estimación de parámetros, sus tipos y propiedades.

Introducción

La *estadística inferencial* tiene como principal finalidad la de facilitar el conocimiento de las poblaciones a partir del estudio de parte de la misma, o muestra, permitiéndonos obtener conclusiones, en términos probabilísticos, acerca de una población, mediante la estimación de los *parámetros* de la población a partir de los índices, o *estadísticos*, obtenidos en las muestras. A las primeras técnicas estadísticas de inferencia utilizadas se las denominó *técnicas paramétricas* porque inferían los parámetros de la población. Estas pruebas se basaban en los supuestos de que las puntuaciones obtenidas en una muestra se distribuían de acuerdo con la curva normal de probabilidades o en el supuesto de que la serie de puntuaciones obtenidas en dos grupos, uno experimental y otro de control del primero, tengan la misma varianza o misma dispersión de sus puntuaciones (Siegel, 1988: 21). Estas técnicas llamadas paramétricas trabajan con los supuestos arriba mencionados, entre otros. Sin embargo, no siempre podemos trabajar asegurando que estos supuestos sean ciertos, no pudiéndose hacer suposiciones severas acerca de los parámetros de la población. Para estos últimos casos se han desarrollado otra serie de técnicas de inferencia que no pretenden hacer suposiciones, ni numerosas, ni severas, acerca de los parámetros de la población. Estas nuevas «distribuciones-libres», diferentes a la distribución de la curva normal, son llamadas *técnicas no paramétricas*.

15.1. La probabilidad

La probabilidad es un concepto imprescindible en estadística. Nunca se puede afirmar que una hipótesis estadística sea cierta o falsa si no es con un margen de error. La atribución de probabilidades ligadas a los sucesos se suele llevar a cabo en forma numérica mediante valores que oscilan entre 0, suceso imposible, y 1, suceso seguro. Así, si tiramos una moneda al aire la probabilidad de que salga cara será $\frac{1}{2}$, pues son dos los resultados posibles, cara o cruz, pero sólo 1 el favorable a cara. La probabilidad de que al tirar un dado salga un 6 es $\frac{1}{6}$; son seis los resultados posibles, pero sólo 1 es favorable. Todos los resultados son igualmente probables.

$$P_{fr6} = \frac{n_6}{n}$$

Siendo:

P_{fr6} : la probabilidad de que salga el 6 (frecuencia relativa, fr_6),

n_6 : número de veces que aparece el resultado deseado,

n : número total de veces que se repite el intento.

Las frecuencias de todos los resultados suman n :

$$n_1 + n_2 + n_3 + \dots + n_N = n$$

Y la suma de todas las frecuencias relativas es 1:

$$\frac{n_1}{n} + \frac{n_2}{n} + \frac{n_3}{n} + \dots + \frac{n_N}{n} = 1$$

Las probabilidades individuales de todos los resultados posibles mutuamente excluyentes han de sumar 1. Las probabilidades pueden expresarse como fracciones, porcentajes o decimales; Por ejemplo, una probabilidad entre cinco puede expresarse: 1/5, 20% o 0,2.

La teoría de la probabilidad, atribuida a Laplace, es conocida como apriorística porque asigna probabilidades basándose en el principio de equiprobabilidad de los acontecimientos o *probabilidad ideal* a la que tienden los fenómenos reales. La *probabilidad real* será aquella determinada de modo empírico a través de un número suficientemente elevado de pruebas, e indica la razón entre las ocasiones en que el fenómeno se da y el número de pruebas realizadas. Cuando el número de pruebas es elevado se da una gran regularidad, estabilizándose la frecuencia relativa en un valor próximo al valor de probabilidad ideal. La probabilidad puede definirse como la razón entre el número de resultados favorables y el número de resultados posibles cuando todos son igualmente probables.

La distribución de frecuencias de los valores obtenidos por un grupo en una determinada variable pueden convertirse en una distribución de probabilidad —o proporciones— sustituyendo la frecuencia absoluta de cada clase —o intervalo— por su frecuencia relativa, dividiendo cada frecuencia por la suma de las frecuencias. Una variable aleatoria es una función de valores numéricos definida sobre un espacio muestral.

15.2. El muestreo

Hemos mencionado, en diversas ocasiones, los términos población y muestra, señalando la importancia del procedimiento de extracción de muestras para el estudio de sus características con la finalidad de inferir sus conclusiones a la totalidad de la población de la que forma parte, y que de otra manera sería difícil de conocer. Por ello, pasamos ahora a estudiar los conceptos y procedimientos relacionados con las poblaciones y las muestras.

15.2.1. Población y muestra

Población es todo conjunto de elementos, objetos o sujetos, con unas características definidas. Las poblaciones pueden ser teóricamente finitas, como por ejemplo, los estudiantes de segundo curso de Bachillerato, e infinitas, como por ejemplo el conjunto de los números naturales. *Población y universo* suelen utilizarse como sinónimos. Los valores de las variables de los estudios estadísticos realizados en las poblaciones se denominan *parámetros*.

Muestra es una parte o subconjunto de una población seleccionada de tal modo que ponga de manifiesto las propiedades de la población a la que pertenece. En estadística, los datos obtenidos a partir del estudio de una determinada característica o variable en las muestras se denominan *estadísticos*, y en estadística inferencial se utilizan como estimadores de los valores de las variables de la población, llamados *parámetros*, cuando son muestras representativas de dicha población.

15.2.2. Aleatoriedad y aleatorización

Aleatoriedad es la base de los métodos de probabilidad y significa que no hay una ley conocida, que pueda ser expresada en palabras, que describa con exactitud o explique los eventos y sus resultados. Decimos que los eventos, o sucesos, ocurren al azar si no podemos predecir sus resultados. Pero, aunque un **evento aleatorio** no se puede predecir individualmente, podemos predecir con gran exactitud los resultados de un gran número de eventos aleatorios *en conjunto*.

La aleatorización es la asignación de objetos —o sujetos, tratamientos, o grupos—, de un universo a subconjuntos del propio universo, en forma tal que para cualquier asignación a un subconjunto, cualquier miembro del universo tiene la misma probabilidad de ser asignado.

La asignación aleatoria es el modo de asignar los tratamientos a los distintos grupos de una investigación cuando no podemos trabajar con muestras aleatorias sino con grupos ya formados, como por ejemplo, los alumnos de una clase, o de un colegio. La asignación aleatoria guarda estrecha relación con la extracción aleatoria de muestras, pudiendo considerarse la asignación aleatoria como una extracción aleatoria de muestras.

15.2.3. El muestreo y sus tipos

El muestreo consiste en tomar cualquier parte de una *población* o universo mediante las técnicas de extracción de muestras, y por *muestreo*

se entiende el procedimiento seguido para la selección de las muestras. Dado que los fenómenos que se pretenden estudiar no siempre puedan analizarse en todos los casos que ocurren, el investigador resuelve este problema seleccionando muestras representativas, lo que le posibilita realizar su estudio en una parte de la población e inferir los resultados encontrados en la muestra a la población.

Para seleccionar una muestra representativa de la población que se quiere estudiar, lo primero que debe hacer el investigador es definir la población, elaborar una lista precisa y completa de las unidades que la componen, extraer de la lista unidades representativas y obtener una muestra lo suficientemente amplia como para que pueda representar las características de la población (Van Dalen, 1981:323). Fox (1972) distingue cinco etapas en el proceso de muestreo:

1. Definición y selección del universo o especificación de los posibles sujetos o elementos de un cierto tipo.
2. Determinación de la población a la que el investigador va a acceder.
3. Selección de la muestra invitada o subconjunto de la población a los que se les pide que participen en la investigación.
4. Muestra aceptante, de la muestra invitada.
5. Muestra productora de datos, de la muestra aceptante a colaborar.

Son dos los tipos de muestreo *probabilístico* y *no probabilístico* o determinístico. El muestreo probabilístico es el único tipo de muestro que nos permite conocer el error posible que cometemos al inferir los resultados procedentes del estudio de una muestra a la población a la que pertenece dicha muestra. Las muestras probabilísticas emplean alguna forma de extracción aleatoria. Las muestras aleatorias están libres de sesgo, y por ello deben ser seleccionadas con métodos aleatorios consistentes en dar a cada miembro de la población las mismas oportunidades de ser incluido en la muestra. Sólo de este modo satisface los supuestos de la teoría de probabilidades y se podrán hacer inferencias científicas.

15.2.3.1. El muestreo aleatorio

El muestreo aleatorio simple es el procedimiento que se utiliza para obtener una muestra de una población de tal modo que cada miembro de la población o universo tenga igual oportunidad de ser incluido en la muestra. En el muestreo aleatorio se establecen las condiciones rigurosas que aseguren que cada unidad de la población tenga iguales probabilidades de estar representada en la muestra, esperando, de este modo,

que estén presentes las características más representativas de la población en una forma aproximadamente igual a como lo están en la población. Los métodos de selección aleatoria no permiten que operen nuestros propios prejuicios o tendencias, ni otros factores de selección sistemática.

Una muestra aleatoria simple es aquella cuyas n observaciones ($X_1, X_2, \dots, X_n$) son independientes. La distribución de las observaciones o valores (X_i) de la variable aleatoria medida en la muestra aleatoria es la distribución de la población $p(x)$, con media μ y varianza σ^2 . Seleccionar una muestra de una población mediante el procedimiento del muestreo aleatorio simple consiste en que los elementos seleccionados de la población para formar la muestra sean devueltos de nuevo a la población para que los siguientes elementos a seleccionar tengan la misma probabilidad de ser elegidos, y se llama *método de muestreo aleatorio con reposición*. *El muestreo aleatorio con reposición es el único muestreo que se puede denominar con precisión muestreo aleatorio*.

Para la selección de una muestra siguiendo el procedimiento de muestreo aleatorio simple se utiliza la tabla de números aleatorios. Lo primero que se hace es asignar números a todos los sujetos de la población N de la que vamos a extraer la muestra aleatoria de tamaño n . Después se abren las páginas de números aleatorios al azar y se puede señalar al azar un número, o el primer número de cualquier página. Si por ejemplo se necesita seleccionar una muestra de $n = 40$ de una población definida de tamaño $N = 240$, seleccionamos de la lista de números aleatorios todos los inferiores a 241, desechando el 0, puesto que la población comienza por el 1, y vamos señalando al azar números, desechando aquellos que ya han sido elegidos y así hasta completar los 40 sujetos que queremos que integren la muestra. A continuación, con los 40 números aleatorios resultantes iríamos a la población y seleccionaríamos a los sujetos cuya asignación numérica coincide con los 40 números aleatorios.

Cuando la extracción se hace de manera aleatoria es predecible la variabilidad del muestreo. No obstante, nunca podemos tener la certeza de que esto ocurra así, por lo que siempre se realizan afirmaciones en términos de probabilidad. En la investigación real extraemos muestras aleatorias siempre que podemos, y cuando no se puede las inferencias probabilísticas se quedan en conjeturas.

Veamos un ejemplo. Supongamos una población de un millón de jóvenes de 21 años cuyas alturas en centímetros oscilan entre 165 y 199. Su amplitud es 35 ($199 - 165 + 1 = 35$). Formamos intervalos para hacer un recuento de frecuencias de la población, considerando que pueden ser 7 los intervalos con una amplitud de 5. Mostramos la tabla resultante con

las marcas de clase de los intervalos, las frecuencias, frecuencias relativas o probabilidad y otros valores que nos permitirán calcular la media, la varianza y la desviación típica poblacionales.

TABLA 15.1.

Estaturas de los jóvenes de 21 años				
Marca de clase (x)	Frecuencia (f)	Frecuencia relativa o probabilidad p(x)	x p(x)	(x - μ) ² p(x)
167	10.000	0,01	1,67	2,25
172	60.000	0,06	10,32	6
177	240.000	0,24	42,48	6
182	380.000	0,38	69,16	0
187	240.000	0,24	44,88	6
192	60.000	0,06	11,52	6
197	10.000	0,01	1,97	2,25
	N=1.000.000	1,00	$\sum x p(x) = \mu$ $\mu = 182,00$	$\sum (x - \mu)^2 p(x) = \sigma^2$ $\sigma^2=28,5$ $\sigma=5,338$

Su representación gráfica mediante un histograma y curva normal queda tal como se muestra:

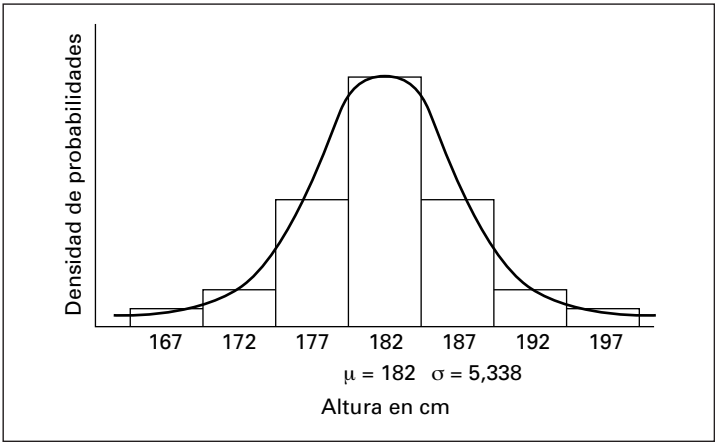


FIGURA 15.1. Histograma y curva normal.

Si la variable aleatoria «Estatura de los jóvenes de 21 años» medida en la población se denomina X, la media de la población se nombra μ de X y la varianza de la población se nombra σ^2 de X. Si quisiéramos realizar un muestreo aleatorio de la población en relación con la variable que esta-

mos estudiando, introduciríamos en un recipiente el millón de números, cada uno impreso en una ficha, y extraeríamos un determinado número de fichas. La primera ficha extraída se denomina X_1 , pudiendo tener cualquiera de los valores de la primera columna de la tabla (marca de clase), y tendrá la probabilidad indicada en la columna tercera, y así seguiríamos extrayendo el resto de fichas que constituirían nuestra muestra n , con la misma distribución de probabilidad que posee la población origen, siendo estas las características del muestreo aleatorio. (Wonnacott y Wonnacott, 1981: 134).

Cuando se trata de poblaciones grandes, millonarias, no incide mucho el hecho de devolver las fichas en la probabilidad de ser elegidas. Pero cuando las poblaciones son menores esto sí es importante, dado que, mediante la reposición de la ficha elegida, se mantiene la independencia en las posibilidades de que cualquier ficha sea elegida, aun cuando otra haya sido elegida con anterioridad. Las elecciones previas no condicionan las siguientes. Sin embargo, cuando se trabaja con poblaciones pequeñas resulta difícil realizar un muestreo mediante la técnica de muestreo aleatorio simple con reposición o reemplazo. Se impone otro procedimiento llamado *muestreo aleatorio sin reposición*, que consiste en tomar un elemento de la población para formar parte de la muestra sabiendo que no tiene más posibilidades de ser elegido.

Dentro del muestreo aleatorio existen otras modalidades como son el muestreo aleatorio sistemático y el muestreo aleatorio estratificado.

El muestreo aleatorio sistemático no llega a ser del todo probabilístico aunque se utiliza con frecuencia. Conocido el tamaño de la población, numerados todos sus elementos y definido el tamaño de la muestra con la que queremos trabajar, el muestreo sistemático consiste en dividir el número total de la población N entre el número que queremos que tenga la muestra, por ejemplo si $N = 200$, decidimos trabajar con una muestra n formada por el 20 % de N , procedemos hallando ese 20%, que nos da un valor de 40. El resultado de la división de N entre n nos indica el valor de la amplitud del intervalo k , para elegir los elementos que formarán parte de la muestra, $200 : 40 = 5$. De cada cinco miembros de la población elegiremos 1, debiendo elegir el primer elemento de la muestra aleatoriamente y a partir de éste se eligen los sucesivos elementos aplicando el intervalo k . Así, si el primer elemento elegido es el 142, el segundo sería el 147, el tercero el 152, así hasta llegar al 137 que sería el cuadragésimo sujeto de la muestra.

El muestreo aleatorio estratificado proporcional utiliza una técnica mediante la cual el investigador trata de evitar que la muestra contenga

proporciones indebidas de los distintos estratos que conforman la población. Dividiendo la población en diferentes estratos, coincidentes con las características que la definen, y luego extrayendo unidades proporcionales de modo aleatorio de cada uno de los diferentes estratos, se asegura la representatividad de la muestra. Por ejemplo, si estamos estudiando la valoración que aficionados y profesionales del cine tienen sobre el cine español, dividiremos al total de asistentes a un famoso festival en profesionales y aficionados, y de cada uno de los grupos resultantes elegimos, sobre el listado numerado, a un 15 % siguiendo el proceso señalado para el muestreo aleatorio simple indicado más arriba. La división por estratos viene determinada pero dentro de cada estrato la selección es proporcional aleatoria.

15.2.3.2. El muestreo determinístico

El muestreo determinístico es un muestreo no probabilístico. No emplea ninguna forma de extracción aleatoria sino que viene condicionado por las necesidades del estudio, y se selecciona de acuerdo con las intenciones del investigador. Alguno de los tipos de muestreo no probabilístico utiliza la selección intencional —en función de un determinado criterio—, selección accidental, selección de casos extremos, selección de caso único, el de grupo homogéneo, el de grupo heterogéneo, la selección por cuota —para seleccionar miembros que sean de determinados estratos—. La selección por cuota no es probabilística pero sí lo es el muestro dentro de cada estrato.

El muestreo no probabilístico no contempla el posible error que permitiría realizar inferencias de la muestra a la población, por lo cual no podemos asegurar que sean muestras representativas que permitan realizar algún tipo de inferencia a la población, aunque puede proporcionar la base para una buena conjetura.

15.2.4. Representatividad y tamaño de la muestra

Una muestra es representativa cuando reúne las características de la población relevantes para los objetivos de la investigación. Cuando un investigador planea realizar una investigación debe tomar dos decisiones en relación con la muestra: una referida a su tamaño y la otra al tipo de muestreo para lograr su representatividad. El tamaño de las muestras se relaciona con el error de estimación de los valores obtenidos en la muestra al inferirlos a la población. Cuanto más grande sea la muestra, mayor

precisión se alcanzará al estimar el valor del parámetro. El tamaño de la muestra debe ser suficiente, de modo que permita incluir los casos extremos, no sólo los casos más comunes o frecuentes.

Se suele considerar que las observaciones realizadas en una variable sobre una muestra de una muestra de tamaño $n = 50$ se distribuye siguiendo el modelo de la curva normal. También se aproxima a este modelo las distribuciones de valores con n superior a 30. En Educación y en Ciencias Sociales no siempre es posible trabajar con muestras superiores a 30, o a 50. Por eso es más frecuente trabajar con series de datos que se distribuyen más siguiendo el modelo de la distribución «t», que el de la distribución normal o «z».

15.3. La distribución muestral

La distribución muestral es la distribución que toma un estadístico estudiado en un número elevado —infinito— de muestras de igual tamaño, extraídas de una misma población. La distribución muestral se caracteriza, como cualquier otro tipo de distribución, por diferentes medidas, especialmente una de tendencia central y otra de variabilidad. Al ser las «n» observaciones de las muestras $X_1, X_2 \dots X_n$, independientes, (elegidas de modo independiente de la población), se trata de muestras aleatorias simples, siendo su distribución la misma distribución que la de la población, con media μ y varianza σ^2 .

La media « μ » y la varianza « σ^2 » poblacional suelen considerarse constantes, a pesar no ser siempre conocidas, al contrario que la media y varianza de las muestras que varían de una muestra a otra con cierta distribución de probabilidad.

15.3.1. Suma, media, varianza y distribución típica de la distribución muestral

La suma de las «n» variables aleatorias pertenecientes a una misma población, y referidas a una determinada variable, viene definida por:

$$S = X_1 + X_2 + \dots + X_N$$

La suma «S» de las «n» variables aleatorias resulta una variable aleatoria de la que podemos conocer su media y su varianza.

La media poblacional « μ » y la varianza « σ^2 » poblacional se definen como la media y varianza de la población conceptual de todas las posibles repeticiones de un experimento cuyas muestras tienen un tamaño defi-

nido. La media de la distribución muestral se deduce a partir de la suma muestral dividida entre el número de muestras:

$$\mu = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{1}{n}S$$

La varianza de la distribución muestral de la media se deduce:

$$\sigma_{\bar{X}}^2 = \left(\frac{1}{n}\right)^2 n\sigma^2 = \frac{\sigma^2}{n}$$

Y la desviación típica, que es la raíz cuadrada de la varianza, queda así:

$$\sigma_{\bar{X}} = \sqrt{\sigma_{\bar{X}}^2} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

15.3.2. El error típico

El error típico o estándar de un estadístico es la desviación típica de los valores que los errores aleatorios toman en una distribución muestral. Es una medida de la variabilidad de una distribución muestral. Se representa por la letra griega σ y tiene como subíndice el estadístico de que se trate. Así, el error típico o desviación típica de la media es, como acabamos de ver, el siguiente:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Como no siempre se puede conocer la varianza de la población « σ^2 », se utiliza la varianza de la muestra « s^2 », que es un buen estimador de la varianza poblacional.

$$\sigma_{\bar{X}} = \frac{s}{\sqrt{n}}$$

En las investigaciones siempre existe la posibilidad de presencia de errores aleatorios que pueden interpretarse como la suma de los errores debidos a la medición y a los errores estocásticos:

- Los errores debidos a la medición son causados por un instrumento impreciso, que no cumple las exigencias técnicas en su fiabilidad y en su validez, o causados por errores introducidos por el propio investigados, sesgos, etc.

- Los errores estocásticos son causados por fenómenos biológicos y sociales irrepetibles, como las diferencias individuales de los sujetos, que son impredecibles, estocásticas.

Ambos tipos de errores pueden reducirse mediante el rigor en la construcción de los instrumentos de medida, y el control experimental sobre la presencia de variables extrañas, tratando de hacerlas constantes. Este último tipo de errores no siempre es posible de controlar en la investigaciones realizadas en Ciencias Sociales, dado que son muchas las variables que escapan al control del investigador. Estos errores contribuyen en grado importante a la varianza total. En los casos en que la distribución muestral sigue el modelo de la curva normal, por tratarse de muestras grandes, la mayoría de estos valores serán pequeños aproximándose al 0. El error típico, como medida de error debido a factores no controlados por el investigador, se emplea a modo de «patrón» con el cual se compara la varianza experimental debida al tratamiento.

15.3.3. El teorema del límite central

El teorema del límite central es clave para la inferencia estadística de muestras cuando su tamaño es grande y dice que cuando una variable se distribuye en la población original siguiendo el modelo de la curva normal, la distribución de la media de n muestras de dicha población se distribuye siguiendo el mismo modelo de curva normal, pero si las n muestras proceden de poblaciones que no tuvieran una distribución normal, a medida que aumenta el tamaño de las n muestras aleatorias extraídas de cualquier población, la distribución de sus medias se aproxima a la distribución normal, con media μ y desviación típica σ .

Veamos el siguiente ejemplo.

Consideremos las calificaciones obtenidas por un grupo numeroso de alumnos de 1.º de Bachillerato en Matemáticas, mediante una prueba calificada de 1 a 10, y que se distribuyen siguiendo el modelo de la curva normal, con media 7,2, y desviación típica 0,9.

1. ¿Cuál será la probabilidad de que un estudiante obtenga una calificación igual o mayor que 7,8?

La probabilidad se calcula transformando las puntuaciones directas en típicas para después buscar en las Tablas de áreas bajo la curva normal la probabilidad asociada a dichas puntuaciones típicas, o puntuaciones « z ». Por lo tanto, transformamos la puntuación directa 7,2, en típica, utilizando la fórmula de z :

$$z = \frac{X_i - \bar{X}}{s} = \frac{7,8 - 7,2}{0,9} = 0,666$$

La probabilidad asociada al valor de $z = 0,666$, consultando la tabla de áreas de probabilidad bajo la curva normal, es $\Pr = 0,2514$, que podemos expresar del siguiente modo:

$$\Pr(X > 7,8) = \Pr\left(\frac{X - \bar{X}}{s} > \frac{7,8 - 7,2}{0,9}\right) = \Pr(z > 0,666) = 0,251$$

Como podemos comprobar existe una probabilidad razonable (1/4), (o un porcentaje del 25,14 %), de que un estudiante obtenga una calificación superior a 7,8.

2. ¿Cuál será la probabilidad de que una muestra de 10 alumnos obtenga una media muestral superior a 7,8?

Por el teorema del límite central sabemos que la distribución muestral es normal, con media poblacional $\mu = 7,2$, y desviación típica:

$$\sigma_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{0,9}{\sqrt{10}} = 0,284$$

La media de la muestra 7,8 representa ahora un valor X_i , dentro de la distribución muestral de la media, con media $\mu = 7,2$ y desviación típica $\sigma = 0,2846$. Entonces, utilizando la misma fórmula pero con términos poblacionales, tenemos que:

$$\Pr(\bar{X} > 7,8) = \Pr\left(\frac{X_i - \mu}{\sigma_{\bar{X}}} > \frac{7,8 - 7,2}{0,284} = \frac{0,6}{0,284}\right) = \Pr(z > 2,11) = 0,0174$$

Por lo que podemos decir que existe una probabilidad muy pequeña, $\Pr = 0,0174$ (o un porcentaje 1,74 %), de que la media muestral de 10 estudiantes obtenga un valor superior a 7,2. Sobre la curva normal podemos ver representados dichos valores:

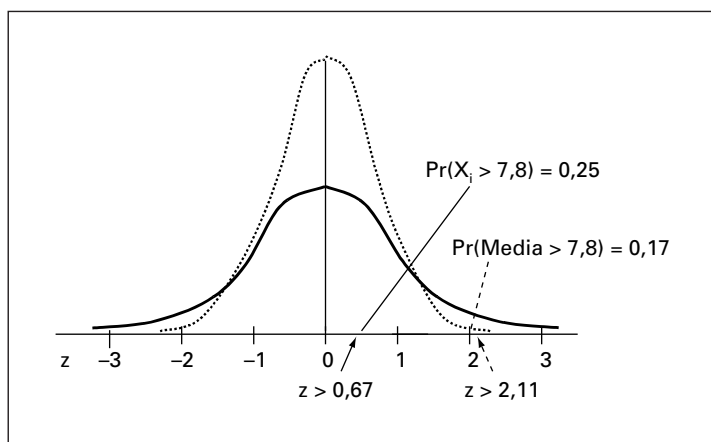


FIGURA 15.2. (Tomada de: Wonnacott y Wonnacott, 1981: 144).

15.4. La estimación de parámetros

La estimación de parámetros es el objetivo de la estadística inferencial: conocer determinados parámetros de la población a partir del estudio de una o varias muestras. El concepto de estimación, por tanto, hace referencia al acto de atribuir valores no medidos a una realidad a partir de otros como resultado de la medición de parte de esa realidad.

Los *parámetros* son valores de la población que casi nunca se pueden calcular directamente por lo inconmensurable de las poblaciones; es por ello que se calculan, o, mejor, se infieren, siguiendo unas determinadas reglas de inferencia, a partir de los *estadísticos*, o medidas obtenidas de las muestras. Los estadísticos son, por tanto, estimadores que permiten *estimar* el valor de los parámetros.

Los *estadísticos* obtenidos de las muestras cumplen la función de estimadores que, salvo errores de cálculo y de medida, son verdaderos para el conjunto de la población de la que se obtuvieron. Los *parámetros* o medidas de toda la población, que difícilmente se pueden obtener en la práctica, suelen ser estimados a partir de los estadísticos muestrales.

MEDIA, VARIANZA Y DISTRIBUCIÓN TÍPICA	
De la muestra	De la población
$\bar{X} = \frac{1}{n} \left(\sum X_1 + X_2 + \dots + X_n \right) = \frac{\sum_{i=1}^n X_i}{n}$ $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X - \bar{X})^2 = \frac{\sum_{i=1}^n (X - \bar{X})^2}{n-1}$ $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X - \bar{X})^2 f_i$ $s = \sqrt{s^2}$	$\mu = \frac{1}{n} (X_1 + X_2 + \dots + X_n) = \frac{1}{n} S$ $\sigma_{\bar{X}}^2 = \left(\frac{1}{n} \right)^2 n \sigma^2 = \frac{\sigma^2}{n}$ $\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$

Un estimador es una función definida sobre una muestra que pretende aproximarse al valor del parámetro desconocido. Sólo si estas muestras son representativas de la población, basándose en el estudio de las mismas, se pueden realizar afirmaciones —inferencias—, dentro de ciertos límites —niveles de confianza— acerca de las características de un parámetro de la población.

15.4.1. La estimación puntual y la estimación por intervalos

Los valores estimados pueden ser de valor único, que casi con toda certeza suelen tener un pequeño error; o puede estimarse el valor en torno a un intervalo que se construye en torno a la media aritmética posible del valor puntual estimado. Por ello, los procedimientos utilizados en la estimación de parámetros son dos:

1. **La estimación puntual:** Es aquella que considera un único valor como estimación del parámetro. Este tipo de estimación no aporta mucha información si no se acompaña de alguna medida del error probable que se comete al realizar la estimación.
2. **La estimación por intervalos:** Es aquella que al estimar el valor de un parámetro no da un único valor para éste sino un conjunto de valores entre los cuales se espera se encuentre el valor verdadero del parámetro. El intervalo, con dos límites, inferior y superior, marca un segmento en el continuo de una escala numérica, en

uno de cuyos puntos se espera la coincidencia con el valor del parámetro.

$$\bar{X} \pm z_{\alpha} \times \sigma_{\bar{X}}$$

15.4.2. *El intervalo de confianza y los límites de confianza para la estimación de parámetros. Estimación de la media poblacional*

El procedimiento más adecuado para realizar inferencias se conoce como *estimación del intervalo de confianza*, mediante el cual se infiere, a partir de un determinado estadístico de la muestra, el valor del parámetro. Conocida la distribución muestral de un estadístico cualquiera es posible estimar, a partir de un valor empíricamente obtenido en el estudio de muestras, el valor del *parámetro* correspondiente. Sin embargo, este valor estimado nunca se consigue que coincida de una manera ajustada con el valor del parámetro, debido a las fluctuaciones del muestreo. Por ejemplo, si el estadístico es la media, se situará en torno a la media poblacional. Para confiar en que la inferencia es correcta debemos hacer una concesión conocida como *intervalo de confianza*. De este modo podemos estimar la media de la población a partir de la media de la muestra sumándole el error de muestreo, lo que dará el valor del parámetro situado dentro de un intervalo de confianza.

El intervalo de confianza, o intervalo de aceptación con una cierta probabilidad, se conoce como los límites entre los cuales se espera se encuentre el valor del parámetro, siendo fundamental conocer la *variabilidad de la distribución*, o lo que es lo mismo, su *error típico*.

15.4.2.1. *Estimación de la media poblacional a partir de una muestra grande*

Si queremos estimar el valor poblacional o, más exactamente, el intervalo confidencial entre cuyos límites se encuentra la media poblacional a partir de la media muestral, necesitamos conocer su error típico y delimitar el nivel de confianza.

Veamos el siguiente ejemplo:

¿Cuál será el intervalo de confianza en el que se encuentre la media de la población a partir de una muestra formada por 100 sujetos, siendo el valor de la media igual a 40, con una desviación típica de 14, al nivel de confianza del 95%?

Lo primero que debemos hacer es hallar el error típico de la media, a partir de su fórmula:

$$\sigma_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{14}{\sqrt{100}} = 1,4$$

El intervalo de la media poblacional viene dado por la siguiente fórmula:

$$\mu = \bar{X} \pm z_{\alpha} \times \sigma_{\bar{X}}$$

El valor de z correspondiente al nivel de confianza del 95 %, o nivel de significación $\alpha = 0,05$, consultando la tabla de probabilidades bajo el área de la curva normal, es $z_{n.c.95\%} = 1,96$, o lo que es lo mismo $z_{\alpha=0,05} = 1,96$.

Y sustituyendo en la fórmula del intervalo para la media poblacional, los símbolos por sus correspondientes valores, tenemos que:

$$\mu = 40 \pm 1,96 \times 1,4 = 37,26 \text{ , } 42,74$$

O lo que es lo mismo:

$$37,26 \leq \mu \leq 42,74$$

Por lo que podemos concluir que la media poblacional se encuentra entre los valores del intervalo (37,26 , 42,74), a un nivel de confianza del 95%, o al nivel de significación $\alpha = 0,05$.

15.4.2.2. *El concepto de grados de libertad*

Los grados de libertad de un estadístico, nombrados como g.l., se definen como el número de observaciones independientes en una muestra, es decir su tamaño n , menos el número de parámetros, k , de la población que deben ser estimados a partir del estudio de la muestra.

$$g.l. = n - k$$

El concepto de grados de libertad se suele explicar, intuitivamente, de modo siguiente: originalmente existen n grados de libertad en una muestra de n observaciones. Como un grado de libertad se utiliza para calcular la media, o la media ocupa el lugar de un grado (o una gradación de la serie de observaciones), deja sólo $n - 1$ grados de libertad a los residuos para calcular la varianza. (Wonnacott y Wonnacott, 1972; 183).

Cuando se estima la media poblacional a partir de muestras pequeñas, dado que la distribución de muestras pequeñas no se distribuye como la

curva normal, debemos utilizar los valores de probabilidad de la distribución de t de Student, debiendo tener en cuenta el concepto de grados de libertad, dado que las distribuciones t son diferentes dependiendo del tamaño de la muestra. La distribución de t no se tabula de acuerdo con el tamaño de la muestra sino conforme al tamaño del divisor de la varianza, que con muestras pequeñas es $n - 1$. A partir de un tamaño superior a 30 o mayor, la distribución de t se asemeja a la distribución de z , o de la curva normal, pasando a utilizar, en estos casos, la tabla de la curva normal.

15.4.2.3. Estimación de la media poblacional a partir de una muestra pequeña

La estimación de la media poblacional a partir del estudio de una muestra pequeña se haría del mismo modo, pero en vez de utilizar los valores de z , se utilizarían los valores de t :

$$\mu = \bar{X} \pm t_{nc} \times \sigma_{\bar{X}} = \bar{X} \pm t_{nc} \times \frac{s}{\sqrt{n}}$$

La distribución de t se utiliza cuando el tamaño de la muestra es pequeño y se desconoce la varianza de la población.

Veamos un ejemplo:

Supongamos que de un grupo elevado de alumnos calificados en una prueba objetiva de Matemáticas formada por 100 ítems, valorado cada ítem como 1 acierto, 0 error, extraemos una muestra de $n = 3$.

TABLA 15.2. Tabla para calcular el intervalo de confianza

X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
47	-11	121
54	-4	16
73	15	225
$\bar{X} = 58$		$s^2 = \frac{362}{3-1} = 181$

El intervalo de confianza para la media de la población será:

$$\mu = \bar{X} \pm t_{nc} \cdot \sigma_{\bar{X}} = \bar{X} \pm t_{0,025,2gl} \cdot \frac{s}{\sqrt{n-1}} = 58 \pm 4,3 \times 9,5 = 58 \pm 40,94 = 17,04, 98,94$$

Dado que la muestra es muy pequeña, el intervalo de confianza es muy impreciso, de tan amplio como es. El error muestral, al tener un denominador pequeño por ser pequeño el tamaño de la muestra, alcanza valores elevados. Y es que las muestras pequeñas aportan poca información que no permiten hacer inferencias precisas, válidas científicamente, dentro de unos márgenes de confianza aceptables. Ampliaremos estas cuestiones en el tema siguiente.

15.4.3. *Propiedades de los estimadores*

Los estimadores deben reunir unas características para poder estimar los parámetros. Los estadísticos, o estimadores, tienen una serie de propiedades, y son las siguientes:

1. **Carencia de sesgo:** Un estimador es insesgado si el promedio de los valores posibles de éste es igual al valor del parámetro. O lo que es lo mismo, si la media de la distribución muestral del estadístico es igual al valor del parámetro. La varianza muestral s^2 es un estimador no sesgado de la varianza poblacional σ^2 . Tanto la media muestral como la mediana son estimadores no sesgados de la media poblacional « μ » en una población normal, pero entre ambos, conviene saber cual de los dos estimadores es más eficiente para estimar μ ; para ello debemos conocer su eficiencia.
2. **Eficiencia relativa.** Cuanto más concentrada esté la distribución de un estimador (de los posibles valores muestrales del estimador del valor del parámetro correspondiente poblacional), es decir, cuanto más próximos estén sus valores, o lo que es lo mismo, cuanto menor sea su varianza, será más eficiente en la estimación del valor del parámetro verdadero. La eficiencia relativa de un estimador se refiere a la precisión en la estimación tomada como cuantía del error en términos de varianza. Esta propiedad indica que se deben preferir los estimadores que tengan la mínima varianza. Y se dice que un buen estimador muestral es aquel que posee un menor error típico, para un « n » dado. El error típico, como sabemos, indica la variabilidad de la distribución muestral. La media muestral es un estimador eficiente de μ . Su varianza es:

$$\sigma_{\bar{x}}^2 = \frac{s^2}{n}$$

Y su error típico, la raíz de la varianza:

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

La mediana muestral tiene una varianza:

$$\sigma_{Me}^2 = \frac{\pi}{2} \times \frac{s^2}{n}$$

π es el símbolo de la proporción de la población, y $\pi/2$ significa el 50% de la población, el punto donde se encuentra la mediana.

Y comparando la varianza de la mediana muestral con la varianza de la media muestral obtenemos la eficiencia relativa:

$$\text{Eficiencia relativa} = \frac{\sigma_{Me}^2}{\sigma_{\bar{X}}^2} = \frac{(\pi/2)(s^2/n)}{(s^2/n)} = \frac{\pi}{2} = 1,57$$

Lo que indica que la media es un 57% más eficiente que la mediana, por lo que se prefiere la media, que proporcionará un valor más preciso, con menor amplitud, más próximo al verdadero valor del parámetro μ .

3. **Consistencia:** Un estimador consistente es aquel que coincide con el valor del parámetro a medida que el tamaño de la muestra aumenta indefinidamente. Dicho de otro modo, un estimador es consistente si a medida que «n» —el tamaño de la muestra—, se aproxima a «N» —el tamaño de la población—, la probabilidad de que su valor sea próximo al parámetro tiende a 1, es decir, el 100%. Un estimador puede ser sesgado pero consistente. La media es un estimador consistente y eficiente de la μ .

Cuando el supuesto de normalidad de una población no se cumple, puede preferirse un estimador del conjunto de la estadística no paramétrica.

15.5. Ejercicios de autoevaluación

15.5.1. Conceptos fundamentales del tema

1. Finalidad de la estadística inferencial.
2. Concepto de población y muestra.
3. Concepto de aleatoriedad.
4. Concepto de muestra aleatoria.
5. Concepto de distribución muestral.

6. Concepto de estimación de parámetros.
7. Concepto de intervalo de confianza.
8. Elabore un cuadro sinóptico sobre el muestreo y sus tipos, incluyendo sus características definitorias.
9. La función de la estadística en la investigación en general, y en particular en educación social.
10. La estimación de parámetros. Tipos y propiedades.

Problema 1

Calcule el intervalo de confianza para la media poblacional de un grupo de 24 alumnos que en una prueba de rendimiento ha obtenido los siguientes resultados globales: $\sum X_i = 390$; $\sum X_i^2 = 6940$.

15.6. Solución a los ejercicios de autoevaluación

Consulte los diferentes epígrafes del tema para las cuestiones de teoría.

Problema 1

Calcule el intervalo de confianza para la media poblacional de un grupo de 24 alumnos que en una prueba de rendimiento ha obtenido los siguientes resultados globales: $\sum X_i = 390$; $\sum X_i^2 = 6940$.

- 1.º Calculamos la media y la varianza (fórmula para muestras pequeñas):

$$\bar{X} = \frac{\sum X_i}{n} = \frac{390}{24} = 16,25;$$

$$s^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}{n-1} = \frac{6940 - \frac{390^2}{24}}{24-1} = \frac{6940 - 6337,5}{23} = 26,2$$

- 2.º A partir de la varianza hallamos la desviación típica y el error típico, con la fórmula para muestras pequeñas:

$$s = \sqrt{s^2} = \sqrt{26,2} = 5,12;$$

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n-1}} = \frac{5,12}{\sqrt{23}} = \frac{5,15}{4,79} = 1,068$$

3. Y estimamos el intervalo de confianza para la media poblacional, con $t_{nc\ 95\%}$ 4,3, hipótesis bilateral.

$$\mu = \bar{X} \pm t_{0,025,2g.l.} \sigma_{\bar{X}}$$

$$\mu = 16,25 \pm 4,3 \times 1,068 = 16,25 \pm 4,59 = 11,66, 20,84$$

La media poblacional se encuentra entre el siguiente intervalo:
 $11,66 < \mu < 20,84$.

15.7. Referencias bibliográficas

- Fox, D. (1972). *El proceso de investigación en educación*. Navarra: EUNSA.
- GARCÍA LLAMAS, J. L., GONZÁLEZ GALÁN, M. A. y BALLESTEROS VELÁZQUEZ, B. (2001). *Introducción a la investigación en Educación*. Unidades Didácticas. Madrid: UNED.
- GARCIA LLAMAS, J. L.; PEREZ JUSTE, R. y RIO SADORNIL, D. DEL (1999). *Problemas y diseños de investigación resueltos*. Ed. Dykinson. Madrid.
- JIMENEZ FERNANDEZ, C. LOPEZBARAJAS ZAYAS, E. y PEREZ JUSTE, R. (1987). *Pedagogía Experimental II*. Madrid: UNED.
- KERLINGER, F. N. (1985). *Investigación del comportamiento. Técnicas y metodología*. Nueva Ed. Interamericana. México.
- MARTÍNEZ MEDIANO, J. M. y otros (1993). *Matemáticas 2. COU*. Madrid: McGraw-Hill.
- VAN DALEN, D. B. y MEYER, W. J. (1979). *Manual de técnicas de investigación educativa*. Buenos Aires: Paidós.
- WONNACOTT, T. H. y WONNACOTT, R. J. (1981). *Introducción a la estadística*. México: Limusa. (1.^a ed.1972, John Wiley and Sons).

TEMA 16

Estimación de la diferencia de dos medias. La prueba de hipótesis

ESQUEMA

- 16.1.** Estimación de la diferencia de dos medias
- 16.2.** La prueba de hipótesis.
 - 16.2.1. El valor de probabilidad o valor crítico.
 - 16.2.2. El valor de probabilidad o valor crítico unilateral para la prueba de hipótesis
- 16.3.** Errores posibles en los que se pueden incurrir en el contraste de hipótesis
- 16.4.** Distribuciones muestrales asociadas a las pruebas de hipótesis
 - 16.4.1. La distribución del estadístico z y del estadístico t de Student
 - 16.4.2. La distribución del estadístico de χ^2
 - 16.4.3. La distribución del estadístico F de Snedecor y la distribución de z
- 16.5.** El modelo estadístico y las pruebas para el contraste de hipótesis
 - 16.5.1. Los modelos estadísticos paramétricos
 - 16.5.2. Los modelos estadísticos no paramétricos
- 16.6.** Ejercicios de autoevaluación
- 16.7.** Solución a los ejercicios de autoevaluación
- 16.8.** Referencias bibliográficas

OBJETIVOS

- ✓ Comprender y valorar la estimación de la diferencia de medias.
- ✓ Comprender y valorar los fundamentos de la prueba de hipótesis.
- ✓ Comprender los errores en los que se pueden incurrir en la prueba de hipótesis.
- ✓ Conocer y comprender las distribuciones muestrales asociadas a las pruebas de hipótesis.
- ✓ Comprender y valorar las características de los modelos estadísticos paramétricos.
- ✓ Comprender y valorar las características de los modelos estadísticos no paramétricos.

Introducción

Ya vimos en el tema anterior que el procedimiento más adecuado para realizar inferencias se conoce como *estimación del intervalo de confianza*, mediante el cual se infiere, a partir de un determinado estadístico de la muestra, el valor del parámetro. Conocida la distribución muestral de un estadístico cualquiera es posible estimar, a partir de un valor empíricamente obtenido en el estudio de muestras, el valor del *parámetro* correspondiente, que nunca se obtiene de una manera exacta, sino dentro de unos límites conocidos como *intervalo de confianza*.

16.1. Estimación de la diferencia de dos medias

La diferencia de medias sirve para comparar medias entre distintas muestras o entre distintas poblaciones. De igual modo que decíamos al hablar de la estimación de parámetros que los valores puntuales suelen ir acompañados de error como consecuencia de las fluctuaciones del muestreo, se suele estimar un intervalo entre dos límites, inferior y superior, entre los cuales se espera se encuentre el verdadero valor del parámetro.

El valor de las diferencias entre dos medias muestrales es un estimador del valor de las diferencias entre dos medias de población.

$$\bar{X}_A - \bar{X}_B$$

En muestras grandes se sabe que la distribución muestral de la media A se distribuye de forma normal con media μ y varianza σ^2/n , y lo mismo ocurre en la distribución muestral de la media B . La independencia de los procedimientos de muestreo en ambas muestras asegurará que las dos variables aleatorias, la distribución muestral de la media de A y de la media de B , sean independientes. El valor estimado de la diferencia de medias es:

$$E(\bar{X}_A - \bar{X}_B) = E(\bar{X}_A) - E(\bar{X}_B) = \mu_A - \mu_B$$

Y el valor estimado de la varianza de las diferencias de las medias es:

$$Es^2_{(\bar{X}_A - \bar{X}_B)} = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}$$

Como sabemos, el error típico es la raíz cuadrada de la varianza:

$$\sigma_{(\bar{X}_A - \bar{X}_B)} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

Aunque las medias de las dos poblaciones difieran, pueden tener una varianza común, y, cuando ese es el caso, la fórmula del error típico de las diferencias de medias de muestras que pertenecen a la misma población o la varianza es común a ambas muestras, es:

$$\sigma_{(\bar{X}_A - \bar{X}_B)} = \sigma \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

Y el intervalo de confianza para la estimación de la diferencia de dos medias será:

$$(\mu_A - \mu_B) = (\bar{X}_A - \bar{X}_B) \pm z_{\alpha} \times \sigma_{(\bar{X}_A - \bar{X}_B)}$$

Cuando la varianza de la población es desconocida hay que calcularla a partir del cálculo de las varianzas de ambas muestras, sumando sus desviaciones cuadráticas y dividiéndola por la suma de los grados de libertad de las dos muestras:

$$s_{pob}^2 = \frac{1}{(n_A + n_B - 2)} \left[\sum (X_A - \bar{X}_A)^2 + \sum (X_B - \bar{X}_B)^2 \right]$$

La distribución de la diferencia de medias será aproximadamente normal, con media $\mu_A - \mu_B$ y varianza $\sigma_A^2/n_A + \sigma_B^2/n_B$.

$$(\bar{X}_A - \bar{X}_B), \left(\mu_A - \mu_B, \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} \right)$$

Su representación gráfica es la siguiente:

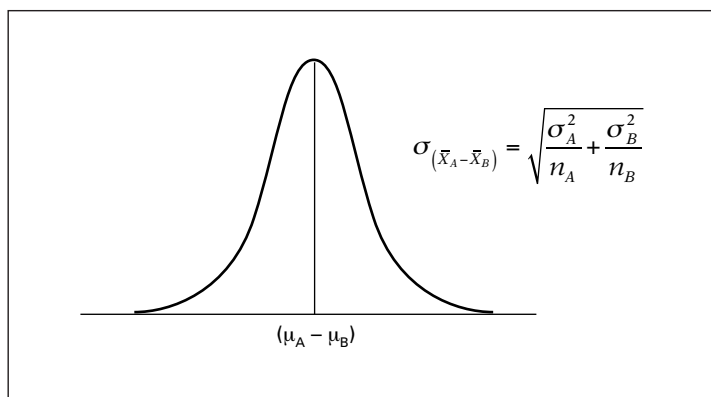


FIGURA 16.1. Distribución de la diferencia de medias. (Tomada de Wonnocott y Wonnocott, 1981: 181).

El intervalo de confianza para la diferencia de medias, al 95%, es:

$$(\mu_A - \mu_B) = (\bar{X}_A - \bar{X}_B) \pm 1,96 \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

Y cuando las muestras proceden de la misma población, o tienen igual varianza:

$$(\mu_A - \mu_B) = (\bar{X}_A - \bar{X}_B) \pm 1,96\sigma \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

Por lo general se desconocen las varianzas de las poblaciones, pudiendo suponerlas a partir de las varianzas observadas en las muestras, dando una aproximación bastante exacta cuando las muestras son grandes (mayores de 30), y por lo tanto su distribución se considera normal.

Veamos un ejemplo para el cálculo de un intervalo de confianza para la estimación de la diferencia de medias de dos muestras pequeñas, independientes, es decir, pertenecientes a dos muestras distintas:

Supongamos que extraemos dos muestras pequeñas de dos grupos de 100 alumnos calificados en la misma prueba de Matemáticas formada por 100 ítems, valorado cada ítem como 1 acierto, 0 error, que utilizamos para estimar el intervalo de confianza de la media de la población (epígrafe 13.4.2.).

TABLA 16.1. Tabla para el cálculo de la varianza

X_A	$(X_A - \bar{X}_A)$	$(X_A - \bar{X}_A)^2$	X_B	$(X_B - \bar{X}_B)$	$(X_B - \bar{X}_B)^2$
55	-15	225	47	-15	225
66	-4	16	54	-8	66
68	-2	4	59	-3	9
72	2	4	64	2	4
75	5	25	67	5	25
84	14	196	70	7	49
			73	11	121
$\bar{X} = \frac{420}{6} = 70$			$\bar{X} = \frac{434}{7} = 62$		
		470			499

Sustituyendo en la fórmula de la varianza los valores de la tabla, tenemos:

$$s_{pob}^2 = \frac{1}{7+6-2}(499+470) = \frac{1}{11}969 = 88,09$$

Y el error típico, (la raíz cuadrada de la varianza) = 9,38

Tomando la fórmula del intervalo de confianza, y sustituyendo los valores, tenemos:

$$(\mu_A - \mu_B) = (\bar{X}_A - \bar{X}_B) \pm t_{0,025,11g.l.} \sigma \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

$$\begin{aligned} (\mu_A - \mu_B) &= (70 - 62) \pm 2,2 \times 9,38 \sqrt{\frac{1}{7} + \frac{1}{6}} = 8 \pm 2,2 \times 5,21 = \\ &= 8 \pm 11,48 = -3,48, 19,48 \end{aligned}$$

La tolerancia del error es un poco elevada debido a que en el ejemplo estamos trabajando con muestras pequeñas. Sin embargo, es menor que el obtenido en el ejemplo del epígrafe 13.4.2 del tema anterior. La muestra en este caso es un poco mayor, y estamos considerando la varianza conjunta, lo que reduce el error muestral. Así en el ejemplo referido teníamos un error muestral de 9,5, y el intervalo era 17,04, 98,94. Ahora la amplitud del intervalo es mucho menor, por lo tanto, más preciso.

El intervalo de confianza es un método para la estimación de parámetros que también se utiliza para el contraste de hipótesis, dado que construimos un intervalo que constituye la región de aceptación de que se encuentre el parámetro de la población a la que pertenece la muestra, con una cierta probabilidad.

Vamos a presentar otro ejemplo sobre el cálculo del intervalo de confianza para la diferencia de medias, pero en esta caso para dos muestras relacionadas. Supongamos que extraemos una muestra de ese grupo calificado en la prueba objetiva de Matemáticas, y que pasado un trimestre le aplicamos la misma prueba:

TABLA 16.2. Tabla para el cálculo del intervalo de confianza. Muestras relacionadas

Sujetos	Primer trimestre X_1	Segundo trimestre X_2	$D = (X_1 - X_2)$	$(D - \bar{D})$	$(D - \bar{D})^2$
1	47	55	12	1	1
2	54	64	10	-1	1
3	55	68	13	2	4
4	64	72	8	-3	9
5	67	75	8	-3	9
6	70	84	14	3	9
7	73	90	17	6	36
			$\bar{D} = \frac{77}{7} = 11$		$s^2 = \frac{69}{7-1} = 11,5$

Y el intervalo de confianza para el parámetro poblacional es:

$$\Delta = \bar{D} \pm t_{n.c.} \sigma_{\bar{D}} = \bar{D} \pm t_{0,025,6 g.l.} \frac{s_D}{\sqrt{n-1}}$$

$$\Delta = 11 \pm 2,45 \frac{3,39}{2,45} = 11 \pm (2,45)(1,38) = 11 \pm 3,39 = 7,61, 14,39$$

Vemos, en este caso, que el intervalo es más ajustado, más próximo, reduciéndose la amplitud del intervalo entre cuyos límites se espera que se encuentre el verdadero valor del parámetro. Su error muestral es, ahora de 1,30, menor que el obtenido anteriormente de 5,21, y el otro de 9,5. Esto nos debe hacer reflexionar sobre la importancia del tamaño de la muestra con la que trabajamos, la misma en ambos casos, los mismos sujetos, por lo que muchas de las variables extrañas que contribuyen a la presencia de errores aleatorios permanecen constantes.

16.2. La prueba de hipótesis

El procedimiento que se sigue para la prueba de hipótesis es equivalente al que se sigue para la estimación del parámetro poblacional a partir de los estadísticos muestrales.

Retomemos el último ejemplo estudiado referido al intervalo de confianza para la diferencia entre dos medias obtenidas por un grupo en una prueba objetiva de Matemáticas en dos momentos diferentes, al finalizar el primer trimestre, y pasado el segundo trimestre del curso. D era la variable a partir de las diferencias entre las puntuaciones del primer tri-

mestre y del segundo trimestre $D = (X_1 - X_2)$, y Δ era el parámetro estimado a partir de la diferencia obtenida en la muestra. El valor del parámetro estimado, al nivel de confianza del 95%, o nivel de significación $\alpha = 0,05$, se encontraba entre el siguiente intervalo:

$$\Delta = 14 \pm 6 = 8, 20 \quad \text{ó} \quad 8 < D < 20,$$

Exceptuando los valores comprendidos dentro de este intervalo, todos los demás valores pueden rechazarse. El intervalo de confianza puede considerarse como el conjunto de hipótesis aceptables del valor del parámetro.

Si el valor del parámetro estimado fuera $\Delta = 0$, significaría la ausencia de diferencias entre las calificaciones promedio obtenidas en la prueba de Matemáticas en el primer y segundo trimestre, coincidiendo con la hipótesis nula H_0 . En nuestro caso el valor de $\Delta = 0$, no está comprendido en el intervalo de confianza obtenido al nivel de confianza del 95%, por lo tanto podemos rechazar la hipótesis nula, pudiendo afirmar que existen una diferencia entre las calificaciones obtenidas entre el primer y segundo trimestre, calificadas de estadísticamente significativas. La significación estadística tiene un carácter técnico y una acepción totalmente diferente a la acepción ordinaria.

16.2.1. El valor de probabilidad o valor crítico para la prueba de hipótesis

Las hipótesis de investigación, o hipótesis alternativas a la hipótesis de nulidad, se prueban mediante el rechazo de la hipótesis de nulidad, por lo que a la estadística se la suele llamar la ciencia de la refutación.

El investigador que trata de mejorar los procedimientos para conseguir mejores rendimientos, mayor eficacia en la consecución de sus objetivos, podría estar convencido de que con los nuevos procesos consigue mejores resultados. Tras la aplicación de los nuevos procedimientos espera que sus resultados medios sean superiores a los obtenidos con los procedimientos antiguos. Sin embargo, para asegurarse debe someter a prueba la hipótesis nula, que dice que las diferencias entre ambos procedimientos, de existir en alguna medida, son fruto del azar, no significativas estadísticamente hablando.

La hipótesis nula diría: Si utilizamos los nuevos procedimientos los nuevos resultados no variarían de los obtenidos con los procedimientos anteriores.

La hipótesis nula formulada estadísticamente quedaría así: $H_0: \bar{X} = \mu$, y se leería: La media muestral no difiere de la media poblacional

La hipótesis alternativa diría: Si utilizamos los nuevos procedimientos los nuevos resultados varían de los obtenidos con los procedimientos anteriores. La hipótesis alternativa formulada estadísticamente quedaría así: $H_1: \bar{X} \neq \mu$, y se leería: La media muestral difiere de la media poblacional.

Tal como está formulada se trata de una hipótesis bilateral, en la que no se muestra la dirección de la diferencia. Sería una hipótesis unilateral si dice, de modo explícito, que los nuevos procedimientos conseguirán mejores resultados que los antiguos ($\bar{X} = \mu$).

Para poner a prueba la hipótesis nula debemos calcular su valor de probabilidad. O dicho de otro modo, debemos comprobar si las diferencias entre la media poblacional y la media muestral difieren en tal grado que sus puntuaciones estandarizadas quedan fuera del valor crítico aceptable científicamente.

El valor crítico, o valor de probabilidad, es el valor límite del intervalo de confianza entre cuyos valores se puede aceptar que la media poblacional y la media muestral difieren significativamente al nivel de significación fijado, aceptable científicamente. Y es que lo importante no es que se alcance o no la significación estadística, sino a qué nivel se alcanza. El nivel que fijamos para decir que no existe significación se llama nivel de confianza crítico (o de importancia máxima) que se corresponde con su nivel crítico α complementario. Ya hemos hablado, en el tema anterior, de niveles de confianza, por ejemplo, del 95 %, del 99 %, y de sus correspondientes niveles de significación, $\alpha = 0,05$, $\alpha = 0,01$.

Un valor de intervalo estimado implica añadirle al valor de la media observada en la muestra el producto de la puntuación z equivalente al nivel de confianza, por el error muestral ($\pm z \cdot \sigma_{\bar{x}}$) que limita a uno y otro lado de la media observada. El intervalo de confianza crítico implica fijar su amplitud:

$$\pm z \frac{\sigma}{\sqrt{n}}, \text{ siendo } z = \frac{|\mu - \bar{X}|}{\sigma_{\bar{x}}} = z_{nc}$$

La puntuación z_{nc} es la probabilidad asociada al valor de z multiplicado por 100. Así, la puntuación $z = 2,58$, tiene una probabilidad asociada de 0,99, que multiplicada por 100, nos da el nivel de confianza del 99 %.

$$100 - n.c. = \alpha \cdot 100 = \text{valor crítico o valor de probabilidad.}$$

El nivel de confianza crítico es el punto límite para aceptar H_0 . Si hablamos del 99 %, el nivel crítico será $\pm 2,58$, a ambos lados de la μ .

Veamos el siguiente ejemplo:

El equipo docente de un centro de Educación Primaria quiere probar la eficacia de un método interactivo con ordenador para la enseñanza del idioma extranjero Inglés. Comunica el proyecto a la Inspección de Educación y solicita su ayuda. Juntos, deciden probar la hipótesis que defiende el equipo de profesores de que se obtendrá mejorar el aprendizaje del Inglés con este nuevo método. Para ello controlarán que en los 6 colegios de la zona se aplique el mismo método de inglés, pero en el centro de estos profesores, en el horario de la clase de Inglés, se va a dedicar 15 minutos de cada sesión de clase, a estudiar el inglés de modo interactivo con ordenador con un grupo de 36 alumnos.

Finalizado el semestre aplican una prueba de Inglés a todos los alumnos de los seis colegios, obteniendo los cinco centros que han utilizado el método tradicional una calificación media de 6,5, y el sexto centro, que ha utilizado el método interactivo, una calificación media de 7, con una desviación típica de 1,3. ¿Se puede afirmar al nivel de confianza del 95 % que existen diferencias significativas entre ambas medias aritméticas, y por tanto, entre ambos métodos de enseñanza del Inglés?

La hipótesis de nulidad defiende que de darse alguna diferencia es fruto del azar, no siendo significativas las diferencias de medias a un nivel de confianza del 95%, quedando el valor de la media de la muestra en la que hemos aplicado el nuevo método, dentro del intervalo de confianza.

$$H_0: \mu = \bar{X}$$

Y la hipótesis alternativa dice que existen diferencias estadísticamente significativas, a dicho nivel de confianza fijado el 95%, entre ambas medias.

$$H_1: \mu \neq \bar{X}$$

Para estimar la media poblacional utilizábamos la media de la muestra, sumándole el error muestral multiplicado por la puntuación de z correspondiente al nivel de confianza fijado:

$$\mu = \bar{X} \pm z_{nc} \times \sigma_{\bar{X}}$$

Pero como conocemos la media de la población, que es 6,5, y la desviación típica del grupo que es la misma que la de la población, por tratarse de la misma población (mismo grupo de edad, la misma zona, las mismas características socioeconómicas, etc.), podemos despreciar de la

ecuación la incógnita que nos queda, que es la puntuación z , para ver si se cumple que queda dentro del intervalo de confianza fijado:

$$z_{nc} = \frac{|\mu - \bar{X}|}{\sigma_{\bar{X}}} = \frac{|6,5 - 7|}{1,3 / \sqrt{36}} = \frac{|0,5|}{0,21} = \pm 2,38$$

Consultadas las tablas de puntuaciones z , vemos que al valor de $z = \pm 2,38$ le corresponde una probabilidad igual a 0,4913 a partir de la media ($z = 0$), hacia ambos lados, que sumados, nos da 0,9826, lo que equivale a un nivel de confianza del 98,26, que es superior al nivel de confianza fijado del 95%.

Veamos la posición del intervalo de confianza en la curva normal y sus valores límites que señalan los valores críticos:

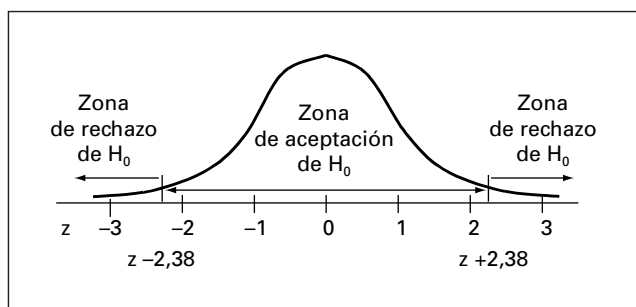


FIGURA 16.2.

Como es una hipótesis bilateral, los valores críticos están a ambos lados de la curva ($z = \pm 2,38$). La hipótesis nula queda dentro de este intervalo de confianza, luego podemos rechazarla y afirmar la hipótesis alternativa. Las diferencias entre la media poblacional y la media de la muestra quedan fuera de la región de aceptación de H_0 . Las diferencias entre ambas son significativas al nivel de significación α del $0,0174 < 0,05$ fijado. El valor de significación obtenido, que queda a ambos lados de la curva por tratarse de una hipótesis bilateral, es de 0,0174, más pequeño que el valor crítico fijado de 0,05. Cuanto mayor es la diferencia entre la media muestral y la media poblacional, menor es el valor de probabilidad o valor crítico. O cuanto mayor es el valor estandarizado de la diferencia de medias (z) menor es el valor de probabilidad y más elevado es el nivel de confianza.

La ciencia se protege contra el azar no aceptando aquellos valores cuya probabilidad de ocurrencia por azar sea superior a un valor cientí-

ficamente aceptado, que el investigador suele fijar con antelación a la realización de su investigación, que suele estar entre el 0,05 o el 0,01. Normalmente, el investigador fija un nivel de confianza del 95 o el 99% para rechazar H_0 , y sólo acepta aquellos valores cuya probabilidad máxima de aparición por azar no supere el 0,05 o el 0,01, respectivamente. Las puntuaciones extremas del intervalo se denominan *límites de confianza*, y el conjunto de valores que los integran se denomina *intervalo de confiabilidad o de confianza*. La determinación de ambos necesita del *conocimiento de la distribución muestral del estadístico* de que se trate, y de su *error típico*.

En el caso de una distribución muestral normal, sabemos que el 95, 98 ó 99 casos de cada 100 se encuentran distribuidos simétricamente alrededor de la media, apartándose de ella 1,96, 2,33, o 2,58 desviaciones típicas, o puntuaciones «z», por encima y por debajo de la media. Como en estos casos σ es el error típico, multiplicado por 1,96, 2,33, o 2,58, según el nivel de confianza fijado, y sumado y restado al estadístico, tendremos los límites de confianza y el correspondiente intervalo de confianza, o confidencial en el que se encontrará el verdadero valor del parámetro. Estos límites son fundamentales para la prueba de hipótesis. Cualquier valor igual o superior a tales límites supone que las diferencias entre los estadísticos obtenidos en dos muestras aleatorias no se deben al azar, dentro de unos límites de confianza, sino a la variable independiente, experimental o tratamiento, introducida en una de las muestras (grupo), permitiendo el rechazo de H_0 , y la aceptación de H_1 .

Veamos algunos valores críticos de z y sus correspondientes niveles de confianza en la figura 16.3:

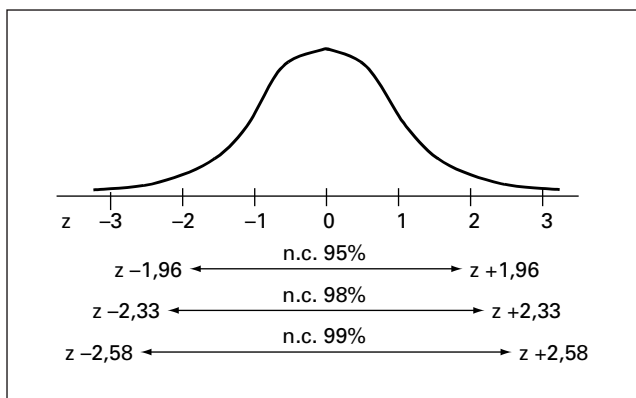


FIGURA 16.3.

16.2.2. *El valor de probabilidad o valor crítico unilateral para la prueba de hipótesis*

En muchos casos, el investigador trabaja con hipótesis unilaterales. Espera que los nuevos procedimientos o métodos con los que experimenta obtendrán mejores resultados que los que venía utilizando anteriormente, por lo tanto, la hipótesis es unilateral y su formulación debe hacerse del siguiente modo:

La hipótesis alternativa unilateral diría: Si utilizamos los nuevos procedimientos los resultados serán superiores a los obtenidos con los procedimientos anteriores. Formulada estadísticamente quedaría así:

$$H_1: \bar{X} > \mu$$

Y se leería: La media muestral es mayor que la media poblacional.

Podría formularse la hipótesis de este otro modo:

$$H_1: \bar{X} \geq \mu$$

Y quiere decir que la media aritmética obtenida por el nuevo procedimiento será igual o mayor que la media aritmética obtenida por el anterior procedimiento.

El cálculo del valor de probabilidad unilateral, o valor crítico unilateral, se obtendría mediante la fórmula:

$$\Pr = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{s/\sqrt{n}} \geq z$$

Y al tratarse de una hipótesis unilateral, o direccional, en este caso unilateral derecha, el valor de z estará representado en el lado derecho de la curva.

16.3. Errores posibles en los que se pueden incurrir en las decisiones estadísticas en el contraste de hipótesis

Los errores posibles en los que se pueden incurrir al decidir aceptar o no aceptar H_0 , en el contraste de hipótesis, son los siguientes:

Cuando	Se decide aceptar H_0	Se decide no aceptar H_0
Siendo H_0 cierta	Decisión correcta (probabilidad = $1 - \alpha$, corresponde al nivel de confianza)	Error tipo I (probabilidad = α , llamada, también, nivel de significación)
Siendo H_0 no cierta (H_1 verdadera)	Error tipo II (probabilidad = β)	Decisión correcta (probabilidad = $1 - \beta$, llamada, también, potencia de la prueba)

α representa la probabilidad de rechazar H_0 cuando es cierta.

β representa la probabilidad de aceptar H_0 cuando no es cierta.

El error más grave que puede cometerse en la decisión estadística en el contraste de hipótesis es rechazar H_0 cuando H_0 es verdadera, su probabilidad es α o nivel de significación. Para evitar los posibles errores se aconseja repetir los experimentos para confirmar o rechazar los resultados. Asimismo aumentar el tamaño de la muestras reduce las fuentes de error. Un n mayor reduce el error típico, y por lo tanto reduce el valor de probabilidad o valor crítico α para el rechazo de H_0 y la aceptación de la H_1 alternativa.

La hipótesis alternativa, o del investigador, se prueba mediante el rechazo de H_0 . Al no confirmarse H_0 se puede rechazar y aceptar la hipótesis alternativa H_1 . Los errores en rechazar H_0 cuando es verdadera pueden deberse a errores en el diseño del experimento, a errores en la construcción de los instrumentos de recogida de datos, a errores debido a la poca experiencia del investigador. Estos errores están relacionados con la validez interna y externa de los diseños de investigación, que fue objeto de estudio en el Tema 3, de la Primera Unidad Didáctica de este texto, dedicado a la investigación experimental.

La valía de toda investigación debe basarse tanto en los aspectos sustantivos de la temática sobre la que se investiga como en los aspectos metodológicos. La calidad en la que ambas circunstancias concurren en una investigación será garantía de que las decisiones adoptadas sean las correctas.

16.4. Distribuciones muestrales asociadas a la prueba de hipótesis

Las principales distribuciones utilizadas en la inferencia estadística son, entre otras, la normal o z , la de t , la de χ^2 y la F . Hasta ahora nos hemos venido refiriendo a las distribuciones z y a la t de Student. La distribución de χ^2 , la hemos mencionado cuando estudiamos las escalas y la significación del coeficiente de contingencia C . Por lo general se considera que cuando el tamaño de la muestra es mayor de 30 sus valores se distribuyen siguiendo el modelo de la curva normal.

16.4.1. La distribución del estadístico z y la del estadístico t de Student

La estimación del valor de la μ en una muestra grande viene determinado por la siguiente fórmula:

$$\mu = \bar{X} \pm z_{\alpha} \times \sigma_{\bar{X}}$$

Conocido el valor de la media muestral, y sabiendo que la distribución muestral de la media en muestras grandes se distribuye siguiendo el modelo de la curva normal, y conocida la desviación típica de la distribución, se puede estandarizar la diferencia de la media muestral y poblacional, mediante la fórmula de las puntuaciones tipificadas « z »:

$$z = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$

Por analogía, para muestras pequeñas la « t » de Student viene definida por siguiente la fórmula:

$$t = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$

Cuando no se puede trabajar con muestras grandes se utiliza la distribución de t en vez de la distribución de z , y al contrario que con la distribución de z , que siempre adopta la misma forma, y por lo tanto es única, la distribución de « t » no es única sino que hay una familia de distribuciones t dependiendo del tamaño de la muestra; a medida que la muestra aumenta, a partir de 30 casos, se aproxima a la distribución normal, y a partir de 50 su aproximación es casi coincidente con la normal. La distribución de t más que con el tamaño de la muestra, se tabula

con el denominador de la desviación típica para muestras menores de 30, que es $n - 1$, y que llama grados de libertad. Si consideramos que de n casos, uno de ellos es la media, el resto son los valores que quedan libres, y por eso se llaman grados de libertad. Cuando las muestras son grandes, el restar 1 al total de casos resulta insignificante.

16.4.2. La distribución del estadístico de χ^2

La distribución de χ^2 se define como la medida que describe la discrepancia entre las frecuencias observadas y las frecuencias esperadas, diciendo la hipótesis de nulidad (H_0) que no existen diferencias entre ambas. Ya vimos el estadístico de χ^2 cuando estudiamos el coeficiente de asociación C de Pearson, para variables medidas a nivel nominal o categorial. Vimos que la significatividad de este coeficiente se basaba en la significatividad del resultado de χ^2 .

La fórmula de χ^2 es la siguiente:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Siendo:

f_e : frecuencias teóricamente esperadas

f_o : frecuencias observadas

La aplicación de χ^2 requiere que las observaciones en los grupos sean independientes. También se define como la suma de cuadrados de puntuaciones z , viniendo la relación entre ambas por la fórmula siguiente:

$$z = \sqrt{\chi^2}$$

Las formas de la distribuciones de χ^2 dependen del tamaño de la muestra, de sus grados de libertad. Hay tantas distribuciones como diferentes sean los grados de libertad. Cuando el tamaño de la muestra es mayor de 30 sus valores se distribuyen siguiendo el modelo de la curva normal.

La aplicación más frecuente del estadístico de χ^2 se hace para probar hipótesis acerca de la forma de las distribuciones, utilizándose como prueba de bondad de ajuste. Los contrastes de hipótesis sobre bondad de ajuste, o pruebas de bondad de ajuste, nos permiten comprobar si la forma de una distribución de probabilidad de una variable se ajusta a una determinada distribución de probabilidad teórica.

16.4.3. La distribución del estadístico F de Snedecor y la distribución de z

La relación entre F y z viene dada por la siguiente ecuación:

$$z = \log_e \sqrt{F}$$

Este estadístico se utilizará en el análisis de la varianza, que veremos en el próximo tema. Cuando el tamaño de la muestra es mayor de 30 sus valores se distribuyen siguiendo el modelo de la curva normal.

16.5. El modelo estadístico y las pruebas para el contraste de hipótesis

Un modelo es una analogía en la que un conjunto de datos representa a otro conjunto por ser isomórficos al existir correspondencia entre todos y cada uno de sus elementos y poseer ciertas estructuras semejantes. La elección del modelo requiere determinar las características de los datos y conocer los modelos disponibles y las características o supuestos subyacentes a cada uno de ellos.

Para la realización de inferencias estadísticas, a partir del estudio de muestras, se han desarrollado dos tipos de modelos: los modelos paramétricos y los modelos no paramétricos.

16.5.1. Los modelos estadísticos paramétricos

La utilización de los modelos estadísticos paramétricos para la realización de inferencias, supone la exigencia de ciertas condiciones, o supuestos, que se deben dar en las muestras que se estudien.

- **Distribución normal.** Una primera condición o supuesto hace referencia a la distribución de los valores, que se pretenden estudiar, en la población de la cual se obtienen las muestras, que deben estar normalmente distribuidos.
- **Muestreo aleatorio.** Otra condición importante, impuesta por la estadística inferencial, es que los datos obtenidos de la muestra deben proceder de muestras seleccionadas aleatoriamente de una población definida.
- **Escala de medición de intervalo o de razón.** Otra de las condiciones se refiere a la escala de medición de los valores de las variables objeto de estudio en la muestra, que deben ser medidos en

escala de intervalo o de razón, que son las que aportan puntuaciones verdaderamente numéricas, condición necesaria para operar algebraicamente con dichos datos asegurando la exactitud de las conclusiones de la prueba.

- ***Igual población o igual varianza.*** Cuando se trata de diseños de grupo para comparar los efectos de un tratamiento para mejorar una condición, fijándose especialmente en el contraste de medias, se exige que los grupos utilizados en la investigación pertenezcan a la misma población o tengan la misma varianza en la distribución de los valores de la variable objeto de estudio.

En resumen, las pruebas estadísticas paramétricas deben cumplir las tres condiciones o supuestos del modelo estadístico paramétrico siguientes:

1. Las características que se estudian en la muestra deben existir en la población.
2. Los valores obtenidos en la variable deben estar distribuidos siguiendo el modelo de la curva normal.
3. Independencia de las observaciones.
4. Medición de la variable dependiente en nivel, al menos, de intervalo.

Los siguientes requisitos son especiales para el contraste de grupos mediante las pruebas estadísticas paramétricas «t» de Student y «F» de Snedecor:

5. Homoscedasticidad o igualdad de las varianzas de las muestras.
6. Linealidad, para la prueba «F». Efectos de la variable independiente sobre la dependiente aditivos, no interactivos ni multiplicativos.

La principal ventaja que aconseja la aplicación de las pruebas estadísticas paramétricas cuando se cumplen sus condiciones o supuestos de aplicación, es que son más potentes que las no paramétricas para rechazar hipótesis nulas cuando la hipótesis alternativa del investigador es verdadera.

16.5.2. Los modelos estadísticos no paramétricos

El modelo estadístico en el que se fundamentan las *pruebas estadísticas no paramétricas* es menos exigente que el de las paramétricas. No parten de supuestos acerca de la población o son muy débiles y operan con datos medidos a escala ordinal o incluso nominal, y la distribución de los

valores de la muestra puede no distribuirse siguiendo el modelo de la curva normal en la población. Estas pruebas estadísticas no paramétricas reciben también el nombre de pruebas de rango, haciendo mención con esto a las características de los datos con los que pueden operar, medidos con escalas de nivel de medida ordinal o nominal o categorial. También se las denomina, en ocasiones, de «distribuciones libres» refiriéndose al hecho de que las observaciones en la variable dependiente pueden no distribuirse siguiendo el modelo de la curva normal. Cuando se trata de comparar los efectos de un tratamiento mediante un diseño de grupos, se centra la diferencia en la mediana, dado que los datos ordinales no permiten operar, solo contar.

El modelo estadístico no paramétrico, en Ciencias Humanas, a veces es el único posible de aplicar, por la dificultad de realizar observaciones que alcancen el nivel de medida de intervalo, trabajar con muestras auténticamente seleccionadas aleatoriamente, o asegurar que las observaciones se distribuyen en la población de acuerdo con el modelo de curva normal, siendo por tanto, en dichas condiciones, el más adecuado a utilizar.

Las condiciones o supuestos del modelo estadístico no paramétricos, ya hemos dicho, son menos exigentes que las del modelo paramétrico.

Supuestos asociados:

1. Independencia de las observaciones.
2. Continuidad básica de la variable.

Ambos tipos de pruebas estadísticas, las paramétricas y las no paramétricas, tienen ventajas e inconvenientes. Entre las ventajas de las pruebas no paramétricas están las siguientes:

- Permiten hacer afirmaciones de probabilidad con un determinado nivel de significación.
- Con n pequeño (4, 5 ó 6), no se pueden aplicar las pruebas paramétricas, salvo si se conoce la naturaleza de su distribución poblacional, pudiendo aplicar en su lugar las pruebas no paramétricas.
- Con observaciones de poblaciones diferentes, en general, no se pueden aplicar las pruebas paramétricas, sólo las no paramétricas que no exigen la igualdad de las varianzas de las muestras (o procedentes de la misma población).
- Son aplicables a datos inherentes a los rangos.
- Si los datos son clasificaciones (escala nominal), sólo pueden aplicarse las pruebas no paramétricas.

En cuanto a los inconvenientes de las pruebas no paramétricas cabe mencionar que su utilización con datos que permitieran la aplicación de pruebas paramétricas supondría un despilfarro de dichos datos. Las pruebas estadísticas no paramétricas se aplican cuando:

1. El nivel de medida es ordinal o nominal, no alcanzando el nivel de intervalo.
2. Cuando en número insuficiente de la muestra no garantiza la continuidad de la variable dependiente medida.

Siempre se recomienda utilizar la prueba no paramétrica que utilice el nivel de medida más exigente, si los datos lo permiten, es decir, el de intervalo sobre el ordinal y nominal, y el ordinal sobre el nominal.

16.6. Ejercicios de autoevaluación

1. ¿Cuál es la finalidad de la estimación de la diferencia de dos medias?
2. Mencione los pasos para la estimación de la diferencia de medias.
3. Defina el concepto de prueba de hipótesis.
4. ¿Por qué se dice que el procedimiento seguido en la prueba de hipótesis es equivalente al que se sigue para estimar la diferencia de medias? Argumente su respuesta.
5. Defina valor de probabilidad o valor crítico para la prueba de hipótesis.
6. Defina el valor de probabilidad o valor crítico unilateral para la prueba de hipótesis.
7. Comente los errores posibles en los que se pueden incurrir en el contraste de hipótesis.
8. Supuestos del modelo estadístico paramétrico para el contraste de hipótesis.
9. Supuestos del modelo estadístico no paramétrico para el contraste de hipótesis.

Problema 1

Si mediante una encuesta de intención de voto realizada a 1.000 posibles votantes conocemos que 610 van a votar a un determinado partido, es posible estimar, o establecer la hipótesis, con un 95% de confianza, que dicho partido obtendrá la mayoría en las elecciones.

16.7. Solución a los ejercicios de autoevaluación

Consulte el tema para comprobar sus respuestas a las preguntas de teoría.

Problema 1

Si mediante una encuesta de intención de voto realizada a 1.000 posibles votantes conocemos que 610 van a votar a un determinado partido, es posible estimar, o establecer la hipótesis, con un 95% de confianza, que dicho partido obtendrá la mayoría en las elecciones.

La hipótesis es que la probabilidad (π) de que gane será superior a 0,5, es decir:

$$H_1 = \pi > 0,50$$

Y la hipótesis nula la contraria:

$$H_0 = \pi < 0,50$$

Para estimar el valor de la probabilidad poblacional (π), la fórmula es:

$$\pi = P \pm z_{\alpha} \sqrt{\frac{P(1-P)}{n}}$$

Es decir, la probabilidad muestral más un intervalo entre el que se espera se encuentre el valor del parámetro. Para ello hallaremos la probabilidad o proporción correspondiente a los 610 votantes, en la muestra aleatoria y suficiente de 1.000 electores, y por lo tanto representativa.

$$P = 610 : 1.000 = 0,61.$$

Y el error típico de P :

$$\sigma_p = \sqrt{\frac{P(1-P)}{n}}$$

Sustituyendo los valores en la fórmula, con $z_{\alpha} = 1,96$, al 95%, tenemos:

$$\pi = 0,61 \pm 1,96 \sqrt{\frac{0,61(1-0,61)}{1000}} = 0,61 \pm 0,03 = 0,58 , 0,64$$

Que nos da el valor del intervalo en el que se encuentra el valor del parámetro, al nivel de confianza del 95%. Queda confirmada la hipótesis H_1 a un nivel de confianza del 95%, o lo que es lo mismo, con una significación de $\alpha = 0,05$, que decía que $\pi > 0,50$.

16.8. Referencias bibliográficas

- CAMPBELL, D. y STANLEY, J. (1982). *Diseños experimentales y cuasi-experimentales en la investigación social*. Amorrortu Editores. Buenos Aires, Argentina.
- FOX, D. (1972). *El proceso de investigación en educación*. Navarra: EUNSA.
- GARCÍA LLAMAS, J. L., GONZÁLEZ GALÁN, M. A. y BALLESTEROS VELÁZQUEZ, B. (2001). *Introducción a la investigación en Educación*. Unidades Didácticas. Madrid: UNED.
- GARCIA LLAMAS, J. L.; PEREZ JUSTE, R. y RIO SADORNIL, D. DEL (1999). *Problemas y diseños de investigación resueltos*. Ed. Dykinson. Madrid.
- JIMENEZ FERNANDEZ, C. LOPEZBARAJAS ZAYAS, E. y PEREZ JUSTE, R. (1987). *Pedagogía Experimental II*. Madrid: UNED.
- KERLINGER, F. N. (1987). *Investigación del comportamiento. Técnicas y metodología*. Nueva Ed. Interamericana. México, 1975.
- VAN DALEN, D. B. y MEYER, W. J. (1979). *Manual de técnicas de investigación educacional*. Buenos Aires: Paidós.
- WONNACOTT, T. H. y WONNACOTT, R. J. (1981). *Introducción a la estadística*. México: Limusa. (1.^a ed. 1972, John Wiley and Sons).

TEMA 17

Resolución de diseños de investigación de dos grupos

ESQUEMA

- 17.1.** Los diseños de investigación de dos grupos
- 17.2.** Tipos de diseños de dos grupos y sus pruebas estadísticas para el contraste de hipótesis
- 17.3.** Resolución de un diseño investigación de dos grupos independientes y nivel de medida en la variable dependiente de intervalo mediante la prueba t de Student
- 17.4.** Resolución de un diseño investigación de dos grupos relacionados y nivel de medida en la variable dependiente de intervalo mediante la prueba t de Student
- 17.5.** Resolución de un diseño investigación de dos grupos independientes y nivel de medida en la variable dependiente ordinal mediante la prueba U de Mann-Whitney
- 17.6.** Resolución de un diseño investigación de dos grupos relacionados y nivel de medida en la variable dependiente ordinal mediante la prueba T de Wilcoxon
- 17.7.** Ejercicios de autoevaluación
- 17.8.** Solución a los ejercicios de autoevaluación
- 17.9.** Referencias bibliográficas

OBJETIVOS

- ✓ Conocer y comprender los procedimientos generales en la resolución de los diseños de investigación de dos grupos.
- ✓ Conocer y comprender los procedimientos seguidos en la resolución de los diseños de investigación de dos grupos independientes y nivel de medida de intervalo.
- ✓ Conocer y comprender los procedimientos seguidos en la resolución de los diseños de investigación de dos grupos independientes y nivel de medida ordinal.
- ✓ Conocer y comprender los procedimientos seguidos en la resolución de los diseños de investigación de dos grupos relacionados y nivel de medida de intervalo.
- ✓ Conocer y comprender los procedimientos seguidos en la resolución de los diseños de investigación de dos grupos relacionados y nivel de medida ordinal.

17.1. Los diseños de dos grupos

El diseño experimental es la estructura elegida por el investigador para resolver el problema o contestar a las preguntas de investigación en condiciones tales que concedan credibilidad y validez a las respuestas alcanzadas. Y no es otra cosa que el principio básico que permite afirmar que los efectos medidos en la variable dependiente se deben a y sólo a las variaciones, o variación, introducidas por la variable independiente. La misión del diseño es probar, manteniendo constantes los elementos o variables extrañas, los efectos del tratamiento o variable independiente experimental. Si no se controlan las demás variables independientes, actuarían como extrañas.

Los diseños de dos grupos son diseños con una sola variable independiente con dos condiciones, tratamiento o niveles; un nivel aplicado a uno de los grupos, y el otro al segundo, o un solo tratamiento que es aplicado al grupo experimental, manteniendo al otro grupo, llamado de control, sin la aplicación de la variable independiente o tratamiento. Al existir dos grupos, si la asignación de los sujetos a los mismos se realiza en condiciones adecuadas, bien sea al azar o bien por técnicas de equiparación como el emparejamiento, el diseño habrá posibilitado el control como medio para asegurar que inicialmente no se dan diferencias significativas entre los grupos, por lo que no se podría afirmar que tales variables hayan sido causa, junto con la variable independiente, de los resultados medidos en la variable dependiente.

Cada uno de los grupos puede ser considerado como control del otro, de modo que todas las posibles variables afectan del mismo modo a ambos grupos, para asegurar que la diferencia en la variable dependiente se debe a la existencia de la variación experimental provocada y controlada por el investigador.

17.2. Tipos de diseños de dos grupos y sus pruebas estadísticas para el contraste de hipótesis

Los tipos de diseños de dos grupos obedecen a la siguiente clasificación, correspondiéndoles las pruebas estadísticas para el contraste de hipótesis que indicamos:

Diseño de dos grupos independientes:

- Nivel de medida:
 - De intervalo:

- distribución normal y homogeneidad de las varianzas: t grupos independientes,
- distribución normal y no homogeneidad de las varianzas: U de Mann-Whitney,
- distribución no normal y homogeneidad de las varianzas: U de Mann-Whitney.
- Ordinal: U de Mann-Whitney.
- Nominal: χ^2

Diseño de dos grupos relacionados:

- Nivel de medida:
 - De intervalo:
 - distribución normal (homogeneidad de las varianzas): t grupos relacionados, $n < 30$; z , $n > 30$; Kruskal-Wallis,
 - distribución no normal: T de Wilcoxon.
 - Ordinal: T de Wilcoxon.

17.3. Resolución de un diseño de investigación de dos grupos independientes y nivel de medida en la variable dependiente de intervalo mediante la prueba t de Student

Un investigador quiere probar la incidencia en el aprendizaje a través de internet con programas que introducen ayudas que favorecen la interactividad del alumno con los contenidos de aprendizaje, la comprobación de sus avances y la ampliación de los contenidos, frente a otros programas que no introducen estas ayudas. Selecciona dos grupos de alumnos, de forma independiente, pertenecientes a la misma población, concretamente del segundo curso de Educación Social de la UNED, con las asignaturas de Metodología de Investigación en Educación aprobadas, lo cual supone que tienen la misma varianza o no difieren significativamente en la variable que vamos a estudiar, el rendimiento en materias de metodología de investigación. La muestra invitada es de 60 alumnos, y a partir de ella formamos dos grupos.

El grupo experimental, formado por 30 alumnos, utilizará un programa de estudio de los temas de inferencia estadística, a través de internet, es decir, virtualizado, con elementos de interactividad para guiar el aprendizaje del alumno. El grupo de control, formado por los otros 30 alumnos seleccionados, utilizará el mismo programa de estudio, a través de internet, pero sin elementos de interactividad.

Se dispone de mes y medio para comprobar los resultados de ambos enfoques, y al finalizar dicho periodo se les envía a los alumnos por correo electrónico una prueba precisa para comprobar el aprendizaje de los contenidos del programa, y que debe ser resuelta y devuelta, por correo electrónico, al profesor en un plazo máximo de dos días después de haberla enviado el profesor. Por diversos motivos los alumnos que contestan a tiempo son 12 del grupo experimental y 10 del grupo de control. Una vez corregidas las pruebas nos aportan los siguientes datos:

TABLA 17.1.

Grupo Experimental: X_i	Grupo Control: X_i
4	3
4	4
5	5
5	5
6	6
6	6
7	6
7	7
7	7
8	7
8	
9	

Para probar la incidencia del método apoyándonos en los resultados muestrales, se procede del siguiente modo:

1.º Esquema y comentario del diseño planteado

Grupos	Selección	Asignación	Pretest	Variable Independiente	Posttest
A Control	Azar	Aleatoria	No	Prog. sin ayudas	T_1
B Experim.	Azar	Aleatoria	No	Prog. con ayudas	T_2

Se trata de un diseño de dos grupos, cuasi-experimental, cuyos sujetos han sido seleccionados al azar y el tratamiento se ha asignado a los grupos aleatoriamente. Las varianzas de la población son desconocidas pero supuestas iguales para ambos grupos, que proceden de la misma población. El tamaño de los grupos es pequeño, y con n diferente: $n_A = 12$ y $n_B = 10$. Sin preset, la variable independiente, en dos niveles, el programa de Internet. La U.

D. se mide mediante una prueba de aprendizaje precisa que alcanza el nivel de medida de intervalo.

2.º Formulamos las hipótesis, tanto literales como las hipótesis estadísticas

Hipótesis textuales:

H_0 : El aprendizaje a través de Internet, mediante programas educativos específicos, no incidirá en el rendimiento de los alumnos.

H_1 : El aprendizaje a través de Internet, mediante programas educativos específicos, incidirá en el rendimiento de los alumnos.

Hipótesis estadísticas:

$$H_0 = \bar{X} = \mu$$

$$H_1 = \bar{X} \neq \mu$$

El nivel de significación fijado para el contraste de hipótesis es $\alpha = 0,05$.

Se trata de una hipótesis bilateral, debido a que el investigador no se decanta abiertamente por una incidencia positiva que le llevaría a formular una hipótesis unilateral del tipo:

$$H_1 = \bar{X} \geq \mu$$

3.º La prueba utilizada para la recogida de los datos

En nuestro caso, para valorar el rendimiento final de los alumnos de ambos grupos, hemos utilizado una prueba precisa de acierto o error. La puntuación total es la suma de las preguntas contestadas acertadamente. El nivel de medida se considera, por tanto, de intervalo, o próximo.

4.º La elección del estadístico para el contraste de hipótesis

De acuerdo con la información comentada, la prueba estadísticas es la t de Student, cuya fórmula es la siguiente:

$$t = \frac{|\bar{X}_A - \bar{X}_B|}{\sqrt{\left[\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2} \right] \times \left[\frac{1}{n_A} + \frac{1}{n_B} \right]}}$$

5.º Aplicación de la prueba estadística para el contraste de hipótesis

Necesitamos, para resolver la ecuación de t , hallar las medias y las varianzas, para lo cual utilizamos la siguiente tabla:

TABLA 17.2.

Grupo experimental (A)			Grupo de control (B)		
X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$	X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
4	-2,33	5,42	3	-2,6	6,76
4	-2,33	5,42	4	-1,6	2,56
5	-1,33	1,77	5	-0,6	0,36
5	-1,33	1,77	-0,6	0,36	
6	-0,33	0,11	6	0,4	0,16
6	-0,33	0,11	6	0,4	0,16
7	0,67	0,45	6	0,4	0,16
7	0,67	0,45	7	1,4	1,96
7	0,67	0,45	7	1,4	1,96
8	1,67	2,79	7	1,4	1,96
8	1,67	2,79			
9	2,67	7,13			
$\bar{X} = \frac{76}{12} = 6,33$		$s^2 = \frac{28,66}{12-1} = 2,60$	$\bar{X} = \frac{56}{10} = 5,6$		$s^2 = \frac{16,3}{10-1} = 1,8$

Y sustituimos los valores obtenidos en cada uno de los grupos en la fórmula:

$$\begin{aligned}
 t &= \frac{|6,33 - 5,6|}{\sqrt{\left[\frac{(11)2,6 + (9)1,81}{12 + 10 - 2} \right] \times \left[\frac{1}{12} + \frac{1}{10} \right]}} = \frac{0,73}{\sqrt{\left[\frac{44,89}{20} \right] \times \left[\frac{22}{120} \right]}} = \\
 &= \frac{0,73}{\sqrt{2,2445 \times 0,183}} = \frac{0,73}{0,64} = 1,14
 \end{aligned}$$

6.º Interpretación del valor empírico obtenido en la prueba estadística y conclusiones

Para ello debemos consultar la tabla de la distribución de t , que para un nivel de significación $\alpha = 0,05$, y grados de libertad $= n_A + n_B - 2 = 12 + 10 - 2 = 20$, e hipótesis bilateral, tenemos un valor de t crítico $= 2,086$.

Como $t_e = 1,14 < t_{0,05,20gl} = 2,086$, no nos permite rechazar la hipótesis de nulidad. Por lo tanto debemos concluir que no existen diferencias estadísticamente significativas, al nivel de confianza del 95%, entre ambos grupos. Ello nos aconseja revisar las características técnicas y pedagógicas del material didáctico utilizado en la enseñanza de los contenidos de estudio a través de Internet.

También debemos cuestionar y mejorar la selección de la muestra, ampliar su tamaño, la duración del experimento, con el fin de mejorar la validez interna del diseño, de modo que podamos asegurar que el no rechazo de la hipótesis de nulidad sea debido al tratamiento y no a otras variables extrañas que contaminen nuestro experimento.

Realizamos los análisis mediante el programa informático SPSS, aportándonos las siguientes tablas de resultados:

TABLA 17.3. Estadísticos descriptivos

Estadísticos de grupo					
	GRUPO	N	Media	Desviación típ.	Error típ. de la media
PUNTUAC	1.00	12	6.3333	1.61433	.46602
	2.00	10	5.6000	1.34990	.42687

TABLA 17.4. Prueba t

Prueba de muestras independientes		Prueba de Levene para la igualdad de varianzas		Prueba t para la igualdad de medias						
		F	Sig.	t	gl	Sig. Bilateral	Difere. de medias	Error típ. de la diferencia	95% Intervalo de confianza para la diferencia	
									Inferior	Superior
PUNTUACIONES	Asumen varianzas iguales	.580	.455	1.14	20	.267	.7333	.64274	-.607	2.0761
	Asumen varianzas diferentes			1.16	19.9	.260	.7333	.63198	-.585	2.0516

La prueba de Levene dice que no existen diferencias significativas en las varianzas ($p = 0,455 > 0,05$) y el valor crítico o de significación para la hipótesis bilateral es muy alto, $p = 0,267 > 0,05$, lo que corresponde a un nivel de confianza del 76,3%, no aceptable científicamente para rechazar la hipótesis de nulidad. Confirma los resultados obtenidos por el sistema no informático.

17.4. Resolución de un diseño de investigación de dos grupos relacionados y nivel de medida en la variable dependiente de intervalo mediante la prueba t de Student

El equipo docente de Ciencias está preocupado por el bajo rendimiento de los alumnos y quiere probar la eficacia de un nuevo método de enseñanza para comprobar su incidencia en el rendimiento de los alumnos. Cuenta con dos grupos de clase de 25 y 27 alumnos cada uno. Asigna al azar el método que quiere probar a uno de los grupos, continuando en el otro con el que venía utilizando hasta ahora. Pasado un trimestre de la experiencia aplica a todos los alumnos una prueba objetiva de rendimiento bastante precisa, y para realizar el análisis de los resultados va a trabajar sólo con los resultados de las diez parejas de alumnos; cada miembro de la pareja procede de uno de los grupos, que obtuvieron en una prueba al inicio del curso los resultados más igualados.

Como vemos, el tratamiento ha sido asignado al azar, pero los alumnos para el análisis de los datos han sido seleccionados de acuerdo con el rendimiento obtenido en una prueba previa en la asignatura de Ciencias Naturales, tomando las 10 parejas cuyas puntuaciones están igualadas, ejerciendo de este modo un control sobre posibles variables extrañas que puedan interferir en nuestro experimento, como son los conocimientos previos en la materia sobre la que estamos investigando.

Pasamos a resolver el diseño de investigación anterior:

1.º Esquema y comentario del diseño planteado

Grupos	Selección	Asignación	Pretest	Variable Independiente	Posttest
A (Control)	Emparejamiento	Azar	No	Método tradic.	T_1
B (Exper.)	Emparejamiento	Azar	No	Método nuevo	T_2

Se trata de un diseño cuasi-experimental de dos grupos relacionados, cuya selección de sujetos se ha hecho mediante una prueba previa para emparejarlos, la asignación de los tratamientos se ha realizado al azar, sin pretest, con una sola variable independiente, el método de aprendizaje, con dos niveles, el que se venía utilizando y el nuevo.

2.º Formulación de las hipótesis

Hipótesis de nulidad textual:

H_0 : El método de aprendizaje no influye en los resultados académicos.

Estadísticamente: $\mu_A = \mu_B$

Hipótesis alternativa textual:

H_1 : El nuevo método de aprendizaje producirá mejores resultados académicos que si se sigue con el método tradicional.

Estadísticamente: $\mu_A \neq \mu_B$

El nivel de significación elegido es $\alpha = 0,01$.

Se trata de un contraste unilateral izquierdo, dado que se espera conseguir mejorar el rendimiento con el nuevo método de enseñanza.

3.º *Justificación de la prueba elegida para el contraste de hipótesis*

Por tratarse de un diseño de dos grupos relacionados, y los datos de la variable dependiente medidos por el postest alcanzan un nivel de medida de intervalo (rendimiento medido a través de una prueba objetiva precisa), y supuestas las condiciones de normalidad y homogeneidad de las varianzas, elegiremos una prueba paramétrica, en este caso para muestras pequeñas y grupos relacionados, la «t de Student» cuya fórmula es la siguiente:

$$t = \frac{|\bar{X}_A - \bar{X}_B|}{\sqrt{\frac{\sum D^2 - \frac{(\sum D)^2}{n}}{n(n-1)}}$$

4.º *Aplicación de la prueba estadística a los datos*

Los resultados obtenidos por las parejas de alumnos en la prueba de rendimiento con la que se ha medido la Variable Dependiente, son los que se presentan en el siguiente cuadro de datos:

TABLA 17.5.

Parejas	Grupo A	Grupo B	D	D ²
1	8	7	1	1
2	6	8	-2	4
3	5	6	-1	1
4	2	4	-2	4
5	3	4	-1	1
6	6	5	1	1
7	7	9	-2	4
8	6	5	1	1
9	4	6	-2	4
10	8	9	-1	1
Totales	55	63	-11 + 3 = -8	22

En la fórmula de t para grupos correlacionados, sustituimos los valores:

$$t = \frac{|\bar{X}_A - \bar{X}_B|}{\sqrt{\frac{\sum D^2 - \frac{(\sum D)^2}{n}}{n(n-1)}}} = \frac{|5,5 - 6,3|}{\sqrt{\frac{22 - \frac{64}{10}}{10(10-1)}}} = \frac{0,8}{\sqrt{\frac{15,6}{90}}} = \frac{0,8}{0,416} = 1,92$$

5.º Interpretación del valor empírico obtenido en la prueba y conclusiones

Contrastamos el valor de t empírico, 1,92, con el valor de t crítico de las tablas para hipótesis de una cola, o unilaterales, y con g.l. = $n - 1 = 10 - 1 = 9$, y nivel de significación $\alpha = 0,01$, siendo $t_c = 2,82$. Como el valor de $t_e = 1,92 < t_c = 2,82$, no podemos rechazar H_0 a un nivel de significación de 0,01.

Realizado el análisis mediante el programa informático SPSS nos aporta las siguientes tablas de resultados:

TABLA 17.6. Descriptivos

Estadísticos de muestras relacionadas					
		Media	N	Desviación típ.	Error típ. de la media
Par 1	Pun.Gr 1	5.50	10	2.01384	.63683
	Pun,Gr 2	6,30	10	1.88856	.59722

TABLA 17.7. Correlaciones

Muestras relacionadas				
		N	Correlación	Sig.
Par 1	Pun.Gr 1 y Pun,Gr 2	10	.774	.009

TABLA 17.8. Resultados prueba t

Prueba t para muestras relacionadas									
		Diferencias relacionadas					t	gl	Sig. (bilateral)
		Media	Desviación típica	Error típ. de la media	95% Intervalo de confianza para la diferencia				
					Inferior	Superior			
Par 1	Pun.Gr 1 Pun,Gr 2	-.8000	1.31656	.41633	-1.7418	.1418	-1.922	9	.087

Los resultados obtenidos con el programa informático SPSS confirman los que ya habíamos obtenido aplicando la prueba directamente. La significatividad de α es de 0,087, lo que equivale a su correspondiente nivel de confianza del 91,3%, no alcanzando el nivel de significación del 0,05%, que es el aceptable científicamente para poder rechazar la hipótesis de nulidad H_0 .

Las reflexiones sobre los resultados aconsejan que se amplíe la duración de la experiencia, un trimestre más, y, sobre todo, el tamaño de la muestra productora de datos, que se procurará ampliar a grupos superiores a 20 alumnos cada uno, dado que es factible puesto que los grupos iniciales dijimos que estaban formados por 25 y 27 alumnos respectivamente.

17.5. Resolución de un diseño de investigación de dos grupos independientes y nivel de medida en la variable dependiente ordinal mediante la prueba U de Mann-Whitney

En un centro de educación de adultos, dependiente de la administración, el equipo de profesores quiere fomentar la afición por la lectura en los adultos. Para ello han pensado realizar una experiencia, centrada en la mejora de la comprensión lectora, y probar sus posibles ventajas en uno de los dos grupos que asisten al centro, elegido al azar.

La experiencia consiste en lo siguiente: dos sesiones de clase por semana se van a dedicar al comentario de libros de lectura. En la primera sesión el profesor les cuenta el resumen de un capítulo, señalando el argumento, el estilo narrativo, su cercanía o lejanía con los ambientes próximos al grupo de alumnos, etc. Los alumnos del grupo se responsabilizan para que, en cada siguiente sesión, cada uno de ellos presente el resumen de un capítulo, o de un libro corto.

Pasado un trimestre realiza una prueba escrita de Verdadero o Falso (V-F) a ambos grupos de alumnos para comprobar como realizan el resumen de un capítulo corto de unas dos, o tres hojas del libro, pidiéndoles que contesten a una prueba para ver su comprensión lectora y capacidad para resumir las ideas del texto. La prueba está formada por 50 preguntas de verdadero o falso, aportando unos resultados de 0 a 50 puntos, con la que se pretende comprobar cuántas de las ideas fundamentales del capítulo han sido entendidas, así como otras cuestiones sobre el interés y gusto por la lectura.

A partir de la corrección de la prueba, el profesor toma los resultados de un pequeño número de sujetos seleccionados de forma aleatoria, de cada uno de los grupos:

Grupo de control (A): 12 13 14 15 16 25 32

Grupo experimental (B): 15 16 20 24 26 30 34 38

Después pasa a analizar la información, las características del diseño y la prueba más adecuada para el contraste de hipótesis, su aplicación e interpretación de los resultados, al nivel de confianza del 95% ($\alpha = 0,05$), siguiendo el esquema utilizado en los ejercicios anteriores:

1.º Esquema y comentario del diseño de investigación

Grupos	Selección	Asignación	Pretest	Variable Independiente	Posttest
A (Control)	Azar	Azar	No	Método tradic.	T_1
B (Exper.)	Azar	Azar	No	Fomento de la lectura	T_2

Estamos ante un diseño cuasi-experimental, de dos grupos independientes. El grupo A actúa como control y el B es el experimental. La selección de los sujetos ha sido al azar, y asimismo la asignación de los tratamientos a los grupos. No hay medida pretest. La V.I. es el nuevo método para el desarrollo de la comprensión verbal e interés por la lectura. La V.D. «comprensión y expresión verbal e interés por la lectura» viene medida mediante una prueba de Verdadero o Falso elaborada por el profesor que no reúne el nivel de medida de intervalo, por realizar preguntas sobre gustos, aficiones, aunque también cuestiones sobre comprensión del texto, pero poco precisas para alcanzar un nivel de intervalo.

2.º Formulación de las hipótesis

Al estar trabajando con un nivel de medida ordinal, no se puede realizar la media, sí la mediana, por lo que las comparaciones entre los grupos se realizará a partir de las medianas.

Hipótesis de nulidad textual:

H_0 : El método aplicado por el profesor no influye en el desarrollo de la comprensión y el interés por la lectura.

Hipótesis estadística:

$$\text{Med}_A = \text{Med}_B.$$

Hipótesis alternativa textual:

H_1 : El método aplicado por el profesor influye en el desarrollo de la comprensión y el interés por la lectura.

Hipótesis estadística:

$$\text{Med}_A \leq \text{Med}_B.$$

Se trata de una hipótesis unilateral, dado que el equipo docente espera mejorar con el nuevo método de trabajo, la comprensión lectora. El nivel de significación fijado es el de $\alpha = 0,05$.

3.º *Cuál sería la prueba estadística más apropiada para el contraste? Justifique su elección*

Por tratarse de un diseño de dos grupos independientes, cuyos datos en el postest (V.D.) alcanzan un nivel de medida ordinal, elegiremos una prueba no paramétrica. En este caso la prueba «U» de Mann-Witney, para grupos independientes y muestras muy pequeñas, menores o iguales a 9, ($n < 9$).

4.º *Aplicación de la prueba seleccionada*

Lo primero que se hace es ordenar todas las puntuaciones de menor a mayor, conservando la identidad del grupo:

TABLA 17.8.

Puntuaciones y grupos															
X_i	12	13	14	15	15	16	16	20	24	25	26	30	32	34	38
Grupo	A	A	A	A	B	A	B	B	B	A	B	B	A	B	B
					x	1				2			3		
							x			1			2		
								x		1			2		
									x	1			2		
											x		1		
												x	1		

Por tratarse de muestras pequeñas, por ser $n \leq 9$, el valor de U se obtiene sumando el número de veces que una puntuación del grupo experimental B precede a una puntuación grupo de control A. De este modo, tenemos que el valor de U es:

$$U = 3 + 2 + 2 + 2 + 1 + 1 = 11$$

5.º *Interpretación de los resultados*

La prueba de U para muestras muy pequeñas ($n \leq 9$) se interpreta buscando en las tablas de U la probabilidad asociada al valor de U empírico obtenido, teniendo en cuenta el número de sujetos de cada uno de los grupos (n_1 y n_2). Si la probabilidad asociada al valor empírico de U , que muestra las tablas es menor que la probabilidad, o valor de significación α fijado, entonces podemos rechazar la hipótesis de nulidad y afirmar que las diferencias entre los grupos se deben a los efectos de la V.I. o tratamiento.

En las tablas de U , con $n_1 = 7$ y $n_2 = 8$, la probabilidad asociada a un valor de U empírico igual a 11, es 0,027, y la probabilidad crítica que hemos determinado para poder rechazar H_0 es de 0,05. Como:

$p = 0,027 < 0,05$, podemos rechazar H_0 y afirmar la hipótesis alternativa.

6.º Conclusiones y recomendaciones

Aunque estos resultados permiten rechazar H_0 para un nivel de confianza del 95%, el equipo docente decide continuar la experiencia para confirmar los resultados.

7.º Revisión del diseño: continuación con la experiencia y ampliación de la muestra

El equipo docente decide continuar con la experiencia durante al menos un trimestre más, y, asimismo, ampliar la muestra para el análisis de los datos. Al finalizar el trimestre, les aplica otra prueba similar, por la que se obtienen los siguientes resultados:

Grupo A: 13 14 15 16 17 18 19 20 22 23 27 30 33 35 37 38 41 42

Grupo B: 16 18 19 22 23 24 25 27 28 33 36 38 40 41 43 44 45 47 48 49

8.º Esquema y comentario del diseño

El esquema del diseño es del mismo tipo que el de la experiencia previa, variando sólo el tamaño de la muestra, por lo que se puede aplicar la misma prueba estadística no paramétrica, pero en este caso para muestras con $n \geq 9$ y ≤ 20 , o de tamaño intermedio.

9.º Aplicación de la prueba U para muestras con $n \geq 9$ y ≤ 20

Como n_2 no es > 20 , utilizamos la ecuación para muestras de tamaño intermedio, que se aplicará a cada uno de los grupos, y cuya fórmula es la siguiente:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1; \quad U' = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2$$

- 1.º Lo primero que hay que hacer es ordenar el total de puntuaciones de ambos grupos, de menor a mayor, sin que cada puntuación pierda la identidad del grupo del que procede, y le asignamos el rango, u orden que le corresponde, del modo siguiente, que podemos ver en la tabla:

TABLA 17.9

X_i	13	14	15	16	16	17	18	18	19	19	20	22	22	23	23	24	25	27	27
Grupo	A	A	A	A	B	A	A	B	A	B	A	A	B	A	B	B	B	A	B
Rango	1	2	3	4,5	4,5	6	7,5	7,5	9,5	9,5	11	12,5	12,5	14,5	14,5	16	17	18,5	18,5
X_i	28	30	33	33	35	36	37	38	38	40	41	41	42	43	44	45	47	48	49
Grupo	B	A	A	B	A	B	A	A	B	B	A	B	A	B	B	B	B	B	B
Rango	20	21	22,5	22,5	24	25	26	27,5	27,5	29	31,5	31,5	32	33	34	35	36	37	38

Cuando aparecen las mismas puntuaciones, como en el caso de 16, 16, a cada puntuación le corresponde un rango, pero al tener el mismo valor se le asignan el valor medio de los rangos que les corresponderían si fueran distintas; en este caso, les correspondería los rangos 4.º y 5.º, que al promediarlo queda en 4,5.º

- 2.º Una vez ordenadas las puntuaciones y asignados los rangos, sumamos los rangos de cada grupo, obteniendo el sumatorio de los rangos del grupo A, y del B:

$$\Sigma R_1 = 1 + 2 + 3 + 4,5 + 4,5 + 6 + 7,5 + 7,5 + 9,5 + 9,5 + 11 + 12,5 + 12,5 + 14,5 + 14,5 + 16 + 17 + 18,5 + 18,5 + 20 + 21 + 22,5 + 22,5 + 24 + 25 + 26 + 27,5 + 27,5 + 29 + 31,5 + 31,5 + 32 + 33 + 34 + 35 + 36 + 37 + 38 = 274,5$$

$$\Sigma R_2 = 4,5 + 7,5 + 9,5 + 12,5 + 14,5 + 16 + 17 + 18,5 + 20 + 22,5 + 25 + 27,5 + 29 + 31,5 + 33 + \dots + 38 = 468,5$$

- 3.º Sustituimos las incógnitas de la ecuación U por sus correspondientes valores numéricos:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - \Sigma R_1; \quad U' = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - \Sigma R_2$$

Conocido el valor de un U, el otro puede calcularse también de este modo:

$$U = n_1 \cdot n_2 - U'$$

$$U = 18 \times 20 + \frac{18(18+1)}{2} - 274,5 = 360 + \frac{342}{2} - 274,5 = 531 - 274,5 = 256,5;$$

$$U' = 18 \times 20 + \frac{20(20+1)}{2} - 468,5 = 360 + \frac{420}{2} - 468,5 = 570 - 468,5 = 101,5$$

10.º Interpretación de los resultados en el segundo caso (nivel de significación fijado: $\alpha = 0,05$)

El valor de U empírico más pequeño se compara con el valor crítico de U de las Tablas; si el valor de $U_e < U_c$, nos permite rechazar H_0 al nivel de significación fijado.

En nuestro caso, para una prueba de una cola, hipótesis unilateral, $\alpha = 0,05$, $n_1 = 18$ y $n_2 = 20$, el valor crítico que nos da la tabla es: $U_c = 123$. Como el valor de U empírico más pequeño es menor que el valor de U crítico:

$$101,5 < 123$$

Podemos rechazar H_0 al nivel de confianza del 95 %.

11.º Conclusiones y recomendaciones

Los resultados apoyan la hipótesis del equipo docente, dado que existen diferencias significativas, al nivel del 95%, a favor de los resultados obtenidos por el grupo experimental en la prueba sobre comprensión lectora y afición por la lectura.

Sobre los resultados obtenidos en el primer supuesto, que permitían rechazar la hipótesis de nulidad, y a la luz de los resultados obtenidos tras la prolongación del nuevo método para la mejora de la comprensión lectora, vemos que ha podido incidir en mejorar los resultados en el grupo experimental, tal como esperaba el equipo docente. La ampliación de la muestra ha confirmado los primeros resultados. Y es que con grupos tan reducidos, tan pequeños, no siempre se pueden realizar análisis estadísticos concluyentes. Es mejor poner en duda la hipótesis del investigador, puesto que se pueden mejorar las condiciones del diseño y replicar la investigación, que rechazar la hipótesis nula cuando es verdadera y dar por válida la hipótesis del investigador. De todos modos, cuando sea posible, es mejor trabajar con grupos que se distribuyan siguiendo el modelo de la curva normal, en torno a 30 sujetos, y con pruebas, o instrumentos de medida, bien elaborados, válidos y fiables, que permitan realizar el contraste de hipótesis con pruebas estadísticas potentes para rechazar hipótesis de nulidad cuando sean falsas.

El caso de muestras con $n > 20$

Cuando se trabaja con muestras grandes ($n > 20$) se calcula el valor de U como en el supuesto anterior (para muestras entre 9 y 20 sujetos):

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1; \quad U' = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2$$

Después se aplica la prueba z para la diferencia de medias, debiendo hallar el valor de la media de U y su desviación, con las fórmulas siguientes:

$$\bar{U} = \frac{n_1 n_2}{2}; \quad s_U = \sqrt{\frac{n_1 \times n_2 (n_1 + n_2 + 1)}{12}}; \quad z_U = \frac{\sum U + \bar{U}}{s_U}$$

Sólo si el valor de la probabilidad asociada a z_U fuera menor o igual que el valor de α fijado, podremos rechazar H_0 .

17.6. Resolución de un diseño de investigación de dos grupos relacionados y nivel de medida en la variable dependiente ordinal mediante la prueba T de Wilcoxon

El equipo docente de la asignatura piensa que la asistencia a las tutorías en los Centros Asociados de la UNED puede ejercer algún tipo de influencia en la actitud de los estudiantes hacia la asignatura. Para comprobarlo decide realizar la siguiente experiencia con dos grupos: uno (A) que actuará como control, apenas mantiene contactos con el profesor tutor, mientras que el otro (B) suele asistir a las tutorías con frecuencia. Ahora bien, con el fin de controlar mejor las variables extrañas se aplica una prueba previa a todos los sujetos de la muestra con cuyos resultados se forman parejas, permaneciendo cada uno de los componentes dentro de su grupo original.

Transcurrido el primer cuatrimestre se recogen datos mediante una hoja, o pauta, de observación del profesor y la aplicación de una escala de actitud hacia la asignatura. Con las nueve parejas con mayor equivalencia, procediendo cada miembro de la pareja de cada uno de los grupos, conforma la siguiente tabla de datos:

TABLA 17.11.

Parejas	X_A	X_B
1.ª	24	25
2.ª	43	40
3.ª	14	18
4.ª	30	35
5.ª	27	20
6.ª	33	41
7.ª	16	25
8.ª	26	36
9.ª	28	40

Para resolver el diseño, seguimos el esquema que venimos utilizando, respondiendo a las siguientes cuestiones:

1.º Diseño planteado: Esquema y comentario

Grupos	Selección	Asignación	Pretest	Variable Independiente	Posttest
A (Control)	Emparejamiento	No azar	No	Asist. infrec.	T_1
B (Exper.)	Emparejamiento	No azar	No	Asist. habitual	T_2

Se trata de un diseño cuasi-experimental de dos grupos (el A actúa como control y el B es el experimental) relacionados (seleccionados mediante emparejamiento), con asignación no aleatoria, puesto que se busca la equivalencia de sujetos procedentes de cada grupo, los que asisten frecuentemente y los que no lo hacen. No existe medida pretest. La Variable Independiente se considera la «asistencia a las tutorías en los Centros Asociados», fomentando un clima de clase participativo en el grupo B. La Variable Dependiente es la «actitud de los alumnos hacia la asignatura» medida a través de un cuestionario sobre la actitud hacia la asignatura y la observación del profesor tutor.

2.º Formulación de las hipótesis

Hipótesis de nulidad textual:

H_0 : La asistencia habitual a las sesiones de tutoría en la asignatura no influye en la actitud de los estudiantes hacia la materia.

Hipótesis estadística:

$$H_0: \text{Med}_A = \text{Med}_B.$$

Hipótesis alternativa textual:

H_1 : La asistencia habitual a las sesiones de tutoría en la asignatura influye en la actitud de los estudiantes hacia la materia.

Hipótesis estadística:

$$H_1: \text{Med}_A \neq \text{Med}_B.$$

Se trata de una hipótesis bilateral, dado que no se explicita una dirección determinada en la hipótesis.

El nivel de significación fijado es el de $\alpha = 0,05$.

3.º *¿Cuál sería la prueba estadística más apropiada para el contraste? Justifique su elección*

Por tratarse de un diseño de dos grupos relacionados, cuyos datos en el postest alcanzan un nivel de medida ordinal, recogidos mediante un cuestionario de opinión y la observación directa del profesor tutor, debemos optar por una prueba no paramétrica, en este caso la prueba «T» de Wilcoxon.

4.º *Aplicación de la prueba seleccionada e interpretación de su valor ($\alpha = 0,05$)*

A partir de la tabla de datos, aplicamos la prueba T de Wilcoxon procediendo del siguiente modo:

- 1.º Ordenación de las parejas en función de su menor diferencia entre las puntuaciones obtenidas, cálculo de la diferencia y signo de la misma, y asignación del rango. Así obtenemos la siguiente tabla:

TABLA 17.12.

Parejas	X_A	X_B	Diferencia y signo	Rango
1.º	24	25	+1	1
2.º	43	40	-3	-2
3.º	14	18	+4	3
4.º	30	35	+5	4
5.º	27	20	-7	-5
6.º	33	41	+8	6
7.º	16	25	+9	7
8.º	26	36	+10	8
9.º	28	40	+12	9

Como podemos comprobar, los pares de datos vienen ordenados atendiendo a la menor diferencia entre ambos miembros de la pareja. Así la primera pareja presenta una diferencia de 1 punto positivo, el segundo miembro de la pareja tiene una puntuación superior en un punto, y la 9.ª de 12, también superior, siendo un valor positivo. Por lo tanto, hallamos la diferencia entre las puntuaciones de las parejas, y asignamos el signo correspondiente, + o -. Indicamos el orden de los rangos conservando el signo, confirmando que las parejas estaban bien ordenadas en función de su menor diferencia de puntuación entre ellas.

- 2.º Aplicamos la prueba T de Wilcoxon, que por tratarse de muestras pequeñas, 9 parejas, $n \leq 25$, se procede a sumar, por un lado, el valor de las diferencias con signo positivo, y, por otro lado, el valor de las diferencias con signo negativo, o lo que es lo mismo el sumatorio de las diferencias positivas y el sumatorio de la diferencias negativas:

$$\Sigma D_+ = 1 + 3 + 4 + 6 + 7 + 8 + 9 = 38;$$

$$\Sigma D_- = 2 + 5 = 7;$$

El valor de T , (o W) viene dado por el valor de la suma de las diferencias menor. Luego:

$$T = 7$$

5.º Interpretación de los resultados

El valor teórico de T en la Tabla para $\alpha = 0,05$, es: $T_t = 6$. Valores empíricos de T iguales o menores que los de las Tablas, permiten el rechazo de H_0 , con riesgo de error tipo I máximo igual al valor del nivel de significación elegido. Como el valor empírico es $T_e = 7 \geq T_t = 6$, no permite rechazar H_0 .

6.º Conclusiones y recomendaciones

Como con estos valores no se puede rechazar H_0 (a un nivel de confianza del 95%), el equipo docente decide continuar con la experiencia un trimestre más, y aumenta la muestra productora de datos hasta 28 parejas. El diseño no varía, salvo en lo que se refiere al tamaño de la muestra y la duración de la experiencia. Vuelve a aplicarles el cuestionario de actitud, que arroja los siguientes resultados, que se muestran ya ordenados en función de la menor diferencia entre las parejas cuyos miembros proceden de ambos grupos de referencia:

TABLA 17.13.

Parejas	X_A	X_B	Parejas	X_A	X_B
1.º	30	30	15.º	20	27
2.º	42	42	16.º	39	31
3.º	20	19	17.º	38	47
4.º	34	35	18.º	15	24
5.º	16	18	19.º	30	39
6.º	43	40	20.º	20	30
7.º	37	34	21.º	25	35
8.º	25	28	22.º	15	26
9.º	14	10	23.º	35	23
10.º	8	13	24.º	23	36
11.º	10	15	25.º	28	42
12.º	36	30	26.º	15	30
13.º	42	36	27.º	32	48
14.º	25	31	28.º	20	40

7.º *Aplicación de la prueba al segundo supuesto, ($\alpha = 0,05$ y $\alpha = 0,01$)*

Si el número de parejas con signo es mayor o igual a 25, ($n \geq 25$), se consideran muestras grandes, debiendo utilizar la ecuación de z para la diferencia de medias, de muestras grandes, cuya distribución se asemeja a la de z :

$$z_T = \frac{T - \bar{T}}{s_T}$$

Siendo T el valor menor de la suma de las diferencias entre las puntuaciones obtenidas, teniendo en cuenta el signo, debiendo hallar el valor de la media de T y su desviación, de acuerdo con las siguientes fórmulas:

$$\bar{T} = \frac{\sum D_+ + \sum D_-}{2} = \frac{n(n+1)}{4}$$

$$s_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Siendo:

ΣD_+ : El sumatorio de los valores de las diferencias entre cada pareja con signo positivo.

ΣD_- : Sumatorio de las parejas con signo negativo.

n : número de parejas con signo.

Procedemos, como antes,

1.º Hallamos las diferencias entre cada miembro de la pareja, y asignamos el signo:

TABLA 17.14.

Parejas	X_A	X_B	Diferencia y signo	Rango y signo	Parejas	X_A	X_B	Diferencia y signo	Rango y signo
1.ª	30	30	0	—	15.ª	20	27	+7	13
2.ª	42	42	0	—	16.ª	39	31	-8	-14
3.ª	20	19	-1	-1,5	17.ª	38	47	+9	16
4.ª	34	35	+1	1,5	18.ª	15	24	+9	16
5.ª	16	18	+2	3	19.ª	30	39	+9	16
6.ª	43	40	-3	-5	20.ª	20	30	+10	18,5
7.ª	37	34	-3	-5	21.ª	25	35	+10	18,5
8.ª	25	28	+3	5	22.ª	15	26	+11	20
9.ª	14	10	-4	-7	23.ª	35	23	-12	-21
10.ª	8	13	+5	8,5	24.ª	23	36	+13	22
11.ª	10	15	+5	8,5	25.ª	28	42	+14	23
12.ª	36	30	-6	-11	26.ª	15	30	+15	24
13.ª	42	36	-6	-11	27.ª	32	48	+16	25
14.ª	25	31	-6	11	28.ª	20	40	+20	26

2.º Hallamos el sumatorio de las diferencias con signo positivo, y el sumatorio de las diferencias con signo negativo:

$$\Sigma D_+ = 1,5 + 3 + 5 + 8,5 + 8,5 + 11 + 13 + 16 + 16 + 18,5 + 18,5 + 20 + 22 + 23 + 24 + 25 + 26 = 275,5$$

$$\Sigma D_- = 1,5 + 5 + 5 + 7 + 11 + 11 + 14 + 21 = 75,5$$

3.º Calculamos la media, la desviación típica y el valor de z:

$$\bar{T} = \frac{n(n+1)}{4} = \frac{26(26+1)}{4} = 175,5$$

$$s_r = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{26(26+1)(2 \times 26+1)}{24}} = 39,37$$

$$z_{TD+} = \frac{T - \bar{T}}{s_r}; z_{TR+} = \frac{275,5 - 175,5}{39,37} = +2,54;$$

$$z_{TD-} = \frac{75,5 - 175,5}{39,37} = -2,54$$

4. Interpretamos los resultados (hipótesis bilateral, significación 0,05 y 0,01):

Consultamos los valores de z , en su tabla:

Para $\alpha = 0,05$, $z = 1,96$; como $2,54 > 1,96$, nos permite rechazar H_0 .

Para $\alpha = 0,01$, $z = 2,58$; como $2,54 < 2,58$, no se puede rechazar H_0 .

8.º Conclusiones y recomendaciones

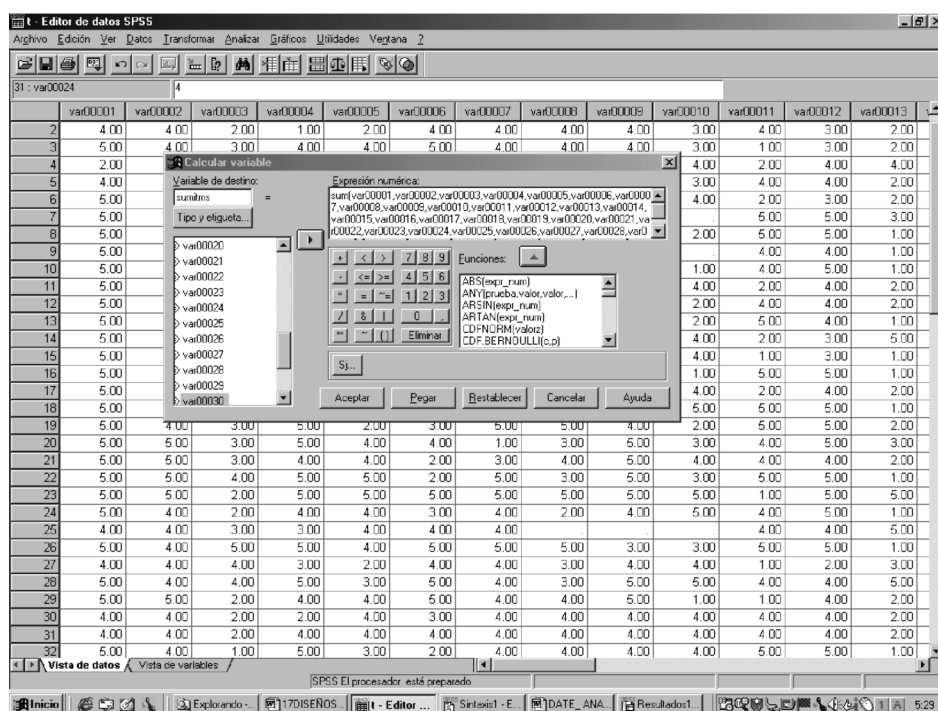
En la experiencia primero no se pudo para rechazar H_0 al nivel de confianza del 95%. Por ello el investigador decidió ampliar la muestra productora de datos, lo que le permitió rechazar H_0 al nivel de confianza del 95% que se considera aceptable en Ciencias Sociales. Se deduce, por tanto, la conveniencia de realizar nuevos estudios tratando de mejorar las condiciones de los diseños; la elaboración de instrumentos más precisos, y la ampliación de los grupos, podrá aportar información más concluyente para poder realizar recomendaciones a los profesores tutores. En cuanto a las recomendaciones, irían en la línea de reforzar las actividades que incentiven la asistencia de los alumnos a las tutorías de la asignatura.

17.7. Ejercicios de autoevaluación

Vamos a mostrar un ejemplo práctico que oriente en el proceso de realización de una intervención educativa, y su valoración, dentro de un proceso de investigación utilizando un diseño cuasi-experimental.

Un orientador de un Instituto de Educación Secundaria educador quiere probar la incidencia de un método para el desarrollo de habilidades sociales en los alumnos de 3.º de ESO. Para ello decide trabajar con uno de los grupos de 3.º, formado por 21 alumnos, utilizando un programa de actividades, centrado en la resolución de problemas de convivencia y respeto a las diferencias de opiniones, valores y comportamientos de los compañeros, de las personas adultas y de los niños, que mediante el role-playing, o desempeño de roles, simula situaciones reales de debates sobre problemas que viven y temas, en los que aparecen las figuras de compañeros, familia, padres y hermanos, y otros personajes representativos de su entorno.

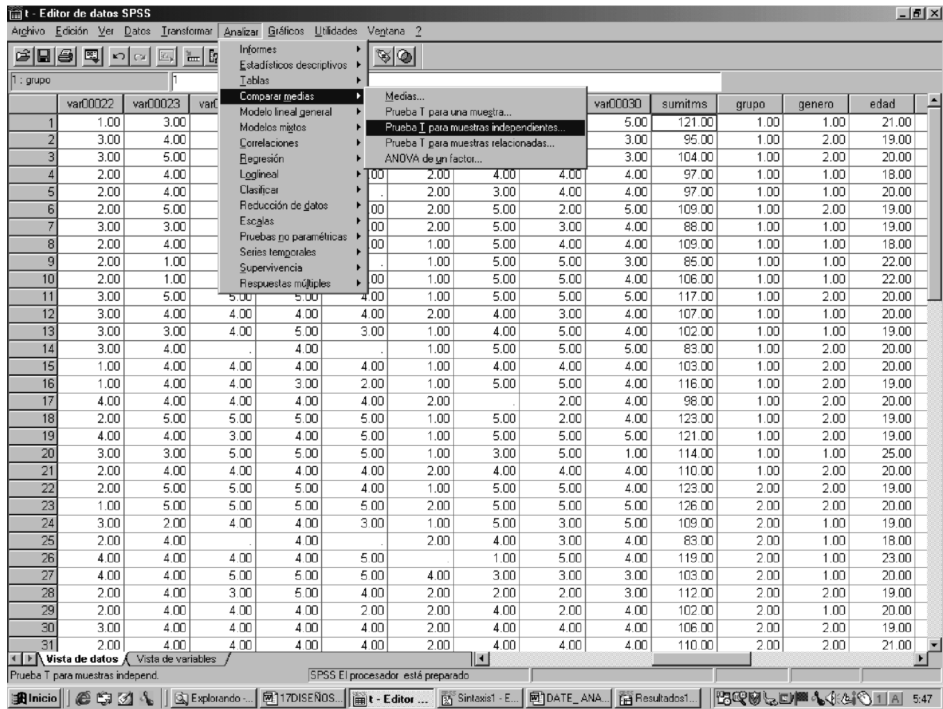
El programa lo aplica una hora semanal, durante los dos primeros meses del curso. Durante la duración del tratamiento el orientador va tomando notas, mediante hojas de registro grupal de observación, durante la hora de actividad. Pasados los dos meses decide aplicarles un cuestionario sobre actitudes y valores integrado por 30 ítems, representativos de los objetivos pretendidos en las actividades realizadas, en un formato de escala tipo Likert, con una respuesta entre 1 a 5, pidiéndoles su grado de acuerdo con el ítem. El mismo cuestionario se les aplica a otro de los grupos de 3.º de ESO, formado por 20 alumnos, que no ha trabajado con el programa. La corrección de los cuestionarios ha proporcionado los datos que introducimos en el editor de datos del SPSS.



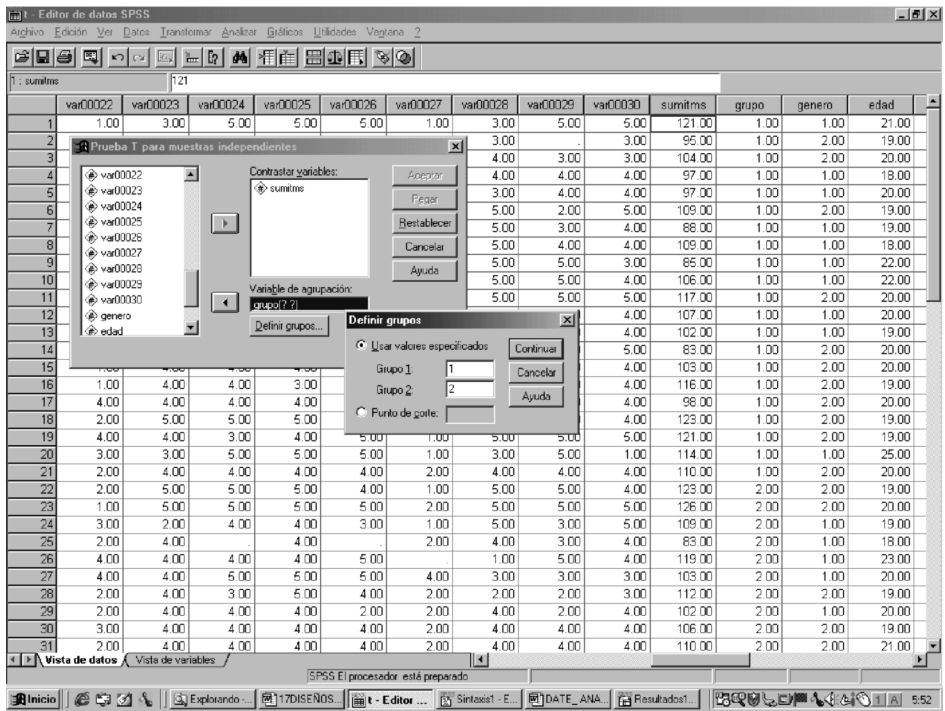
Como puede observarse hemos introducido los datos obtenidos por cada uno de los sujetos en cada uno de los 30 ítems, y en la barra de herramientas hemos pulsado transformar>calcular, y nos da el cuadro de diálogo que se ve dentro de la pantalla del fichero de datos. Se trata de obtener la suma total obtenida por cada uno de los alumnos, o casos (filas) en el cuestionario. Para ello nombramos en el cuadro de diálogo la variable destino, que llamamos sumitms, y en el cuadro expresión numérica selec-

cionamos la operación que queremos hacer, que viene sintetizada como SUM, e introducimos las variables (columnas, valor de cada ítem) que queremos sumar, separadas por comas cada una de ellas, y encerradas en un paréntesis. Quedaría así, SUM(var0001,var0002,...,var0031) como puede observarse en el cuadro ‘Expresión numéricas’. Después pulsamos la tecla aceptar y obtenemos la nueva variable sumitms con la suma de los ítems de cada sujeto.

Los nuevos valores son los resultados obtenidos por cada uno de los alumnos, que como sabemos, los 21 primeros pertenecen al grupo de control y los 20 restantes, al grupo experimental. Ahora, en la columna siguiente a la de sumitms, indicamos el grupo al que cada puntuación pertenece, mediante 1, al de control, y 2 al experimental. Y marcando en la barra de herramientas Analizar>comparar mediar>T grupos independientes, nos da el siguiente cuadro:



Que nos aportará otro de diálogo:



Marcamos la variable sumitms que va a ser objeto del análisis, y la de grupo, indicando sus valores, 1 (que era el de control) y 2 (que era el experimental), y pulsamos continuar. Y aceptar, dándonos en el archivo de resultados las siguientes tablas:

Estadísticos de Grupo					
	GRUPO	N	Media	Desviación típica	Error típico de la Media
SUMITMS	1.00	21	105.0000	11.63615	2.53922
	2.00	20	110.0000	10.61776	2.37420

Observando los estadísticos descriptivos, vemos que la media del grupo de control es 105, y la del grupo experimental de 110. No parece existir mucha diferencia.

		Prueba de Levene para la igualdad de las varianzas		Prueba t para la igualdad de medias						
		F	Sig.	t	df	Sig. (2-colas)	Diferencia de medias	Error típico de la diferencia	Intervalo de confianza para la diferencia del 95%	
									Por debajo	Sobre
SUMITMS	Equal variances assumed	.408	.527	-1.435	39	.159	-5.0000	3.48422	-12.04750	2.04750
	Equal variances no assumed			-1.438	38.933	.158	-5.0000	3.47267	-12.03181	2.03181

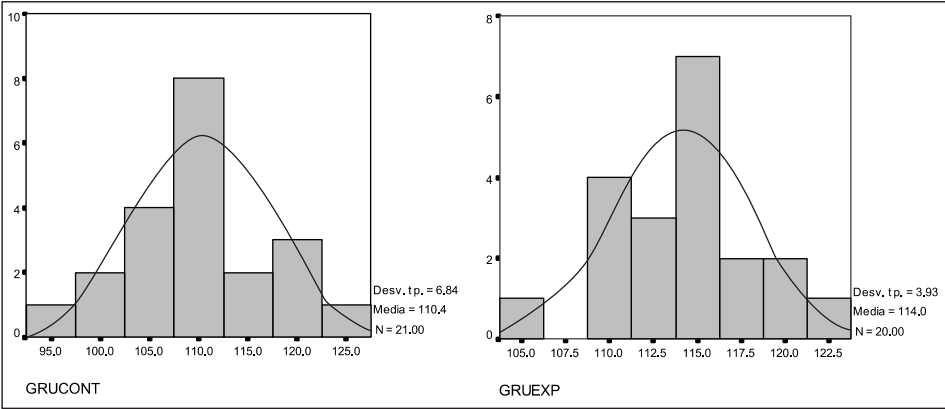
La prueba de Levene nos indica que no existen diferencias significativas entre las varianzas de los grupos, lo cual apoya la estrategia de aplicar la prueba paramétrica t, que exige igualdad de las varianzas entre los grupos. Sin embargo, comprobamos que no existen diferencias significativas entre los grupos en la prueba t, la probabilidad asociada al valor de t empírico es $p = 0,159 > 0,05$. Por ello se decide continuar con la experiencia durante el segundo trimestre, y se les vuelve a aplicar un cuestionario muy similar a ambos grupos.

Aplicada la prueba t se obtienen los siguientes resultados:

Estadísticos de Grupo					
	GRUPO	N	Media	Desviación típica	Error típico de la Media
SUMITMS	1.00	21	110.3810	6.84453	1.49360
	2.00	20	113.9500	3.92663	.87802

Los descriptivos indican poca diferencia entre las medias, pero sin embargo si se da diferencias en las desviaciones típicas en el error típico, mucho menores en el grupo experimental.

Vamos a realizar unos gráficos de las puntuaciones de cada uno de los grupos, concretamente el histograma con curva normal. Para ello hemos formado dos nuevas variables de cada grupo, con los resultados totales de los ítems, y hemos pulsado en la barra de herramientas Gráficos>histograma y marcada curva normal.



Los histogramas nos dan una información que enriquece la dada por los estadísticos descriptivos.

		Prueba de Levene para la igualdad de las varianzas		Prueba t para la igualdad de medias						
		F	Sig.	t	df	Sig. (2- colas)	Diferencia de medias	Error típico de la diferencia	Intervalo de confianza para la diferencia del 95%	
									Por debajo	Sobre
SUMITMS	Equal variances assumed	3.980	.053	-2.034	39	.049	-3.5690	1.75457	-7.11800	-.02010
	Equal variances no assumed			-2.060	32.168	.048	-3.5690	1.73256	-7.09743	-.04066

La prueba de Levene nos indica que no existen diferencias significativas entre las varianzas de los grupos, $p = 0,53 > 0,05$. Y en la prueba t, comprobamos que existen diferencias significativas $0,49 < 0,05$.

Complete el alumno los aspectos que faltan de la investigación, como son formular con precisión el problema, la formulación de las hipótesis, la concreción de las variables, la discusión crítica del diseño, la valoración pedagógica de los resultados y las posibles recomendaciones.

17.8. Solución a los ejercicios de autoevaluación

Consulte los temas del libro para contestar a las preguntas.

17.9. Referencias bibliográficas

JIMÉNEZ FERNÁNDEZ, M. C.; LÓPEZ-BARAJAS, E. y PÉREZ JUSTE, R. (1987). *Pedagogía Experimental II*. Madrid: UNED.

WONNACOTT, T. H. y WONNACOTT, R. J. (1981). *Introducción a la estadística*. México: Limusa.

TEMA 18

Resolución de diseños de investigación de dos y más grupos. El análisis de varianza

ESQUEMA

- 18.1.** El concepto de varianza y sus tipos
 - 18.1.1. Definición de varianza
 - 18.1.2. La varianza de la población y la varianza de las muestras
 - 18.1.3. Tipos de varianza
- 18.2.** La resolución del análisis de la varianza
 - 18.2.1. Ecuaciones para la resolución del análisis de varianza
 - 18.2.2. Interpretación del valor de «F»
 - 18.2.3. Las pruebas de Tukey y Scheffé para el contraste a posteriori
- 18.3.** Resolución de los diseños de investigación de tres grupos independientes y nivel de medida en la variable dependiente de intervalo mediante el análisis de la varianza
- 18.4.** Resolución de un diseño de investigación de tres grupos independientes y nivel de medida en la variable dependiente ordinal, mediante la prueba de Kruskal-Wallis
- 18.5.** Ejercicios de autoevaluación
- 18.6.** Solución a los ejercicios de autoevaluación
- 18.7.** Referencias bibliográficas

OBJETIVOS

- ✓ Conocer y comprender el concepto de varianza y sus tipos.
- ✓ Conocer y comprender los procedimientos generales en la resolución de los diseños de investigación de tres grupos.
- ✓ Conocer y comprender los procedimientos seguidos en la resolución de los diseños de investigación de tres grupos independientes y nivel de medida de intervalo.
- ✓ Conocer y comprender los procedimientos seguidos en la resolución de los diseños de investigación de tres grupos independientes y nivel de medida ordinal.
- ✓ Conocer y comprender los procedimientos seguidos en la resolución de los diseños de investigación de tres grupos relacionados y nivel de medida de intervalo.
- ✓ Conocer y comprender los procedimientos seguidos en la resolución de los diseños de investigación de tres grupos relacionados y nivel de medida ordinal.

18.1. El concepto de varianza y sus tipos

Estudiamos la varianza, junto con las otras medidas de variabilidad, como la desviación típica, en el tema referido a la estadística descriptiva. En este apartado vamos a revisar diferentes tipos de varianza necesarios para la resolución de diseños de investigación de dos o más grupos mediante la prueba estadística del análisis de varianza.

18.1.1. Definición de varianza

La varianza se puede definir como la suma de las puntuaciones diferencias con respecto a su media al cuadrado dividida entre el número de casos de dicho conjunto, o dicho de otro modo, como el promedio de todas las puntuaciones diferenciales al cuadrado:

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

También se puede definir la varianza como el promedio del sumatorio de las puntuaciones individuales al cuadrado menos la media al cuadrado:

$$s^2 = \frac{\sum X_i^2}{n} - \bar{X}^2$$

La varianza es una medida de la dispersión del conjunto de puntuaciones de un grupo, y es un indicador de la homogeneidad de las puntuaciones del grupo. Para comparar las varianzas de dos grupos se divide la varianza mayor entre la varianza menor. Si el cociente resultante es bajo (1 o cercano a 1) las diferencias de los grupos en cuanto a variabilidad es baja, y a la inversa la variabilidad será alta si se aleja mucho del valor 1.

18.1.2. La varianza de la población y la varianza de las muestras

La varianza de la población es la correspondiente al universo o población de mediciones. Podrá conocerse si se conoce el total de las puntuaciones de la población. Si no se conocen todas estas puntuaciones, esta varianza se estimará al calcular la varianza de una o más muestras de dicho universo o población. La varianza de la población se representa con el símbolo σ^2 , que es la desviación típica de la población al cuadrado.

Vimos en el tema dedicado a la estimación de parámetros que los valores de los parámetros poblacionales se representan con letras griegas, para diferenciarlos de los estadísticos de las muestras que se representan con sus equivalentes letras latinas.

La varianza de las muestras se obtiene a partir de los análisis estadísticos de las muestras. Si las muestras son amplias y aleatorias los estadísticos hallados en ellas serán buenos estimadores de los parámetros de la población.

18.1.3. Tipos de varianza

Pasamos a definir los diversos tipos de varianza:

La varianza sistemática es la variación en los resultados de las mediciones en la variable dependiente por influencias conocidas, o desconocidas, que hacen que éstos se inclinen más hacia una dirección que hacia otra.

La varianza experimental es la varianza originada por la manipulación activa de las variables independientes por parte del experimentador.

La varianza intergrupo es fundamentalmente equivalente a la varianza experimental y surge de las diferencias entre grupos cubriendo tanto el caso de las diferencias de varianza de tipo experimental como no experimental. *La varianza intergrupo no experimental* es la que se da, por ejemplo, entre grupos de niños escolarizados en zonas urbanas muy desarrolladas y los escolarizados en zonas rurales poco desarrolladas de un país, que aunque estén escolarizados en el mismo sistema educativo, las influencias por las características específicas, por ejemplo, sociofamiliares, introducen variaciones diferentes en unos y otros grupos de alumnos debidas a este tipo de variables que están presentes e intervienen en los resultados del estudio. *La varianza intergrupo experimental* es la que se da comparando por ejemplo, tres grupos de un mismo contexto, o centro educativo, tras la aplicación de tres métodos de enseñanza diferentes. Se halla a partir de las medias de cada uno de los grupos, calculando los valores diferenciales de cada uno de los sujetos con respecto a su media grupal y promediándolos. Sería el sumatorio de las puntuaciones diferenciales al cuadrado dividido por el número de casos, g , que en este caso son grupos.

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{g}$$

Toda variable activa, manipulada por el experimentador, guarda relación con la varianza sistemática. Las mediciones de los resultados del grupo experimental, llamadas también medidas criterio por la influencia de la variable activa varían sistemáticamente, ya sea en sentido ascendente o descendente, en tanto que las mediciones del grupo de control se mantienen, por lo general, en el mismo nivel. El experimentador trata de incrementar la varianza entre los grupos como consecuencia de la manipulación de alguna variable en el grupo experimental.

La varianza error es la variación en las medidas debidas a fluctuaciones, por lo regular pequeñas y autocompensatorias, de las medidas. La varianza de error es una varianza aleatoria y en gran medida queda fuera de toda identificación o control.

Controlando la varianza sistemática identificamos y controlamos parte de la varianza error, pero siempre queda parte de la varianza sin explicar, por lo que la varianza error es aquella que persiste en un conjunto de mediciones después de que se eliminan todas las causas conocidas de varianza sistemática.

La varianza total (s_T^2) es igual a la suma de la varianza experimental (intergrupos causada por el efecto del tratamiento en uno de los grupos, que llamaremos varianza entre grupos —inter o entre— s_E^2) y la varianza error (o varianza intragrupos, medida en cada uno de los grupos que llamaremos varianza dentro de los grupos —intra o dentro— s_D^2).

$$s_T^2 = s_E^2 + s_D^2$$

Estas varianzas son aditivas, por lo que es posible analizar un conjunto de puntuaciones para identificar la varianza sistemática y la varianza error. O la varianza experimental sistemática, la varianza sistemática debida a diferencias individuales y la varianza aleatoria o error. Una vez que hemos identificado en la varianza total la explicación de una porción de la misma (varianza experimental o intergrupo) el resto de la varianza se supone que depende de factores aleatorios y la denominamos varianza error y suponemos que está igualmente distribuida en los distintos grupos cuando se seleccionan al azar y se asignan los tratamientos igualmente al azar.

En cualquier experimento o estudio la variable o variables independientes son una fuente de varianza sistemática, o deberían serlo. El investigador desea que los grupos experimentales difieran sistemáticamente de los de control, buscando llevar al máximo dicha varianza en tanto que controla, para que interfieran lo mínimo, otras fuentes de varianza, sean sistemática o de error.

Veamos el siguiente ejemplo de varianza sistemática y varianza error en dos grupos, a través del siguiente caso:

Un profesor de Matemáticas quiere conocer la influencia de dos estilos distintos en la corrección de exámenes tiene en el rendimiento de sus alumnos. Un estilo estaría relacionado con la introducción de comentarios realizados en la corrección, además de la calificación numérica. El otro modo se basaría sólo en la calificación numérica de los ejercicios. Decide poner a prueba su hipótesis, en dos grupos de alumnos del mismo curso integrados por 35 y 37 alumnos respectivamente.

Llegada la fecha para realizar el examen del primer parcial les aplica una prueba de Matemáticas, y al corregirlos decide utilizar ambos estilos de corrección. Uno de los grupos el A, llamado grupo de control, no recibirá comentarios en sus exámenes, sólo su calificación numérica. El otro grupo, el B, denominado grupo experimental, recibirá, además de su calificación numérica, explicaciones sobre sus aciertos y errores. Decide aleatoriamente cual de los grupos será el A y cual el B. Pasado un mes les aplica a ambos grupos otra prueba de Matemáticas, de características similares a la anterior, sobre los nuevos contenidos explicados en la materia.

A partir de los resultados obtenidos en esta segunda prueba selecciona aleatoriamente de cada uno de los grupos a seis sujetos, con los que va a realizar el estudio de las varianzas, y los refleja en la siguiente tabla de puntuaciones:

TABLA 18.1.

Grupo A (control)	Grupo B (experimental)
8	9
7	8
5	6
6	7
6	6
4	8
Media = 6	Media = 7,33

La media de B es superior a la media de A, lo cual indica que hay variaciones intergrupos. Vamos a comprobarlo calculando el valor de las variaciones intergrupos. Para ello tratamos las dos medias como si fueran puntuaciones directas individuales:

TABLA 18.2.

X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
7,33	0,635	0,403
6	-0,665	0,442
$\bar{X} = 6,66$		$\Sigma = 0,845$

$$s_E^2 = \frac{\sum (X_i - \bar{X})^2}{n} = \frac{0,845}{2} = 0,423$$

Si ordenamos las puntuaciones de ambos grupos, tenemos la siguiente tabla que nos va a facilitar el cálculo de la varianza de todos los sujetos, o varianza total:

TABLA 18.3.

X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
9	2,33	5,429
8	1,33	1,769
6	-0,66	0,436
7	0,33	0,109
6	0,66	0,435
8	1,33	1,769
8	1,33	1,769
7	0,33	0,109
5	-1,66	2,755
6	0,66	0,435
6	0,66	0,435
4	-2,66	7,075
$\bar{X} = 6,66$		$\Sigma = 22,567$

$$s_T^2 = \frac{\sum (X_i - \bar{X})^2}{n} = \frac{22,567}{12} = 1,88$$

La varianza total abarca todas las fuentes de variación de las puntuaciones, siendo una de ellas la varianza intergrupos, o entre grupos, que ya hemos hallado, y cuyo valor es:

$$s_E^2 = 0,4227$$

A partir de estos datos podemos calcular otro tipo de varianza, la varianza intragrupos, o la varianza dentro de los grupos. Para ello calcu-

lamos primero la varianza de un grupo y luego la del otro, y su promedio será la varianza intragrupos (s_D^2).

TABLA 18.4.

X_A	$(X_A - \bar{X})$	$(X_A - \bar{X})^2$	X_i	$(X_B - \bar{X})$	$(X_B - \bar{X})^2$
9	1,66	2,755	8	2	4
8	0,66	0,435	7	2	1
6	-1,33	1,769	5	-1	1
7	-0,33	0,109	6	0	0
6	-1,33	1,769	6	0	0
8	0,66	0,435	4	-2	4
$\bar{X} = \frac{44}{6} = 7,33$		$s_A^2 = \frac{7,29}{6} = 1,22$	$\bar{X} = \frac{36}{6} = 6$	$s_B^2 = \frac{10}{6} = 1,66$	

La varianza intragrupos (s_D^2) vendrá dada por el promedio de las varianzas grupales:

$$s_D^2 = \frac{1,215 + 1,666}{2} = 1,44$$

Y como la varianza total está formada por la varianza intragrupos (dentro) más la varianza intergrupos (entre), tenemos que:

$$s_T^2 = s_E^2 + s_D^2$$

Y sustituyendo en la ecuación los valores obtenidos en el ejemplo:

$$1,8806 = 0,4227 + 1,4408$$

Lo que viene a demostrar que las varianzas son aditivas.

Recordemos que la varianza intragrupos, o dentro de los grupos, también se denomina varianza error, s_e^2 , luego:

$$s_T^2 = s_E^2 + s_e^2$$

Y que la varianza intergrupo, o entre los grupos, s_E^2 , representa la medida del efecto del tratamiento experimental o diferencia de los grupos en función del tratamiento en uno de ellos.

La varianza de error representa la varianza promedio de cada uno de los grupos debida a los efectos de factores aleatorios o no controlados por el investigador; como pueden ser las diferencias individuales como por ejemplo la inteligencia o la madurez de cada uno de los sujetos. Dado que se utilizan técnicas de control como la aleatorización, consistente en

este caso en la asignación por parte del profesor de los sujetos al grupo que iba a recibir el tratamiento experimental, y al grupo que no lo iba a recibir; al azar, se supone que estas diferencias están distribuidas por igual en ambos grupos, y que se autocompensarán.

18.2. La resolución del análisis de la varianza

Ya hemos visto en el ejemplo del apartado anterior, que la aplicación más simple del análisis de varianza es probar la diferencia entre las medias de dos grupos. También hemos estudiado la prueba t , para probar si la diferencia entre las medias de dos grupos independientes, o seleccionados al azar, al que uno de ellos se le había aplicado un tratamiento diferente, era o no estadísticamente significativa a un nivel de significación determinado. Los mismos resultados o equivalentes podrían haberse obtenido si en vez de la prueba t de Student, o la prueba z , se hubiera utilizado el Análisis de varianza, aplicado al diseño de dos grupos. Sin embargo, lo contrario no podría haberse hecho, es decir, aplicar la prueba t , o la z , a un diseño de más de dos grupos.

Si dibujamos en un eje de coordenadas la distribución de las puntuaciones de los dos grupos, obtendremos aproximadamente dos distribuciones semejantes a la curva normal. Un grupo habrá obtenido menores puntuaciones que el otro, y por lo tanto tendrá una media más baja, y todos sus miembros se situarán más cerca del cero en los valores de la variable dependiente. En el eje horizontal se sitúan los valores obtenidos en la variable dependiente por los sujetos, y en el eje vertical se sitúan las frecuencias, o número de sujetos que obtienen los distintos valores. El segundo grupo obtiene valores más altos en la variable dependiente, situándose más alejado del cero, y tendrá una media superior.

La pregunta que nos hacíamos en la prueba t y que hay que hacerse ahora es: ¿Se dan diferencias estadísticamente significativas entre las medias de estos grupos? Hay que señalar que para responder a esta pregunta debemos calcular la Suma de Cuadrados Total, que es un valor que resulta de tomar todos los sujetos que forman parte del experimento como un todo, con independencia de que se les halla administrado o no la variable independiente o tratamiento.

La suma de cuadrados total puede ser dividida en dos componentes principales: la suma de cuadrados intergrupo, y la suma de cuadrados intragrupo. La suma de cuadrados intergrupo nos informa del grado en que las medias de cada grupo difieren: recordemos que la media del grupo que obtuvo puntuaciones más bajas será menor que la media del

grupo que obtuvo puntuaciones más altas. Cuanto mayor sea la diferencia entre las medias mayor será la suma de cuadrado intergrupos.

La suma de cuadrados intragrupos viene determinada por el grado en que los sujetos difieren entre sí en cada uno de los grupos. Si los sujetos del primer grupo difieren mucho entre sí, o lo hacen los del segundo grupo, nos dará una suma de cuadrados intragrupos grande. Cuanto mayor sea la diferencia entre el sujeto que obtiene la puntuación menor, y el sujeto que obtiene la puntuación mayor, mayor será la amplitud, y mayor será la diferencia que obtenga cada sujeto con respecto a la media, mayor será su desviación típica, y su varianza, más heterogéneo será el grupo, o, dicho de otros modo, menos homogéneo será en la variable dependiente medida. Esto podría deberse a la influencia de otras variables experimentales desconocidas, no controladas por el investigador, aumentando la varianza de error, repercutiendo en la menor validez del diseño y de sus conclusiones. Cuanto mayor sea la suma de cuadrados dentro de los grupos —o intragrupos—, mayor será el «error» en el experimento. Así, si partiéramos de un grupo de sujetos iguales, especialmente en aquellas variables que guardan una relación muy elevada con la variable dependiente, obtendríamos una suma de cuadrados para este grupo muy próxima al 0; dado que los sujetos no suelen ser iguales entre sí, debido a las diferencias individuales, este es un valor muy difícil de obtener.

Recordemos que cuando usábamos la prueba t , el numerador de la ecuación estaba formado por la diferencia entre las medias de los dos grupos, y el denominador por una medida del error del experimento, análogo a la suma de cuadrados intragrupos de ahora. En el análisis de la varianza, el numerador es una medida de la diferencia entre las medias de los grupos, y el denominador es la suma de cuadrados intragrupos, como medida de error. Cuanto mayores sean las diferencias entre los grupos, que va en el numerador, y menor sean las diferencias dentro de los grupos, que va en el denominador, el valor resultante será mayor, acentuando la probabilidad de obtener diferencias estadísticas significativas.

18.2.1. Ecuaciones para la resolución del análisis de varianza

Las ecuaciones que ahora presentamos están basadas en el siguiente razonamiento:

- 1.º La media total, o general, se calcula basándose en todos los valores de la variable dependiente del experimento sin tener en cuenta el

hecho de que algunos sujetos estaban bajo la influencia de la variable independiente y otros no. Entonces, la suma de cuadrados total mide la desviación de todas las calificaciones en relación con esta media general.

- 2.º La suma de cuadrados entre los grupos es una medida de desviación de las medias de los diversos grupos en relación con la media general de todos los grupos.
- 3.º La suma de cuadrados intragrupos, o dentro de los grupos, es una suma combinada de cuadrados basada en la desviación de las calificaciones de cada grupo en relación con la media de dicho grupo.

1. Primera ecuación: Suma de cuadrados entre

$$SC_E = \sum_{g=1}^G \frac{\left(\sum_{i=1}^n X_i \right)^2}{n} - \frac{\left(\sum_{i=1}^N X_i \right)^2}{N}$$

Siendo la suma de cuadrados entre igual al sumatorio del cuadrado de la suma de las puntuaciones de cada uno de los grupos dividido entre el número de casos de su grupo, menos el sumatorio del cuadrado de la suma de todas las puntuaciones de los grupos dividido entre el total de casos.

2. Segunda ecuación: Suma de cuadrados total

$$SC_T = \sum_{i=1}^N X_i^2 - \frac{\left(\sum_{i=1}^N X_i \right)^2}{N}$$

Siendo la suma de cuadrados total igual al sumatorio de todas las puntuaciones individuales al cuadrado, menos el sumatorio del cuadrado de la suma de todas las puntuaciones de los grupos dividido entre el total de casos. La suma de cuadrados total tiene dos componentes principales, el entre o intergrupos y el intragrupos.

$$SC_T = SC_E - SC_D$$

3. Tercera ecuación: Suma de cuadrados dentro

Una vez halladas las sumas de cuadrados total y entre el componente de la suma de cuadrados intragrupos puede calcularse

restando a la suma de cuadrados total la suma de cuadrado entre los grupos.

$$SC_D = SC_T - SC_E$$

A partir de las sumas de cuadrados hallamos las medias de cuadrados. Estamos calculando los valores de las muestras como estimaciones de los valores de la población. La regla para calcular la media de los cuadrados, o media cuadrática, es sencilla: se divide la suma de cuadrados dada entre la cantidad de grados de libertad correspondiente, para lo cual debemos hallar los grados de libertad (g.l.).

4. *Cálculo de los grados de libertad*

Grados de libertad total: $N - 1$

Grados de libertad entre: $G - 1$

Grados de libertad intra: $N - G$

Siendo g.l. total = g.l. entre + g.l. intra; $N - 1 = (G - 1) + (N - G)$

5. *Cálculo de las medias de cuadrados:*

La suma de cuadrados total está integrada por dos componentes: suma de cuadrados entre y suma de cuadrados intra. Necesitamos calcular la media de cuadrados entre y la media de cuadrados intra correspondientes a cada fuente de variación, entre-grupos e intra-grupos. Para ello se divide cada SC entre su correspondientes grados de libertad.

Entre los grupos:

$$MC_E = \frac{SC_E}{G-1}$$

Dividiendo la suma de cuadrados entre, entre los grados de libertad $G - 1$.

Dentro de los grupos:

$$MC_D = \frac{SC_D}{N-G}$$

Dividiendo la suma de cuadros dentro, entre los grados de libertad $N - G$.

Si la media de cuadrados entre es considerablemente mayor que la media de cuadrados intra, podemos concluir que los valores

de la variable dependiente para los grupos son distintos; pero ¿Qué tan distintos como para que podamos concluir que la variable independiente ha sido efectiva y no se deben esas diferencias a fluctuaciones del azar? Para contestar a esta pregunta debemos aplicar la prueba estadística para el contraste de hipótesis adecuada, que es la proporción F o razón F de Snedecor, en el caso del Análisis de la Varianza que estamos estudiando, desarrollada por el estadístico Fischer, en cuyo honor lleva este nombre, F . El alumno deberá recordar, y repasar, algunos conceptos importantes vistos en temas anteriores, como las características de los modelos paramétricos y no paramétricos, y las exigencias para la aplicación de la prueba F .

6. Cálculo de « F »empírico

$$F = \frac{MC_E}{MC_D}$$

La prueba F se halla dividiendo la media cuadrática entre, entre la media cuadrática dentro. Es la razón entre la media de cuadrados entre (o diferencias debidas supuestamente a los efectos del tratamiento o V.I.) y las medias de cuadrados dentro (o intra) (supuestamente producidas por las diferencias iniciales que introducen errores en cada grupo).

Podemos comprobar la similitud entre la prueba F y la prueba t . El numerador indica las diferencias entre los grupos (además del error experimental o varianza de error del experimento), y el denominador es una estimación de la varianza del error.

El valor de F refleja el efecto de la variable independiente para producir diferencias entre las medias de los diferentes grupos. Supongamos que la variable independiente es totalmente ineficaz para influir en la variable dependiente, entonces, el numerador —o la MC_E — vendría dada por una estimación del error de la varianza, similar a la estimación del error de la varianza del denominador, por lo que el valor de F estaría muy próximo al valor 1. Por lo tanto, cuando obtengamos un valor de F cercano al 1, podemos casi asegurar que la variable independiente no produjo ninguna diferencia en las medias de la variable dependiente medida en los grupos. En otros casos, el numerador puede ser mayor que el denominador, resultando un valor de F mayor, pero, la pregunta es: ¿Qué tan grande deber ser el valor de F para concluir que podemos rechazar la hipótesis de nulidad al nivel de significación fijado?

Para contestar a la anterior pregunta tenemos que determinar el valor de la probabilidad (P) asociada al valor obtenido de F . Si fijamos el nivel de significación de 0,05, y el valor de F tiene una probabilidad menor de 0,05 de que ocurra al azar, podemos rechazar H_0 , y afirmar la hipótesis alternativa, que defendía que sí existen diferencias significativas entre las medias de los grupos debidas a la influencia del tratamiento diferencial que ha actuado en ellos.

Para determinar la probabilidad asociada al valor de F empírico, debemos consultar la Tabla de F . En la Tabla de F , en la línea superior se sitúan los grados de libertad asociados con el numerador, y en la columna de la izquierda se sitúan los grados de libertad asociados con el denominador. Por cada cruce aparecen cuatro valores en función de cuatro niveles de significación distintos: 0,01, 0,05, 0,1 y 0,2. Comprobamos el punto de cruce de nuestros grados de libertad y nos situamos en el nivel de significación previamente fijado. Si nuestra F empírica no sobrepasa el valor crítico, o teórico, de F , que aparece en las Tablas de F , concluiremos que no hay diferencias entre los grupos, no pudiendo rechazar la hipótesis de nulidad; pero si nuestra F empírica sobrepasa el valor de la F de las tablas, podemos rechazar la hipótesis de nulidad.

18.2.2. Interpretación del valor de F

Cuando el valor empírico de F nos permite rechazar la hipótesis de nulidad, y tenemos más de dos grupos, nuestra siguiente pregunta es: ¿Entre cuáles de los diferentes grupos podemos rechazar la hipótesis de nulidad? Y es que hay que saber en dónde está la diferencia en el caso de más de dos grupos. ¿Entre el grupo A y el B, entre el A y el C, entre el B y el C? Para ello debemos realizar los contrastes a posteriori entre cada dos grupos, mediante las pruebas de Tukey, la de Scheffé, o la de Bonferroni, que veremos y aplicaremos en los ejemplos que a continuación se muestran.

La prueba paramétrica «F» del Análisis de la Varianza (ANAVA) debe cumplir los supuestos del modelo estadístico paramétrico, que, como hemos estudiado, básicamente son los siguientes:

- Las muestras deben ser aleatorias simples de sus respectivas poblaciones por lo que existirá independencia de las puntuaciones obtenidas en la variable dependiente.
- La variable estudiada en las muestras debe distribuirse siguiendo el modelo de curva normal que podemos comprobar mediante

pruebas de bondad de ajuste de χ^2 de Pearson y de Kolmogorov-Smirnov.

- Homogeneidad de las varianzas de los grupos, consideradas aproximadamente iguales al proceder los sujetos de muestras extraídas de la misma población, o de poblaciones de igual varianza en la variable estudiada, a comprobar mediante el test de Cochran o el test de Bartlett o Levene.
- Linealidad.

Cuando el valor de la razón F es inferior a 1 ($F < 1$) puede deberse a variaciones del muestreo al trabajar con muestras reales, por lo que no se considera grave. Sin embargo debemos asegurarnos que no se deba al hecho de no cumplir los supuestos exigidos para la aplicación del Análisis de varianza.

18.2.3. Las pruebas de Tukey y Scheffé para el contraste a posteriori

Las pruebas de Tukey y Scheffé se utilizan una vez que el investigador ha aplicado el análisis de varianza y ha podido confirmar que existen diferencias significativas entre los grupos, al resultar ser la razón F significativa al nivel de confianza fijado, con la finalidad de conocer entre qué grupos se dan dichas diferencias significativas.

18.2.3.1. La prueba de Tukey

La prueba elaborada por Tukey tiene la finalidad de realizar comparaciones entre pares de medias, cuando el n de los grupos es equilibrado ($n_1 = n_2 = n_3$), siguiendo el nivel de error tipo I fijado en el α . La prueba trata de obtener un valor considerado como la diferencia mínima entre las medias muestrales dos a dos para comprobar si existen diferencias con respecto a la media poblacional, al nivel de significación fijado en el análisis de varianza realizado previamente.

La fórmula para aplicar la prueba de Tukey es:

$$\frac{|\bar{X}_A - \bar{X}_B|}{\sqrt{\frac{MC_D}{n}}}$$

Siendo:

MC_D : la media cuadrática dentro de los grupos, o intragrupos,

n : el número de sujetos en cada grupo, supuestos iguales en todos los grupos.

$$si \frac{|\bar{X}_A - \bar{X}_B|}{\sqrt{\frac{MC_D}{n}}} > {}_{(1-\alpha)}q_{G,(N-G)}$$

entonces existen diferencias estadísticamente significativas al nivel de significación α fijado, siendo q el valor crítico de la tabla de Percentiles de las distribuciones por rangos, q , de Student', para la probabilidad $(1 - \alpha)$ y G y $(N - G)$ grados de libertad.

18.2.3.2. La prueba de Scheffé

La prueba de Scheffé es menos potente que la prueba de Tukey para rechazar hipótesis nulas cuando no son ciertas, pero sin embargo permite su aplicación cuando los grupos no son equilibrados, es decir, tienen un número de sujetos desigual. Utiliza el mismo nivel de confianza que viene utilizando el análisis de varianza previamente realizado, como en el caso de Tukey. Se basa en la comparación de las medias de los grupos dos a dos para precisar entre cuáles de los grupos se dan las diferencias significativas al nivel de significación α con el que trabajamos.

La formula para aplicar la prueba es la siguiente:

$$si \frac{|\bar{X}_A - \bar{X}_B|}{\sqrt{MC_D \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}} > s = \sqrt{(g-1) \cdot F_c} \Rightarrow \text{exi existen diferencias significativas a favor del grupo con media mayor}$$

18.3 Resolución de un diseño de investigación de tres grupos independientes y nivel de medida en la variable dependiente de intervalo mediante el análisis de la varianza

En un Centro de Enseñanza Primaria, el equipo docente de Ciencias Sociales desea comprobar de forma empírica si la utilización del método de enseñanza incide en el rendimiento académico de sus alumnos. Para llevar a cabo la experiencia, deciden trabajar con tres grupos de 6.º curso,

y aleatoriamente asignan los distintos métodos de enseñanza a los grupos: en el grupo «A» se utilizará la metodología tradicional consistente en la explicación del profesor de los temas y realización de controles evaluativos escritos cada dos semanas. En el grupo «B» se utilizará exclusivamente una metodología basada en la realización de trabajos por parte de los alumnos. Y en el grupo «C», se utilizará una metodología mixta, añadiendo a las explicaciones del profesor con controles quincenales, la realización de trabajos en equipo mediante la consulta de manuales por parte de los alumnos.

Aunque trabaja con tres grupos del mismo nivel, selecciona de forma aleatoria a siete sujetos de cada grupo a los que aplica, al final de curso, una prueba de rendimiento bastante precisa, cuyos datos se muestran en la tabla adjunta:

TABLA 18.1.

Sujetos	Grupo A	Grupo B	Grupo C
1	4	5	7
2	6	7	9
3	3	4	8
4	5	8	9
5	7	6	7
6	2	5	8
7	4	3	6
ΣX_i	31	38	54
ΣX_i^2	155	224	424

Utilizamos pocos casos para facilitar las operaciones, pero la prueba de ANAVA se utiliza con muestras grandes $n > 30$, aunque también se aplica en casos con $n < 30$.

1. Esquema y comentario del diseño de investigación

Lo primero que debemos hacer es analizar la información que tenemos para saber con qué diseño nos encontramos para poder decidir la prueba estadística de contraste adecuada a la información.

Este sería el esquema del diseño:

<i>Grupos</i>	<i>Selección</i>	<i>Asignación</i>	<i>Pretest</i>	<i>Variable Independiente</i>	<i>Postest</i>
A (Control)	Azar	Azar	No	Métodología tradic.	Sí
B (Exper. 1)	Azar	Azar	No	Met. trabajos en grupo	Sí
C (Exper. 2)	Azar	Azar	No	Metodología mixta	Sí

Se trata de un diseño cuasi-experimental, puesto que se desarrolla en contextos naturales, en el propio centro educativo, y en los grupos de clase. Son tres grupos independientes, y los alumnos que han aportados los datos para el análisis estadístico se han selecciona al azar. El grupo A actúa como control, el B y el C son los grupos experimentales con dos niveles de la variable independiente (VI). La asignación de los tratamientos a los grupos se ha realizado al azar. No hay medida pretest. La VI es la metodología de enseñanza de clase; así en el grupo A se utiliza una metodología tradicional, en el B una metodología basada en el trabajo de los alumnos en equipo y en el C se utiliza una metodología mixta (del A y B). Los resultados en el postest se miden mediante una prueba objetiva de rendimiento bastante precisa.

2. *Formulación de las hipótesis*

Hipótesis textual:

H_0 : La metodología de enseñanza no incide en el rendimiento de los alumnos.

H_1 : La metodología de enseñanza sí incide en el rendimiento de los alumnos.

Hipótesis estadística:

H_0 : $\mu_A = \mu_B = \mu_C$

H_1 : $\mu_A < \mu_B < \mu_C$

Se trata de una hipótesis unilateral derecha.

El nivel de significación fijado es $\alpha = 0,01$.

3. *Elección de la prueba estadística más apropiada para el contraste*

Por tratarse de un diseño de tres grupos independientes y cuyos resultados del postest (Variable Dependiente) han sido medidos

mediante una prueba objetiva de rendimiento bastante precisa, por lo que se considera que alcanzan un nivel de medida de intervalo, y supuestas las condiciones de normalidad en la distribución de los datos, homogeneidad de varianzas de los grupos y linealidad, elegiremos una prueba paramétrica que en este caso es el Análisis de la varianza, ANAVA. Si «F» resulta significativo al nivel de confianza de 99 %, y puesto que el tamaño de los grupos es igual, $n_A = n_B = n_C$, aplicaremos la prueba de Tukey para el contraste «a posteriori», o «post hoc», entre los grupos.

4. Aplicación de la prueba seleccionada

Tomamos las fórmulas para realizar el ANAVA y sustituimos las incógnitas por los datos de la tabla.

a) Sumas de cuadrados entre, total y dentro:

$$SC_E = \sum_{g=1}^G \frac{\left(\sum_{i=1}^n X_i \right)^2}{n} - \frac{\left(\sum_{i=1}^N X_i \right)^2}{N} = \frac{31^2 + 38^2 + 54^2}{7} - \frac{123^2}{21} = 39,7$$

$$SC_T = \sum_{i=1}^N X_i^2 - \frac{\left(\sum_{i=1}^N X_i \right)^2}{N} = (155 + 224 + 424) - \frac{15129}{21} = 82,58$$

$$SC_D = SC_T - SC_E = 82,58 - 39,72 = 42,86$$

b) Cálculo de los grados de libertad:

Grados de libertad total: $N - 1 = 21 - 1 = 20$

Grados de libertad entre: $G - 1 = 3 - 1 = 2$

Grados de libertad intra, o dentro: $N - G = 21 - 3 = 18$

c) Cálculo de las medias de cuadrados:

Entre los grupos:

$$MC_E = \frac{SC_E}{G - 1} = \frac{39,72}{2} = 19,86$$

Dentro de los grupos:

$$MC_D = \frac{SC_D}{N - G} = \frac{42,86}{18} = 2,38$$

d) Cálculo de «F»empírico:

$$F = \frac{MC_E}{MC_D} = \frac{19,86}{2,38} = 8,34$$

e) Interpretación de «F»empírico:

Debemos consultar las tablas de la distribución de F:

$$(1 - \alpha) F_{G-1, N-G} \text{ , } 0,99 F_{2,18 \text{ gl}} = 6,01$$

Como el valor empírico es mayor que el valor teórico:

$$8,34 > 6,01$$

podemos rechazar H_0 , al nivel de significación del 0,01, lo que nos permite aceptar la hipótesis alternativa, H_1 , a un nivel de confianza del 99 %.

5. Contraste a posteriori para ver entre qué grupos se dan las diferencias

Para comprobar entre qué grupos se dan las diferencias necesitamos realizar las pruebas *a posteriori* o «pos hoc». Realizaremos de la de Tukey, para cuando el tamaño, n , de los grupos es igual.

$$\text{Si } \frac{|\bar{X}_A - \bar{X}_B|}{\sqrt{\frac{MC_D}{n}}} > q \Rightarrow \text{ existen diferencias significativas; para}$$

$$\alpha = 0,01, \text{ } 0,99 q_{3,18 \text{ gl}} = 4,703$$

Siendo:

$$\bar{X}_A = 4,43; \quad \bar{X}_B = 5,43; \quad \bar{X}_C = 7,7; \quad \sqrt{\frac{MC_D}{n}} = \sqrt{\frac{2,38}{7}} = 0,58$$

$$\text{Entre A y B } \frac{|4,43 - 5,43|}{0,58} = \frac{1}{0,58} = 1,726 < 4,703 \Rightarrow \text{ no existen diferencias significativas.}$$

$$\text{Entre B y C } \frac{|5,43 - 7,7|}{0,58} = 3,93 < 4,703 \Rightarrow \text{ no existen diferencias significativas.}$$

Entre A y C $\frac{|4,43 - 7,7|}{0,58} = 5,65 < 4,703 \Rightarrow$ sí existen diferencias significativas al nivel de confianza del 99%, a favor del grupo C.

Ahora vamos a hacer los cálculos con el programa informático SPSS. Las tablas de resultados aportan en primer lugar los análisis descriptivos de los tres grupos, de modo individual y del total.

TABLA 18.2.

DESCRIPTIVOS								
	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
1.00	7	4.4286	1.71825	.64944	2.8395	6.0177	2.00	7.00
2.00	7	5.4286	1.71825	.64944	3.8395	7.0177	3.00	8.00
3.00	7	7.7143	1.11270	.42056	6.6852	8.7434	6.00	9.00
Total	21	5.8571	2.03189	.44339	4.9322	6.7820	2.00	9.00

Después realiza la prueba de Levene para ver la homogeneidad de la varianzas. Y nos indica que son homogéneas al dar una diferencia significativa entre ellas de 0,502, es decir, muy elevada para aceptar que existen diferencias significativas entre ellas, lo que indica que son homogéneas.

TABLA 18.3.

PRUEBA DE HOMOGENEIDAD DE VARIANZAS			
Estadístico de Levene	gl1	gl2	Sig.
.717	2	18	.502

Y la siguiente tabla de resultados es el análisis de varianza para un factor.

TABLA 18.4.

ANOVA DE UN FACTOR					
	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	39.714	2	19.857	8.340	.003
Intra-grupos	42.857	18	2.381		
Total	82.571	20			

El análisis de la varianza realizado mediante el programa SPSS concluye que existen diferencias significativas entre los grupos a un nivel de significación del 0,003, más pequeño que el 0,01, lo cual indica un nivel de confianza del 99,7 %.

Las pruebas «post hoc» realizadas, mediante la prueba de Tukey, indican que existen diferencias significativas entre el grupo A y el grupo C, al nivel de confianza del 99 %, a favor del grupo C.

TABLA 18.5.

Pruebas <i>post hoc</i> . Comparaciones múltiples. HSD de Tukey						
(I) VAR02	(J) VAR02	Diferencia de medias (I - J)	Error típico	Sig.	Intervalo de confianza al 99%	
					Límite inferior	Límite superior
1.00	2.00	-1.0000	.82479	.461	-3.7431	1.7431
	3.00	-3.2857(*)	.82479	.002	-6.0288	-.5427
2.00	1.00	1.0000	.82479	.461	-1.7431	3.7431
	3.00	-2.2857	.82,479	.032	-5.0288	.4573
3.00	1.00	3.2857(*)	.82479	.002	.5427	6.0288
	2.00	2.2857	.82479	.032	-.4573	5.0288

* La diferencia entre las medias es significativa al nivel .01.

6. Conclusiones y recomendaciones

Al realizar las conclusiones de la investigación debemos ser críticos con todo el proceso. Conviene destacar que los resultados del contraste de hipótesis nos permiten defender la mayor eficacia del método utilizado en el Grupo Experimental B, a un nivel de confianza del 99 %, lo cual permite realizar recomendaciones para

la utilización de una metodología mixta, de explicaciones, evaluaciones periódicas y realización de trabajos en equipo por los alumnos, en la enseñanza de las Ciencias Sociales. Sin embargo, el hecho de haber realizado los análisis con muestras de alumnos muy pequeñas, 7 alumnos por grupo, aconsejan la realización de nuevos experimentos ampliando el tamaño de la muestra, con el fin de confirmar los hallazgos.

18.4. Resolución de un diseño de investigación de tres grupos independientes y nivel de medida en la variable dependiente ordinal, mediante la prueba de Kruskal Wallis

El equipo docente de Química de un Centro de Educación Secundaria, preocupado por el escaso interés de los alumnos de 4.º de ESO por la Química, decide llevar a cabo el siguiente plan para tratar de incrementar el interés por la Química. Con uno de los grupos (A) continúa con el método que venía utilizando hasta ahora: las explicaciones del libro de texto y las prácticas de laboratorio una hora por semana. Con otro de los grupos (B) añadirá al método que venía utilizando, el estudio de revistas científicas especializadas en Química de Secundaria, en las que se difunden los últimos avances en este campo. Y a un tercer grupo (C), añadirá, además de la revistas científicas, una visita mensual a una empresa química en la que se les explica a los alumnos las funciones que realizan en dicha empresa y la utilidad de los mismos.

Pasado el primer cuatrimestre se les aplica a todos los alumnos un cuestionario de actitud hacia la Química, consistente en una escala de 15 ítems valorada de 1 a 4, que ha elaborado el equipo docente. Los datos obtenidos por unas submuestras elegidas aleatoriamente de cada uno de los grupos, se muestran en la siguiente tabla de datos:

TABLA 18.6.

Grupo A	Grupo B	Grupo C
9	13	13
10	19	20
13	20	32
17	28	40
18	30	42
20	38	44
25	40	45
30	41	46
35	44	49
38		50

Para analizar la información es necesario contestar a las siguientes cuestiones:

1. Diseño planteado: Esquema y comentario.
2. Formulación de las hipótesis.
3. Selección de la prueba estadística más apropiada para el contraste.
4. Aplicación e interpretación de la prueba al nivel de significación de $\alpha = 0,01$.
5. Conclusiones y recomendaciones, derivadas del estudio, incluido un análisis crítico de los objetivos pretendidos, del planteamiento y de la realización de la investigación.

Que pasamos a resolver:

1. Esquema y comentario del diseño planteado

Grupos	Selección	Asignación	Pretest	Tratamiento: V. I.	Postest VD
A (Control)	Azar	Azar	No	Métodología tradic.	Sí
B (Exper. 1)	Azar	Azar	No	Nuevos elementos didácticos	Sí
C (Exper. 2)	Azar	Azar	No	Nuevos elem. didáct. + visitas	Sí

Estamos ante un diseño cuasi-experimental de tres grupos; el A actúa como control y el B y C como experimentales. Tanto la selección de los sujetos como la asignación a los tratamientos se ha realizado al azar. No se ha aplicado ninguna medida pretest. La V.I. es la metodología didáctica para fomentar el interés hacia la Química, con tres niveles de tratamiento: En el A no se introducen novedades, en el B el uso de revistas científicas y el C, además de las revistas, visitas a empresas químicas.

Los resultados del postest (V.D.) se han evaluado mediante la aplicación de una prueba de actitud (escala de 15 ítems, con una valoración de 1 a 4, que oscila entre una puntuación total de 0-50 puntos) elaborada por el equipo docente.

2. *Formulación de las hipótesis*

Hipótesis Textuales:

H_0 : La metodología didáctica utilizada en la asignatura de Química no incide en el interés de los alumnos hacia la Química.

H_1 : La metodología didáctica utilizada en la asignatura de Química incide en el interés de los alumnos hacia la Química.

Hipótesis estadísticas:

H_0 : $\text{Med}_A = \text{Med}_B = \text{Med}_C$

H_1 : $\text{Med}_A < \text{Med}_B < \text{Med}_C$

Se trata de una hipótesis unilateral derecha.

El nivel de significación fijado por el investigador es el de $\alpha = 0,01$.

3. *Cuál sería la prueba estadística más apropiada para el contraste?*

Dado que se trata de un diseño de más de dos grupos independientes, cuyos resultados en el postest han sido medidos mediante una escala de actitud que alcanza un nivel de medida ordinal, la prueba estadística para el contraste de hipótesis debe ser no paramétrica, siendo la más apropiada el Análisis de Varianza por Rangos o prueba «H» de Kruskal-Wallis, cuya fórmula es la siguiente:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^G \frac{R_i^2}{n_i} - 3(N+1)$$

Siendo:

N : el número total de sujetos

R : los rangos

ΣR_i : sumatorio de los rangos de cada grupo

n : numero de caso en cada grupo.

4. Aplicación de la prueba seleccionada

Como paso previo al cálculo de «H» debemos transformar las puntuaciones en rangos, considerando los tres grupos como uno solo:

TABLA 18.7.

G.Control (A)	G. Exp.(B)	G. Exp.(C)
9 (1)	13 (4)	13 (4)
10 (2)	19 (8)	20 (10)
13 (4)	20 (10)	32 (16)
17 (6)	28 (13)	40 (20,5)
18 (7)	30 (14,5)	42 (23)
20 (10)	38 (18,5)	44 (24,5)
25 (12)	40 (20,5)	45 (26)
30 (14,5)	41 (22)	46 (27)
35 (17)	44 (24,5)	49 (28)
38 (18,5)		50 (29)
$\Sigma R_i = 92$	$\Sigma R_i = 135$	$\Sigma R_i = 208$

Y ahora sustituimos los sumarios de los rangos de cada grupo en la fórmula, los elevamos al cuadrado y los dividimos entre el número de sujetos de cada grupo.

$$\begin{aligned}
 H &= \frac{12}{29(29+1)} \left(\frac{92^2}{10} + \frac{135^2}{9} + \frac{208^2}{10} \right) - 3(29+1) = \\
 &= \frac{12}{870} (846,4 + 2025 + 4326,4) - 90 = 9,28
 \end{aligned}$$

5. Interpretación de su valor ($\alpha = 0,01$)

El valor empírico obtenido es de 9,28, y se interpreta comparándolo con la distribución de χ^2 . Tabla de χ^2 (Apéndice), con g.l. = $G - 1 = 2$ y $\alpha = 0,01$, el valor teórico asociado es $_{0,99}\chi^2_{2gl} = 9,210$.

Como $H_e = 9,28 > H_c = 9,210$, nos permite rechazar H_0 y por lo tanto aceptar la hipótesis alternativa del investigador, H_1 , al nivel de confianza del 99%, que postulaba la existencia de diferencias significativas entre los grupos.

Ahora hacemos los cálculos de la Prueba estadística no paramétrica de Kruskal-Wallis mediante el programa SPSS, que nos aporta los siguientes resultados:

TABLA 18.8.

Estadísticos descriptivos					
	N	Media	Desviación típica	Mínimo	Máximo
VAR001	29	29.9655	12.99308	9.00	50.00
VAR002	29	2.0000	.84515	1.00	3.00

TABLA 18.9.

Rangos			
	GRUPOS	N	Rango promedio
PUNTUAC	1.00	10	9.20
	2.00	9	15.00
	3.00	10	20.80
	Total	29	

TABLA 18.10.

Estadísticos de contraste(a,b) Prueba de Kruskal-Wallis	
	PUNTUAC
Chi-cuadrado	9.308
gl	2
Sig. asintót.	.010
a Prueba de Kruskal-Wallis b Variable de agrupación: GRUPOS	

que confirma los anteriores resultados.

6. Conclusiones y recomendaciones, derivadas del estudio, referidas tanto a los objetivos pretendidos por el investigador, como al diseño de investigación

Hemos podido rechazar H_0 , a un nivel de confianza del 99 % (significación $\alpha = 0,01$), lo cual nos permite confirmar la hipótesis alternativa H_1 , lo que viene a apoyar que la metodología didáctica que acerca a la realidad, tanto en los avances de investigación como a la práctica en las empresas fomenta el interés hacia la disciplina, en este caso, a la Química.

18.5. Ejercicios de autoevaluación

Problema:

Se quiere comprobar si existen diferencias significativas entre los resultados obtenidos en una prueba de actitud frente al estudio mediante el ordenador, aplicada a tres grupos de estudiantes universitarios del mismo curso y carrera, obteniendo los siguientes resultados:

Sujetos	U1	U2	U3
1	30	41	62
2	35	32	57
3	60	70	35
4	61	56	42
5	42	80	78
6		43	

18.6. Solución a los ejercicios de autoevaluación

Solución al problema:

Pruebas no paramétricas: Prueba de Kruskal-Wallis

Rangos			
	VAR00001	N	Rango promedio
VAR00002	1.00	5	6.80
	2.00	6	9.00
	3.00	5	9.60
	Total	16	

Estadísticos de contraste(a,b)	
	VAR00002
Chi-cuadrado	.973
gl	2
Sig. asintót.	.615
a Prueba de Kruskal-Wallis	
b Variable de agrupación: VAR00001	

No existen diferencias significativas entre los grupos, por lo que se consideran homogéneos en cuanto a la variable estudiada: «actitud frente al estudio mediante el ordenador».

Utilizando la prueba de la mediana obtenemos los mismos resultados:

Prueba de la mediana

Frecuencias				
		VAR00001		
		1.00	2.00	3.00
VAR00002	> Mediana	2	3	3
	<= Mediana	3	3	2

Estadísticos de contraste(b)	
	VAR00002
N	16
Mediana	49.5000
Chi-cuadrado	.400(a)
gl	2
Sig. asintót.	.819
a 6 casillas (100.0%) tienen frecuencias esperadas menores que 5. La frecuencia de casilla esperada mínima es 2.5.	
b Variable de agrupación: VAR00001	

que confirman los anteriores resultados:

18.7. Referencias bibliográficas

- GARCÍA LLAMAS, J. L., GONZÁLEZ GALÁN, M. A. y BALLESTEROS VELÁZQUEZ, B. *Introducción a la investigación en Educación*. Unidades Didácticas. Madrid: UNED. 2001.
- GARCIA LLAMAS, J. L.; PEREZ JUSTE, R. y RIO SADORNIL, D. DEL. *Problemas y diseños de investigación resueltos*. Ed. Dykinson. Madrid, 1999.
- JIMENEZ FERNANDEZ, C. LOPEZBARAJAS ZAYAS, E. y PEREZ JUSTE, R. (1987). *Pedagogía Experimental II*. Madrid: UNED.
- KERLINGER, F. N. (1987). *Investigación del comportamiento. Técnicas y metodología*. Nueva Ed. Interamericana. México, 1975.

APÉNDICE

TABLAS ESTADÍSTICAS

Tabla 1. Áreas ordenadas de la curva de distribución normal en función x/σ	507-514
Tabla 2. Valores críticos de t	515
Tabla 3. Percentiles de distribución de ji-cuadrado	516
Tabla 4. Transformación z de Fisher de r_{xy}	517
Tabla 5. Valores absolutos críticos del coeficiente de correlación de Spearman, por rangos ordenados, r_s , para el contraste de hipótesis nula de existencia de correlación mediante el contraste de dos colas.....	518
Tabla 6. Valores críticos de r_s , el coeficiente de correlación de rangos de Spearman.....	519
I. Probabilidades asociada con valores tan pequeños como los valores observados de U en la prueba de Mann-Whitney....	520-522
II. Valores críticos de U en la prueba de Mann-Whitney	523-526
Tabla 7. Valores críticos de T , en la prueba de los rangos señalados de pares igualados de Wilcoxon	527
Tabla 8. Percentiles de distribución F	528-534
Tabla 9. Percentiles de distribuciones por rangos q , de Student para J y v grados de libertad	535-539
Tabla 10. Probabilidades asociadas con valores tan grandes como valores observados de H en el análisis de varianza de una clasificación por rangos de Kruskal-Wallis	540-541

TABLA 1. Áreas y ordenadas de la curva de distribución normal en función x/σ

(1) <i>z</i> Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) <i>A</i> Área desde la media a $\frac{x}{\sigma}$	(3) <i>B</i> Área de la parte mayor	(4) <i>C</i> Área de la parte menor	(5) <i>y</i> Ordenada en $\frac{x}{\sigma}$
0.00	.0000	.5000	.5000	.3989
0.01	.0040	.5040	.4960	.3989
0.02	.0080	.5080	.4920	.3989
0.03	.0120	.5120	.4880	.3988
0.04	.0160	.5160	.4840	.3986
0.05	.0199	.5199	.4801	.3984
0.06	.0239	.5239	.4761	.3982
0.07	.0279	.5279	.4721	.3980
0.08	.0319	.5319	.4681	.3977
0.09	.0359	.5359	.4641	.3973
0.10	.0398	.5398	.4602	.3970
0.11	.0438	.5438	.4562	.3965
0.12	.0478	.5478	.4522	.3961
0.13	.0517	.5517	.4483	.3956
0.14	.0557	.5557	.4443	.3951
0.15	.0596	.5596	.4404	.3945
0.16	.0636	.5636	.4364	.3939
0.17	.0675	.5675	.4325	.3932
0.18	.0714	.5714	.4286	.3925
0.19	.0753	.5753	.4247	.3918
0.20	.0793	.5793	.4207	.3910
0.21	.0832	.5832	.4168	.3902
0.22	.0871	.5871	.4129	.3894
0.23	.0910	.5910	.4090	.3885
0.24	.0948	.5948	.4052	.3876
0.25	.0987	.5987	.4013	.3867
0.26	.1026	.6026	.3974	.3857
0.27	.1064	.6064	.3936	.3847
0.28	.1103	.6103	.3897	.3836
0.29	.1141	.6141	.3859	.3825
0.30	.1179	.6179	.3821	.3814
0.31	.1217	.6217	.3783	.3802
0.32	.1255	.6255	.3745	.3790
0.33	.1293	.6293	.3707	.3778
0.34	.1331	.6331	.3669	.3765
0.35	.1368	.6368	.3632	.3752
0.36	.1406	.6406	.3594	.3739
0.37	.1443	.6443	.3557	.3725
0.38	.1480	.6480	.3520	.3712
0.39	.1517	.6517	.3483	.3697
0.40	.1554	.6554	.3446	.3683
0.41	.1591	.6591	.3409	.3668
0.42	.1628	.6628	.3372	.3653
0.43	.1664	.6664	.3336	.3637
0.44	.1700	.6700	.3300	.3621

Tomada del *Formulario y tablas de Pedagogía Experimental*, UNED, 1983, págs. 119-126 (ambas inclusive).

(Continúa)

Copyright © 2014. UNED - Universidad Nacional de Educación a Distancia. All rights reserved.

(Continuación)

(1) <i>z</i> Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) <i>A</i> Área desde la media a $\frac{x}{\sigma}$	(3) <i>B</i> Área de la parte mayor	(4) <i>C</i> Área de la parte menor	(5) <i>y</i> Ordenada en $\frac{x}{\sigma}$
0.45	.1736	.6736	.3264	.3605
0.46	.1772	.6772	.3228	.3589
0.47	.1804	.6808	.3192	.3572
0.48	.1844	.6844	.3156	.3555
0.49	.1879	.6879	.3121	.3538
0.50	.1915	.6915	.3085	.3521
0.51	.1950	.6950	.3050	.3503
0.52	.1985	.6985	.3015	.3485
0.53	.2019	.7019	.2981	.3467
0.54	.2054	.7054	.2946	.3448
0.55	.2088	.7088	.2912	.3429
0.56	.2123	.7123	.2877	.3410
0.57	.2157	.7157	.2843	.3391
0.58	.2190	.7190	.2810	.3372
0.59	.2224	.7224	.2776	.3352
0.60	.2257	.7257	.2743	.3332
0.61	.2291	.7291	.2709	.3312
0.62	.2324	.7324	.2676	.3292
0.63	.2357	.7357	.2643	.3271
0.64	.2389	.7389	.2611	.3251
0.65	.2422	.7422	.2578	.3230
0.66	.2454	.7454	.2546	.3209
0.67	.2486	.7486	.2514	.3187
0.68	.2517	.7517	.2483	.3166
0.69	.2549	.7549	.2451	.3144
0.70	.2580	.7580	.2420	.3123
0.71	.2611	.7611	.2389	.3101
0.72	.2642	.7642	.2358	.3079
0.73	.2673	.7673	.2327	.3056
0.74	.2704	.7704	.2296	.3034
0.75	.2734	.7734	.2266	.3011
0.76	.2764	.7764	.2236	.2989
0.77	.2794	.7794	.2206	.2966
0.78	.2823	.7823	.2177	.2943
0.79	.2852	.7852	.2148	.2920
0.80	.2881	.7881	.2119	.2897
0.81	.2910	.7910	.2090	.2874
0.82	.2939	.7939	.2061	.2850
0.83	.2967	.7967	.2033	.2827
0.84	.2995	.7995	.2005	.2803
0.85	.3023	.8023	.1977	.2780
0.86	.3051	.8051	.1949	.2756
0.87	.3078	.8078	.1922	.2732
0.88	.3106	.8106	.1894	.2709
0.89	.3133	.8133	.1867	.2685

(Continúa)

(Continuación)

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Área desde la media a $\frac{x}{\sigma}$	(3) B Área de la parte mayor	(4) C Área de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
0.90	.3159	.8159	.1841	.2661
0.91	.3186	.8186	.1814	.2637
0.92	.3212	.8212	.1788	.2613
0.93	.3238	.8238	.1762	.2589
0.94	.3264	.8264	.1736	.2565
0.95	.3289	.8289	.1711	.2541
0.96	.3315	.8315	.1685	.2516
0.97	.3340	.8340	.1660	.2492
0.98	.3365	.8365	.1635	.2468
0.99	.3389	.8389	.1611	.2444
1.00	.3413	.8413	.1587	.2420
1.01	.3438	.8438	.1562	.2396
1.02	.3461	.8461	.1539	.2371
1.03	.3485	.8485	.1515	.2347
1.04	.3508	.8508	.1492	.2323
1.05	.3531	.8531	.1469	.2299
1.06	.3554	.8554	.1446	.2275
1.07	.3577	.8577	.1423	.2251
1.08	.3599	.8599	.1401	.2227
1.09	.3621	.8621	.1379	.2203
1.10	.3643	.8643	.1357	.2179
1.11	.3665	.8665	.1335	.2155
1.12	.3686	.8686	.1314	.2131
1.13	.3708	.8708	.1292	.2107
1.14	.3729	.8729	.1271	.2083
1.15	.3749	.8749	.1251	.2059
1.16	.3770	.8770	.1230	.2036
1.17	.3790	.8790	.1210	.2012
1.18	.3810	.8810	.1190	.1989
1.19	.3830	.8830	.1170	.1965
1.20	.3849	.8849	.1151	.1942
1.21	.3869	.8869	.1131	.1919
1.22	.3888	.8888	.1112	.1895
1.23	.3907	.8907	.1093	.1872
1.24	.3925	.8925	.1075	.1849
1.25	.3944	.8944	.1056	.1826
1.26	.3962	.8962	.1038	.1804
1.27	.3980	.8980	.1020	.1781
1.28	.3997	.8997	.1003	.1758
1.29	.4015	.9015	.0985	.1736
1.30	.4032	.9032	.0968	.1714
1.31	.4049	.9049	.0951	.1691
1.32	.4066	.9066	.0934	.1669
1.33	.4082	.9082	.0918	.1647
1.34	.4099	.9099	.0901	.1626

(Continúa)

(Continuación)

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Área desde la media a $\frac{x}{\sigma}$	(3) B Área de la parte mayor	(4) C Área de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
1.35	.4115	.9115	.0885	.1604
1.36	.4131	.9131	.0869	.1582
1.37	.4147	.9147	.0853	.1561
1.38	.4162	.9162	.0838	.1539
1.39	.4177	.9177	.0823	.1518
1.40	.4192	.9192	.0808	.1497
1.41	.4207	.9207	.0793	.1476
1.42	.4222	.9222	.0778	.1456
1.43	.4236	.9236	.0764	.1435
1.44	.4251	.9251	.0749	.1415
1.45	.4265	.9265	.0735	.1394
1.46	.4279	.9279	.0721	.1374
1.47	.4292	.9292	.0708	.1354
1.48	.4306	.9306	.0694	.1334
1.49	.4319	.9319	.0681	.1315
1.50	.4332	.9332	.0668	.1295
1.51	.4345	.9345	.0655	.1276
1.52	.4357	.9357	.0643	.1257
1.53	.4370	.9370	.0630	.1238
1.54	.4382	.9382	.0618	.1219
1.55	.4394	.9394	.0606	.1200
1.56	.4406	.9406	.0594	.1182
1.57	.4418	.9418	.0582	.1163
1.58	.4429	.9429	.0571	.1145
1.59	.4441	.9441	.0559	.1127
1.60	.4452	.9452	.0548	.1109
1.61	.4463	.9463	.0537	.1092
1.62	.4474	.9474	.0526	.1074
1.63	.4484	.9484	.0516	.1057
1.64	.4495	.9495	.0505	.1040
1.65	.4505	.9505	.0495	.1023
1.66	.4515	.9515	.0485	.1006
1.67	.4525	.9525	.0475	.0989
1.68	.4535	.9535	.0465	.0973
1.69	.4545	.9545	.0455	.0957
1.70	.4554	.9554	.0446	.0940
1.71	.4564	.9564	.0436	.0925
1.72	.4573	.9573	.0427	.0909
1.73	.4582	.9582	.0418	.0893
1.74	.4591	.9591	.0409	.0878
1.75	.4599	.9599	.0401	.0863
1.76	.4608	.9608	.0392	.0848
1.77	.4616	.9616	.0384	.0833
1.78	.4625	.9625	.0375	.0818
1.79	.4633	.9633	.0367	.0804

(Continúa)

(Continuación)

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Área desde la media a $\frac{x}{\sigma}$	(3) B Área de la parte mayor	(4) C Área de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
1.80	.4641	.9641	.0359	.0790
1.81	.4649	.9649	.0351	.0775
1.82	.4656	.9656	.0344	.0761
1.83	.4664	.9664	.0336	.0748
1.84	.4671	.9671	.0329	.0734
1.85	.4678	.9678	.0322	.0721
1.86	.4686	.9686	.0314	.0707
1.87	.4693	.9693	.0307	.0694
1.88	.4699	.9699	.0301	.0681
1.89	.4706	.9706	.0294	.0669
1.90	.4713	.9713	.0287	.0656
1.91	.4719	.9719	.0281	.0644
1.92	.4726	.9726	.0274	.0632
1.93	.4732	.9732	.0268	.0620
1.94	.4738	.9738	.0262	.0608
1.95	.4744	.9744	.0256	.0596
1.96	.4750	.9750	.0250	.0584
1.97	.4756	.9756	.0244	.0573
1.98	.4761	.9761	.0239	.0562
1.99	.4767	.9767	.0233	.0551
2.00	.4772	.9772	.0228	.0540
2.01	.4778	.9778	.0222	.0529
2.02	.4783	.9783	.0217	.0519
2.03	.4788	.9788	.0212	.0508
2.04	.4793	.9793	.0207	.0498
2.05	.4798	.9798	.0202	.0488
2.06	.4803	.9803	.0197	.0478
2.07	.4808	.9808	.0192	.0468
2.08	.4812	.9812	.0188	.0459
2.09	.4817	.9817	.0183	.0449
2.10	.4821	.9821	.0179	.0440
2.11	.4826	.9826	.0174	.0431
2.12	.4830	.9830	.0170	.0422
2.13	.4834	.9834	.0166	.0413
2.14	.4838	.9838	.0162	.0404
2.15	.4842	.9842	.0158	.0396
2.16	.4846	.9846	.0154	.0387
2.17	.4850	.9850	.0150	.0379
2.18	.4854	.9854	.0146	.0371
2.19	.4857	.9857	.0143	.0363
2.20	.4861	.9861	.0139	.0355
2.21	.4864	.9864	.0136	.0347
2.22	.4868	.9868	.0132	.0339
2.23	.4871	.9871	.0129	.0332
2.24	.4875	.9875	.0125	.0325

(Continúa)

(Continuación)

(1) <i>z</i> Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) <i>A</i> Área desde la media a $\frac{x}{\sigma}$	(3) <i>B</i> Área de la parte mayor	(4) <i>C</i> Área de la parte menor	(5) <i>y</i> Ordenada en $\frac{x}{\sigma}$
2.25	.4878	.9878	.0122	.0317
2.26	.4881	.9881	.0119	.0310
2.27	.4884	.9884	.0116	.0303
2.28	.4887	.9887	.0113	.0297
2.29	.4890	.9890	.0110	.0290
2.30	.4893	.9893	.0107	.0283
2.31	.4896	.9896	.0104	.0277
2.32	.4898	.9898	.0102	.0270
2.33	.4901	.9901	.0099	.0264
2.34	.4904	.9904	.0096	.0258
2.35	.4906	.9906	.0094	.0252
2.36	.4909	.9909	.0091	.0246
2.37	.4911	.9911	.0089	.0241
2.38	.4913	.9913	.0087	.0235
2.39	.4916	.9916	.0084	.0229
2.40	.4918	.9918	.0082	.0224
2.41	.4920	.9920	.0080	.0219
2.42	.4922	.9922	.0078	.0213
2.43	.4925	.9925	.0075	.0208
2.44	.4927	.9927	.0073	.0203
2.45	.4929	.9929	.0071	.0198
2.46	.4931	.9931	.0069	.0194
2.47	.4932	.9932	.0068	.0189
2.48	.4934	.9934	.0066	.0184
2.49	.4936	.9936	.0064	.0180
2.50	.4938	.9938	.0062	.0175
2.51	.4940	.9940	.0060	.0171
2.52	.4941	.9941	.0059	.0167
2.53	.4943	.9943	.0057	.0163
2.54	.4945	.9945	.0055	.0158
2.55	.4946	.9946	.0054	.0154
2.56	.4948	.9948	.0052	.0151
2.57	.4949	.9949	.0051	.0147
2.58	.4951	.9951	.0049	.0143
2.59	.4952	.9952	.0048	.0139
2.60	.4953	.9953	.0047	.0136
2.61	.4955	.9955	.0045	.0132
2.62	.4956	.9956	.0044	.0129
2.63	.4957	.9957	.0043	.0126
2.64	.4959	.9959	.0041	.0122
2.65	.4960	.9960	.0040	.0119
2.66	.4961	.9961	.0039	.0116
2.67	.4962	.9962	.0038	.0113
2.68	.4963	.9963	.0037	.0110
2.69	.4964	.9964	.0036	.0107

(Continúa)

(Continuación)

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Área desde la media a $\frac{x}{\sigma}$	(3) B Área de la parte mayor	(4) C Área de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
2.70	.4965	.9965	.0035	.0104
2.71	.4966	.9966	.0034	.0101
2.72	.4967	.9967	.0033	.0099
2.73	.4968	.9968	.0032	.0096
2.74	.4969	.9969	.0031	.0093
2.75	.4970	.9970	.0030	.0091
2.76	.4971	.9971	.0029	.0088
2.77	.4972	.9972	.0028	.0086
2.78	.4973	.9973	.0027	.0084
2.79	.4974	.9974	.0026	.0081
2.80	.4974	.9974	.0026	.0079
2.81	.4975	.9975	.0025	.0077
2.82	.4976	.9976	.0024	.0075
2.83	.4977	.9977	.0023	.0073
2.84	.4977	.9977	.0023	.0071
2.85	.4978	.9978	.0022	.0069
2.86	.4979	.9979	.0021	.0067
2.87	.4979	.9979	.0021	.0065
2.88	.4980	.9980	.0020	.0063
2.89	.4981	.9981	.0019	.0061
2.90	.4981	.9981	.0019	.0060
2.91	.4982	.9982	.0018	.0058
2.92	.4982	.9982	.0018	.0056
2.93	.4983	.9983	.0017	.0055
2.94	.4984	.9984	.0016	.0053
2.95	.4984	.9984	.0016	.0051
2.96	.4985	.9985	.0015	.0050
2.97	.4985	.9985	.0015	.0048
2.98	.4986	.9986	.0014	.0047
2.99	.4986	.9986	.0014	.0046
3.00	.4987	.9987	.0013	.0044
3.01	.4987	.9987	.0013	.0043
3.02	.4987	.9987	.0013	.0042
3.03	.4988	.9988	.0012	.0040
3.04	.4988	.9988	.0012	.0039
3.05	.4989	.9989	.0011	.0038
3.06	.4989	.9989	.0011	.0037
3.07	.4989	.9989	.0011	.0036
3.08	.4990	.9990	.0010	.0035
3.09	.4990	.9990	.0010	.0034
3.10	.4990	.9990	.0010	.0033
3.11	.4991	.9991	.0009	.0032
3.12	.4991	.9991	.0009	.0031
3.13	.4991	.9991	.0009	.0030
3.14	.4992	.9992	.0008	.0029

(Continúa)

(Continuación)

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Área desde la media a $\frac{x}{\sigma}$	(3) B Área de la parte mayor	(4) C Área de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
3.15	.4992	.9992	.0008	.0028
3.16	.4992	.9992	.0008	.0027
3.17	.4992	.9992	.0008	.0026
3.18	.4993	.9993	.0007	.0025
3.19	.4993	.9993	.0007	.0025
3.20	.4993	.9993	.0007	.0024
3.21	.4993	.9993	.0007	.0023
3.22	.4994	.9994	.0006	.0022
3.23	.4994	.9994	.0006	.0022
3.24	.4994	.9994	.0006	.0021
3.30	.4995	.9995	.0005	.0017
3.40	.4997	.9997	.0003	.0012
3.50	.4998	.9998	.0002	.0009
3.60	.4998	.9998	.0002	.0006
3.70	.4999	.9999	.0001	.0004

Tomada de DOWDIE, N. M. y HEATH, R. W.: *Métodos Estadísticos Aplicados*, Ed. del Castillo, Madrid, 1979, págs. 320-327.

Tabla 2. Valores críticos de t^*

gl	Nivel de significación para prueba de una cola					
	.10	.05	.025	.01	.005	.0005
	Nivel de significación para prueba de dos colas					
	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Tomada de SIEGEL, S.: *Estadística no paramétrica*, México, Trillas, 1980, pág. 282.

TABLA 3. Percentiles de distribución ji-cuadrado *

gl	Percentil														
	1	2	5	10	20	30	50	70	80	90	95	98	99	99.9	
1	.0002	.0006	.00393	.0158	.0642	.148	.455	1.074	1.642	2.706	3.841	5.412	6.635	10.827	
2	.0201	.0404	.103	.211	.446	.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.815	
3	.115	.185	.352	.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.341	16.268	
4	.297	.429	.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277	18.465	
5	.554	.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086	20.517	
6	.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22.457	
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322	
8	1.646	2.032	2.733	3.490	4.549	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.125	
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877	
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588	
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264	
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909	
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528	
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123	
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697	
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252	
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790	
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312	
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820	
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315	
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797	
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268	
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728	
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980	51.179	
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314	52.620	
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642	54.052	
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963	55.476	
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278	56.893	
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588	58.302	
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59.703	

* Si χ^2 es una variable con más de 30 gl, entonces $z = \sqrt{2\chi^2} - \sqrt{2gl} - \frac{1}{\sqrt{2gl}}$ se distribuye casi normalmente con una media de 0 y una desviación típica de 1.

Tomada de DOMENICH I MASSONS, J.: *Biostatística*, Herder, Barcelona, 1980, págs. 612-614 (ambas inclusive).

TABLA 4. Transformación z de Fisher de r_{xy}

r	z_r	r	z_r	r	z_r	r	z_r	r	z_r
.000	.000	.200	.203	.400	.424	.600	.693	.800	1.099
.005	.005	.205	.208	.405	.430	.605	.701	.805	1.113
.010	.010	.210	.213	.410	.436	.610	.709	.810	1.127
.015	.015	.215	.218	.415	.442	.615	.717	.815	1.142
.020	.020	.220	.224	.420	.448	.620	.725	.820	1.157
.025	.025	.225	.229	.425	.454	.625	.733	.825	1.172
.030	.030	.230	.234	.430	.460	.630	.741	.830	1.188
.035	.035	.235	.239	.435	.466	.635	.750	.835	1.204
.040	.040	.240	.245	.440	.472	.640	.758	.840	1.221
.045	.045	.245	.250	.445	.478	.645	.767	.845	1.238
.050	.050	.250	.255	.450	.485	.650	.775	.850	1.256
.055	.055	.255	.261	.455	.491	.655	.784	.855	1.274
.060	.060	.260	.266	.460	.497	.660	.793	.860	1.293
.065	.065	.265	.271	.465	.504	.665	.802	.865	1.313
.070	.070	.270	.277	.470	.510	.670	.811	.870	1.333
.075	.075	.275	.282	.475	.517	.675	.820	.875	1.354
.080	.080	.280	.288	.480	.523	.680	.829	.880	1.376
.085	.085	.285	.293	.485	.530	.685	.838	.885	1.398
.090	.090	.290	.299	.490	.536	.690	.848	.890	1.422
.095	.095	.295	.304	.495	.543	.695	.858	.895	1.447
.100	.100	.300	.310	.500	.549	.700	.867	.900	1.472
.105	.105	.305	.315	.505	.556	.705	.877	.905	1.499
.110	.110	.310	.321	.510	.563	.710	.887	.910	1.528
.115	.116	.315	.326	.515	.570	.715	.897	.915	1.557
.120	.121	.320	.332	.520	.576	.720	.908	.920	1.589
.125	.126	.325	.337	.525	.583	.725	.918	.925	1.623
.130	.131	.330	.343	.530	.590	.730	.929	.930	1.658
.135	.136	.335	.348	.535	.597	.735	.940	.935	1.697
.140	.141	.340	.354	.540	.604	.740	.950	.940	1.738
.145	.146	.345	.360	.545	.611	.745	.962	.945	1.783
.150	.151	.350	.365	.550	.618	.750	.973	.950	1.832
.155	.156	.355	.371	.555	.626	.755	.984	.955	1.886
.160	.161	.360	.377	.560	.633	.760	.996	.960	1.946
.165	.167	.365	.383	.565	.640	.765	1.008	.965	2.014
.170	.172	.370	.388	.570	.648	.770	1.020	.970	2.092
.175	.177	.375	.394	.575	.655	.775	1.033	.975	2.185
.180	.182	.380	.400	.580	.662	.780	1.045	.980	2.298
.185	.187	.385	.406	.585	.670	.785	1.058	.985	2.443
.190	.192	.390	.412	.590	.678	.790	1.071	.990	2.647
.195	.198	.395	.418	.595	.685	.795	1.085	.995	2.994

Tomada de GLASS, G. V. y otro: *Métodos Estadísticos Aplicados a las Ciencias Sociales*, México, Prentice-Hall, Madrid, 1980, p. 532.

TABLA 5. Valores absolutos críticos del coeficiente de correlación de Spearman por rangos ordenados r_s , para el contraste de hipótesis nula de inexistencia de correlación mediante el contraste de dos colas*

n	$\alpha = .10$	$\alpha = .05$	$\alpha = .02$	$\alpha = .01$
5	0.900	—	—	—
6	0.829	0.886	0.943	—
7	0.714	0.786	0.893	—
8	0.643	0.738	0.833	0,881
9	0.600	0.683	0.783	0,833
10	0.564	0.648	0.745	0,818
11	0.523	0.623	0.736	0,794
12	0.497	0.591	0.703	0,780
13	0.475	0.566	0.673	0,745
14	0.457	0.545	0.646	0,716
15	0.441	0.525	0.623	0,689
16	0.425	0.507	0.601	0,666
17	0.412	0.490	0.582	0,645
18	0.399	0.476	0.564	0,625
19	0.388	0.462	0.549	0,608
20	0.377	0.450	0.534	0,591
21	0.368	0.438	0.521	0,576
22	0.359	0.428	0.508	0,562
23	0.351	0.418	0.496	0,549
24	0.343	0.409	0.485	0,537
25	0.336	0.400	0.475	0,526
26	0.329	0.392	0.465	0,515
27	0.323	0.385	0.456	0,505
28	0.317	0.377	0.448	0,496
29	0.311	0.370	0.440	0,487
30	0.305	0.364	0.432	0,478

* En la tabla aparecen los valores absolutos del contraste de dos colas para valores críticos. Por ejemplo, los valores críticos de r_s , con $n = 10$ y $\alpha = 0,10$, son $+0,564$ y $-0,564$.

Tomada de GLASS, G. V. y otro: O. c., p. 537.

TABLA 6. Valores críticos de r_s , el coeficiente de correlación de rangos de Spearman*

N	Nivel de significación (prueba de una cola)	
	.05	.01
4	1.000	
5	.900	1.000
6	.829	.943
7	.714	.893
8	.643	.833
9	.600	.783
10	.564	.746
12	.506	.712
14	.456	.645
16	.425	.601
18	.399	.564
20	.377	.534
22	.359	.508
24	.343	.485
26	.329	.465
28	.317	.448
30	.306	.432

* Tomada de OLDS, E. G. 1938. Distribuciones de la sumas de los cuadrados de diferencias de rangos para números pequeños de individuos. *Ann. Math. Statist.*, 9, 133-148, y de OLDS, E. G. 1949. Los niveles de significación del 5% para las sumas de los cuadrados de diferencias de rangos y una corrección. *Ann. Math. Statist.*, 20, 117-118, con el amable permiso del autor y el editor.

Tomada de SIEGEL, S.: *Estadística no paramétrica*, O. c., p, 318.

I. Probabilidades asociadas con valores tan pequeños como los valores observados de U en la prueba de Mann-Whitney*

		$n_2 = 3$					$n_2 = 4$			
n_1		1	2	3	n_1		1	2	3	4
U					U					
0		.250	.100	.050	0		.200	.067	.028	.014
1		.500	.200	.100	1		.400	.133	.057	.029
2		.750	.400	.200	2		.600	.267	.114	.057
3			.600	.350	3			.400	.200	.100
4				.500	4			.600	.314	.171
5				.650	5				.429	.243
					6				.571	.343
					7					.443
					8					.557

		$n_2 = 5$							$n_2 = 6$					
n_1		1	2	3	4	5	n_1		1	2	3	4	5	6
U							U							
0		.167	.047	.018	.008	.004	0		.143	.036	.012	.005	.002	.001
1		.333	.095	.036	.016	.008	1		.286	.071	.024	.010	.004	.002
2		.500	.190	.071	.032	.016	2		.428	.143	.048	.019	.009	.004
3		.667	.286	.125	.056	.028	3		.571	.214	.083	.033	.015	.008
4			.429	.196	.095	.048	4			.321	.131	.057	.026	.013
5			.571	.286	.143	.075	5			.429	.190	.086	.041	.021
6				.393	.206	.111	6			.571	.274	.129	.063	.032
7				.500	.278	.155	7				.357	.176	.089	.047
8				.607	.365	.210	8				.452	.238	.123	.066
9					.452	.274	9				.548	.305	.165	.090
10					.548	.345	10					.381	.214	.120
11						.421	11					.457	.268	.155
12						.500	12					.545	.331	.197
13						.579	13						.396	.242
							14						.465	.294
							15						.535	.350
							16							.409
							17							.469
							18							.531

* Reproducida de MANN, H. B., y WHITNEY, D. R., 1947. En una prueba de si una o dos variables aleatorias es estocásticamente mayor que la otra. *Ann Math. Statist.*, 18, 52-54, con el amable permiso de los autores y el editor.

Tomada de SIEGEL, S.: O. c., pp. 305-307.

(Continúa)

(Continuación)

		$n_2 = 7$						
U	n_1	1	2	3	4	5	6	7
0		.125	.028	.008	.003	.001	.001	.000
1		.250	.056	.017	.006	.003	.001	.001
2		.375	.111	.033	.012	.005	.002	.001
3		.500	.167	.058	.021	.009	.004	.002
4		.625	.250	.092	.036	.015	.007	.003
5			.333	.133	.055	.024	.011	.006
6			.444	.192	.082	.037	.017	.009
7			.556	.258	.115	.053	.026	.013
8				.333	.158	.074	.037	.019
9				.417	.206	.101	.051	.027
10				.500	.264	.134	.069	.036
11				.583	.324	.172	.090	.049
12					.394	.216	.117	.064
13					.464	.265	.147	.082
14					.538	.319	.183	.104
15						.378	.223	.130
16						.438	.267	.159
17						.500	.314	.191
18						.562	.365	.228
19							.418	.267
20							.473	.310
21							.527	.355
22								.402
23								.451
24								.500
25								.549

* Reproducida de MANN, H. B., y WHITNEY, D. R., 1947. En una prueba de si una o dos variables aleatorias es estocásticamente mayor que la otra. *Ann Math. Statist.*, 18, 52-54, con el amable permiso de los autores y el editor.

Tomada de SIEGEL, S.: O. c., p. 284.

(Continúa)

(Continuación)

		$n_2 = 8$								
U	n_1	1	2	3	4	5	6	7	8	t Normal
0		.111	.022	.006	.002	.001	.000	.000	.000	3.308 .001
1		.222	.044	.012	.004	.002	.001	.000	.000	3.203 .001
2		.333	.089	.024	.008	.003	.001	.001	.000	3.098 .001
3		.444	.133	.042	.014	.005	.002	.001	.001	2.993 .001
4		.556	.200	.067	.024	.009	.004	.002	.001	2.888 .002
5			.267	.097	.036	.015	.006	.003	.001	2.783 .003
6			.356	.139	.055	.023	.010	.005	.002	2.678 .004
7			.444	.188	.077	.033	.015	.007	.003	2.573 .005
8			.556	.248	.107	.047	.021	.010	.005	2.468 .007
9				.315	.141	.064	.030	.014	.007	2.363 .009
10				.387	.184	.085	.041	.020	.010	2.258 .012
11				.461	.230	.111	.054	.027	.014	2.153 .016
12				.539	.285	.142	.071	.036	.019	2.048 0.20
13					.341	.177	.091	.047	.025	1.943 0.26
14					.404	.217	.114	.060	.032	1.838 .033
15					.467	.262	.141	.076	.041	1.733 0.41
16					.533	.311	.172	.095	.052	1.628 0.52
17						.362	.207	.116	.065	1.523 .064
18						.416	.245	.140	.080	1.418 .078
19						.472	.286	.168	.097	1.313 .094
20						.528	.331	.198	.117	1.208 .113
21							.377	.232	.139	1.102 .135
22							.426	.268	.164	.998 .159
23							.475	.306	.191	.893 .185
24							.525	.347	.221	.788 .215
25								.389	.253	.683 .247
26								.433	.287	.578 .282
27								.478	.323	.473 .318
28								.522	.360	.368 .356
29									.399	.263 .396
30									.439	.158 .437
31									.480	.052 .481
32									.520	

* Reproducida de MANN, H. B., y WHITNEY, D. R., 1947. En una prueba de si una o dos variables aleatorias es estocásticamente mayor que la otra. *Ann Math. Statist.*, 18, 52-54, con el amable permiso de los autores y el editor.

II. Valores críticos de *U* en la prueba de Mann-Whitney*

Valores críticos de *U* para una prueba de una cola en $\alpha = 0.001$
o para una prueba de dos colas en $\alpha = 0.002$.

$n_2 \backslash n_1$	9	10	11	12	13	14	15	16	17	18	19	20
1												
2												
3									0	0	0	0
4		0	0	0	1	1	1	2	2	3	3	3
5	1	1	2	2	3	3	4	5	5	6	7	7
6	2	3	4	4	5	6	7	8	9	10	11	12
7	3	5	6	7	8	9	10	11	13	14	15	16
8	5	6	8	9	11	12	14	15	17	18	20	21
9	7	8	10	12	14	15	17	19	21	23	25	26
10	8	10	12	14	17	19	21	23	25	27	29	32
11	10	12	15	17	20	22	24	27	29	32	34	37
12	12	14	17	20	23	25	28	31	34	37	40	42
13	14	17	20	23	26	29	32	35	38	42	45	48
14	15	19	22	25	29	32	36	39	43	46	50	54
15	17	21	24	28	32	36	40	43	47	51	55	59
16	19	23	27	31	35	39	43	48	52	56	60	65
17	21	25	29	34	38	43	47	52	57	61	66	70
18	23	27	32	37	42	46	51	56	61	66	71	76
19	25	29	34	40	45	50	55	60	66	71	77	82
20	26	32	37	42	48	54	59	65	70	76	82	88

* Tomada de SIEGEL, S.: O. c., pp. 308-311.

(Continúa)

(Continuación)

Valores críticos de U para una prueba de una cola en $\alpha = 0.01$
o para una prueba de dos colas en $\alpha = 0.02$.

$n_2 \backslash n_1$	9	10	11	12	13	14	15	16	17	18	19	20
1												
2					0	0	0	0	0	0	1	1
3	1	1	1	2	2	2	3	3	4	4	4	5
4	3	3	4	5	5	6	7	7	8	9	9	10
5	5	6	7	8	9	10	11	12	13	14	15	16
6	7	8	9	11	12	13	15	16	18	19	20	22
7	9	11	12	14	16	17	19	21	23	24	26	28
8	11	13	15	17	20	22	24	26	28	30	32	34
9	14	16	18	21	23	26	28	31	33	36	38	40
10	16	19	22	24	27	30	33	36	38	41	44	47
11	18	22	25	28	31	34	37	41	44	47	50	53
12	21	24	28	31	35	38	42	46	49	53	56	60
13	23	27	31	35	39	43	47	51	55	59	63	67
14	26	30	34	38	43	47	51	56	60	65	69	73
15	28	33	37	42	47	51	56	61	66	70	75	80
16	31	36	41	46	51	56	61	66	71	76	82	87
17	33	38	44	49	55	60	66	71	77	82	88	93
18	36	41	47	53	59	65	70	76	82	88	94	100
19	38	44	50	56	63	69	75	82	88	94	101	107
20	40	47	53	60	67	73	80	87	93	100	107	114

* Tomada abreviadamente de las tablas 1, 3, 5 y 7 de AUBLE, D. 1953. Tablas extendidas para la estadística de Mann-Whitney. *Bulletin of the Institute of Educational Research at Indiana University*, 1, núm. 2, con el amable permiso del autor y el editor.

(Continúa)

(Continuación)

Valores críticos de U para una prueba de una cola en $\alpha = 0.025$
o para una prueba de dos colas en $\alpha = 0.05$.

$n_2 \backslash n_1$	9	10	11	12	13	14	15	16	17	18	19	20
1												
2	0	0	0	1	1	1	1	1	2	2	2	2
3	2	3	3	4	4	5	5	6	6	7	7	8
4	4	5	6	7	8	9	10	11	11	12	13	13
5	7	8	9	11	12	13	14	15	17	18	19	20
6	10	11	13	14	16	17	19	21	22	24	25	27
7	12	14	16	18	20	22	24	26	28	30	32	34
8	15	17	19	22	24	26	29	31	34	36	38	41
9	17	20	23	26	28	31	34	37	39	42	45	48
10	20	23	26	29	33	36	39	42	45	48	52	55
11	23	26	30	33	37	40	44	47	51	55	58	62
12	26	29	33	37	41	45	49	53	57	61	65	69
13	28	33	37	41	45	50	54	59	63	67	72	76
14	31	36	40	45	50	55	59	64	67	74	78	83
15	34	39	44	49	54	59	64	70	75	80	85	90
16	37	42	47	53	59	64	70	75	81	86	92	98
17	39	45	51	57	63	67	75	81	87	93	99	105
18	42	48	55	61	67	74	80	86	93	99	106	112
19	45	52	58	65	72	78	85	92	99	106	113	119
20	48	55	62	69	76	83	90	98	105	112	119	127

* Tomada abreviadamente de las tablas 1, 3, 5 y 7 de AUBLE, D. 1953. Tablas extendidas para la estadística de Mann-Whitney. *Bulletin of the Institute of Educational Research at Indiana University*, 1, núm. 2, con el amable permiso del autor y el editor.

(Continúa)

(Continuación)

Valores críticos de U para una prueba de una cola en $\alpha = 0.05$ o para una prueba de dos colas en $\alpha = 0.10$.

$n_1 \backslash n_2$	9	10	11	12	13	14	15	16	17	18	19	20
1											0	0
2	1	1	1	2	2	2	3	3	3	4	4	4
3	3	4	5	5	6	7	7	8	9	9	10	11
4	6	7	8	9	10	11	12	14	15	16	17	18
5	9	11	12	13	15	16	18	19	20	22	23	25
6	12	14	16	17	19	21	23	25	26	28	30	32
7	15	17	19	21	24	26	28	30	33	35	37	39
8	18	20	23	26	28	31	33	36	39	41	44	47
9	21	24	27	30	33	36	39	42	45	48	51	54
10	24	27	31	34	37	41	44	48	51	55	58	62
11	27	31	34	38	42	46	50	54	57	61	65	69
12	30	34	38	42	47	51	55	60	64	68	72	77
13	33	37	42	47	51	56	61	65	70	75	80	84
14	36	41	46	51	56	61	66	71	77	82	87	92
15	39	44	50	55	61	66	72	77	83	88	94	100
16	42	48	54	60	65	71	77	83	89	95	101	107
17	45	51	57	64	70	77	83	89	96	102	109	115
18	48	55	61	68	75	82	88	95	102	109	116	123
19	51	58	65	72	80	87	94	101	109	116	123	130
20	54	62	69	77	84	92	100	107	115	123	130	138

* Tomada abreviadamente de las tablas 1, 3, 5 y 7 de AUBLE, D. 1953. Tablas extendidas para la estadística de Mann-Whitney. *Bulletin of the Institute of Educational Research at Indiana University*, 1, núm. 2, con el amable permiso del autor y el editor.

TABLA 7. Valores críticos de T , en la prueba de los rangos señalados de pares igualados de Wilcoxon*

N	<i>Nivel de significación para prueba de una cola</i>		
	<i>.025</i>	<i>.01</i>	<i>.005</i>
	<i>Nivel de significación para prueba de dos colas</i>		
	<i>.05</i>	<i>.02</i>	<i>.01</i>
6	0	—	—
7	2	0	—
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

* Tomada de SIEGEL, S.: O. c., p. 288.

TABLA 8. Percentiles de distribución F

Percentiles 75

$\frac{n_1}{n_2}$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.41	9.49	9.58	9.63	9.67	9.71	9.76	9.80	9.85
2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.41	3.43	3.43	3.44	3.45	3.46	3.47	3.48
3	2.02	2.28	2.36	2.41	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47
4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.88	1.87	1.87	1.87
6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.76	1.76	1.75	1.75	1.75	1.74	1.74	1.74
7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69	1.69	1.68	1.68	1.67	1.67	1.66	1.66	1.65	1.65	1.65
8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63	1.62	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58
9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.54	1.53	1.53
10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.53	1.52	1.52	1.51	1.51	1.50	1.49	1.48
11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47	1.46	1.45
12	1.46	1.56	1.55	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.42
13	1.45	1.55	1.55	1.54	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40
14	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.41	1.40	1.39	1.38
15	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45	1.44	1.43	1.41	1.41	1.40	1.39	1.38	1.37	1.36
16	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.44	1.43	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
17	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33
18	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32
19	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.30
20	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.29
21	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28
22	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.37	1.36	1.34	1.33	1.32	1.31	1.30	1.29	1.28
23	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28	1.27
24	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.36	1.35	1.33	1.32	1.31	1.30	1.29	1.28	1.26
25	1.39	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.32	1.31	1.29	1.28	1.27	1.25
26	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.35	1.34	1.32	1.31	1.30	1.29	1.28	1.26	1.25
27	1.38	1.46	1.45	1.43	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.33	1.32	1.31	1.30	1.28	1.27	1.26	1.24
28	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.31	1.30	1.29	1.28	1.27	1.25	1.24
29	1.38	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.32	1.31	1.30	1.29	1.27	1.26	1.25	1.23
30	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.34	1.32	1.30	1.29	1.28	1.27	1.26	1.24	1.23
40	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.31	1.30	1.28	1.26	1.25	1.24	1.22	1.21	1.19
60	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.27	1.25	1.24	1.22	1.21	1.19	1.17	1.15
120	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.26	1.24	1.22	1.21	1.19	1.18	1.16	1.13	1.10
∞	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.22	1.19	1.18	1.16	1.14	1.12	1.08	1.00

Tomada de GLASS, G. V. y otro: O. c., pp. 520-526.

Percentiles 90

$n_1 \backslash n_2$		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33	
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49	
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.22	5.20	5.18	5.17	5.16	5.15	5.14	5.13	
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76	
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.24	3.19	3.17	3.16	3.14	3.12	3.10	
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72	
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47	
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29	
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16	
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.24	2.20	2.18	2.16	2.13	2.11	2.08	
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.17	2.12	2.10	2.08	2.05	2.03	2.00	
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.10	2.06	2.04	2.01	1.99	1.96	1.93	
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.05	2.01	1.98	1.96	1.93	1.90	1.85	
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	2.01	1.96	1.94	1.91	1.89	1.86	1.80	
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.76	
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.72	
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.69	
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.94	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.66	
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.88	1.83	1.79	1.76	1.73	1.70	1.67	1.63	
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.61	
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.59	
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.57	
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.55	
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.53	
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.52	
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.50	
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.49	
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.48	
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.47	
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.47	
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38	
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29	
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.21	
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.50	1.44	1.38	1.34	1.30	1.24	1.17	1.00	

(Continúa)

(Continuación) n_1 n_2		Percentiles 95																		(Continúa) ∞ 120 60 40 30 24 20 15 12 10 9 8 7 6 5 4 3 2 1	
1	161.4	199.55	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3		
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50		
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53		
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63		
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36		
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67		
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23		
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93		
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71		
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54		
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40		
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30		
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21		
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13		
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07		
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01		
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96		
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92		
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88		
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84		
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81		
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78		
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76		
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73		
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71		
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69		
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67		
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65		
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64		
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62		
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51		
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39		
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25		
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00		

$\frac{n_1}{n_2}$		Percentiles 97.5																		
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018	
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.47	39.47	39.48	39.49	39.50	
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90	
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26	
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02	
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85	
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14	
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67	
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33	
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08	
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88	
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72	
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60	
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49	
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40	
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32	
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25	
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19	
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13	
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09	
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04	
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00	
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97	
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94	
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91	
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88	
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85	
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83	
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81	
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79	
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64	
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48	
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31	
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00	

Continúa

(Continúa)

(Continuación)

</	
---	--

(Continuación)

Percentiles 99.5

n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
n_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24426	24630	24836	24940	25044	25148	25253	25359	25465
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88	43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	41.83
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14	20.97	20.70	20.44	20.17	20.3	19.89	19.75	19.61	19.47	19.32
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	12.14
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.18	7.97	7.75	7.65	7.53	7.42	7.31	7.19	7.08
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.44	4.34	4.23
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88	3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81	3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75	3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64	3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60	3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56	3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.29
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52	3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.25
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48	3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.21
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	1.69
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	1.43
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00

(Continúa)

(Continuación)

Percentiles 99.9

$\frac{n_1}{n_2}$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4053*	5000*	5404*	5625*	5764*	5859*	5929*	5981*	6023*	6056*	6107*	6158*	6209*	6235*	6261*	6287*	6313*	6340*	6366*
2	998.5	999.0	999.2	999.2	999.3	999.3	999.4	999.4	999.4	999.4	999.4	999.4	999.4	999.5	999.5	999.5	999.5	999.5	999.5
3	167.0	148.5	141.1	137.1	134.6	132.8	131.6	130.6	129.9	129.2	128.3	127.4	126.4	125.9	125.4	125.0	124.5	124.0	123.5
4	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47	48.05	47.41	46.76	46.10	45.77	45.43	45.09	44.75	44.40	44.05
5	47.18	37.12	33.20	31.09	29.75	28.84	28.16	27.64	27.24	26.92	26.42	25.91	25.39	25.14	24.87	24.60	24.33	24.06	23.79
6	35.51	27.00	23.70	21.92	20.81	20.03	19.46	19.03	18.69	18.41	17.99	17.56	17.12	16.89	16.67	16.44	16.21	15.99	15.75
7	29.25	21.69	18.77	17.19	16.21	15.52	15.02	14.63	14.33	14.08	13.71	13.32	12.93	12.63	12.33	12.03	11.71	11.51	11.20
8	25.42	18.49	15.83	14.39	13.49	12.86	12.40	12.04	11.77	11.54	11.19	10.84	10.48	10.30	10.11	9.92	9.73	9.53	9.33
9	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11	9.89	9.57	9.24	8.90	8.72	8.55	8.37	8.19	8.00	7.81
10	21.04	14.91	12.55	11.28	10.48	9.92	9.52	9.20	8.96	8.75	8.45	8.13	7.80	7.64	7.47	7.30	7.12	6.94	6.76
11	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12	7.92	7.63	7.32	7.01	6.85	6.68	6.52	6.35	6.17	6.00
12	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48	7.29	7.00	6.71	6.40	6.25	6.09	5.93	5.76	5.59	5.42
13	17.81	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98	6.80	6.52	6.23	5.93	5.78	5.63	5.47	5.30	5.14	4.97
14	17.14	11.78	9.73	8.62	7.92	7.43	7.08	6.80	6.58	6.40	6.13	5.85	5.56	5.41	5.25	5.10	4.94	4.77	4.60
15	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.81	5.54	5.25	5.10	4.95	4.80	4.64	4.47	4.31
16	16.12	10.97	9.00	7.94	7.27	6.81	6.46	6.19	5.98	5.81	5.55	5.27	4.99	4.85	4.70	4.54	4.39	4.23	4.06
17	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75	5.58	5.32	5.05	4.78	4.63	4.48	4.33	4.18	4.02	3.85
18	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56	5.39	5.13	4.87	4.59	4.43	4.30	4.15	4.00	3.84	3.67
19	15.08	10.16	8.28	7.26	6.62	6.16	5.85	5.59	5.39	5.22	4.97	4.70	4.43	4.29	4.14	3.99	3.84	3.68	3.51
20	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.82	4.56	4.29	4.15	4.00	3.86	3.70	3.54	3.38
21	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11	4.95	4.70	4.44	4.17	4.03	3.88	3.74	3.58	3.42	3.26
22	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99	4.83	4.58	4.33	4.06	3.92	3.78	3.63	3.48	3.32	3.15
23	14.19	9.47	7.67	6.69	6.08	5.65	5.33	5.09	4.89	4.73	4.48	4.23	3.97	3.82	3.68	3.53	3.38	3.22	3.05
24	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80	4.64	4.39	4.14	3.87	3.74	3.59	3.45	3.29	3.14	2.97
25	13.88	9.22	7.45	6.49	5.88	5.46	5.15	4.91	4.71	4.56	4.31	4.06	3.79	3.66	3.52	3.37	3.22	3.06	2.89
26	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64	4.48	4.24	3.99	3.72	3.59	3.44	3.30	3.15	2.99	2.82
27	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57	4.41	4.17	3.92	3.66	3.52	3.38	3.23	3.08	2.92	2.75
28	13.50	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50	4.35	4.11	3.86	3.60	3.46	3.32	3.18	3.02	2.86	2.69
29	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45	4.29	4.05	3.80	3.54	3.41	3.27	3.12	2.97	2.81	2.64
30	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.00	3.75	3.49	3.36	3.22	3.07	2.92	2.76	2.59
40	12.61	8.25	6.60	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.64	3.40	3.15	3.01	2.87	2.73	2.57	2.41	2.23
60	11.97	7.76	6.17	5.31	4.76	4.37	4.09	3.87	3.69	3.54	3.31	3.08	2.83	2.69	2.55	2.41	2.25	2.08	1.89
120	11.38	7.32	5.79	4.95	4.42	4.04	3.77	3.55	3.38	3.24	3.02	2.78	2.53	2.40	2.26	2.11	1.95	1.76	1.54
∞	10.83	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10	2.96	2.74	2.51	2.27	2.13	1.99	1.84	1.66	1.45	1.00

* Multiplicar por 100 estas entradas.

Tomada de GASS, G. V. y STANLEY, J. C., O., C., pp. 520-524.

TABLA 9. Percentiles de las distribuciones por rangos, q , de Student para J y v grados de libertad*

Percentiles 90									
$J \backslash v$	2	3	4	5	6	7	8	9	10
1	8.929	13.44	16.36	18.49	20.15	21.51	22.64	23.62	24.48
2	4.130	5.733	6.773	7.538	8.139	8.633	9.049	9.049	9.725
3	3.328	4.467	5.199	5.738	6.162	6.511	6.806	7.062	7.287
4	3.015	3.976	4.586	5.035	5.388	5.679	5.926	6.139	6.327
5	2.850	3.717	4.264	4.664	4.979	5.238	5.458	5.648	5.816
6	2.748	3.559	4.065	4.435	4.726	4.966	5.168	5.344	5.499
7	2.680	3.451	3.931	4.280	4.555	4.780	4.972	5.137	5.283
8	2.630	3.374	3.834	4.169	4.431	4.646	4.829	4.987	5.126
9	2.592	3.316	3.761	4.084	4.337	4.545	4.721	4.873	5.007
10	2.563	3.270	3.704	4.018	4.264	4.465	4.636	4.783	4.913
11	2.540	3.234	3.658	3.965	4.205	4.401	4.568	4.711	4.838
12	2.521	3.204	3.621	3.922	4.156	4.349	4.511	4.652	4.776
13	2.505	3.179	3.589	3.885	4.116	4.305	4.464	4.602	4.724
14	2.491	3.158	3.563	3.854	4.081	4.267	4.424	4.560	4.680
15	2.479	3.140	3.540	3.828	4.052	4.235	4.390	4.524	4.641
16	2.469	3.124	3.520	3.804	4.026	4.207	4.360	4.492	4.608
17	2.460	3.110	3.503	3.784	4.004	4.183	4.334	4.464	4.579
18	2.452	3.098	3.488	3.767	3.984	4.161	4.311	4.440	4.554
19	2.445	3.087	3.474	3.751	3.966	4.142	4.290	4.418	4.531
20	2.439	3.078	3.462	3.736	3.950	4.124	4.271	4.398	4.510
24	2.420	3.047	3.423	3.692	3.900	4.070	4.213	4.336	4.445
30	2.400	3.017	3.386	3.648	3.851	4.016	4.155	4.275	4.381
40	2.381	2.988	3.349	3.605	3.803	3.963	4.099	4.215	4.317
60	2.363	2.959	3.312	3.562	3.755	3.911	4.042	4.155	4.254
120	2.344	2.930	3.276	3.520	3.707	3.859	3.987	4.096	4.191
∞	2.326	2.902	3.240	3.478	3.661	3.808	3.931	4.037	4.129

$J \backslash v$	11	12	13	14	15	16	17	18	19
1	25.24	25.92	26.54	27.10	27.62	28.10	28.54	28.96	29.35
2	10.01	10.26	10.49	10.70	10.89	11.07	11.24	11.39	11.54
3	7.487	7.667	7.832	7.982	8.120	8.249	8.368	8.479	8.584
4	6.495	6.645	6.783	6.909	7.025	7.133	7.233	7.327	7.414
5	5.966	6.101	6.223	6.336	6.440	6.536	6.626	6.710	6.789
6	5.637	5.762	5.875	5.979	6.075	6.164	6.247	6.325	6.398
7	5.413	5.530	5.637	5.735	5.826	5.910	5.988	6.061	6.130
8	5.250	5.362	5.464	5.558	5.644	5.724	5.799	5.869	5.935
9	5.127	5.234	5.333	5.423	5.506	5.583	5.655	5.723	5.786
10	5.029	5.134	5.229	5.317	5.397	5.472	5.542	5.607	5.668
11	4.951	5.053	5.146	5.231	5.309	5.382	5.450	5.514	5.573
12	4.886	4.986	5.077	5.160	5.236	5.308	5.374	5.436	5.495
13	4.832	4.930	5.019	5.100	5.176	5.245	5.311	5.372	5.429
14	4.786	4.882	4.970	5.050	5.124	5.192	5.256	5.316	5.373
15	4.746	4.841	4.927	5.006	5.079	5.147	5.209	5.269	5.324
16	4.712	4.805	4.890	4.968	5.040	5.107	5.169	5.227	5.282
17	4.682	4.774	4.858	4.935	5.005	5.071	5.133	5.190	5.244
18	4.655	4.746	4.829	4.905	4.975	5.040	5.101	5.158	5.211
19	4.631	4.721	4.803	4.879	4.948	5.012	5.073	5.129	5.182
20	4.609	4.699	4.780	4.855	4.924	4.987	5.047	5.103	5.155
24	4.541	4.628	4.708	4.780	4.847	4.909	4.966	5.021	5.071
30	4.474	4.559	4.635	4.706	4.770	4.830	4.886	4.939	4.988
40	4.408	4.490	4.564	4.632	4.695	4.752	4.807	4.857	4.905
60	4.342	4.421	4.493	4.558	4.619	4.675	4.727	4.775	4.821
120	4.276	4.353	4.422	4.485	4.543	4.597	4.647	4.694	4.738
∞	4.211	4.285	4.351	4.412	4.468	4.519	4.568	4.612	4.654

Tomada de GLASS, G. V. y otro: O. c., pp. 527-531.
 * En un análisis de varianza de un factor con n observaciones en los J grupos, $v = J(n - 1)$. En general, v es el número de grados de libertad para la media de cuadrados «intra» del análisis de varianza.

(Continúa)

(Continuación)

Percentiles 95

$\begin{matrix} J \\ v \end{matrix}$	2	3	4	5	6	7	8	9	10
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
2	6.085	8.331	9.798	10.88	11.74	12.44	13.03	13.54	13.99
3	4.501	5.910	6.825	7.502	8.037	8.478	8.853	9.177	9.462
4	3.927	5.040	5.757	6.287	6.707	7.053	7.347	7.602	7.826
5	3.635	4.602	5.218	5.673	6.033	6.330	6.582	6.802	6.995
6	3.461	4.339	4.896	5.305	5.628	5.895	6.122	6.319	6.493
7	3.344	4.165	4.681	5.060	5.359	5.606	5.815	5.998	6.158
8	3.261	4.041	4.529	4.886	5.167	5.399	5.597	5.767	5.918
9	3.199	3.949	4.415	4.756	5.024	5.244	5.432	5.595	5.739
10	3.151	3.877	4.327	4.654	4.912	5.124	5.305	5.461	5.599
11	3.113	3.820	4.256	4.574	4.823	5.028	5.202	5.353	5.487
12	3.082	3.773	4.199	4.508	4.751	4.950	5.119	5.265	5.395
13	3.055	3.735	4.151	4.453	4.690	4.885	5.049	5.192	5.318
14	3.033	3.702	4.111	4.407	4.639	4.829	4.990	5.131	5.254
15	3.014	3.674	4.076	4.367	4.595	4.782	4.940	5.077	5.198
16	2.998	3.649	4.046	4.333	4.557	4.741	4.897	5.031	5.150
17	2.984	3.628	4.020	4.303	4.524	4.705	4.858	4.991	5.108
18	2.971	3.609	3.997	4.277	4.495	4.673	4.824	4.956	5.071
19	2.960	3.593	3.977	4.253	4.469	4.645	4.794	4.924	5.038
20	2.950	3.578	3.958	4.232	4.445	4.620	4.768	4.896	5.008
24	2.919	3.532	3.901	4.166	4.373	4.541	4.684	4.807	4.915
30	2.888	3.486	3.845	4.102	4.302	4.464	4.602	4.720	4.824
40	2.858	3.442	3.791	4.039	4.232	4.389	4.521	4.635	4.735
60	2.829	3.399	3.737	3.977	4.163	4.314	4.441	4.550	4.646
120	2.800	3.356	3.685	3.917	4.096	4.241	4.363	4.468	4.560
∞	2.772	3.314	3.633	3.858	4.030	4.170	4.286	4.387	4.474

$\begin{matrix} J \\ v \end{matrix}$	11	12	13	14	15	16	17	18	19
1	50.59	51.96	53.20	54.33	55.36	56.32	57.22	58.04	58.83
2	14.39	14.75	15.08	15.38	15.65	15.91	16.14	16.37	16.57
3	9.717	9.946	10.15	10.35	10.53	10.69	10.84	10.98	11.11
4	8.027	8.208	8.373	8.525	8.664	8.794	8.914	9.028	9.134
5	7.168	7.324	7.466	7.596	7.717	7.828	7.932	8.030	8.122
6	6.649	6.789	6.917	7.034	7.143	7.244	7.338	7.426	7.508
7	6.302	6.431	6.550	6.658	6.759	6.852	6.939	7.020	7.097
8	6.054	6.175	6.287	6.389	6.483	6.571	6.653	6.729	6.802
9	5.867	5.983	6.089	6.186	6.276	6.359	6.437	6.510	6.579
10	5.722	5.833	5.935	6.028	6.114	6.194	6.269	6.339	6.405
11	5.605	5.713	5.811	5.901	5.984	6.062	6.134	6.202	6.265
12	5.511	5.615	5.710	5.798	5.878	5.953	6.023	6.089	6.151
13	5.431	5.533	5.625	5.711	5.789	5.862	5.931	5.995	6.055
14	5.364	5.463	5.554	5.637	5.714	5.786	5.852	5.915	5.974
15	5.306	5.404	5.493	5.574	5.649	5.720	5.785	5.846	5.904
16	5.256	5.352	5.439	5.520	5.593	5.662	5.727	5.786	5.843
17	5.212	5.307	5.392	5.471	5.544	5.612	5.675	5.734	5.790
18	5.174	5.267	5.352	5.429	5.501	5.568	5.630	5.688	5.743
19	5.140	5.231	5.315	5.391	5.462	5.528	5.589	5.647	5.701
20	5.108	5.199	5.282	5.357	5.427	5.493	5.553	5.610	5.663
24	5.012	5.099	5.179	5.251	5.319	5.381	5.439	5.494	5.545
30	4.917	5.001	5.077	5.147	5.211	5.271	5.327	5.379	5.429
40	4.824	4.904	4.977	5.044	5.106	5.163	5.216	5.266	5.313
60	4.732	4.808	4.878	4.942	5.001	5.056	5.107	5.154	5.199
120	4.641	4.714	4.781	4.842	4.898	4.950	4.998	5.044	5.086
∞	4.552	4.622	4.685	4.743	4.796	4.845	4.891	4.934	4.974

(Continúa)

(Continuación)

Percentiles 97.5

$\begin{matrix} J \\ \backslash \\ v \end{matrix}$	2	3	4	5	6	7	8	9	10
1	35.99	54.00	65.69	74.22	80.87	86.29	90.85	94.77	98.20
2	8.776	11.94	14.01	15.54	16.75	17.74	18.58	19.31	19.95
3	5.907	7.661	8.808	9.660	10.34	10.89	11.37	11.78	12.14
4	4.943	6.244	7.088	7.716	8.213	8.625	8.976	9.279	9.548
5	4.474	5.558	6.257	6.775	7.186	7.527	7.816	8.068	8.291
6	4.199	5.158	5.772	6.226	6.586	6.884	7.138	7.359	7.554
7	4.018	4.897	5.455	5.868	6.194	6.464	6.695	6.895	7.072
8	3.892	4.714	5.233	5.616	5.919	6.169	6.382	6.568	6.732
9	3.797	4.578	5.069	5.430	5.715	5.950	6.151	6.325	6.479
10	3.725	4.474	4.943	5.287	5.558	5.782	5.972	6.138	6.285
11	3.667	4.391	4.843	5.173	5.433	5.648	5.831	5.989	6.130
12	3.620	4.325	4.762	5.081	5.332	5.540	5.716	5.869	6.004
13	3.582	4.269	4.694	5.004	5.248	5.449	5.620	5.769	5.900
14	3.550	4.222	4.638	4.940	5.178	5.374	5.540	5.684	5.811
15	3.522	4.182	4.589	4.885	5.118	5.309	5.471	5.612	5.737
16	3.498	4.148	4.548	4.838	5.066	5.253	5.412	5.550	5.672
17	3.477	4.118	4.512	4.797	5.020	5.204	5.361	5.496	5.615
18	3.458	4.092	4.480	4.761	4.981	5.162	5.315	5.448	5.565
19	3.442	4.068	4.451	4.728	4.945	5.123	5.275	5.405	5.521
20	3.427	4.047	4.426	4.700	4.914	5.089	5.238	5.368	5.481
24	3.381	3.983	4.347	4.610	4.816	4.984	5.126	5.250	5.358
30	3.337	3.919	4.271	4.523	4.720	4.881	5.017	5.134	5.238
40	3.294	3.858	4.197	4.439	4.627	4.780	4.910	5.022	5.120
60	3.251	3.798	4.124	4.356	4.536	4.682	4.806	4.912	5.006
120	3.210	3.739	4.053	4.276	4.447	4.587	4.704	4.805	4.894
∞	3.170	3.682	3.984	4.197	4.361	4.494	4.605	4.700	4.784

$\begin{matrix} J \\ \backslash \\ v \end{matrix}$	11	12	13	14	15	16	17	18	19
1	101.3	104.0	106.5	108.8	110.8	112.7	114.5	116.2	117.7
2	20.52	21.03	21.49	21.91	22.30	22.67	23.01	23.32	23.62
3	12.46	12.75	13.01	13.26	13.48	13.69	13.88	14.06	14.23
4	9.788	10.01	10.20	10.39	10.55	10.71	10.85	10.99	11.11
5	8.490	8.670	8.834	8.984	9.124	9.253	9.374	9.486	9.593
6	7.729	7.887	8.031	8.163	8.286	8.399	8.506	8.605	8.698
7	7.230	7.373	7.504	7.624	7.735	7.839	7.935	8.025	8.111
8	6.879	7.011	7.132	7.244	7.347	7.443	7.532	7.616	7.695
9	6.617	6.742	6.856	6.961	7.058	7.148	7.232	7.311	7.385
10	6.416	6.534	6.643	6.742	6.834	6.920	7.000	7.075	7.146
11	6.256	6.369	6.473	6.568	6.657	6.739	6.815	6.887	6.955
12	6.125	6.235	6.335	6.427	6.512	6.591	6.665	6.734	6.799
13	6.017	6.123	6.220	6.309	6.392	6.468	6.539	6.607	6.670
14	5.926	6.029	6.123	6.210	6.290	6.364	6.434	6.499	6.560
15	5.848	5.949	6.041	6.125	6.203	6.276	6.344	6.407	6.467
16	5.781	5.879	5.969	6.052	6.128	6.199	6.265	6.328	6.386
17	5.722	5.818	5.907	5.987	6.062	6.132	6.197	6.258	6.315
18	5.670	5.765	5.852	5.931	6.004	6.073	6.137	6.197	6.253
19	5.624	5.718	5.803	5.881	5.954	6.020	6.083	6.142	6.198
20	5.583	5.675	5.759	5.836	5.907	5.974	6.036	6.093	6.148
24	5.455	5.543	5.623	5.697	5.764	5.827	5.886	5.941	5.994
30	5.330	5.414	5.490	5.560	5.624	5.684	5.740	5.792	5.841
40	5.208	5.288	5.360	5.426	5.487	5.544	5.597	5.646	5.693
60	5.089	5.164	5.232	5.295	5.352	5.406	5.456	5.503	5.546
120	4.972	5.043	5.107	5.166	5.221	5.271	5.318	5.362	5.403
∞	4.858	4.925	4.985	5.041	5.092	5.139	5.183	5.224	5.262

(Continúa)

(Continuación)

Percentiles 99

$\begin{matrix} J \\ v \end{matrix}$	2	3	4	5	6	7	8	9	10
1	90.03	135.0	164.3	185.6	202.2	215.8	227.2	237.0	245.6
2	14.04	19.02	22.29	24.72	26.63	28.20	29.53	30.68	31.69
3	8.261	10.62	12.17	13.33	14.24	15.00	15.64	16.20	16.69
4	6.512	8.120	9.173	9.958	10.58	11.10	11.55	11.93	12.27
5	5.702	6.976	7.804	8.421	8.913	9.321	9.669	9.972	10.24
6	5.243	6.331	7.033	7.556	7.973	8.318	8.613	8.869	9.097
7	4.949	5.919	6.543	7.005	7.373	7.679	7.939	8.166	8.368
8	4.746	5.635	6.204	6.625	6.960	7.237	7.474	7.681	7.863
9	4.596	5.428	5.957	6.348	6.658	6.915	7.134	7.325	7.495
10	4.482	5.270	5.769	6.136	6.428	6.669	6.875	7.055	7.213
11	4.392	5.146	5.621	5.970	6.247	6.476	6.672	6.842	6.992
12	4.320	5.046	5.502	5.836	6.101	6.321	6.507	6.670	6.814
13	4.260	4.964	5.404	5.727	5.981	6.192	6.372	6.528	6.667
14	4.210	4.895	5.322	5.634	5.881	6.085	6.258	6.409	6.543
15	4.168	4.836	5.252	5.556	5.796	5.994	6.162	6.309	6.439
16	4.131	4.786	5.192	5.489	5.722	5.915	6.079	6.222	6.349
17	4.099	4.742	5.140	5.430	5.659	5.847	6.007	6.147	6.270
18	4.071	4.703	5.094	5.379	5.603	5.788	5.944	6.081	6.201
19	4.046	4.670	5.054	5.334	5.554	5.735	5.889	6.022	6.141
20	4.024	4.639	5.018	5.294	5.510	5.688	5.839	5.970	6.087
24	3.956	4.546	4.907	5.168	5.374	5.542	5.685	5.809	5.919
30	3.889	4.455	4.799	5.048	5.242	5.401	5.536	5.653	5.756
40	3.825	4.367	4.696	4.931	5.114	5.265	5.392	5.502	5.599
60	3.762	4.282	4.595	4.818	4.991	5.133	5.253	5.356	5.447
120	3.702	4.200	4.497	4.709	4.872	5.005	5.118	5.214	5.299
∞	3.643	4.120	4.403	4.603	4.757	4.882	4.987	5.078	5.157

$\begin{matrix} J \\ v \end{matrix}$	11	12	13	14	15	16	17	18	19
1	253.2	260.0	266.2	271.8	277.0	281.8	286.3	290.4	294.3
2	32.59	33.40	34.13	34.81	35.43	36.00	36.53	37.03	37.50
3	17.13	17.53	17.89	18.22	18.52	18.81	19.07	19.32	19.55
4	12.57	12.84	13.09	13.32	13.53	13.73	13.91	14.08	14.24
5	10.48	10.70	10.89	11.08	11.24	11.40	11.55	11.68	11.81
6	9.301	9.485	9.653	9.808	9.951	10.08	10.21	10.32	10.43
7	8.548	8.711	8.860	8.997	9.124	9.242	9.353	9.456	9.554
8	8.027	8.176	8.312	8.436	8.552	8.659	8.760	8.854	8.943
9	7.647	7.784	7.910	8.025	8.132	8.232	8.325	8.412	8.495
10	7.356	7.485	7.603	7.712	7.812	7.906	7.993	8.076	8.153
11	7.128	7.250	7.362	7.465	7.560	7.649	7.732	7.809	7.883
12	6.943	7.060	7.167	7.265	7.356	7.441	7.520	7.594	7.665
13	6.791	6.903	7.006	7.101	7.188	7.269	7.345	7.417	7.485
14	6.664	6.772	6.871	6.962	7.047	7.126	7.199	7.268	7.333
15	6.555	6.660	6.757	6.845	6.927	7.003	7.074	7.142	7.204
16	6.462	6.564	6.658	6.744	6.823	6.898	6.967	7.032	7.093
17	6.381	6.480	6.572	6.656	6.734	6.806	6.873	6.937	6.997
18	6.310	6.407	6.497	6.579	6.655	6.725	6.792	6.854	6.912
19	6.247	6.342	6.430	6.510	6.585	6.654	6.719	6.780	6.837
20	6.191	6.285	6.371	6.450	6.523	6.591	6.654	6.714	6.771
24	6.017	6.106	6.186	6.261	6.330	6.394	6.453	6.510	6.563
30	5.849	5.932	6.008	6.078	6.143	6.203	6.259	6.311	6.361
40	5.686	5.764	5.835	5.900	5.961	6.017	6.069	6.119	6.165
60	5.528	5.601	5.667	5.728	5.785	5.837	5.886	5.931	5.974
120	5.375	5.443	5.505	5.562	5.614	5.662	5.708	5.750	5.790
∞	5.227	5.290	5.348	5.400	5.448	5.493	5.535	5.574	5.611

(Continúa)

(Continuación)

Percentiles 99.5

$\begin{matrix} J \\ v \end{matrix}$	2	3	4	5	6	7	8	9	10
1	180.1	270.1	328.5	371.2	404.4	431.6	454.4	474.0	491.1
2	19.93	26.97	31.60	35.02	37.73	39.95	41.83	43.46	44.89
3	10.55	13.50	15.45	16.91	18.06	19.01	19.83	20.53	21.15
4	7.916	9.814	11.06	11.99	12.74	13.35	13.88	14.33	14.74
5	6.751	8.196	9.141	9.847	10.41	10.88	11.28	11.63	11.93
6	6.105	7.306	8.088	8.670	9.135	9.522	9.852	10.14	10.40
7	5.699	6.750	7.429	7.935	8.339	8.674	8.961	9.211	9.433
8	5.420	6.370	6.981	7.435	7.797	8.097	8.354	8.578	8.777
9	5.218	6.096	6.657	7.074	7.405	7.680	7.915	8.120	8.303
10	5.065	5.888	6.412	6.800	7.109	7.365	7.584	7.775	7.944
11	4.945	5.727	6.222	6.588	6.878	7.119	7.325	7.505	7.664
12	4.849	5.597	6.068	6.416	6.693	6.922	7.118	7.288	7.439
13	4.770	5.490	5.943	6.277	6.541	6.760	6.947	7.111	7.255
14	4.704	5.401	5.838	6.160	6.414	6.626	6.805	6.962	7.101
15	4.647	5.325	5.750	6.061	6.308	6.511	6.685	6.837	6.971
16	4.599	5.261	5.674	5.977	6.216	6.413	6.582	6.729	6.859
17	4.557	5.205	5.608	5.903	6.136	6.329	6.493	6.636	6.763
18	4.521	5.156	5.550	5.839	6.067	6.255	6.415	6.554	6.678
19	4.488	5.113	5.500	5.783	6.005	6.189	6.346	6.482	6.603
20	4.460	5.074	5.455	5.732	5.951	6.131	6.285	6.418	6.537
24	4.371	4.955	5.315	5.577	5.783	5.952	6.096	6.221	6.332
30	4.285	4.841	5.181	5.428	5.621	5.780	5.914	6.031	6.135
40	4.202	4.731	5.053	5.284	5.465	5.614	5.739	5.848	5.944
60	4.122	4.625	4.928	5.146	5.316	5.454	5.571	5.673	5.762
120	4.045	4.523	4.809	5.013	5.172	5.301	5.410	5.504	5.586
∞	3.970	4.424	4.694	4.886	5.033	5.154	5.255	5.341	5.418

$\begin{matrix} J \\ v \end{matrix}$	11	12	13	14	15	16	17	18	19
1	506.3	520.0	532.4	543.6	554.0	563.6	572.5	580.9	588.7
2	46.16	47.31	48.35	49.30	50.17	50.99	51.74	52.45	53.12
3	21.70	22.20	22.66	23.08	23.46	23.82	24.15	24.46	24.76
4	15.10	15.42	15.72	15.99	16.24	16.48	16.70	16.90	17.09
5	12.21	12.46	12.69	12.90	13.09	13.27	13.44	13.60	13.75
6	10.63	10.83	11.02	11.20	11.36	11.51	11.65	11.78	11.90
7	9.632	9.812	9.977	10.13	10.27	10.40	10.52	10.64	10.75
8	8.955	9.117	9.265	9.401	9.527	9.644	9.754	9.857	9.953
9	8.466	8.614	8.749	8.874	8.990	9.097	9.198	9.292	9.381
10	8.096	8.234	8.360	8.476	8.583	8.683	8.777	8.865	8.947
11	7.807	7.937	8.055	8.164	8.265	8.359	8.447	8.530	8.608
12	7.575	7.697	7.810	7.914	8.009	8.099	8.183	8.261	8.335
13	7.384	7.502	7.609	7.708	7.800	7.886	7.965	8.040	8.111
14	7.225	7.338	7.442	7.537	7.625	7.707	7.784	7.856	7.924
15	7.091	7.200	7.300	7.392	7.477	7.556	7.630	7.699	7.765
16	6.976	7.081	7.178	7.267	7.349	7.426	7.498	7.566	7.629
17	6.876	6.979	7.072	7.159	7.239	7.314	7.384	7.449	7.511
18	6.788	6.888	6.980	7.064	7.142	7.215	7.283	7.347	7.407
19	6.711	6.809	6.898	6.981	7.057	7.128	7.195	7.257	7.316
20	6.642	6.738	6.826	6.907	6.981	7.051	7.116	7.177	7.235
24	6.431	6.520	6.602	6.677	6.747	6.812	6.872	6.930	6.983
30	6.227	6.310	6.387	6.456	6.521	6.581	6.638	6.691	6.741
40	6.030	6.108	6.179	6.244	6.304	6.360	6.412	6.461	6.507
60	5.841	5.913	5.979	6.039	6.094	6.146	6.194	6.239	6.281
120	5.660	5.726	5.786	5.842	5.893	5.940	5.984	6.025	6.064
∞	5.485	5.546	5.602	5.652	5.699	5.742	5.783	5.820	5.856

TABLA 10. Probabilidades asociadas con valores tan grandes como valores observados de H en el análisis de varianza de una clasificación por rangos de Kruskal-Wallis*

Tamaño de muestras			H	p	Tamaño de muestras			H	p
n_1	n_2	n_3			n_1	n_2	n_3		
2	1	1	2.7000	.500	4	3	2	6.4444 6.3000 5.4444 5.4000 4.5111 4.4444	.008 .011 .046 .051 .098 .102
2	2	1	3.6000	.200					
2	2	2	4.5714 3.7143	.067 .200					
3	1	1	3.2000	.300					
3	2	1	4.2857 3.8571	.100 .133	4	3	3	6.7455 6.7091 5.7909 5.7273 4.7091 4.7000	.010 .013 .046 .050 .092 .101
3	2	2	5.3572 4.7143 4.5000 4.4643	.029 .048 .067 .105					
3	3	1	5.1429 4.5714 4.0000	.043 .100 .129	4	4	1	6.6667 6.1667 4.9667 4.8667 4.1667	.010 .022 .048 .054 .082
3	3	2	6.2500 5.3611 5.1389 4.5556 4.2500	.011 .032 .061 .100 .121					
3	3	3	7.2000 6.4889 5.6889 5.6000 5.0667 4.6222	.004 .011 .029 .050 .086 .100	4	4	2	7.0364 6.8727 5.4545 5.2364 4.5545 4.4455	.006 .011 .046 .052 .098 .103
4	1	1	3.5714	.200	4	4	3	7.1439 7.1364 5.5985 5.5758 4.5455 4.4773	.010 .011 .049 .051 .099 .102
4	2	1	4.8214 4.5000 4.0179	.057 .076 .114	4	4	4	7.6538 7.5385 5.6923 5.6538 4.6539 4.5001	.008 .011 .049 .054 .097 .104
4	2	2	6.0000 5.3333 5.1250 4.4583 4.1667	.014 .033 .052 .100 .105	5	1	1	3.8571	.143
4	3	1	5.8333 5.2083 5.0000 4.0556 3.8889	.021 .050 .057 .093 .129	5	2	1	5.2500 5.0000 4.4500 4.2000 4.0500	.036 .048 .071 .095 .119

* Tomada de SIEGEL, S.: O. c., pp. 316-17.

(Continúa)

(Continuación)

Tamaño de muestras			H	p	Tamaño de muestras			H	p
n ₁	n ₂	n ₃			n ₁	n ₂	n ₃		
5	2	2	6.5333	.008				4.5487	.099
			6.1333	.013				4.5231	.103
			5.1600	.034					
			5.0400	.056	5	4	4	7.7604	.009
			4.3733	.090				7.7440	.011
			4.2933	.122				5.6571	.049
								5.6176	.050
5	3	1	6.4000	.012				4.6187	.100
			4.9600	.048				4.5527	.102
			4.8711	.052					
			4.0178	.095	5	5	1	7.3091	.009
			3.8400	.123				6.8364	.011
								5.1273	.046
5	3	2	6.9091	.009				4.9091	.053
			6.8218	.010				4.1091	.086
			5.2509	.049				4.0364	.105
			5.1055	.052					
			4.6509	.091	5	5	2	7.3385	.010
			4.4945	.101				7.2692	.010
								5.3385	.047
5	3	3	7.0788	.009				5.2462	.051
			6.9818	.011				4.6231	.097
			5.6485	.049				4.5077	.100
			5.5152	.051					
			4.5333	.097	5	5	3	7.5780	.010
			4.4121	.109				7.5429	.010
								5.7055	.046
5	4	1	6.9545	.008				5.6264	.051
			6.8400	.011				4.5451	.100
			4.9855	.044				4.5363	.102
			4.8600	.056					
			3.9873	.098	5	5	4	7.8229	.010
			3.9600	.102				7.7914	.010
								5.6657	.049
5	4	2	7.2045	.009				5.6429	.050
			7.1182	.010				4.5529	.099
			5.2727	.049				4.5200	.101
			5.2682	.050					
			4.5409	.098	5	5	5	8.0000	.009
			4.5182	.101				7.9800	.010
5	4	3	7.4449	.010				5.7800	.049
			7.3949	.011				5.6600	.051
			5.6564	.049				4.5600	.100
			5.6308	.050				4.5000	.102

* Versión abreviada de KRUSKAL, W. H., y WALLIS, W. A. 1952. Uso de los rangos en el análisis de varianza de un criterio. *J. Amer. Statist. Ass.*, 47, 614-617. Con el amable permiso de los autores y editores. (Las correcciones de esta tabla dadas por los autores en Errata. *J. Amer. Statist., Ass.*, 48, 910, se han incorporado.)

Se trata de un libro centrado en los métodos de investigación en educación, en cuya primera unidad didáctica se revisan los tipos de investigación y los cuatro métodos fundamentales como son: el método descriptivo para las investigaciones tipo encuesta; el método experimental y cuasi-experimental, para los diseños de grupo y para las pruebas de hipótesis; los métodos cualitativos para la comprensión de los problemas en sus entornos naturales, y los métodos evaluativos con el predominio de la complementariedad de métodos y su aplicación en los modelos de calidad. En la segunda unidad didáctica, dedicada a la medida y a las técnicas de recogida de información, se desarrollan temas sobre medida, la construcción de escalas, las características técnicas de los instrumentos de recogida de información, y los procedimientos de recogida de información, tanto desde las perspectivas de los enfoques de investigación positivista, como son la observación sistemática, el cuestionario o las pruebas de rendimiento, como desde la perspectiva cualitativa: la observación participante, la técnica delphi o los grupos de discusión. En la tercera y última parte de la obra se presentan los contenidos dedicados al análisis de la información, desde la estadística descriptiva y la inferencial. Es una obra amplia en cuanto a los contenidos que abarca, aunque trata de exponerlos de un modo sencillo y con ejemplos prácticos. Está orientada a los alumnos del título de Educación Social, y para todos aquellos interesados en los temas de investigación en Educación y en Ciencias Sociales.

