Benjamín Garza Olvera

Estadística y probabilidad

ALWAYS LEARNING

PEARSON

Estadística y probabilidad

Primera edición

BENJAMÍN GARZA OLVERA

Revisión técnica

Físico. Guillermo Govea Anaya
por la Universidad Nacional Autónoma de México

Ing. Ind. Teresita Betancourt Pérez
Centro de Enseñanza Técnica Industrial campus Olmos

PEARSON

Datos de catalogación

Autor: Garza, Benjamín.

Estadística y probabilidad
Primera edición

Pearson Educación de México, S. A. de C. V., 2014
ISBN: 978-607-32-2783-4

Área: Bachillerato/Matemáticas

Formato: 21 x 27 cm Páginas: 296

Estadística y probabilidad

1a. edición

Libro del estudiante

El proyecto didáctico Estadística y probabilidad, es una obra colectiva creada por encargo de la editorial Pearson Educación de México, S.A. de C.V, por un equipo de profesionales en distintas áreas, que trabajaron siguiendo los lineamientos y estructuras establecidos por el departamento pedagógico de Pearson Educación de México, S.A. de C.V.

■ **Dirección general:** Phillip De la Vega ■ **Dirección K-12:** Santiago Gutiérrez ■ **Gerencia editorial K-12:** Jorge Luis Íñiguez ■ **Coordinación de bachillerato y custom:** Lilia Moreno ■ **Edición sponsor:** Berenice Torruco ■ **Coordinación de arte y diseño:** Mónica Galván ■ **Supervisión de arte y diseño:** José Hernández ■ **Edición de desarrollo:** José Huerta ■ **Asistencia editorial:** Mirna González ■ **Corrección de estilo:** Mireille Bravo ■ **Lectura de pruebas:** Carlos del Razo ■ **Diseño de portada:** Pulso Comunicación ■ **Diagramación:** Ediciones OVA.

ISBN LIBRO IMPRESO: 978-607-32-2783-4
ISBN E-BOOK: 978-607-32-2784-1
ISBN E-CHAPTER: 978-607-32-2785-8

Atlacomulco 500, 5° piso
Col. Industrial Atoto, C.P. 53519
Naucalpan de Juárez, Edo. de México
Cámara Nacional de la Industria Editorial Mexicana Reg. Núm. 1031

Impreso en México. Printed in Mexico.
1 2 3 4 5 6 7 8 9 0 – 17 16 15 14

PEARSON

www.pearsonenespañol.com

Contenido

Presentación

Este libro se escribió pensando en estudiantes que cursan el sexto semestre de bachillerato tecnológico. El estudio de la estadística y probabilidad, hoy en día, es indispensable en todos los campos de estudio, como: sociología, economía, política, educación, negocios, medicina, industria, informática, ingeniería, investigación, etcétera. El objetivo de este texto es proporcionar el conocimiento, comprensión y dominio en los aspectos fundamentales y básicos de la estadística y probabilidad. El objetivo principal, fue escribir un libro que ustedes los estudiantes pudieran leer, entender y disfrutar. A lo largo del libro se utiliza un lenguaje claro y preciso que propicia la generación de conocimientos que, por lo general, resultan difíciles de entender y aprender. Se utilizan oraciones cortas, explicaciones claras y muchos ejemplos resueltos a detalle.

La didáctica que se desarrolla en el texto se fundamenta en una exposición de conceptos introductorios y ejemplos demostrativos, así como, una diversificación en el planteamiento del problema. Donde las técnicas empleadas en la solución tienen por objeto desarrollar el razonamiento reflexivo y la destreza en el estudiante. Los problemas y ejercicios que se desarrollan a lo largo de las dos unidades que presenta el libro, utilizan distintos tipos de reactivos, lo cual permite tener una evaluación continua del proceso enseñanza-aprendizaje. Se hace énfasis en el incremento gradual de la complejidad de cada ejercicio hasta lograr el cambio de la memorización por un razonamiento más analítico en el planteamiento y desarrollo del proceso de solución de un problema determinado, en estadística se hace uso de representaciones gráficas de los resultados, las cuales, apoyan a la obtención de un análisis orientado a la selección de criterios y evaluación de estrategias para la toma de decisión; todo ello fortaleciendo la garantía de una enseñanza permanente.

Es importante recalcar que la consecución de los temas tratados en el libro dependerá del trabajo estrecho entre docente y alumnos, de manera tal que el libro sea sólo un apoyo para reforzar lo que se trabaja en el salón de clases.

Agradezco el apoyo de mis compañeros de la Academia Local, Estatal y Nacional, quienes, confiaron en mi esfuerzo y dedicación para lograr contenidos de calidad; también al editor por su esmerada atención a la realización de esta obra.

Por último, a mis ex alumnos y, en especial, a mi familia a quienes distingo con este mensaje filosófico: "El éxito es un volcán en cuyo cráter no crece la hierba del fracaso; un poco de confianza y un poco de suerte son necesarias en la vida ya que la grandeza de un ideal no es alcanzarlo, sino luchar por él. Alcanzarlo es sólo una recompensa para continuar..."

¡Ten fe y triunfarás!

EL AUTOR
Q. I. y Lic. Benjamín Garza Olvera

Metodología para el trabajo con este material

El material está dividido en dos unidades, cada una con dos temas principales, donde se desarrollan los contenidos actuales del programa general de bachillerato tecnológico. Cada unidad cuenta con una evaluación diagnóstica, el desarrollo de los diversos temas y una autoevaluación.

Evaluación diagnóstica

Es una serie de ejercicios que sirven como repaso operativo, pero en general se busca desarrollar habilidades de lógica, aritmética y álgebra.

Cuadros de competencias genéricas y disciplinares

Se localiza en cada una de las actividades que favorecen el logro de competencias; en este apartado el alumno, con la mediación del maestro, deberá determinar cuáles son las competencias genéricas y las competencias disciplinares que está desarrollando y escribir en el cuadro las que sean pertinentes.

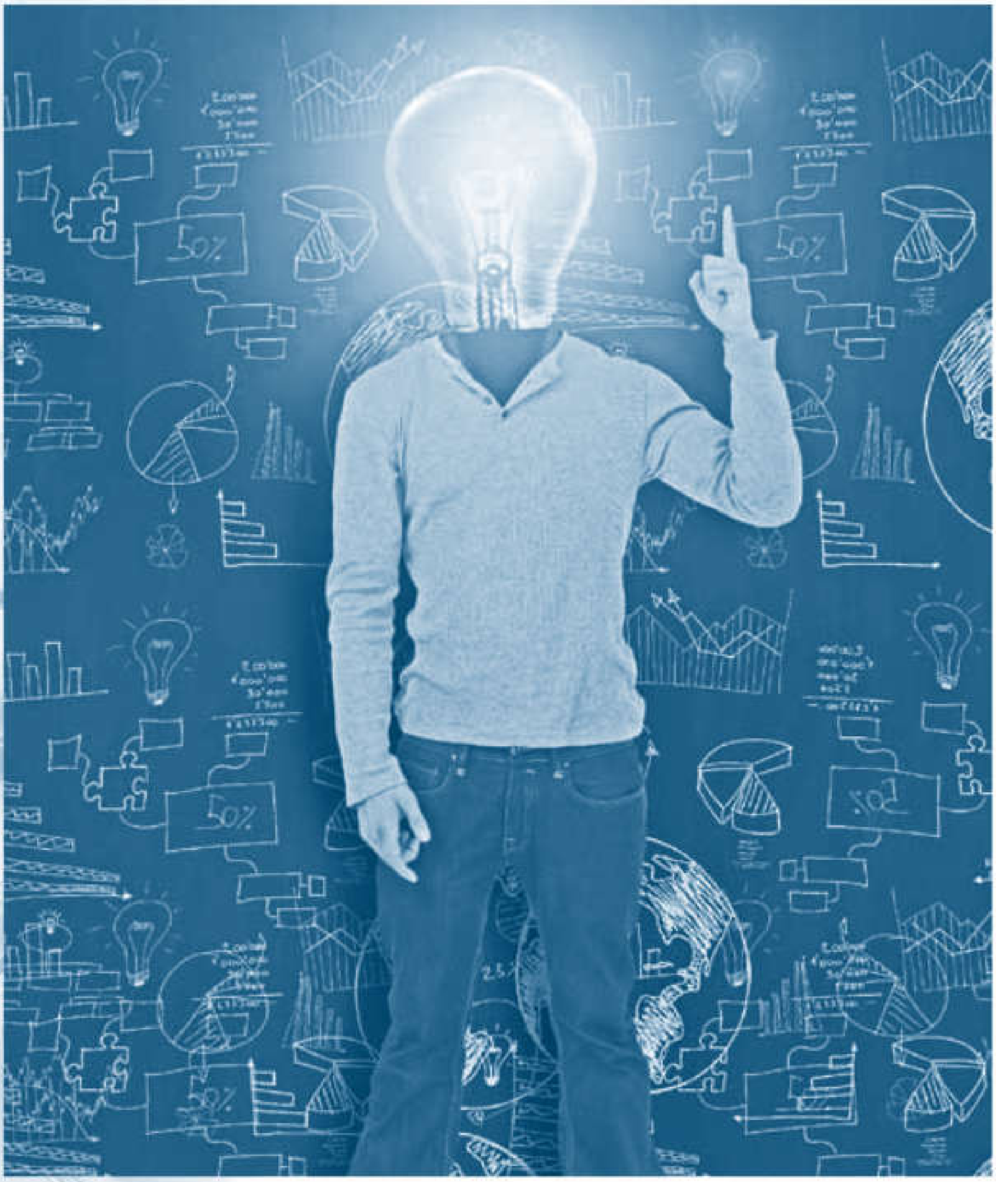
Autoevaluación

Es una colección de ejercicios que ayudan a reforzar el trabajo desarrollado a lo largo de la unidad.

Competencias genéricas

Categorías	Competencias
Se autodetermina y cuida de sí	1. Se conoce y valora a sí mismo y aborda problemas y retos teniendo en cuenta los objetivos que persigue. 2. Es sensible al arte y participa en la apreciación e interpretación de sus expresiones en distintos géneros. 3. Elige y practica estilos de vida saludables.
Se expresa y se comunica	4. Escucha, interpreta y emite mensajes pertinentes en distintos contextos mediante la utilización de medios, códigos y herramientas apropiados.
Piensa crítica y reflexivamente	5. Desarrolla innovaciones y propone soluciones a problemas a partir de métodos establecidos. 6. Sustenta una postura personal sobre temas de interés y relevancia general, considerando otros puntos de vista de manera crítica y reflexiva.
Aprende de forma autónoma	7. Aprende por iniciativa e interés propio a lo largo de la vida.
Trabaja en forma colaborativa	8. Participa y colabora de manera efectiva en diversos equipos.
Participa con responsabilidad en la sociedad	9. Participa con una conciencia cívica y ética en la vida de su comunidad, región, México y el mundo. 10. Mantiene una actitud respetuosa hacia la interculturalidad y la diversidad de creencias, valores, ideas y prácticas sociales. 11. Contribuye al desarrollo sustentable de manera crítica con acciones responsables.

UNIDAD 1



Estadística

Evaluación diagnóstica

Realiza lo que se indica a continuación.

1. Indica que variables son cualitativas y cuales cuantitativas:

   a) Comida Favorita.

   b) Profesión que te gusta.

   c) Número de penales anotados por ti en la última temporada.

   d) Número de sillas en tu escuela.

   e) El color de los autos del estacionamiento de tu escuela.

2. Clasificar las siguientes variables en cualitativas y cuantitativas discretas o continuas.

   a) La nacionalidad de una persona.

   b) Número de litros de agua de la reserva de tu casa.

   c) Número de medicamentos en un estante de farmacia.

   d) La profesión de una persona.

   e) El área de un edificio.

3. A un conjunto de 5 números cuya media es 7.31 se le añaden los números 4.47 y 10.15. ¿Cuál es la media del nuevo conjunto de números?

4. Un pediatra obtuvo la siguiente tabla sobre los meses de edad de 50 niños de su consulta en el momento de andar por primera vez. Calcular la moda, la mediana, la media y la varianza.

Meses	Niños
9	1
10	4
11	9
12	16
13	11
14	8
15	1

TEMA 1 Manejo de la información

Propósito del tema

Que el estudiante:

- Analice fenómenos sociales o naturales, utilizando las herramientas básicas de la estadística descriptiva para muestrear, procesar y comunicar información social y científica, para la toma de decisiones.
- Exprese razones para fundamentar una respuesta y obtenga conclusiones pertinentes a partir de datos estadísticos.
- Identifique e interprete los datos datos a partir de una representación gráfica de datos.

Competencias disciplinares

1. Construye e interpreta modelos matemáticos mediante la aplicación de procedimientos aritméticos, algebraicos, geométricos y variacionales, para la comprensión y análisis de situaciones reales, hipotéticas o formales.
3. Explica e interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos o situaciones reales.
5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.
8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

Contenidos que aborda el tema

Contenidos conceptuales

- Definición de los elementos básicos de la estadística.
- Concepto de frecuencia y distribución de frecuencias.
- Representación gráfica e interpretación de datos estadísticos.

Contenidos procedimentales

- Usará la terminología para datos esatdísticos.
- Interpretará y argumentará los resultados de una gráfica de barras.
- Resolverá problemas de aplicación a partir de datos estadísticos.
- Usará diversas estrategias de resolución de problemas.

Contenidos actitudinales

- Expresará sus ideas mediante los elementos básicos de la estadística.
- Trabajará en equipo y respetará a sus compañeros al resolver problemas.
- Aprenderá a valorar el trabajo de sus compañeros al resolver problemas.
- Estructurará ideas y argumentos de manera clara, coherente y sintética.

Desarrollo histórico de la estadística

Los primeros conocimientos estadísticos surgen en la Antigüedad cuando Moisés contó el número de soldados israelitas menores de veinte años. Tiempo después, los egipcios emplearon la estadística para calcular los impuestos recaudables de las cosechas; sin embargo, en aquella época jamás se llegó a realizar un análisis estadístico.

En el siglo X, el gobierno de Guillermo I de Inglaterra ordenó elaborar un censo de población y propiedades, es decir, de los recursos económicos y de las tierras que las personas poseían. Mientras que en el siglo XVI, el inglés John Graunt utilizó la estadística para referir las defunciones producidas por la peste en Londres y España. A principios del siglo XVII, los jugadores profesionales del azar solicitaron a los matemáticos el desarrollo de principios que les permitieran aumentar las posibilidades de ganar con los naipes y los dados.

El primer y principal estudio acerca de la probabilidad fue realizado por Bernovilli y De Moivre, siendo este último quien desarrolló la ecuación de la curva de distribución normal en 1730. Se considera que el fundador de la estadística fue Gottfried Achenwall, quien en 1748 realizó estudios de población, que después fueron complementados con la teoría de las probabilidades y que conjuntamente se aplicaron para actualizar aspectos sociales tales como la natalidad, criminalidad, mortalidad, educación, enfermedades, etcétera.

Durante el siglo XVIII, la estadística se aplicó en acciones políticas y gubernamentales, asociadas con las necesidades de los habitantes como hospitales, escuelas, plazas cívicas, servicios de agua, electricidad, transporte y vías de comunicación. En 1829, el estadista belga Adolfo Quetelet fue el primero en aplicar los métodos estadísticos en la investigación de problemas educativos y sociales. Contribuyó en la elaboración de los primeros censos oficiales europeos, así como en el desarrollo de la igualdad y similitud de datos estadísticos entre naciones; razón por la cual, se considera el padre de la estadística moderna.

Por la misma época, destacó Florence Nightingale, quien sostuvo que todo hombre de negocios debía guiarse por el conocimiento estadístico para triunfar, y que los políticos fracasaban casi siempre por su incapacidad para aplicar los métodos estadísticos. También sobresalió Francis Galton, quien introdujo y empleó la estadística en estudios de herencia, eugenesia (aplicación de las leyes biológicas al perfeccionamiento de la especie humana), psicología y antropología criminal.

A fines del siglo XVIII el francés Pierre Laplace y, a principios del siglo XIX, el alemán Karl Friedrich Gauss, realizaron trabajos sobre el cálculo de probabilidades al evaluar los errores de observación en astronomía, de lo que resultó una curva de distribución en forma de campana en donde la simetría se denomina **distribución de Gauss para errores**; ésta también se aplica en el estudio y análisis de los errores en las medidas que presentan los productos.

En el siglo XX, William S. Gosset contribuyó en el desarrollo de la estadística con su famosa teoría de las pequeñas muestras, deduciendo la distribución *t*. Por otra parte, el inglés Ronald A. Fisher trabajó en estudios sobre agricultura y biología de productos del campo por medio de métodos estadísticos, logrando hallar la distribución *F*.

La sistematización y el perfeccionamiento actual de la estadística permite intervenir en todos los campos y actividades del ser humano, como instrumento indispensable para la toma de decisiones que permitan estructurar la sociedad, que cada vez es más compleja.

EJERCICIO 1

I. Realiza lo que se te indica en cada caso.

1. ¿Cómo surgió la estadística?
2. ¿Cuál es el nombre del fundador de la estadística?
3. ¿Por qué se considera a Quetelet el padre de la estadística moderna?
4. ¿Cómo surgió la distribución Gaussiana?

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

5. Explica brevemente la contribución de los siguientes estadistas.

   a) John Graunt
   b) De Moivre
   c) Gottfried Achenwall
   d) Florence Nightingale
   e) Laplace y Gauss
   f) Ronald A. Fisher

II. Investiga en equipo, lo que se indica a continuación y discute en plenaria tus resultados.

1. ¿Qué es un estadista?
2. ¿Cuáles son las principales fuentes históricas de la estadística?
3. Explica cómo se abusa de la estadística en el mundo social.

Verifica tus resultados en la sección de soluciones correspondiente.

Definición y aplicaciones de la estadística

Para iniciar con el estudio de los temas de esta unidad, te presentamos algunas definiciones complementarias de la estadística.

a) Es la ciencia o arte de reunir y analizar datos e inferir consecuencias a partir de estos elementos.
b) Es la rama de las matemáticas que aborda los datos numéricos o cuantitativos y los relaciona con el método científico en la toma, organización, recopilación, presentación y análisis de los mismos, con el fin de tomar decisiones razonables.
c) Es la ciencia que se ocupa de la reunión o recopilación de todos los hechos que se pueden valorar mecánicamente para hacer comparaciones entre las cifras y obtener conclusiones, aplicando la teoría de las probabilidades.
d) Es el conjunto de métodos o técnicas que estudia y analiza los datos que son susceptibles de expresión numérica para llegar a conclusiones que permitan tomar decisiones y pronosticar las consecuencias de las mismas.
e) Es el estudio de los fenómenos del azar, que agrupa, clasifica y ordena experiencias y observaciones sobre la manifestación de hechos, para deducir las leyes o los principios que los rigen.

Estas definiciones proporcionan una visión más amplia sobre lo que puede hacerse con la estadística y ayudan a comprender cómo aprovechar al máximo la información de los datos estadísticos.

Aplicaciones de la estadística

Quienes se dedican a la pedagogía, sociología, medicina, finanzas, comercio, etc., recopilan gran número de datos diferentes. Puesto que proceden de instrumentos de medición o de conteos, el conocimiento de la estadística es imprescindible para la interpretación y el análisis de dichos datos, pues al investigador en sus estudios, le permite encontrar aplicaciones más útiles y prácticas.

En la actualidad, se aplica en las ciencias sociales, en las ciencias naturales (físicas, meteorológicas), en la industria (producción y control de calidad), en la administración industrial (recursos humanos, materiales, económicos, de tiempos y movimientos), en la física atómica (fisión y fusión nuclear), en la astronomía, finanzas (bienes y raíces, inversiones, bolsa de valores), en la agricultura (períodos de siembra, calendarios de lluvias), en el comercio (estudios de mercado, análisis de oferta y demandas), en la educación (niveles académicos, presentación de servicios), en la medicina (controles de peso, embarazo, cardiológico), en las ciencias políticas, en la computación, etcétera.

EJERCICIO 2

I. Realiza en equipo las siguientes actividades.

1. Formula, con tus propias palabras, una definición de la estadística.
2. Explica detalladamente la aplicación de la estadística en actividades de tu especialidad.
3. Cita un ejemplo que ilustre la aplicación de la estadística en cada una de las siguientes actividades industriales.
   a) Dirección de personal
   b) Departamento de producción
   c) Departamento de ventas
   d) Consejo administrativo
   e) Control de calidad
   f) Departamento de diseño
4. Escribe cuál crees que es la contribución de la estadística en tu formación como bachiller.

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

Verifica tus resultados en la sección de soluciones correspondiente.

División de la estadística

Para su estudio, la estadística se divide en dos ramas, **la estadística descriptiva** y **la estadística inferencial**. Tanto una como la otra son funciones del análisis estadístico.

Estadística descriptiva

Es la parte de la estadística que también se conoce con el nombre de **estadística deductiva**, ya que trata solamente de describir y analizar un grupo de datos dados, los cuales se representan con tablas, gráficas, cuadros e índices. No realiza inferencias —y por tanto, no obtiene conclusiones— a partir de un grupo mayor de datos.

Ejemplos

Cuando un investigador conduce un estudio, reúne gran cantidad de información numérica o datos acerca del problema en cuestión. Los datos pueden tener formas variadas, por ejemplo: *datos de frecuencia* (recuentos de votantes que prefieren uno u otro candidato político) o *datos escolares* (los cocientes intelectuales de un grupo de estudiantes, los pesos netos de los ingredientes para hacer un pastel). Para realizar la función descriptiva, el estadista formula reglas y procedimientos para la presentación de los datos en una forma más útil y significativa, como los gráficos y los parámetros. Otro ejemplo es cuando el jefe de personal de una empresa aplica una prueba de habilidades y destrezas a un grupo de trabajadores; utilizando estadística descriptiva en las puntuaciones obtenidas, los datos se tienen que ordenar, clasificar, calcular promedios, identificar las cualidades más típicas, construir tablas, gráficas y cuadros para comparaciones, predecir el rendimiento, etcétera.

Estadística inferencial

Es la parte de la estadística que también se llama **estadística inductiva**, ya que trata las condiciones bajo las cuales tales inferencias son válidas; al no poder estar absolutamente seguros de la corrección de tales inferencias, con frecuencia se utiliza el término de **probabilidad** en las conclusiones. También se considera como la técnica con la que se obtienen conclusiones o generalizaciones acerca del parámetro de una **población** que se basa en una **muestra** de dicha población.

Ejemplos

Sabemos que una baraja normal consta de 52 naipes. Cada uno tiene la misma probabilidad de ser seleccionado cuando se hace una extracción de los mismos; entonces, se puede concluir que el evento *obtener un rey* ocurre con probabilidad de $\frac{4}{52}=\frac{1}{13}$, puesto que en un mazo hay 4 reyes; así, a partir del conocimiento de la población se realizó una predicción respecto a una muestra (evento).

Otro ejemplo es cuando se investiga el aprendizaje de matemáticas de los estudiantes del bachillerato tecnológico del país (población); si se elige una representación de alumnos (muestra), las observaciones arrojan un **promedio** o **media aritmética** sobre dicha investigación.

Diferencia entre estadística y probabilidad

La estadística implica un razonamiento a partir de una muestra para llegar al conocimiento de la población, mientras que la probabilidad se fundamenta en el razonamiento de una población hasta la muestra. Por lo anterior, se concluye que la estadística no es más que el trabajo preliminar de la probabilidad.

EJERCICIO 3

I. Realiza lo que se indica.

1. ¿Cuál es la división de la estadística para su estudio?
2. ¿Qué es la estadística descriptiva?
3. Desarrolla un ejemplo de aplicación de la estadística descriptiva.
4. ¿Qué es la estadística inferencial?
5. Desarrolla un ejemplo de aplicación de la estadística inferencial.
6. ¿Cuál es la diferencia entre estadística y probabilidad? ¿Cuál es su relación?

II. Realiza una presentación y discute con el resto del grupo.

1. ¿Por qué la estadística inferencial cumple una función de vital importancia en las empresas científicas?
2. ¿Cómo se aplican la estadística descriptiva y la estadística inferencial en tu vida diaria?

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

Verifica tus resultados en la sección de soluciones correspondiente.

Elementos de la estadística

Entre los elementos básicos de la estadística, se encuentran fundamentalmente la **población** o el **universo**, la **variable** (continua y discreta) y los **tipos de datos**; todos ellos con el único fin de fortalecer la comprensión de los procedimientos estadísticos.

Población o universo

Se define como la fuente de observaciones o medidas que describen detalladamente a un conjunto de individuos u objetos.

Si se considera una investigación sobre la preferencia de los lectores para un determinado periódico, la población consistiría en todas las suscripciones de los lectores; ahora, si se observa el número de automóviles que circulan y que no circulan en el Distrito Federal en un día determinado, entonces la población estaría constituida por el número total de automóviles.

Existen dos clases de población:

Población finita. Es aquella que consta de un número finito de elementos, por ejemplo:

- La fabricación de todos los televisores por una industria en un determinado turno de trabajo.
- El tiraje de periódicos hechos por una empresa en un determinado día.
- El número de teléfonos instalados en la ciudad de León, Guanajuato.

Población infinita. Es aquella que consta de un número infinito de elementos, por ejemplo:

- Aquella que se forma con todos los posibles sucesos (pares e impares) en tiradas sucesivas de un dado.
- La que se constituye como todos los posibles números primos del conjunto números naturales.
- La que se obtiene por todos los posibles sucesos de sacar un naipe con figura o uno con número en extracciones sucesivas de una baraja.

Si la población es abundante, a veces es imposible o inusual observar la totalidad de los elementos que la conforman, por lo que se recomienda analizar una parte representativa de dicho conjunto, la cual se denomina **muestra**.

Muestra

Se define como el conjunto de observaciones que representan la totalidad de las características a examinar de una población. Así, una muestra contiene un subconjunto de elementos de la población a la que pertenece.

Variables

Son cantidades a las que se asigna un número ilimitado de valores. Estadísticamente, las variables se identifican como **continuas** y **discretas**.

Variable continua. Es aquella que puede tomar cualquier valor entre dos valores dados, es decir, puede adquirir valores enteros, decimales o fracciones. Toda variable continua tiene un límite inferior y un límite superior.

Ejemplos

1. El cambio de peso en la mujer durante su embarazo.
2. El tiempo de recorrido de tu casa a la escuela.
3. Los centímetros de precipitaciones pluviales en una región durante el año.
4. Período de duración de las barras luminosas.
5. La cantidad de gasolina que se vende en un mes.

Variable discreta. Es aquella que no cumple con las condiciones de una variable continua, es decir, el valor de una variable discreta se expresa en números enteros.

Ejemplos

1. El número de libros en un estante.
2. El número de niños nacidos en diferentes horas del día.
3. El número de goles anotados por Hugo Sánchez en el Torneo de Copa 1990.
4. El número de estudiantes de una preparatoria.
5. El número de neumáticos en un automóvil.

Por lo anterior, se concluye que los datos que provienen de una variable continua y de una discreta se llaman datos continuos y discretos, respectivamente; así, las **medidas** dan lugar a **datos continuos**, mientras que los **conteos** o **enumeraciones** dan lugar a **datos discretos**.

Recopilación de datos

Los **datos** son situaciones o hechos que se representan numéricamente y que algunas veces forman parte de la vida cotidiana y otras, se encuentran en libros porque han sido recopilados anteriormente por otras personas. Por lo tanto, los tipos de datos pueden ser:

Datos originales. Son aquellos que son recopilados por el propio investigador; por lo tanto, son comprobables en forma rigurosa.

Datos indirectos. Son aquellos que son recopilados de enciclopedias, libros de registro, sucesos grabados en audio y video, etcétera.

Para que la estadística sea exacta y verdadera, la recopilación de datos debe ser cuidadosa y precisa, y usar los medios, recursos y procedimientos que faciliten objetivamente su obtención; por ejemplo:

1. Cuestionarios y entrevistas realizadas por personas competentes y profesionales para dar lugar a datos originales.
2. Consultas en fuentes originales y fieles para dar lugar a datos indirectos.

Manejo de datos

Una vez hecha la recopilación de los datos, el siguiente paso es su correcta organización para que brinden una información fiel y de gran utilidad, considerando su naturaleza de acuerdo con los propósitos para los que fueron recopilados.

Si las observaciones se hicieron atendiendo a ciertas cualidades o atributos de una población, se les denomina **datos cualitativos**; si, por el contrario, las observaciones se hicieron atendiendo a características que pueden representarse numéricamente, es decir, que provienen de variables continuas y discretas de una población, se les denomina **datos cuantitativos**.

Redondeo de datos

El redondeo de datos es una técnica útil, ya que tiene la finalidad de reducir al mínimo los errores acumulados por dicha práctica, sobre todo cuando la cantidad de datos es muy grande.

Para redondear un número a cualquier cifra en particular, es necesario saber qué tan cerca se encuentra dicho número de aquellos que se localizan antes y después de él.

Ejemplos

1. Si se tiene el número 54.8 y se solicita redondearlo al entero más próximo, resulta que es 55, debido a que el número dado está más cerca de 55 que de 54.
2. Si se tiene el número 127.83 y se solicita redondearlo al número decimal más próximo con una décima, resulta que es 127.8, debido a que el número dado está más cerca de 127.8 que de 127.9.
3. Si se tiene el número 6.472 y se solicita redondearlo al número decimal más próximo con una centésima, resulta que es 6.47, debido a que el número dado está más cerca de 6.47 que de 6.48.
4. Si se tiene el número 91.8627 y se solicita redondearlo al número decimal más próximo con una milésima, resulta que es 91.863, debido a que el número dado está más cerca de 91.863 que de 91.862.

En caso en que el número dado esté justo a la mitad (que termine en 5) del recorrido entre dos números, se acostumbra en tales situaciones redondear al número par más próximo que antecede o sea consecuente del 5.

Ejemplos

1. Si se tiene el número 76.5 y se solicita redondearlo al entero más próximo, resulta que es 76, debido a que el número dado está más cerca de 76, que es el número par más próximo que antecede al 5.
2. Si se tiene el número 1.35 y se solicita redondearlo al número decimal más próximo con una décima, resulta que es 1.4, debido a que el número dado está más cerca de 1.4, que es el número par más próximo que es consecuente del 5.

3. Si se tiene el número 385.275 y se solicita redondearlo al número decimal más próximo con una centésima, resulta que es 385.28, debido a que el número dado está más cerca de 385.28, que es el número par más próximo que es consecuente del 5.
4. Si se tiene el número 0.4685 y se solicita redondearlo al número decimal más próximo con una milésima, resulta que es 0.468, debido a que el número dado está más cerca de 0.468, que es el número par más próximo que antecede al 5.

EJEMPLO

Ejemplo 1 Suma los datos 7.75, 9.85, 4.55, 13.65, 8.45, 10.35, 12.15 en las siguientes formas:

a) Directa.
b) Redondeando por el criterio del número par más próximo.
c) Redondeando los datos por exceso.

Solución

a)	*b*)	*c*)
7.75	7.8	7.8
9.85	9.8	9.9
4.55	4.6	4.6
13.65	13.6	13.7
8.45	8.4	8.5
10.35	10.4	10.4
12.15	12.2	12.2
66.75	66.8	67.1

Se hace notar que el resultado en *b*) es más exacto que el que se obtiene en *c*), ya que se acerca más al resultado directo en *a*); también porque la **acumulación de error** por redondeo es mucho menor al aplicar el criterio del número par más próximo que antecede o es consecuente del 5.

Notación con índice o subíndice

En el lenguaje simbólico, la representación X_k —que se lee **X subíndice k** o también **X sub k**— indica cualquiera de los n valores $X_1, X_2, X_3, X_4, \ldots, X_n$ que se le pueden asignar a la variable X. La letra k en X_k se denomina **índice** o **subíndice** ya que representa cualquiera de los números 1, 2, 3, 4,…, n.

Como subíndice se puede utilizar cualquier letra diferente de k, por ejemplo: Y_z, S_p, T_m, V_l, Z_q, W_j.

Notación de suma

La notación $\sum_{k=1}^{n} X_k$ simboliza la operación de la suma de toda las X desde $k = 1$ hasta $k = n$, es decir:

$$\sum_{k=1}^{n} X_k = X_1 + X_2 + X_3 + X_4 + \ldots + X_n$$

La letra griega mayúscula Σ, llamada **sigma**, matemáticamente simboliza a la suma. Por esta razón la notación de suma también suele llamarse notación sigma.

EJEMPLOS

1 Escribe los sumandos de cada una de las siguientes sumas indicadas por el lenguaje simbólico.

a) $\sum_{k=1}^{8} X_k$ *b)* $\sum_{k=1}^{5} (Z_k - 1)^3$ *c)* $\sum_{k=1}^{4} (Y_k - b)$

Solución

a) $\sum_{k=1}^{8} X_k = X_1 + X_2 + X_3 + X_4 + \ldots + X_8$

b) $\sum_{k=1}^{5} (Z_k - 1)^3 = (Z_1 - 1)^3 + (Z_2 - 1)^3 + (Z_3 - 1)^3 + (Z_4 - 1)^3 + (Z_5 - 1)^3$

c) $\sum_{k=1}^{4} (Y_k - b) = (Y_1 - b) + (Y_2 - b) + (Y_3 - b) + (Y_4 - b) = Y_1 + Y_2 + Y_3 + Y_4 - 4b$

2 Aplicando el lenguaje simbólico de la notación de suma, expresa cada una de las siguientes sumas.

a) $(s_1 - t_1) + (s_2 - t_2) + (s_3 - t_3) + \ldots + (s_7 - t_7)$

b) $(f_1 z_1^2 + f_2 z_2^2 + f_3 z_3^2 + f_4 z_4^2 + \ldots + f_{12} z_{12}^2)$

c) $(m_1 n_1 + m_2 n_2 + m_3 n_3 + m_4 n_4 + \ldots + m_8 n_8)$

Solución

La notación de suma es:

a) $\sum_{k=1}^{7} (s_k - y_k)$ *b)* $\sum_{k=1}^{12} f_k z_k^2$ *c)* $\sum_{k=1}^{8} m_k n_k$

3 Dos variables X y Y toman los valores de:

$X_1 = 4, X_2 = -10, X_3 = 2, X_4 = -6$

$Y_1 = -6, Y_2 = -16, Y_3 = 5, Y_4 = 9$

Calcula:

a) $\sum_{k=1}^{4} X_k$ *b)* $\sum_{k=1}^{4} X_k Y_k$ *c)* $\sum_{k=1}^{4} Y_k^2$ *d)* $\sum_{k=1}^{4} (X_k + Y_k)(X_k + Y_k)$

Solución

Al escribir los sumandos para la suma indicada por el lenguaje simbólico se tiene:

a) $\sum_{k=1}^{4} X_k = X_1 + X_2 + X_3 + X_4$; al sustituir sus valores resulta:

$$\sum_{k=1}^{4} X_k = 4 + (-10) + 2 + (-6) = 4 - 10 + 2 - 6 = -10$$

b) $\sum_{k=1}^{4} X_k Y_k = X_1 Y_1 + X_2 Y_2 + X_3 Y_3 + X_4 Y_4$; al sustituir sus valores resulta:

$$\sum_{k=1}^{4} X_k Y_k = (4)(-6) + (-10)(-16) + (2)(5) + (-6)(9) = -24 + 160 + 10 - 54 = 92$$

c) $\sum_{k=1}^{4} Y_k^2 = Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2$; al sustituir sus valores resulta:

$$\sum_{k=1}^{4} Y_k^2 = (-6)^2 + (-16)^2 + (5)^2 + (9)^2 = 36 + 256 + 25 + 81 = 398$$

d) $\sum_{k=1}^{4} (X_k + Y_k)(X_k - Y_k) = \sum_{k=1}^{4} (X_k^2 - Y_k^2) = (X_1^2 - Y_1^2) + (X_2^2 - Y_2^2) + (X_3^2 - Y_3^2) + (X_4^2 - Y_4^2)$;

al sustituir sus valores resulta:

$$\sum_{k=1}^{4} (X_k^2 - Y_k^2) = [(4)^2 - (-6)^2] + [(-10)^2 - (-16)^2] + [(2)^2 - (5)^2] + [(-6)^2 - (9)^2]$$

$$\sum_{k=1}^{4} (X_k^2 - Y_k^2) = (16 - 36) + (100 - 256) + (4 - 25) + (36 - 81)$$

$$\sum_{k=1}^{4} (X_k^2 - Y_k^2) = -20 - 156 - 21 - 45 = -242$$

EJERCICIO 4

I. Realiza las siguientes actividades.

1. Enuncia los elementos básicos de la estadística.
2. Define el concepto población o universo.
3. Enuncia el concepto muestra.
4. Cita cinco ejemplos de población con su respectiva muestra.
5. Lista y describe las clases de población.
6. Escribe cinco ejemplos de población finita y cinco de población infinita.
7. Define qué es una variable.
8. Define variable continua y variable discreta.
9. Describe qué son los datos continuos.
10. Enuncia el procedimiento para obtener los datos discretos.
11. Explica la diferencia entre datos originales y datos indirectos.
12. Explica la diferencia entre datos cualitativos y datos cuantitativos.
13. Escribe cuál es la función del redondeo de datos.
14. Explica en qué consiste el criterio número par más próximo.
15. Enuncia cómo se llama la representación "$\sum$" y qué simboliza matemáticamente.

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

II. Indica si los siguientes enunciados representan una variable continua o una variable discreta.

1. El número de votos del partido demócrata en las elecciones de 1988. _______________
2. La cantidad de dinero ganado diariamente por un comerciante. _______________
3. La equivalencia de un pie lineal a pulgadas. _______________
4. El número de horas que tiene un día en total. _______________
5. El número de juegos ganados por Fernando Valenzuela en el béisbol de Grandes Ligas en la temporada de 1990. _______________
6. Las temperaturas registradas cada media hora en una caldera. _______________
7. Periodos de duración de los neumáticos de un automóvil. _______________
8. Cantidad de agua contenida en una lavadora de ropa. _______________
9. El número de municipios del estado de Campeche. _______________
10. La edad cronológica promedio de los seres humanos. _______________
11. El número de estaciones del año meteorológico. _______________
12. Los diámetros de las flechas medidos con un micrómetro. _______________
13. El número de errores tipográficos en un periódico. _______________
14. La cantidad de tornillos empacados por caja. _______________
15. Las longitudes de los dedos de sus manos. _______________

III. Redondea cada uno de los siguientes números a la aproximación indicada y discute tus resultados.

1.	2.065	a centésimas	_______________
2.	59.21	a décimas	_______________
3.	14.03359	a diezmilésimas	_______________
4.	99.599	a enteros	_______________
5.	67.7605	a milésimas	_______________
6.	17.55	a décimas	_______________
7.	125.4575	a milésimas	_______________
8.	2.05	a enteros	_______________
9.	278.660	a décimas	_______________
10.	0.555	a centésimas	_______________
11.	0.088	a décimas	_______________
12.	48.6682	a centésimas	_______________
13.	5.78500	a centésimas	_______________
14.	52.7	a enteros	_______________
15.	103.0501	a décimas	_______________
16.	341.5995	a milésimas	_______________
17.	0.7385	a centésimas	_______________
18.	3.14155	a diezmilésimas	_______________
19.	12.99995	a milésimas	_______________
20.	8.985985985	a cienmilésimas	_______________

IV. Resuelve los siguientes problemas.

1. Suma los datos 136.5, 32.675, 43.095, 18.6565, 84.455, 64.7385, 73.375 de las siguientes formas:
   a) Directa.
   b) Redondeando por el criterio del número par más próximo.
   c) Redondeando los datos por exceso.
2. Suma los datos 6.735, 11.965, 3.75, 8.535, 2.995, 3.6875, 2.11825, 14.895, 1.125 de las siguientes formas:
   a) Directa.
   b) Redondeando por el criterio del número par más próximo.
   c) Redondeando los datos por exceso.
3. Escribe los sumandos de cada una de las siguientes sumas indicadas con lenguaje simbólico.

   a) $\sum_{p=1}^{11} Z_p$ *c*) $\sum_{k=1}^{3}(3-Y_k)^2$ *e*) $\sum_{i=1}^{10} F_i X_i$ *g*) $\sum_{k=1}^{4}(X_k+Y_k)^2$

   b) $\sum_{q=1}^{7} U_q^2$ *d*) $\sum_{r=1}^{5} M_r$ *f*) $\sum_{j=1}^{6}(Z_j-2)$ *h*) $\sum_{p=1}^{2}(2T_p-1)$
4. Aplicando el lenguaje simbólico de la notación de suma, expresa cada una de las siguientes sumas.
   a) $Z_1^3 + Z_2^3 + Z_3^3 + Z_4^3$
   b) $g_1t_1^a + g_2t_2^a + \ldots + g_8t_8^a$
   c) $(2X_1 - Y_1 + 3Z_1) + (2X_2 - Y_2 + 3Z_2) + \ldots + (2X_7 - Y_7 + 3Z_7)$
   d) $X_1Y_1 + X_2Y_2 + X_3Y_3 + \ldots + X_6Y_6$
   e) $(X_1^3 + a) + (X_1^3 + a) + \ldots + (X_5^3 + a)$
5. Las variables S y T toman los valores:

   $S_1 = 3, S_2 = -7, S_3 = 4, S_4 = -1, S_5 = 8, S_6 = -9$

   $T_1 = -5, T_2 = -3, T_3 = 6, T_4 = 2, T_5 = -4, T_6 = -2$

   Calcula:

   a) $\sum_{k=1}^{6} S_k$ *d*) $\sum_{k=1}^{6} S_k{}^2$ *g*) $\sum_{k=1}^{6}(S_k+T_k)^2$

   b) $\sum_{k=1}^{6} T_k$ *e*) $\sum_{k=1}^{6} T_k^3$ *h*) $\sum_{k=1}^{6}(2S_k^2-3T_k^2)$

   c) $\sum_{k=1}^{6} S_kT_k$ *f*) $\sum_{k=1}^{6}(S_k^2-T_k^2)$ *i*) $\sum_{k=1}^{6}(S_k+T_k)^3$
6. Si $Y_k = -8$ y $Y_k^2 = 16$ determina:

   a) $\sum_{k=1}^{8}(3Y_k+1)$ *b*) $\sum_{k=1}^{8} Y_k(Y_k-2)$ *c*) $\sum_{k=1}^{8}(Y_k-3)^2$

Verifica tus resultados en la sección de respuestas.

Toma, ordenación y distribución de frecuencias de datos

Toma de datos

Consiste en recolectar una serie de datos que deben ser organizados con base en un orden numérico y en subgrupos, de acuerdo con las características comunes que presentan.

Ordenación estadística de los datos

Es el método o técnica que se usa para reunir datos numéricos, los cuales se pueden ordenar en forma creciente (ascendente) o decreciente (descendente).

Recorrido o rango

Es la diferencia que existe entre el mayor y el menor de los datos (Rango $= D_{mayor} - D_{menor}$).

EJEMPLOS

Ejemplos

1 Ordena y calcula el rango para los siguientes datos: 8, 13, 29, 16, 32, 24, 17, 10, 21, 28, 35.

Solución

(Ascendente)

Ordenación: 8, 10, 13, 16, 17, 21, 24, 28, 29, 32, 35

(Descendente)

El rango se determina por la ecuación:

$$\text{Rango} = D_{mayor} - D_{menor}$$
$$\text{Rango} = 35 - 8$$
$$\text{Rango} = 27$$

2 Las calificaciones finales de matemáticas de 40 estudiantes del CBTIS 30, son:

95	70	96	85	100	49	83	89
55	55	65	77	80	70	92	93
74	66	95	65	87	100	45	77
60	75	69	52	82	68	78	92
58	56	70	70	74	98	75	64

Determina su ordenación creciente y su rango.

Solución

La ordenación creciente es:

Dato menor	**45**	56	65	70	75	80	89	95	
	49	58	66	70	75	82	92	96	
	52	60	68	70	77	83	92	98	
	55	61	69	74	77	85	93	100	
	55	65	70	74	78	87	95	**100**	Dato mayor

Su rango es:

$$\text{Rango} = D_{\text{mayor}} - D_{\text{menor}}$$
$$\text{Rango} = 100 - 45$$
$$\text{Rango} = 55$$

EJERCICIO 5

I. Contesta las siguientes preguntas.

1. ¿En qué consiste la toma de datos?
2. ¿Qué es la ordenación estadística de los datos?
3. ¿Cómo se llama la forma de ordenar los datos de menor a mayor?
4. ¿Cómo se llama la forma de ordenar los datos de mayor a menor?
5. ¿Cómo se define el recorrido o rango de los datos?

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

II. Desarrolla en equipo los siguientes ejercicios y discute en plenaria los resultados.

1. Ordena y calcula el rango para los siguientes datos.
   a) 68, 60, 74, 94, 75, 78, 88, 72, 63, 78, 85, 95, 77, 62, 93, 90.
   b) 19, 54, 36, 29, 10, 48,16, 52, 39, 24.
   c) 2, 7, 20, 13, 1, 9, 16, 11, 5, 18.
2. Un cuestionario de 60 preguntas con el tema "Clásicos de la música", aplicado a 50 ciudadanos comunes, dio lugar a los siguientes aciertos.

   39, 44, 28, 53, 37, 46, 51, 33, 29, 48, 46, 57, 55, 38, 31, 49, 52, 56, 29, 55, 46, 58,

   55, 43, 36, 34, 41, 40, 44, 38, 36, 51, 50, 30, 35, 39, 49, 53, 19, 24, 37, 46, 31, 26,

   48, 44, 39, 52, 40, 35.

   Considera los datos anteriores y determina:
   a) Su ordenación decreciente.
   b) Su rango.
   c) Las cantidades de las 3 personas con el menor número de aciertos.

d) Las cantidades de las 3 personas con el mayor número de aciertos.

e) Las cantidades del primero al décimo lugar con mayor puntuación.

f) Cuántas personas obtuvieron más de 40 aciertos.

g) Cuántas personas obtuvieron menos de 30 aciertos.

h) La cantidad del vigésimo lugar en orden ascendente.

Verifica tus resultados en la sección de respuestas.

Distribución de frecuencias de datos

Cuando se tiene una gran cantidad de datos, se recomienda distribuirlos en clases o categorías y determinar con precisión el número de datos pertenecientes a cada clase. Este procedimiento se denomina **frecuencia de clase** o también, **frecuencia absoluta de clase.**

Lo anterior es de gran utilidad ya que permite analizar con mayor facilidad un grupo de datos sin que se tenga que considerar individualmente cada uno de ellos.

La orientación tabular de los datos en clases o categorías en la cual se conjunta cada clase con su respectiva frecuencia se denomina **distribución de frecuencias** o **tabla de frecuencias.**

Ejemplo

La tabla siguiente es una distribución de frecuencias de los salarios diarios de 72 profesionales de la industria petrolera.

Tabla de frecuencias

Clases o categorías

Salarios (dólares)	Número de profesionales
30 - 39	7
40 - 49	12
50 - 59	19
60 - 69	16
79 - 79	10
80 - 89	6
90 - 99	2
	72

Frecuencias de clase

La segunda clase o categoría comprende los salarios de 40 a 49 dólares y se representa (40 - 49), dado que doce profesionales tienen un salario perteneciente a esta clase, la respectiva frecuencia absoluta de clase es 12.

Cuando los datos se ordenan y se resumen en una tabla o distribución de frecuencias, éstos se denominan **datos agrupados**.

Intervalo de clase

Si se observa la tabla del ejemplo anterior, la cuarta clase o categoría se denota por los números "60 - 69". El conjunto de valores que abarcan desde 60 hasta 69 recibe el nombre de **intervalo de clase**. Cada intervalo de clase delimita los valores que ésta puede poseer y siempre consta de un límite inferior y de un límite superior.

Los números extremos, 60 y 69, son los **límites de clase**; el número menor (60) es el **límite inferior de la clase** y el mayor (69) es el **límite superior de la clase**.

Intervalo de clase abierto

Se define como el intervalo de clase a aquel que, al menos teóricamente, no tiene límite inferior o límite superior.

Ejemplo

Al referirse a las estaturas de grupos de individuos, los intervalos de clase que contienen estaturas mayores de 180 centímetros o menores de 150 centímetros, respectivamente se consideran un intervalos de clase abierto, es decir:

	Límite inferior		
Intervalo de clase abierto	Mayores de 180 cm	——	No tiene límite superior
Intervalo de clase abierto	No tiene límite inferior	——	Menores de 150 cm
			Límite superior

Límites reales de clases

Considerando la tabla del ejemplo sobre los salarios diarios que se registran con aproximación de dólar, la tercera clase (50 - 59) incluye a todos los salarios desde 49.50 a 59.50 dólares.

La representación exacta de los datos anteriores, 49.5 y 59.5, se denomina **límites reales de clase** o **límites verdaderos de clase**; el menor de ellos (49.5) se identifica como **límite real inferior** y el mayor de ellos (59.5) se identifica como **límite real superior**.

Límite real inferior

Es el límite que se determina sumando el límite inferior más el límite superior de la clase contigua anterior y dividiendo entre dos.

Límite real superior

Es el límite que se determina sumando el límite superior más el límite inferior de la clase contigua siguiente superior y dividiendo entre dos.

Ejemplo

Considerando los intervalos de clases de la tabla de los salarios diarios del ejemplo anterior.

Salarios (dólares)		
Límites inferiores		Límites superiores
*	————	29
30	————	39
40	————	49
50	————	59
60	————	69
70	————	79
80	————	89
90	————	99
100	————	*

Intervalos de clases

Al calcular los límites reales de clases para el primer intervalo de clase resulta:

$$\text{Límite real inferior } = \frac{30+29}{2} = 29.5$$

$$\text{Límite real superior } = \frac{39+40}{2} = 39.5$$

Para el segundo intervalo de clase resulta:

$$\text{Límite real inferior } = \frac{40+39}{2} = 39.5$$

$$\text{Límite real superior } = \frac{49+50}{2} = 49.5$$

Para el tercer intervalo de clase resulta:

$$\text{Límite real inferior } = \frac{50+49}{2} = 49.5$$

$$\text{Límite real superior } = \frac{59+60}{2} = 59.5$$

Por lo anterior, se concluye que los límites reales de clases para la tabla del ejemplo son:

Salarios (dólares)		
Límites reales inferiores		Límites reales superiores
29.5	————	39.5
39.5	————	49.5
49.5	————	59.5
59.5	————	69.5
69.5	————	79.5
79.5	————	89.5
89.5	————	99.5

Límites reales de clases

La tabla anterior resulta ser incierta, ya que los límites reales de clases no son coincidentes con las observaciones reales; por ejemplo, si una observación fuese 49.5, no es posible definir si pertenece al intervalo de clase (39.5 - 49.5) o al intervalo de clase (49.5 - 59.5). Lo anterior da lugar a que en algunas ocasiones los límites reales de clases sean utilizados únicamente como símbolos de las clases.

Tamaño, anchura o longitud de un intervalo de clase

Se define como la diferencia entre los límites reales de clase que forman el intervalo de clase.

La anchura común se presenta cuando todos los intervalos de clase de una distribución de frecuencias tienen igual tamaño o longitud, la cual se simboliza por ***C*** y matemáticamente es igual a la diferencia entre dos límites reales de clase inferiores o superiores sucesivos.

Ejemplo

Si consideramos los intervalos de clases de la tabla de los salarios diarios de 72 profesionales de la industria petrolera, tenemos que la anchura respectiva es:

$$\left.\begin{aligned} C &= 39.5 - 29.5 = 10 \\ C &= 49.5 - 39.5 = 10 \\ C &= 59.5 - 49.5 = 10 \end{aligned}\right\} \textbf{Anchura común}$$

Por lo anterior, se concluye que los intervalos de clase de dicha tabla presentan una anchura común de 10 unidades.

Marca de clase

Se define como el punto medio de un intervalo de clase y se representa por (x).

La marca de clase también se denomina **punto medio de la clase**, y matemáticamente se determina por la suma de los límites inferior y superior del intervalo de clase, dividida entre dos.

Lo anterior se simboliza por la ecuación:

$$\text{Marca de clase } (x) = \frac{\text{Límite inferior } (L_I) + \text{Límite superior } (L_S)}{2}$$

Ejemplo

Al considerar los intervalos de clases de la tabla de los salarios diarios del ejemplo anterior, tenemos que la marca de clase del primer intervalo es:

$$x = \frac{L_I + L_S}{2} = \frac{30+39}{2} = \frac{69}{2} = 34.5$$

Para el tercer intervalo de clase es:

$$x = \frac{L_I + L_S}{2} = \frac{50+59}{2} = \frac{109}{2} = 54.5$$

Para el quinto intervalo de clase es:

$$x = \frac{L_I + L_S}{2} = \frac{70+79}{2} = \frac{149}{2} = 74.5$$

Por lo anterior, se concluye que la marca de clase para la tabla del ejemplo es:

Salarios (dólares)		Marca de clase (x)
Límites inferiores	Límites superiores	
30	39	34.5
40	49	44.5
50	59	54.5
60	69	64.5
70	79	74.5
80	89	84.5
90	99	94.5

En todos los análisis estadísticos, se supone que el total de los elementos pertenecientes a un intervalo de clase son coincidentes con el valor de la marca de clase; por ejemplo, todos los salarios del segundo intervalo de clase (40 – 49) se consideran de 44.5 dólares.

Reglas generales para formar las distribuciones de frecuencias

1. Determinar el recorrido o rango entre los datos registrados.
2. Dividir el rango en un número razonable de intervalos de clase que tengan el mismo tamaño o anchura. Si lo anterior no es posible, se aconseja emplear intervalos de clase de diferente tamaño o anchura; también se recomienda utilizar intervalos de clase abiertos. Por lo general, el número de intervalos de clase se selecciona entre 5 y 20, dependiendo de la cantidad de datos registrados. Los intervalos de clase se eligen también de manera que las marcas de clase o puntos medios sean coincidentes con los datos observados realmente, lo cual permite disminuir el **error de agrupamiento**.
3. Determinar las frecuencias de clase, es decir, contar el número de observaciones que caen dentro de cada intervalo de clase.

EJEMPLOS

Ejemplos

1 En la siguiente tabla se presentan los pesos de 50 estudiantes que se registraron con aproximación de una libra. Construye la tabla de distribución de frecuencias.

128	154	140	122	154	135	159	167	142	144
136	148	130	137	146	143	162	154	146	147
158	116	179	141	139	129	164	175	149	128
136	163	132	137	145	144	150	145	170	181
151	135	125	132	160	156	155	138	153	147

Solución

a) Al ordenar los datos en forma creciente tenemos:

Dato menor	116	129	135	138	143	146	149	154	159	167	
	122	130	136	139	144	146	150	154	160	170	
	125	132	136	140	144	147	151	155	162	175	
	128	132	137	141	145	147	153	156	163	179	
	128	135	137	142	145	148	154	158	164	181	Dato mayor

b) Su rango es:

$$\text{Rango} = D_{\text{mayor}} - D_{\text{menor}}$$
$$\text{Rango} = 181 - 116$$
$$\text{Rango} = 65$$

c) Selección del número de intervalo de clase por medio de la ecuación:

$$\begin{array}{c}\text{Tamaño o anchura}\\ \text{de la clase}\end{array} = \frac{\text{Rango}}{\text{Número de intervalos de clase}}$$

Si se elige el 5 como el número de intervalos de clase, el tamaño o anchura de clase será de:

$$\frac{65}{5} = 13$$

Si se elige el 9 como el número de intervalos de clase, el tamaño o anchura de clase será de:

$$\frac{65}{9} = 7.22 = 7$$

Si se elige el 15 como el número de intervalos de clase, el tamaño o anchura de clase será de:

$$\frac{65}{15} = 4.33 = 4$$

Si se elige el 20 como el número de intervalos de clase, el tamaño o anchura de clase será de:

$$\frac{65}{20} = 3.25 = 3$$

Una selección razonable para el tamaño del intervalo de clase es de 7 libras, así los intervalos de clase pueden ser:

115 - 121

122 - 128

129 - 135

136 - 142

143 - 149

150 - 156

157 - 163

164 - 170

171 - 177

178 - 184

De igual manera, las marcas de clase pueden ser: 118, 125, 132, 139, 146, 153, 160, 167, 174 y 181 libras.

Con la elección anterior, los límites reales de clase serán: 114.5, 121.5, 128.5, 135.5, 142.5, 149.5, 156.5, 163.5, 170.5, 177.5, 184.5 libras, que no son coincidentes con los datos observados.

d) Si se construye la tabla de distribución de frecuencias tenemos:

Intervalos de clase	Marca de clase (*x*)	Columna de conteo	Frecuencia de clase
115 - 121	118	|	1
122 - 128	125	||||	4
129 - 135	132	𝍸 |	6
136 - 142	139	𝍸 ||||	9
143 - 149	146	𝍸 𝍸 |	11
150 - 156	153	𝍸 |||	8
157 - 163	160	𝍸	5
164 - 170	167	|||	3
171 - 177	174	|	1
178 - 184	181	||	2
			50

Es posible construir otras tablas de distribuciones de frecuencias utilizando otros criterios sobre la selección del número de intervalos de clase.

Si se establece que la anchura del intervalo de clase es de 13 libras, la tabla de distribución de frecuencias resultante es:

Intervalos	Marca de clase (*x*)	Columna de conteo	Frecuencia de clase
115 - 127	121	|||	3
128 - 140	134	𝍸 𝍸 𝍸	15
141 - 153	147	𝍸 𝍸 𝍸 |	16
154 - 166	160	𝍸 𝍸 |	11
167 - 179	173	||||	4
180 - 192	186	|	1
			50

Nota: la columna de conteo se emplea para facilitar la tabulación de las frecuencias de clase de los datos dados y siempre se omite formalmente en la tabla de distribución de frecuencias.

2 Los porcentajes de consumo nacional de la producción de trigo durante los últimos 60 años se registran en la siguiente tabla.

72.7	72.5	71.8	77.9	69.8	73.0	88.9	78.5	73.2	58.7
67.3	67.6	69.6	76.8	72.6	84.6	91.4	69.9	61.4	62.3
69.3	70.1	77.3	80.4	76.5	66.7	83.9	78.6	69.2	68.5
74.6	68.2	73.9	79.1	78.1	87.4	78.4	57.9	65.3	71.7
66.1	75.3	74.7	79.8	82.3	61.6	84.1	62.4	54.8	74.2
65.6	73.4	71.2	81.6	64.2	79.2	89.7	63.7	59.2	64.9

Construye una distribución de frecuencias de los porcentajes, utilizando intervalos de clase adecuados.

Solución

a) Después de ordenar los datos en forma creciente, tenemos:

Dato menor	**54.8**	62.3	65.6	68.5	70.1	72.7	74.6	77.9	79.2	84.1	
	57.9	62.4	66.1	69.2	71.2	73.0	74.7	78.1	79.8	84.6	
	58.7	63.7	66.7	69.3	71.7	73.2	75.3	78.4	80.4	87.4	
	59.2	64.2	67.3	69.6	71.8	73.4	76.5	78.5	81.6	88.9	
	61.4	64.9	67.6	69.8	72.5	73.9	76.8	78.6	82.3	89.7	
	61.6	65.3	68.2	69.9	72.6	74.2	77.3	79.1	83.9	**91.4**	Dato mayor

b) Su rango es:

$$\text{Rango} = D_{\text{mayor}} - D_{\text{menor}}$$
$$\text{Rango} = 91.4 - 54.8$$
$$\text{Rango} = 36.6$$

c) Al seleccionar convenientemente diez intervalos de clase, su tamaño o anchura es:

$$\begin{array}{c}\text{Tamaño o anchura}\\ \text{de la clase}\end{array} = \frac{\text{Rango}}{\text{Núm. de intervalos de clase}} = \frac{36.6}{11} = 3.327 \approx 3.33$$

Con la elección anterior ya podemos establecer los intervalos de clase, las marcas de clase y los límites reales de clase.

d) Al construir la tabla de distribución de frecuencias, tenemos:

Intervalos de clase	Marca de clase (*x*)	Frecuencia de clase
52.7 - 56.3	54.5	1
56.4 - 60.0	58.2	3
60.1 - 63.7	61.9	5
63.8 - 67.4	65.6	7
67.5 - 71.1	69.3	9
71.2 - 74.8	73.0	13
74.9 - 78.5	76.7	8
78.6 - 82.2	80.4	6
82.3 - 85.9	84.1	4
86.0 - 89.6	87.8	2
89.7 - 93.3	91.5	2
		60

3 ●●• La siguiente tabla muestra los diámetros interiores de 40 inyectores para motor diesel producidos por una compañía; éstos están registrados con aproximación de milésimas de pulgada. Construye la tabla de distribución de frecuencias.

0.438	0.441	0.435	0.431	0.426	0.437	0.428	0.424
0.429	0.444	0.440	0.436	0.436	0.439	0.439	0.437
0.443	0.436	0.446	0.433	0.442	0.434	0.435	0.430
0.440	0.434	0.439	0.432	0.437	0.438	0.437	0.442
0.436	0.432	0.434	0.430	0.435	0.433	0.431	0.438

Solución

a) Al ordenar los datos en forma creciente, se tiene:

0.424	0.430	0.433	0.435	0.436	0.437	0.439	0.442
0.426	0.431	0.433	0.435	0.436	0.438	0.439	0.442
0.428	0.431	0.434	0.435	0.437	0.438	0.440	0.443
0.429	0.432	0.434	0.436	0.437	0.438	0.440	0.444
0.430	0.432	0.434	0.436	0.437	0.439	0.441	**0.446**

b) Su rango es:

$$\text{Rango} = D_{\text{mayor}} - D_{\text{menor}}$$
$$\text{Rango} = 0.446 - 0.424$$
$$\text{Rango} = 0.022$$

Después de seleccionar convenientemente siete intervalos de clase, su tamaño o anchura es:

$$\begin{matrix}\text{Tamaño o anchura}\\ \text{de la clase}\end{matrix} = \frac{\text{Rango}}{\text{Num. de intervalos de clase}} = \frac{0.022}{8} = 0.00275 \approx 0.003$$

Con la elección anterior, estamos en condición de establecer los intervalos de clase, las marcas de clase y los límites reales de clase.

d) Al construir la tabla de distribución de frecuencias, se obtiene:

Intervalos de clase	Marca de clase (x)	Frecuencia de clase
0.423 - 0.425	0.424	1
0.426 - 0.428	0.427	2
0.429 - 0.431	0.430	5
0.432 - 0.434	0.433	7
0.435 - 0.437	0.436	11
0.438 - 0.440	0.439	8
0.441 - 0.443	0.442	4
0.444 - 0.446	0.445	2
		40

EJERCICIO 6

I. Realiza lo que se te indica a continuación.

1. ¿A qué se le denomina frecuencia de clase o frecuencia absoluta de clase?
2. ¿Cuál es la ventaja de distribuir los datos en clases o categorías con su frecuencia de clase?
3. ¿Qué es una distribución de frecuencias o tabla de frecuencias?

4. ¿Cómo se define un intervalo de clase?
5. ¿Qué es un intervalo de clase abierto?
6. Cita dos ejemplos que presenten intervalos de clase abiertos.
7. ¿Qué son los límites reales de clases?
8. ¿Cómo se obtienen los límites reales inferior y superior?
9. Define el tamaño, anchura o longitud de un intervalo de clase.
10. ¿Qué es la anchura común y cómo se simboliza?
11. Define el concepto marca de clase.
12. Matemáticamente, ¿cómo se determina la marca de clase?
13. Describe las reglas generales para formar las distribuciones de frecuencias.
15. Los estadistas recomiendan que el número de intervalos de clase para agrupar datos no sea menor de _______________ ni mayor de _______________.
16. ¿Qué es una clase o categoría de datos?

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

II. En pareja, resuelve los siguientes ejercicios y realiza un mapa conceptual con el proceso de solución.

1. La siguiente tabla presenta una distribución de frecuencias de los saltos de longitud en metros de 40 atletas del Comité Olímpico Mexicano.

Longitud del salto (metros)	Número de atletas
5.30 - 5.70	1
5.80 - 6.20	2
6.30 - 6.70	5
6.80 - 7.20	7
7.30 - 7.70	11
7.80 - 8.20	8
8.30 - 8.70	4
8.80 - 9.20	2
	40

Utiliza estos datos para determinar:

a) El número de clases o categorías.
b) El intervalo de la tercera clase.
c) El límite inferior de la quinta clase.
d) El límite superior de la segunda clase.
e) Los límites reales de la séptima clase.
f) El tamaño o anchura del cuarto intervalo de clase.
g) La marca de clase para todos los intervalos de la tabla.
h) La frecuencia de la tercera clase.
i) El intervalo que tiene la mayor frecuencia de clase.

Escribe los números correspondientes

Competencias genéricas

Competencias disciplinares

2. La siguiente tabla presenta una distribución de frecuencias de la duración en horas de la carga eléctrica de 150 pilas.

Duración (horas)	Número de pilas
30 - 39	7
40 - 49	16
50 - 59	21
60 - 69	32
70 - 79	43
80 - 89	28
90 o más	3
	150

Utiliza la información anterior para determinar:

a) El número de intervalos de clase.
b) El intervalo de la segunda clase.
c) El límite inferior de la cuarta clase.
d) El límite superior de la sexta clase.
e) Los límites reales de la tercera clase.
f) El tamaño o anchura del quinto intervalo de clase.
g) El tamaño o anchura del séptimo intervalo de clase.
h) El límite superior de la séptima clase.
i) La marca de clase para todos los intervalos de clase.
j) El total de frecuencias hasta la cuarta clase.

III. Con ayuda de tu profesor construye la tabla de distribución de frecuencias para los siguientes problemas.

1. Las siguientes son las puntuaciones finales que obtuvieron 48 estudiantes en un examen de química orgánica.

63	73	40	77	65	70	58	75
86	90	55	72	73	56	76	69
94	66	84	53	75	89	42	64
65	47	62	88	48	84	77	89
76	86	75	60	64	67	65	69
33	79	69	76	35	49	96	73

2. Las horas-clase que 80 docentes de una universidad tuvieron frente a grupo en el semestre de febrero-junio de 2013 son:

24	21	13	24	19	20	30	18	16	8
10	23	26	13	15	22	18	19	20	21
20	**25**	**17**	**23**	**21**	**36**	**24**	**38**	**12**	**18**
15	**18**	**20**	**15**	**18**	**12**	**28**	**17**	**18**	**30**
12	16	22	26	27	29	18	22	28	16
16	22	18	25	34	25	32	15	22	32
32	19	20	17	15	18	20	21	15	18
28	23	26	14	24	23	16	19	24	40

3. La masas en kilogramos de 50 reses destinadas al rastro de Monterrey, Nuevo León son:

336	392	315	318	321	337	365	380	375	345
319	365	381	329	387	358	370	387	370	356
327	353	310	326	318	382	394	358	400	388
346	326	359	360	305	362	376	343	395	363
376	335	317	302	346	379	324	354	335	348

4. Los siguientes son los diámetros, en pulgadas, de 42 tubos requeridos para perforar en primera fase, igual número de pozos petroleros en el sureste de la República Mexicana:

23.3	22.0	22.6	25.9	25.6	23.8	26.0
24.2	23.5	23.7	21.8	23.8	22.7	25.2
25.4	25.2	24.1	26.3	24.2	22.9	23.4
24.6	25.8	22.3	25.1	25.4	23.6	24.1
26.7	24.4	25.6	24.8	24.9	24.7	25.2
23.6	26.1	24.8	25.4	23.9	22.0	22.6

5. Los diámetros interiores de 63 arandelas fabricadas por una compañía, que se miden en pulgada, son:

0.583	0.571	0.574	0.582	0.565	0.574.	0.580	0.549	0.548
0.582	0.568	0.573	0.553	0.584	0.568	0.576	0.552	0.572
0.554	0.592	0.593	0.571	0.577	0.546	0.554	0.553	0.586
0.548	0.573	0.581	0.564	0.563	0.568	0.575	0.582	0.576
0.563	0.563	0.593	0.572	0.558	0.589	0.591	0.555	0.544
0.542	0.580	0.564	0.589	0.590	0.593	0.563	0.559	0.560
0.592	0.566	0.552	0.590	0.580	0.584	0.584	0.557	0.586

6. La siguiente tabla registra el número de días lluviosos durante los últimos 64 años en el norte de Texas.

106	130	117	143	132	106	87	117
146	111	128	107	105	110	139	132
122	118	99	138	113	124	119	128
118	123	134	124	102	136	82	156
124	89	141	79	126	128	145	149
114	96	109	133	88	98	97	136
108	124	112	85	139	101	152	124
130	102	126	121	94	87	120	103

Verifica tus resultados en la sección de respuestas.

Representaciones gráficas de los datos de una tabla de frecuencias

Introducción

La representación de información por medio de **gráficos** es realmente una exposición artística que no tiene sólo la finalidad de presentarlos datos, sino también enfocar las ideas y propósitos deseados.

Los estadistas manifiestan su imaginación en la forma de expresar gráficamente el mensaje que toda información debe proyectar, tratando de cumplir con el nivel de interpretación y comprensión de las personas a quienes va dirigido.

La mayoría de las representaciones gráficas de datos estadísticos son relaciones existentes entre dos variables, que permiten proyectar una **curva descriptiva** fácil de asimilar; entre estos tipos de curvas, tenemos las gráficas de líneas, las gráficas de barras, los pictogramas, los gráficos circulares, los histogramas, los polígonos de frecuencia, etcétera.

Gráficos de línea o diagramas lineales

Son gráficos que se emplean para representar sucesiones cronológicas y distribuciones de frecuencia; se pueden representar de dos formas:

Para datos acumulativos. Ejemplos de este tipo de datos son las producciones industriales (automóviles, artículos eléctricos, deportivos, etc.), producciones agrícolas (cosechas de trigo, maíz, frijol, etc.), producción ganadera (sementales, vacuno, porcino, bovino, etc.), ventas de toda índole, población (trabajadora, estudiantil, con servicio militar, etcétera).

Para datos instantáneos. Ejemplos de este tipo de datos son los controles de almacenes e inventarios de refaccionarias, tiendas de ropa y comestibles, madererías, farmacias, ferreteras, librerías, etc.; control de temperaturas en máquinas automáticas de soldar, calderas de vapor, etc.; control de presiones de bombeo de gas, inyectores de aire para equipo neumático, etc.; cotizaciones de mercado de la moneda, de acciones, de productos bursátiles, bienes y raíces, etcétera.

Los gráficos o curvas de distribuciones de frecuencia se presentan para datos agrupados como los pesos y estaturas de estudiantes, los salarios profesionales de cierto tipo de trabajadores, control ambiental, etcétera.

Recomendaciones para trazar gráficos lineales

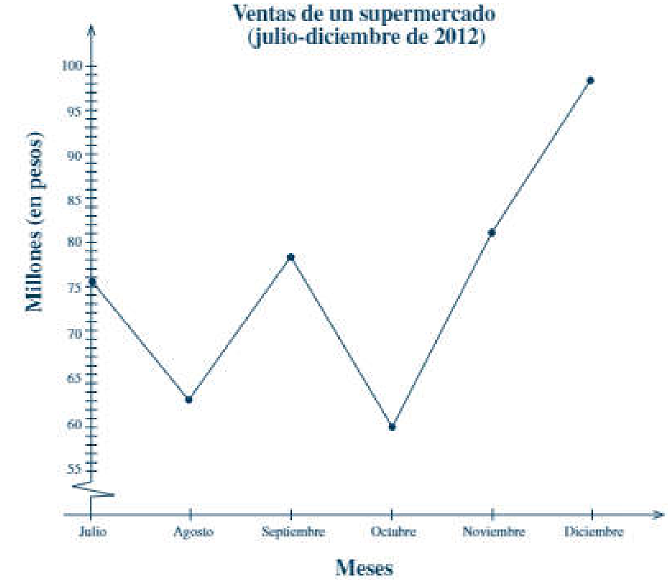
1. En las gráficas de sucesiones cronológicas la variable que representa el tiempo se ubica sobre el eje horizontal, la otra variable se coloca en el eje vertical. La relación de ambas variables se hace mediante el trazo de dos líneas rectas perpendiculares que cruzan en un punto perteneciente a la curva descriptiva.

   Se advierte que el **cero** se ubica en el eje vertical y siempre debe representarse. Si es necesario interrumpir la sucesión normal de valores de alguno de los ejes, esto se hace notar por una línea en zig-zag.
2. Las unidades de las variables deben sobresalir claramente y la curva se traza más gruesa que los ejes, para que resalte.
3. La longitud de los ejes se selecciona de manera que la gráfica quede equilibrada a lo largo y ancho.
4. Los títulos se escriben en la parte superior del gráfico, los letreros y notas se escriben por abajo del eje horizontal; si hay que destacar puntos específicos de la curva, éstos deben indicarse con notas al pie del eje horizontal.
5. Siempre debe citarse la fuente informativa.

EJEMPLOS

Ejemplos

1 Las ventas de un supermercado en el periodo julio-diciembre de 2012 se registraron de la siguiente manera: julio, $75 632 000.00; agosto, $63 584 600.00; septiembre, $78 325 200.00; octubre, $60 752 000.00; noviembre, $81 270 300.00; diciembre, $98 460 700.00; construye el gráfico lineal correspondiente.

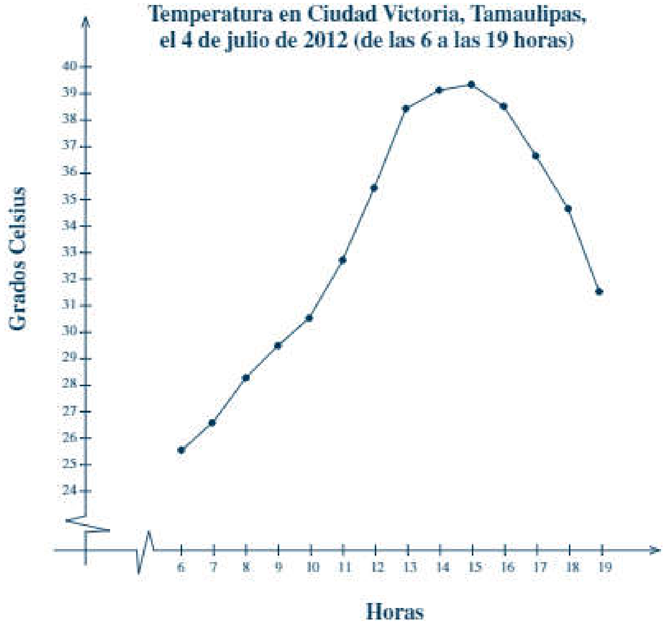


Fuente: Departamento de Ventas.

2. Las lecturas de temperatura observadas en Ciudad Victoria, Tamaulipas, el 4 de julio de 2012 entre las 6 y 19 horas, registradas con un termómetro comercial en unidades de grados Celsius cada hora, son:

6 horas — 25.3 °C	11 horas — 32.7 °C	16 horas — 38.6 °C
7 horas — 26.7 °C	12 horas — 35.4 °C	17 horas — 36.8 °C
8 horas — 28.2 °C	13 horas — 38.3 °C	18 horas — 34.7 °C
9 horas — 29.4 °C	14 horas — 39.2 °C	19 horas — 31.4 °C
10 horas — 30.6 °C	15 horas — 39.5 °C	

Construye el gráfico lineal correspondiente.



3. Traza el gráfico lineal correspondiente para la siguiente tabla de distribución de frecuencias

Intervalos de clase	Marca de clase (*x*)	Frecuencia de clase
30 - 39	34.5	9
40 - 49	44.5	32
50 - 59	54:5	43
60 - 69	64.5	21
70 - 79	74.5	11
80 - 89	84.5	3
90 - 99	94.5	1
		120

Para trazar el gráfico lineal se relacionan las marcas de clase que se ubican en el eje horizontal y las correspondientes frecuencias de clase que se colocan sobre el eje vertical, es decir:

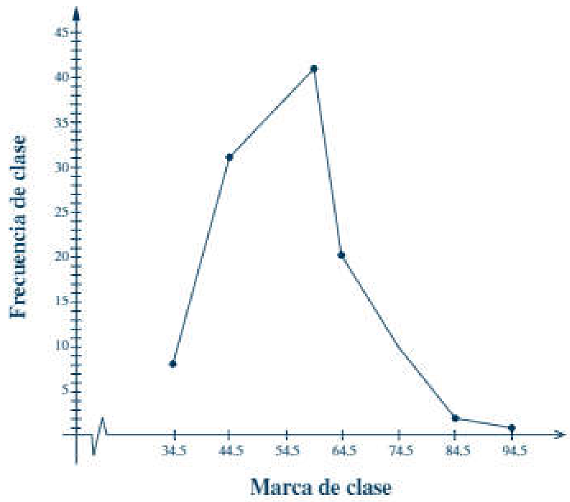


Gráfico o diagrama de barras

Los gráficos de barras proporcionan una mayor claridad y permiten una excelente interpretación de la información estadística.

Los diagramas de barras son muy diversos y numerosos, por lo se que utilizan para variables ordinales (datos ordenados o sucesivos, como *primero*, *segundo*, *tercero*, etc.); variables cardinales (datos que expresan cantidades de conteo).

Recomendaciones para trazar gráficos de barras

1. Las variable que representan las situaciones nominales, ordinales y cardinales se colocan sobre el eje horizontal y se relacionan con los valores correspondientes que se ubican en el eje vertical.
2. Se debe equilibrar el largo y ancho de cada barra.
3. Siempre se deja un espacio entre las barras, el cual no debe ser menor que la mitad del ancho de cada barra.
4. Si el diagrama consta de muchas barras, lo mejor es reemplazarlo por un gráfico de líneas.
5. El gráfico de barras puede ser **compuesto** con el fin de ahorrar espacio en el eje horizontal, pero las barras no deben saturarse al tratar de expresar muchos tipos de datos en cada una de ellas.

EJEMPLOS

Ejemplos

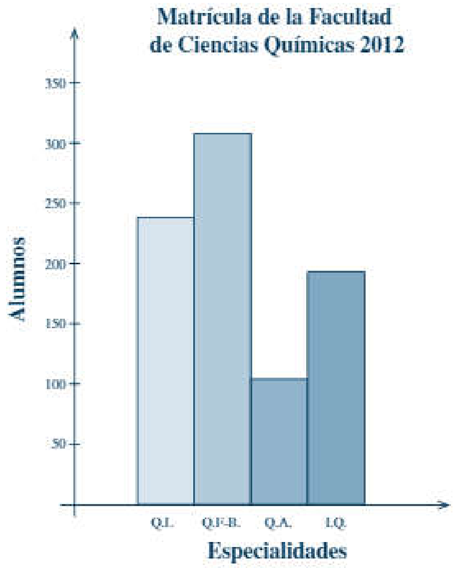
1 La matrícula de la Facultad de Ciencias Químicas de una universidad en sus diferentes especialidades (Químico industrial, Químico farmaco-biólogo, Químico administrador e Ingeniero químico) para el semestre agosto-diciembre de 2012, fue:

Especialidad	Matrícula	Hombres	Mujeres
Q.I.	242	178	64
Q.F-B.	318	98	220
Q.A.	104	59	45
I.Q.	182	165	17

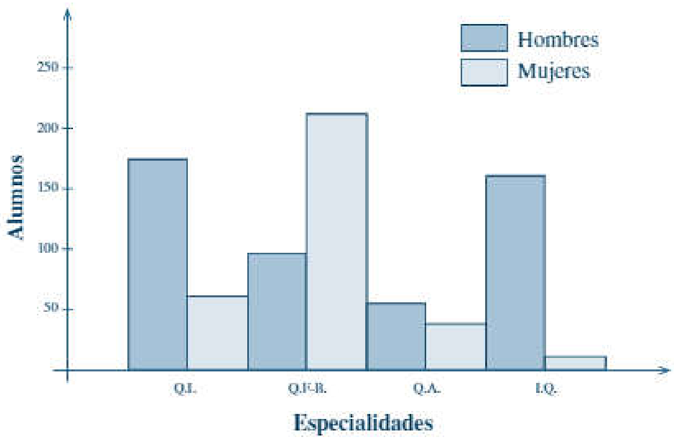
Construye el gráfico de barras para dicha información.

Solución

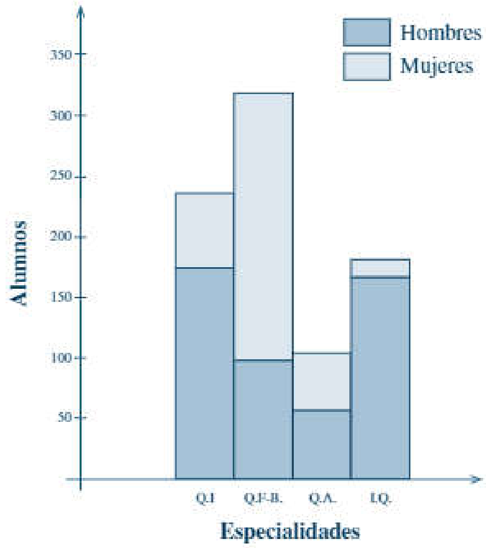
a) Si consideramos la matrícula en general para cada especialidad, se obtiene:



b) Al analizar la matrícula de acuerdo con el sexo, tenemos:



c) Al representar la matrícula con base en el sexo en un gráfico de barras compuesto, se obtiene:



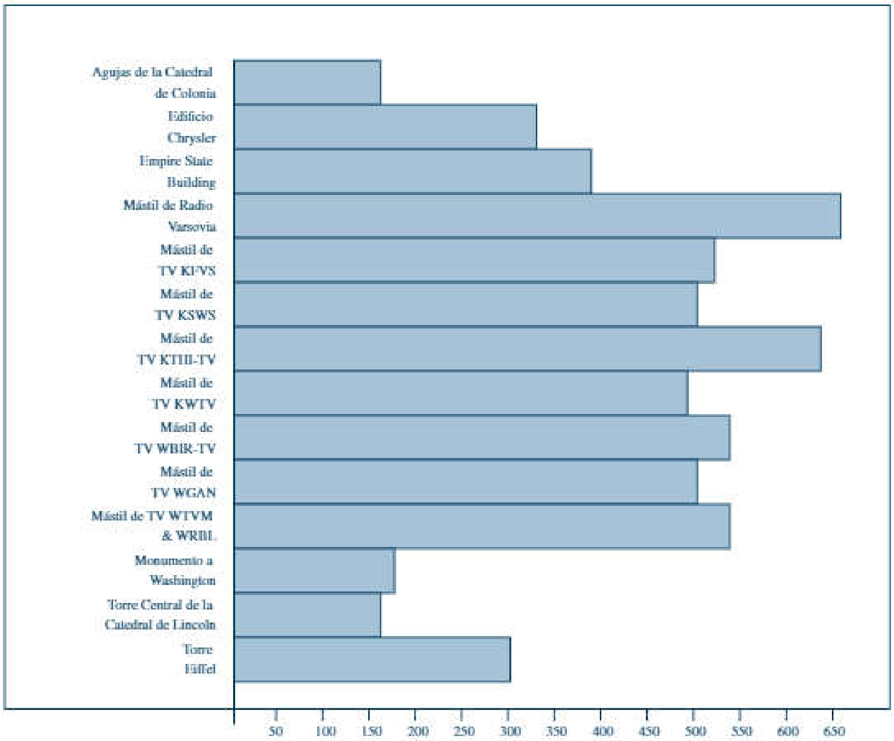
2 ●●•La siguiente tabla registra las 14 estructuras más altas del mundo hasta 1989.

Estructura	Altura (metros)	Localización
Agujas de la Catedral de Colonia	156	Colonia (Alemania)
Edificio Chrysler	318	Nueva York (EE.UU.)
Empire State Building	381	Nueva York (EE.UU.)
Mástil de Radio Varsovia	646	Plock (Polonia)
Mástil de TV KFVS	510	Cape Girardeau (EE.UU.)
Mástil de TV KSWS	490	Roswell (EE.UU.)
Mástil de TV KTHI-TV	628	Fargo (EE.UU.)
Mástil de TV KWTV	479	Oklahoma (EE.UU.)
Mástil de TV WBIR-TV	533	Knoxville (EE.UU.)
Mástil de TV WGAN	493	Portland (EE.UU.)
Mástil de TV WTVM & WRBL	533	Columbus (EE.UU.)
Monumento a Washington	169	Washington D. C. (EE.UU.)
Torre Central de la Catedral de Lincoln	160	Lincoln (Inglaterra)
Torre Eiffel	300	París (Francia)

Construye el gráfico de barras para dicha información.

Solución

Estructuras más altas del mundo hasta 1989 (datos suministrados por el libro *Guinness* de los Récords, edición 1989)



El gráfico de barras anterior se presenta con las barras horizontales en lugar de verticales (forma normal), por ser una modalidad que nos permite aprovechar mayor el papel sobre el que se traza la gráfica; si se desea, la gráfica de barras puede presentarse para los datos ordenados en forma creciente o decreciente.

Pictogramas o pictógrafos

El pictograma o pictógrafo se define como la representación de datos por medio de símbolos, en la que la forma de ésta insinúa la especie del dato estadístico. La originalidad en el arte de presentar datos estadísticos por medio de pictogramas, tiene como finalidad captar la atención del observador-analista y hacer comprensible e impactante la idea estadística.

Como generalmente los pictogramas se utilizan para exponer comparaciones de datos en magnitud y tiempo, siempre se presenta cierta indecisión en el análisis de alturas, áreas y volúmenes, razón por la cual los pictógrafos se deben indicar con un mayor o menor número de figuras del mismo tamaño, las aproximaciones se representan por fracciones de la misma figura. En el pictograma se debe indicar la equivalencia cuantitativa de la figura, tal y como trata las escalas en un dibujo técnico.

EJEMPLOS

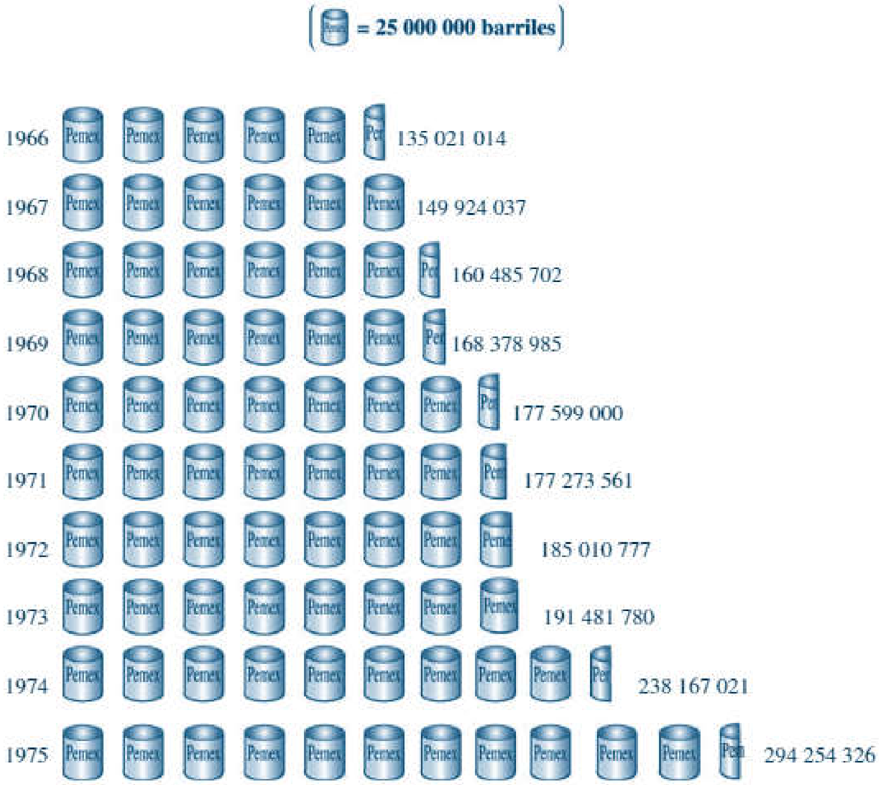
Ejemplos

1 La siguiente tabla registra la producción de petróleo crudo, condensado y líquidos de absorción en la República Mexicana en la década 1966-1975.

Año	Producción (barriles)
1966	135 021 014
1967	149 924 037
1968	160 485 702
1969	168 378 985
1970	177 599 000
1971	177 273 561
1972	185 010 777
1973	191 481 780
1974	238 167 021
1975	294 254 326

Solución

Producción de petróleo crudo, condensado y líquidos de absorción en la República Mexicana (1966-1975)



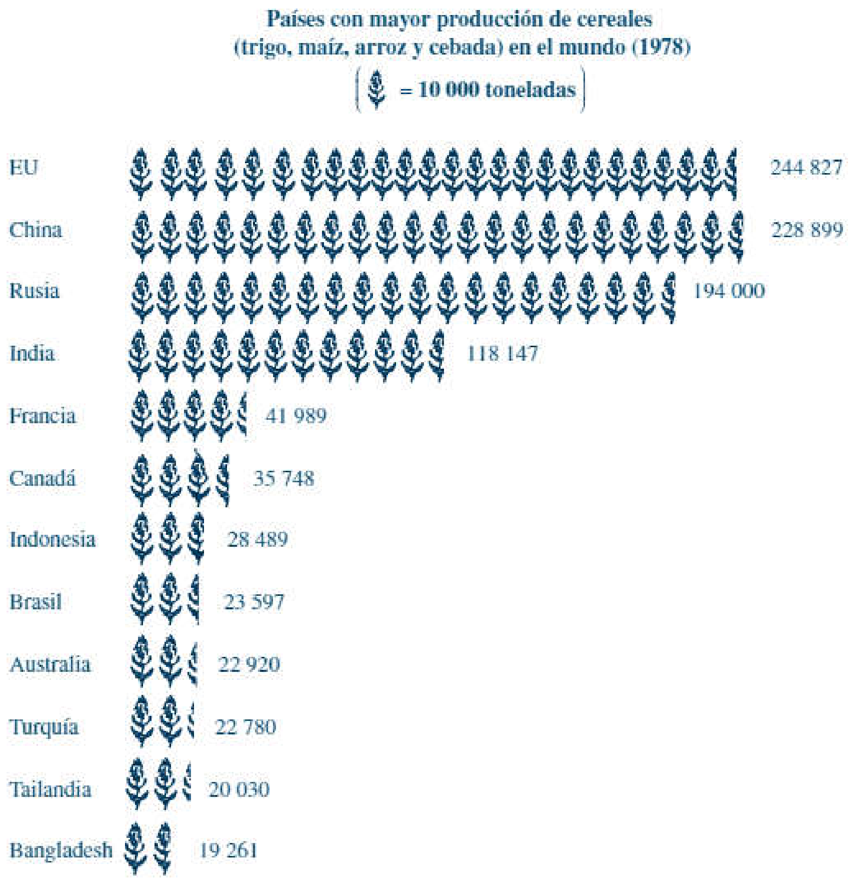
Fuente: Informes anuales de Petróleos Mexicanos.

2 La siguiente tabla registra los 12 países con mayor producción de cereales (trigo, maíz, arroz y cebada) en el mundo hasta 1978.

País	Millones de toneladas
EE.UU.	244 827
China	228 899
Rusia	194 000
India	118 147
Francia	41 989
Canadá	35 748

País	Millones de toneladas
Indonesia	28 489
Brasil	23 597
Australia	22 920
Turquía	22 780
Tailandia	20 030
Bangladés	19 261

Solución



Fuente: Anuario Estadístico de Naciones Unidas.

Gráfica circular o diagrama de pastel

Una gráfica circular o diagrama de pastel se define como la representación de datos distribuidos en forma porcentual, es decir, el círculo se divide en sectores (rebanadas de pastel) que son equivalentes en tamaño a las frecuencias porcentuales correspondientes.

La construcción de este tipo de gráficos tiene como punto inicial la consideración de que el área total corresponde a 360° equivalentes al 100% del círculo; cada porción de área corresponde a una determinada clase de datos, es decir, es un sector que representa un tanto por ciento igual a la razón entre el ángulo que forman los radios que lo limitan y los 360° de la circunferencia. Mediante el apoyo del transportador se trazan las porciones resultantes.

Este gráfico también se utiliza para mostrar secuencias cronológicas de datos, para ello es necesario dibujar círculos de igual radio, uno por cada año, en donde cada círculo representa la distribución porcentual correspondiente a los datos dados.

EJEMPLOS

Ejemplos

1 La siguiente tabla registra las superficies de los cinco continentes que conforman el mundo, en millones de kilómetros cuadrados.

Continente	Superficie (km^2)
África	30 224 000
América	42 198 760
Asia	44 180 000
Europa	10 000 000
Oceanía	8 970 000
Total	135 572 760 km^2

Construye el gráfico circular correspondiente.

Solución

Al transformar las superficies dadas en grados sexagesimales, tenemos:

$$\begin{aligned} \text{Si } 360° &\longrightarrow 135\,572\,760 \text{ km}^2 \\ x &\longrightarrow 30\,224\,000 \text{ km}^2 \end{aligned} \qquad x = 80°\ 16'$$

Si transformamos las superficies dadas en tanto por ciento, se obtiene:

$$\begin{aligned} \text{Si } 100\% &\longrightarrow 135\,572\,760 \text{ km}^2 \\ x &\longrightarrow 30\,224\,000 \text{ km}^2 \end{aligned} \qquad x = 22.3\%$$

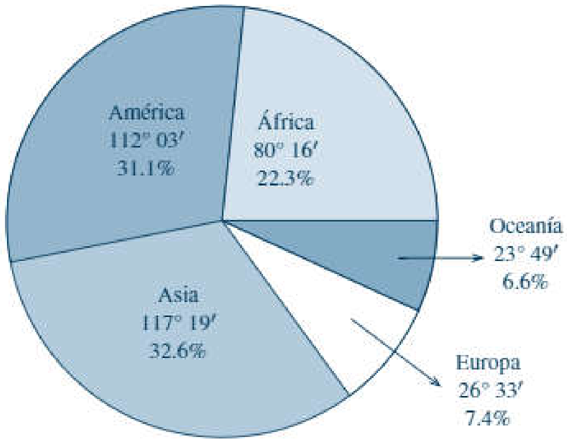
Al aplicar ambas operaciones para las demás superficies resulta:

Continente	Superficie (km^2)	Grados sexagesimales	%
África	30 224 000	80° 16′	22.3
América	42 198 760	112° 03′	31.1
Asia	44 180 000	117° 19′	32.6
Europa	10 000 000	26° 33′	7.4
Oceanía	8 970 000	23° 49′	6.6
Totales	135 572 760 km^2	360°	100%

Se hace notar que la suma total de los grados sexagesimales y de los porcentajes deben ser respectivamente igual a 360° y 100%.

Con ayuda del transportador, trazamos las porciones resultantes, dando lugar al gráfico circular correspondiente.

Superficies de los cinco continentes que conforman el mundo (millones de kilómetros cuadrados)



Fuente: Organización de las Naciones Unidas (ONU).

2 La capacidad eléctrica instalada en la República Mexicana en 1991 era de 26 793 Megawatts (Mw). Dicha capacidad era generada por distintas plantas, como se puede observar en la siguiente tabla.

Base	Capacidad de generación (Mw)
Carboeléctricas	1 205.7
Geotermoeléctricas	723.4
Hidrocarburos	16 263.4
Hidroeléctricas	7 930.7
Nucleoeléctricas	669.8
Total	26 793

Construye el gráfico circular de acuerdo con los datos dados.

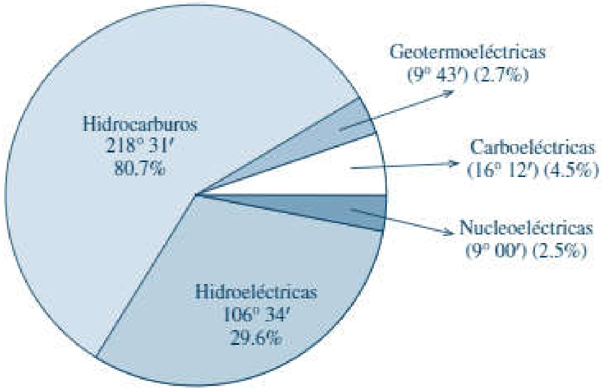
Solución

Después de transformar las capacidades de generación dadas en grados sexagesimales y en porcentaje, resulta:

Base generadora	Capacidad de generación (Mw)	Grados sexagesimales	%
Carboeléctricas	1 205.7	16° 12′	4.5
Geotermoeléctricas	723.4	9° 43′	2.7
Hidrocarburos	16 263.4	218° 31′	60.7
Hidroeléctricas	7 930.7	106° 34′	29.6
Nucleoeléctricas	669.8	9° 00′	2.5
Totales	26 793	360°	100%

Si trazamos las porciones resultantes, tenemos el gráfico circular correspondiente.

Capacidad eléctrica instalada en la República Mexicana en 1991



Fuente: Comisión Federal de Electricidad (CFE).

Gráfica de distribución de frecuencia

La construcción de una distribución de frecuencias tiene como objetivo principal condensar grandes conjuntos de datos y hacer notar sus propiedades en forma gráfica.

Algunos de los medios más comunes de mostrar gráficamente las distribuciones de frecuencias son los histogramas y polígonos de frecuencias; otras formas son las distribuciones de frecuencias porcentuales o relativas y las distribuciones de frecuencias acumuladas u ojivas.

Histograma

Se define como la forma de representar gráficamente una distribución de frecuencias; básicamente consta de una sucesión de rectángulos cuyas bases se ubican sobre el eje horizontal y cuya longitud es igual a la anchura de los intervalos de clase; sus alturas son proporcionales a las frecuencias de clase que se ubican sobre el eje vertical.

Para construir un histograma se marcan sucesivamente sobre el eje horizontal las anchuras de los intervalos de clase (el límite inferior y el límite superior de cada intervalo de clase), con su respectiva marca o punto medio de la clase; de esta manera se obtienen las bases de los rectángulos.

Sobre el eje vertical se marcan las frecuencias de clase, para dar lugar a las alturas de los rectángulos.

Si la anchura de los intervalos de clase es del mismo tamaño para todos los datos, las superficies de los rectángulos serán proporcionales a las frecuencias de clase; en caso contrario, las áreas de los rectángulos deberán ser calculadas.

A pesar de que un histograma es muy similar a los diagramas de barras, conceptualmente se hacen notar las siguientes diferencias:

a) En un gráfico de barras, las alturas de las mismas guardan relación con la variable ubicada sobre el eje vertical, mientras que en los histogramas las superficies de los rectángulos (barras) son proporcionales a las frecuencias de clase.

b) En un diagrama de barras éstas se grafican separadas, es decir, dejando espacios entre cada una de ellas; en los histogramas los rectángulos (barras) se representan en forma consecutiva.

Polígono de frecuencia

Es otra forma gráfica de mostrar las distribuciones de frecuencia y que fundamentalmente consta en que las frecuencias de clase se indican en las marcas o puntos medios de clase, resultando una serie de puntos que se conectan por medio de líneas rectas.

Para construir un polígono de frecuencias, se trazan sucesivamente sobre el eje horizontal las marcas de clase y las frecuencias correspondientes sobre el eje vertical, luego los puntos que resultan se unen por segmentos de recta; se hace notar que se deben agregar clases con frecuencia cero en ambos extremos de la distribución con el fin de enlazar el diagrama al eje horizontal.

El polígono de frecuencias se considera derivado del histograma. La suma de las áreas de los rectángulos del histograma es igual al área total limitada por el polígono de frecuencia y el eje horizontal.

EJEMPLOS

Ejemplos

1 Dada la siguiente tabla de distribución de frecuencias de los pesos de 50 estudiantes, construye:

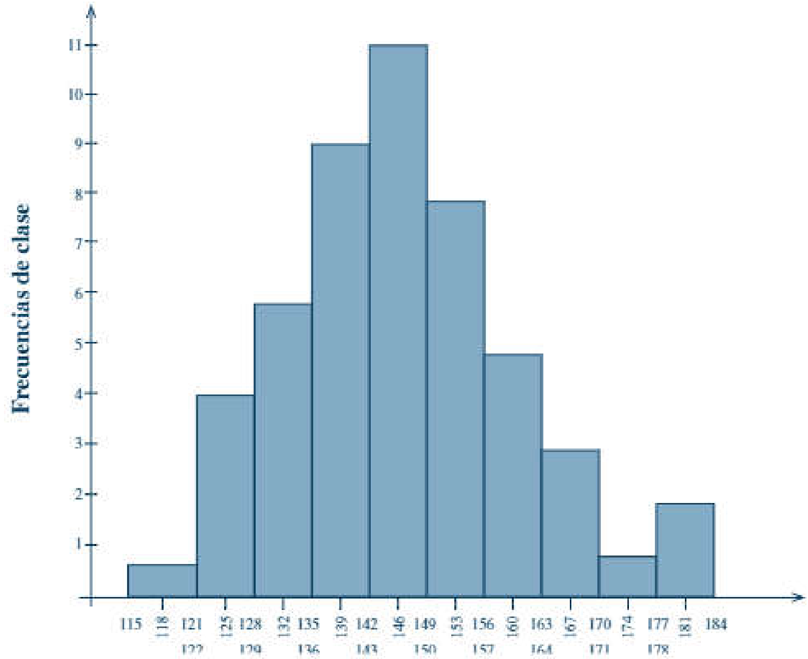
a) Su histograma.
b) Su polígono de frecuencias.
c) Su histograma y polígono de frecuencias en una misma representación gráfica.

Intervalos de clase	Marca de clase (*x*)	Frecuencia de clase
115 - 121	118	1
122 - 128	125	4
129 - 135	132	6
136 - 142	139	9
143 - 149	146	11
150 - 156	153	8
157 - 163	160	5
164 - 170	167	3
171 - 177	174	1
178 - 184	181	2
		50

Solución

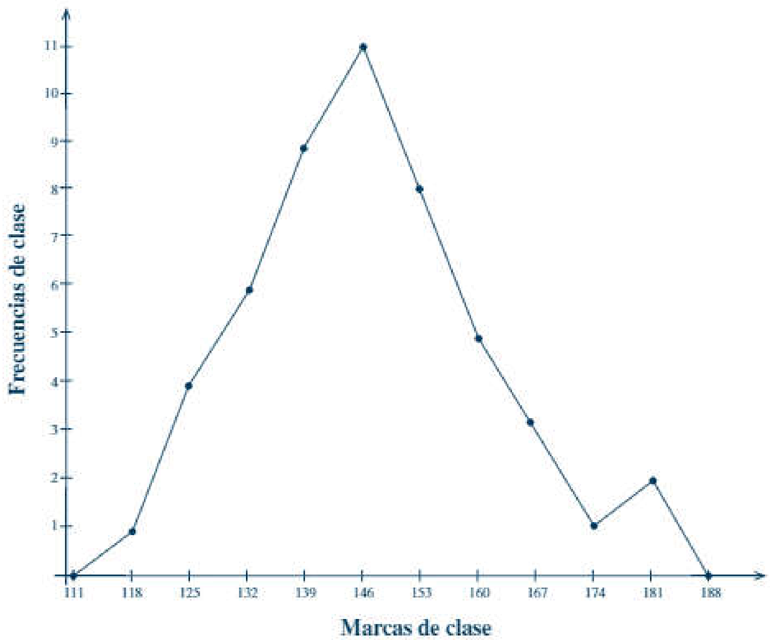
a) El histograma se construye marcando sucesivamente sobre el eje horizontal las anchuras de los intervalos de clase, con su respectiva marca de clase; de esta manera se obtienen las bases de los rectángulos. Sobre el eje vertical se marcan las frecuencias de clase, para dar lugar a las alturas de los rectángulos.

Se recomienda usar papel milimétrico.

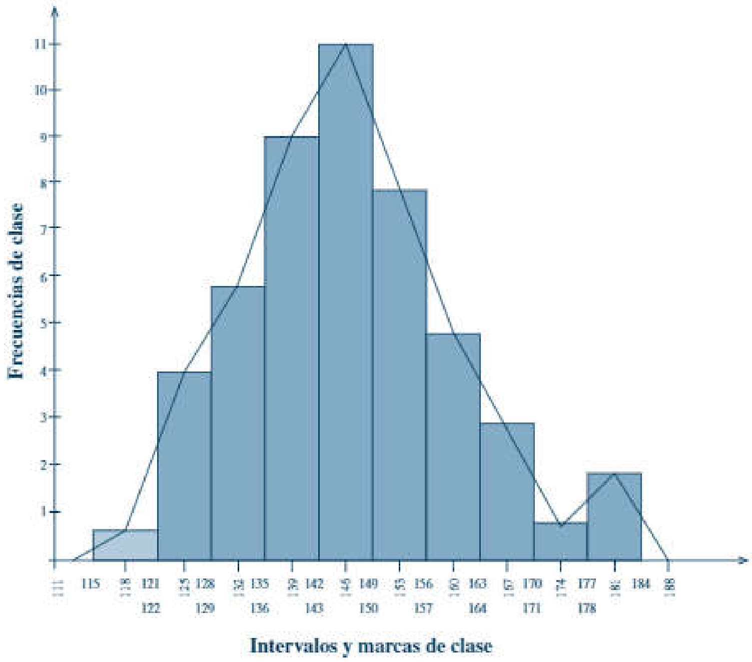


b) El polígono de frecuencias se construye trazando sucesivamente sobre el eje horizontal las marcas de clase, y sobre el eje vertical las frecuencias de clase correspondientes; luego, los puntos resultantes se unen con segmentos de recta; se deben agregar clases con frecuencia cero en ambos extremos de la distribución para conectar el gráfico al eje horizontal.

Se recomienda usar papel milimétrico.



c) El polígono de frecuencias es considerado un gráfico que se deriva del histograma, ya que resulta del trazado de segmentos de recta que conectan los puntos medios de los **techos** de los rectángulos en el histograma.

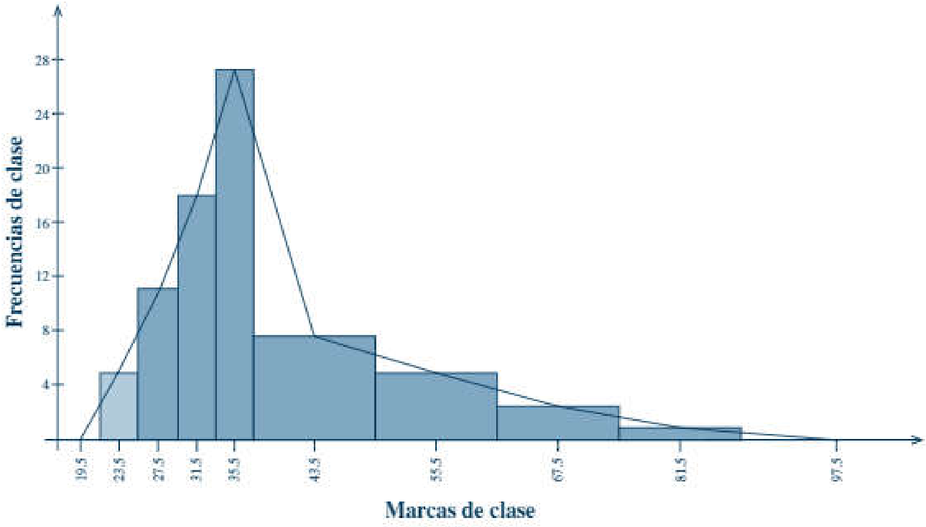


2 Traza el histograma y el polígono de frecuencias para la siguiente tabla de distribución de frecuencias.

Intervalos de clase	Marca de clase (x)	Frecuencia de clase
22 - 25	23.5	5
26 - 29	27.5	11
30 - 33	31.5	18
34 - 37	35.5	27
38 - 49	43.5	23
50 - 61	55.5	16
62 - 72	67.5	9
73 - 89	81.5	3
		112

Solución

Para construir el histograma y el polígono de frecuencias, partimos del principio que establece que el área es proporcional a la frecuencia de clase.



Se observa que las bases del quinto, sexto y séptimo rectángulo (tamaño de la clase igual a 12), al igual que la del último (tamaño de la clase igual a 16), son tres y cuatro veces mayores que las bases de los primeros rectángulos (tamaño de la clase igual a 4), respectivamente.

Si suponemos que el segundo rectángulo con frecuencia 11 y el quinto rectángulo con frecuencia 23, se relacionan con su respectiva área, resulta:

11 frecuencia $\longrightarrow$ 1 unidad de área
23 frecuencia $\longrightarrow$ x

$x = 2.0909$ unidades de área

Lo anterior, indica que el área del quinto rectángulo es 2.0909 veces mayor que el área del segundo rectángulo.

Dado que la base del quinto rectángulo es tres veces mayor que la base del segundo rectángulo, establecemos que su altura será 1/3 de la que tiene el segundo rectángulo, es decir:

unidad de altura $\longrightarrow$ 2.0909 unidad de área
1/3 unidad de altura $\longrightarrow$ x

$x = 0.696969$ unidades de área

Transformando el área anterior en unidades de frecuencia, tenemos:

1 unidad de área $\longrightarrow$ 11 frecuencia
0.696969 unidad de área $\longrightarrow$ x

$x = 7.666$ frecuencia

Por lo anterior se deduce que la altura del quinto rectángulo es de 7.666 unidades de frecuencia.

Si relacionamos el área del primer rectángulo con frecuencia 5 con el sexto rectángulo con frecuencia 16, obtenemos que el área del sexto rectángulo es 3.2 veces mayor que la del primer rectángulo.

Dado que la base del sexto rectángulo es 3 veces mayor que la base del primer rectángulo, establecemos que su altura será 1/3 de la que tiene el primer rectángulo, resultando 1.06 unidades de área y que transformándola en unidades de frecuencia, se tiene que la altura del sexto rectángulo es de 5.3 unidades.

Si relacionamos el área del tercer rectángulo con frecuencia 18 con la del séptimo rectángulo con frecuencia 9, obtenemos que el área del séptimo rectángulo es 0.5 veces menor que la del tercer rectángulo.

Dado que la base del séptimo rectángulo es tres veces mayor que la base del tercer rectángulo, establecemos que su altura será 1/3 de la que tiene el tercer rectángulo, resultando 0.1666 unidades de área y que transformándola en unidades de frecuencia, se tiene que la altura del séptimo rectángulo es de 2.999 unidades.

Si relacionamos el área del tercer rectángulo con frecuencia 18 con el último rectángulo con frecuencia 3, obtenemos que el área del último rectángulo es 0.1666 veces menor que la del tercer rectángulo.

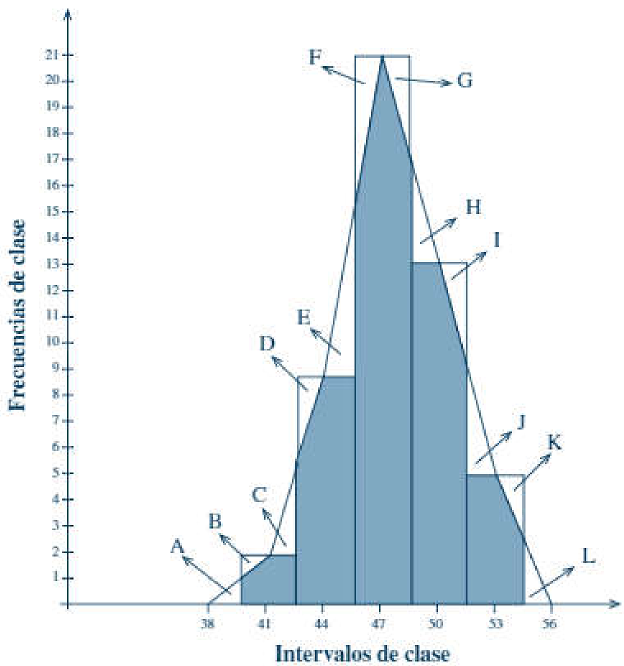
Dado que la base del último rectángulo es cuatro veces mayor que la base del tercer rectángulo, establecemos que su altura será 1/4 de la que tiene el tercer rectángulo, resultando 0.0416 unidades de área y que transformándola en unidades de frecuencia, se tiene que la altura del último rectángulo es de 0.75 unidades.

3 Construye el histograma y el polígono de frecuencias para la siguiente tabla de distribución de frecuencias. Comprueba que el área total de los rectángulos del histograma es igual al área total limitada por el polígono de frecuencias y el eje horizontal.

Intervalos de clase	Marca de clase (x)	Frecuencia de clase
40 - 42	41	2
43 - 45	44	9
46 - 48	47	21
49 - 51	50	13
52 - 54	53	5
		50

Solución

Si trazamos el histograma y el polígono de frecuencias correspondiente, tenemos:



Para la comprobación del área total de cada gráfico, se hace notar que:

Área (A) = Área (B) Área (G) = Área (H)

Área (C) = Área (D) Área (I) = Área (J)

Área (E) = Área (F) Área (K) = Área (L)

Por lo anterior, tenemos:

$$\begin{array}{c}\text{Área total}\\ \text{de los rectángulos}\\ \text{(histograma)}\end{array} = \begin{array}{c}\text{Área}\\ \text{sombreada}\end{array} + \text{área }(B) + \text{área }(D) + \text{área }(F) + \text{área }(G) + \text{área }(I) + \text{área }(K)$$

$$\begin{array}{c}\text{Área total}\\ \text{de los rectángulos}\\ \text{(histograma)}\end{array} = \begin{array}{c}\text{Área}\\ \text{sombreada}\end{array} + \text{área }(A) + \text{área }(C) + \text{área }(E) + \text{área }(H) + \text{área }(J) + \text{área }(L)$$

$$\begin{array}{c}\text{Área total}\\ \text{de los rectángulos}\\ \text{(histograma)}\end{array} = \begin{array}{c}\text{Área total limitada por}\\ \text{el polígono de frecuencia}\\ \text{y el eje horizontal}\end{array}$$

Distribuciones de frecuencia relativa

Una tabla de distribución de frecuencia relativa o porcentual resulta al sustituir las frecuencias absolutas de las clases por las frecuencias relativas respectivas.

La frecuencia relativa de una clase se obtiene al dividir la frecuencia de clase entre el total de frecuencias de todas las clases.

$$\begin{array}{c}\text{Frecuencia relativa}\\ \text{de una clase (FRC)}\end{array} = \frac{\text{Frecuencia de clase}}{\text{Frecuencia total}} = (100\%)$$

Para expresar la frecuencia relativa en porcentaje, es posible obtener la frecuencia relativa porcentual al multiplicar la frecuencia relativa por 100.

$$\begin{array}{c}\text{Frecuencia relativa porcentual}\\ \text{de una clase}\end{array} = \frac{\text{Frecuencia de clase}}{\text{Frecuencia total}}\,(100)$$

Al sumar las frecuencias relativas de todas las clases, obtenemos que es 1; si las frecuencias relativas están expresadas como porcentaje, el resultado es 100%.

Las gráficas de distribuciones de frecuencia relativa se pueden derivar del histograma y del polígono de- frecuencias. Lo anterior se obtiene al cambiar en el eje vertical la frecuencia de clase por frecuencia relativa, sin alterar exactamente el mismo gráfico. Los gráficos resultantes se denominan **histogramas porcentuales** o **de frecuencias relativas** y **polígonos porcentuales** o **de frecuencias relativas**, respectivamente.

Distribuciones de frecuencia acumulada (ojivas)

Una tabla de distribución de frecuencia acumulada resulta al sustituir las frecuencias de las clases por las frecuencias acumuladas.

La frecuencia acumulada hasta un determinado intervalo de clase se define como la frecuencia total de todos los valores menores que el límite real superior de clase del intervalo de clase considerado.

Un diagrama que represente las frecuencias acumuladas menores que cualquier límite real superior de clase dibujado sobre los límites reales superiores de clase, se denomina **polígono de frecuencias acumuladas u ojivas**; también se le llama **distribución acumulada "menor que"**.

Al considerar un gráfico que muestre las frecuencias acumuladas de todos los valores mayores o iguales al límite real inferior de clase dibujado sobre los límites reales inferiores de clase, se denomina **polígono de frecuencias acumuladas** u **ojivas** o **distribución acumulada "mayor que"**.

Distribuciones de frecuencias relativas acumuladas (ojivas porcentuales)

La frecuencia relativa acumulada se determina al dividir la frecuencia acumulada entre la frecuencia total. La frecuencia porcentual acumulada se determina al multiplicar la frecuencia relativa acumulada por 100.

Al sustituir en una tabla de distribuciones y en un gráfico las frecuencias acumuladas por las frecuencias relativas acumuladas, los resultados se denominan **distribuciones de frecuencia relativas acumuladas** o **porcentual acumulada** y **polígonos de frecuencias relativas acumuladas** u **ojivas porcentuales**, respectivamente.

EJEMPLOS

Ejemplos

1 Observa la tabla de distribuciones de frecuencia para los diámetros interiores de 40 inyectores para motores diesel producidos por una compañía.

Tabla de distribuciones de frecuencias

Intervalos de clase	Marca de clase (x)	Frecuencia de clase
0.423 - 0.425	0.424	1
0.426 - 0.428	0.427	2
0.429 - 0.431	0.430	5
0.432 - 0.434	0.433	7
0.435 - 0.437	0.436	11
0.438 - 0.440	0.439	8
0.441 - 0.443	0.442	4
0.444 - 0.446	0.445	2
		40

Construye:

a) Una distribución de frecuencias relativas o porcentuales.
b) Un histograma de frecuencias relativas.
c) Un polígono de frecuencias relativas.
d) Una distribución de frecuencias acumuladas y distribución de frecuencias acumuladas porcentual.
e) Una ojiva y una ojiva porcentual.

Solución

a) La distribución de frecuencia relativa o porcentual se obtiene dividiendo las frecuencias de cada clase entre la frecuencia total. El resultado se puede expresar como porcentaje; por ejemplo, para la primera frecuencia de clase, tenemos:

$$\begin{matrix}\text{Frecuencia relativa}\\ \text{de una clase (FRC)}\end{matrix} = \frac{\text{Frecuencia de clase}}{\text{Frecuencia total}}\ (100\%)$$

$$(\text{FRC}) = \frac{1}{40}(100\%) = 2.5\%$$

Para la segunda frecuencia de clase, tenemos:

$$(\text{FRC}) = \frac{2}{40}(100\%) = 5\%$$

Para la tercera frecuencia de clase, tenemos:

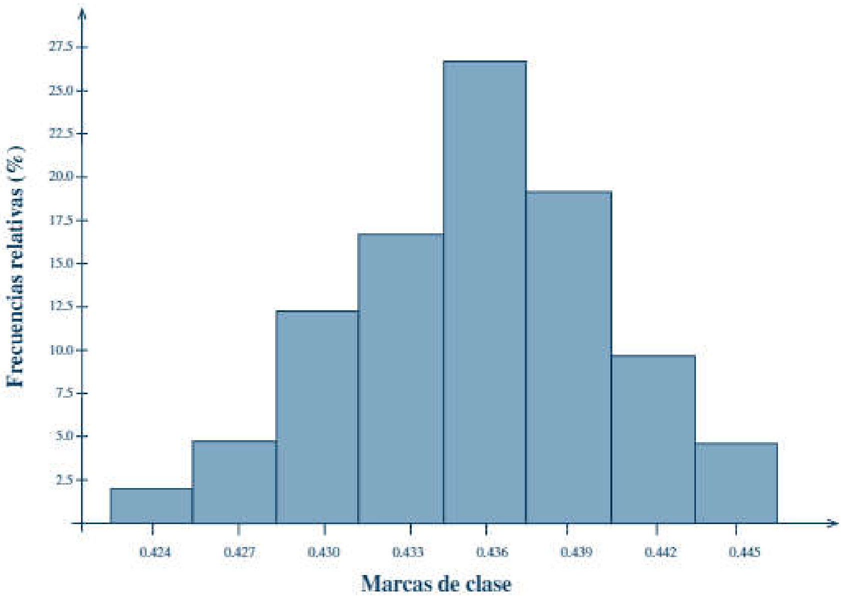
$$(\text{FRC}) = \frac{5}{40}(100\%) = 12.5\%$$

Aplicando el mismo procedimiento para las demás frecuencias de clase, se obtiene la distribución de frecuencia relativa o porcentual.

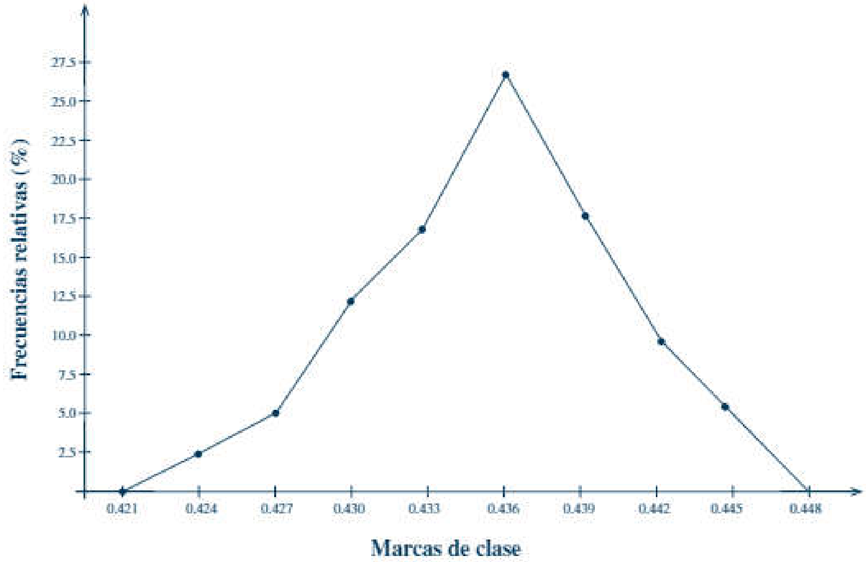
Tabla de distribución de frecuencia relativa o porcentual

Intervalos de clase	Marca de clase	Frecuencia de clase	Frecuencia relativa de clase (FRC %)
0.423 - 0.425	0.424	1	2.5%
0.426 - 0.428	0.427	2	5.0%
0.429 - 0.431	0.430	5	12.5%
0.432 - 0.434	0.433	7	17.5%
0.435 - 0.437	0.436	11	27.5%
0.438 - 0.440	0.439	8	20.0%
0.441 - 0.443	0.442	4	10.0%
0.444 - 0.446	0.445	2	5.0%
		40	100%

b) Para construir el histograma de frecuencias relativas, sólo es necesario marcar sobre la escala vertical las frecuencias relativas, es decir:



c) Para construir el polígono de frecuencias relativas, sólo es necesario marcar sobre la escala vertical las frecuencias relativas, es decir:



d) La distribución de frecuencias acumuladas se obtiene al sustituir las frecuencias de las clases por las frecuencias acumuladas.

La frecuencia acumulada hasta el primer intervalo de clase es la frecuencia total de todos los valores menores que el límite real superior de dicho intervalo de clase, es decir, su valor es cero; para el segundo intervalo de clase es $(0 + 1 = 1)$; para el tercer intervalo de clase es $(0 + 1 + 2 = 3)$; para el cuarto intervalo de clase es $(0 + 1 + 2 + 5 = 8)$, etcétera.

La frecuencia relativa acumulada o porcentual acumulada se obtiene al dividir la frecuencia acumulada entre la frecuencia total.

$$\begin{matrix}\text{Frecuencia acumulada}\\ \text{porcentual (FAP)}\end{matrix} = \frac{\text{Frecuencia acumulada}}{\text{Frecuencia total}}(100\%)$$

La frecuencia acumulada porcentual de la primera frecuencia acumulada es:

$$(\text{FAP}) = \frac{0}{40}(100\%) = 0.0\%$$

Para la segunda frecuencia acumulada es:

$$(\text{FAP}) = \frac{1}{40}(100\%) = 2.5\%$$

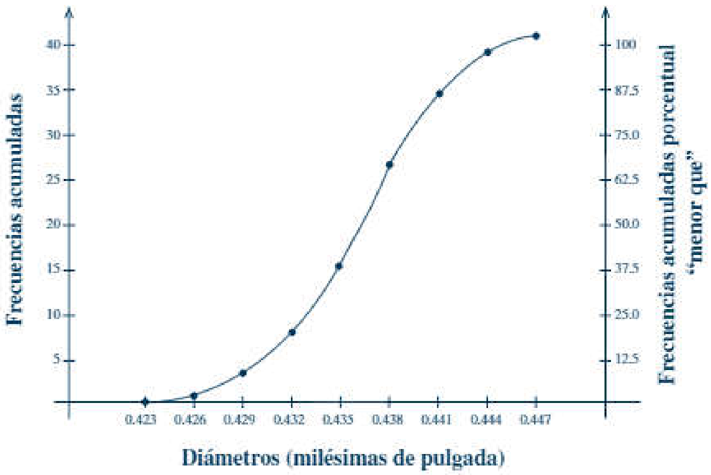
Para la tercera frecuencia acumulada es:

$$(\text{FAP}) = \frac{3}{40}(100\%) = 7.5\%$$

Tabla de distribución de frecuencias acumuladas y distribución de frecuencias acumuladas porcentual

Diámetros	Frecuencias acumuladas	Frecuencias acumuladas porcentual
Menor que 0.423	0	0.0
Menor que 0.426	1	2.5
Menor que 0.429	3	7.5
Menor que 0.432	8	20.0
Menor que 0.435	15	37.5
Menor que 0.438	26	65.0
Menor que 0.441	34	85.0
Menor que 0.444	38	95.0
Menor que 0.447	40	100.0

e) Para construir la ojiva o polígono de frecuencias acumuladas, sólo es necesario marcar sobre la escala vertical izquierda las frecuencias acumuladas porcentual; por lo anterior resulta:



2 La siguiente tabla de distribuciones de frecuencias registra las edades de 100 jefes de familia de una comunidad indígena del norte de Chihuahua, México; a partir de los datos construye:

a) Una distribución de frecuencias acumuladas mayor que y una porcentual o más.
b) Una ojiva mayor que y una ojiva porcentual mayor que.

Intervalo de clase (edad)	Frecuencia de clase
52 - 54	5
55 - 57	12
58 - 60	31
61 - 63	24
64 - 66	17
67 - 69	9
70 - 72	2
	100

Solución

a) La distribución de frecuencias acumuladas mayor que se obtiene al sustituir las frecuencias de las clases por las frecuencias acumuladas mayor que, por ejemplo:

La frecuencia acumulada mayor que para el primer intervalo de clase es la frecuencia total de todos los valores mayores que el límite real superior de dicho intervalo de clase, es decir, su valor es 100; para el segundo intervalo de clase es $(100 - 5 = 95)$; para el tercer intervalo de clase es $(100 - 5 - 12 = 83)$; para el cuarto intervalo de clase es $(100 - 5 - 12 - 31 = 52)$; etcétera.

La frecuencia relativa acumulada o porcentual acumulada mayor que, se obtiene al dividir la frecuencia acumulada mayor que entre la frecuencia total.

$$\begin{matrix}\text{Frecuencia acumulada} \\ \text{porcentual (mayor que)}\end{matrix} = \frac{\text{Frecuencia acumulada (mayor que)}}{\text{Frecuencia total}}(100\%)$$

La frecuencia acumulada porcentual mayor que, de la primera frecuencia acumulada mayor que es:

$$\text{F. A. P. (mayor que)} = \frac{100}{100}(100\%) = 100\%$$

Para la segunda frecuencia acumulada mayor que es:

$$\text{F.A.P. (mayor que)} = \frac{95}{100}(100\%) = 95\%$$

Para la tercera frecuencia acumulada mayor que es:

$$\text{F.A.P. (mayor que)} = \frac{83}{100}(100\%) = 83\%$$

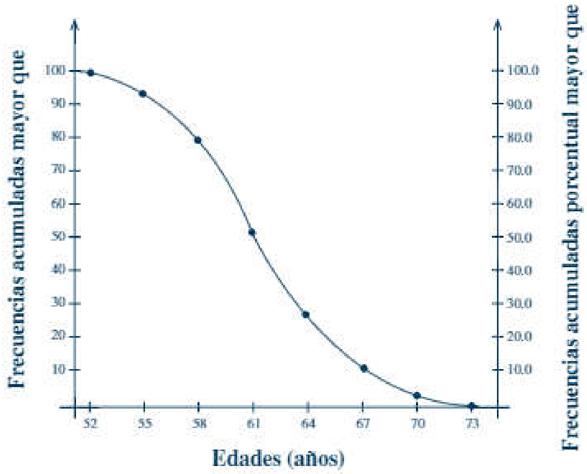
Por todo lo anterior, resulta:

Edades	Frecuencias acumuladas mayor que	Frecuencias acumuladas porcentual mayor que
mayor que 52	100	100.0
mayor que 55	95	95.0
mayor que 58	83	83.0
mayor que 61	52	52.0
mayor que 64	28	28.0
mayor que 67	11	11.0
mayor que 70	2	2.0
mayor que 73	0	0.0

b) Para construir la ojiva o polígono de frecuencias acumuladas mayor que, sólo es necesario marcar sobre la escala vertical izquierda las frecuencias acumuladas mayor que.

Para construir la ojiva porcentual o polígono de frecuencias acumuladas porcentuales mayor que, sólo es necesario marcar sobre la escala vertical derecha las frecuencias acumuladas porcentual mayor que.

Por todo lo anterior, resulta:



Curvas de frecuencias (ojivas suavizadas)

Si consideramos que un conjunto de datos pertenecen a una muestra que se extrae de una población excesivamente grande, como pueden realizarse muchas y muy variadas observaciones, es necesario que teóricamente para datos continuos (que resultan de mediciones) se seleccionen intervalos de clase muy pequeños y todavía tener un número adecuado de observaciones dentro de cada clase.

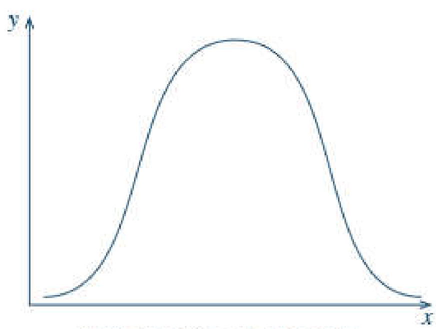
Con base en lo anterior, el polígono de frecuencias o el de frecuencias relativas para poblaciones excesivamente grandes se puede formar por muchos pequeños segmentos rectos que se asemejan a una curva la cual se denomina **curvas de frecuencias o curvas de frecuencias relativas**, respectivamente.

Es de esperarse que tales curvas teóricas resulten de la suavización de los polígonos de frecuencias o de los polígonos de frecuencias relativas de la muestra, la semejanza de la curva es más exacta conforme aumenta el tamaño de la muestra, por lo que se denomina **polígono de frecuencias suavizadas**.

Tipos de curvas de frecuencia

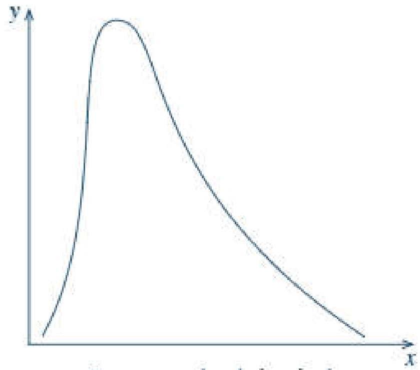
Las curvas de frecuencia se presentan con determinadas formas características que las distinguen entre sí; por ejemplo:

a) Las curvas de frecuencia denominadas **simétricas o bien formadas** se distinguen por el hecho de que las observaciones que equidistan del máximo central tienen la misma frecuencia; por ejemplo la **curva normal** pertenece a este tipo de curvas.

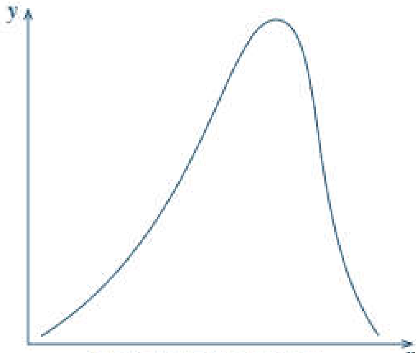


Curva simétrica o bien formada

b) Las curvas de frecuencia denominadas **moderadamente asimétricas o sesgadas** se distinguen por el hecho de que la **cola** de la curva a un lado del máximo central es mayor que al otro lado. Si la **cola mayor** se muestra a la derecha de la curva, se llama **sesgada a la derecha** o de **sesgo positivo**; si la **cola mayor** se muestra a la izquierda de la curva, se llama **sesgada la izquierda** o de **sesgo negativo**.

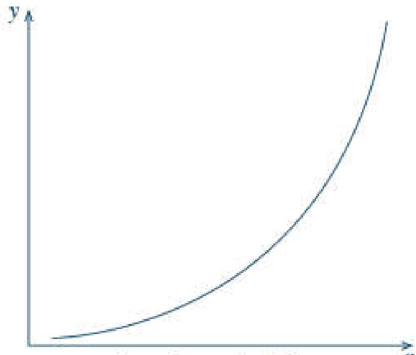


Curva sesgada a la izquierda
(sesgo negativo)

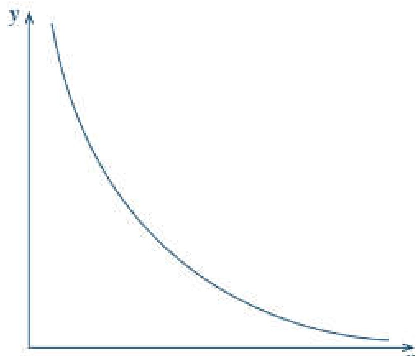


Curva sesgada la derecha
(sesgo positivo)

c) Las curvas de frecuencia denominadas **forma de J** *y* **forma de J invertida** se distinguen por el hecho de que el máximo se muestra en un extremo.

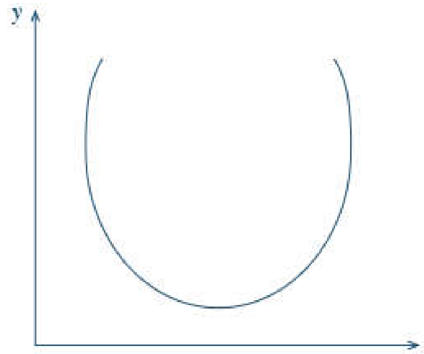


Curva forma de "J"



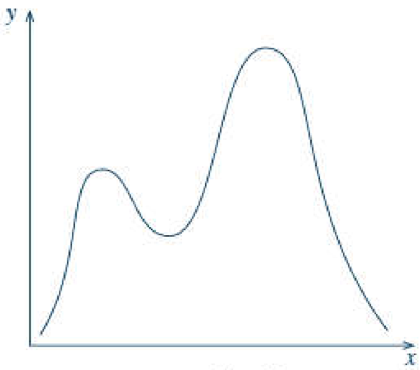
Curva forma de "J" invertida

d) Las curvas de frecuencia denominadas **forma de U** se distinguen por el hecho de que tienen el máximo en ambos extremos.



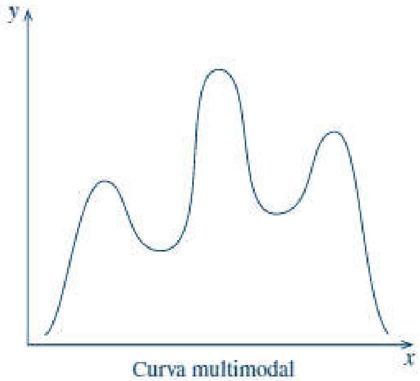
Curva forma de "U"

e) Las curvas de frecuencia del tipo **bimodal** se distinguen por el hecho de que tienen dos máximos.



Curva bimodal

f) Las curvas de frecuencia del tipo **multimodal** se distinguen por el hecho de que tienen más de dos máximos.



Curva multimodal

EJEMPLOS

Ejemplos

1 La siguiente tabla de distribuciones de frecuencias, registran los pesos en kilogramos de 1 000 estudiantes del primer semestre de una especialidad del área físico-matemáticas en la facultad de ingeniería de una Universidad en México; las cuales realmente constituyen una muestra que forma parte de un total de 12 748 estudiantes en dicha facultad. Con la información anterior:

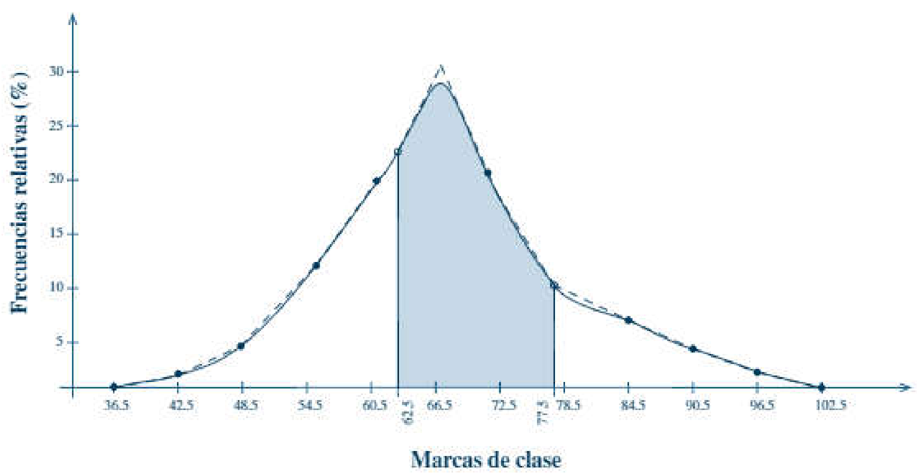
a) Construir un polígono de frecuencias porcentual suavizado (curva de frecuencias).
b) Construir una ojiva porcentual **menor que** suavizada.
c) Estimar el número de los estudiantes en la facultad que tienen peso entre 63 y 77 kilogramos. ¿Qué hipótesis deberá hacerse?
d) De acuerdo con los resultados, ¿pueden utilizarse para estimar la proporción de los estudiantes en la República Mexicana que tengan peso entre 63 y 77 kilogramos?

Intervalos de clase	Frecuencia de clase
40 - 45	19
46 - 51	48
52 - 57	87
58 - 63	156
64 - 69	294
70 - 75	189
76 - 81	93
82 - 87	72
83 - 93	35
94 - 99	7
	1000

Solución

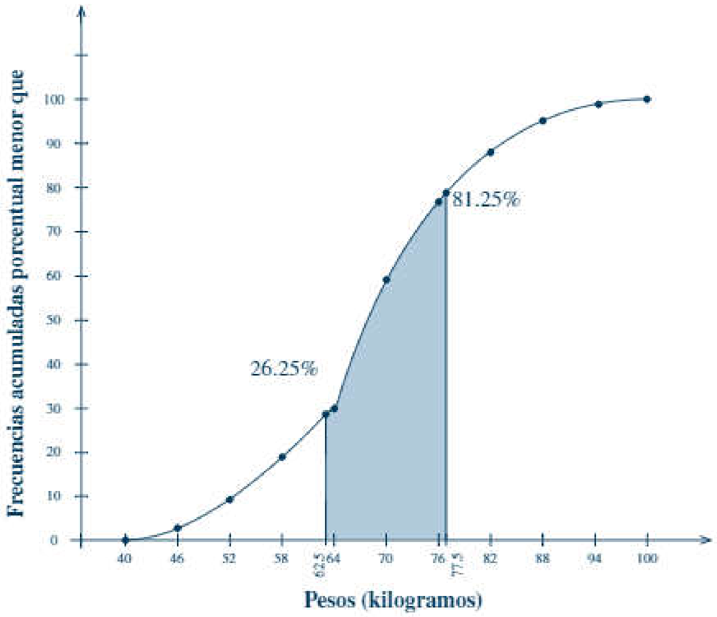
a) Para construir un polígono de frecuencias porcentual suavizado (curva de frecuencias), es necesario obtener primeramente la tabla de distribución de frecuencias relativas o porcentual; en segundo lugar construir la curva de frecuencias y por último suavizarla mediante curvas suavizadas; resultando:

Intervalos de clase	Marca de clase (*x*)	Frecuencia de clase	Frecuencia relativa de clase (FRC %)
40 - 45	42.5	19	1.9%
46 - 51	48.5	48	4.8%
52 - 57	54.5	87	8.7%
58 - 63	60.5	156	15.6%
64 - 69	66.5	294	29.4%
70 - 75	72.5	189	18.9%
76 - 81	78.5	93	9.3%
82 - 87	84.5	72	7.2%
88 - 93	90.5	35	3.5%
94 - 99	96.6	7	0.7%
		1000	100.0%



b) Para construir una ojiva porcentual **menor que** suavizada, es necesario obtener primeramente la falta de distribución de frecuencias acumuladas menor que; en segunda parte construir la ojiva porcentual y por último suavizarla mediante curvas suavizadas; resultando:

Pesos	Frecuencias acumuladas	Frecuencias acumuladas porcentual
Menor que 40	0	0.0
Menor que 46	19	1.9
Menor que 52	67	6.7
Menor que 58	154	15.4
Menor que 64	310	31.0
Menor que 70	604	60.4
Menor que 76	793	79.3
Menor que 82	886	88.6
Menor que 88	958	95.8
Menor que 94	993	99.3
Menor que 100	1 000	100.0



Si la muestra de 1 000 estudiantes es representativa de la población 12 748 estudiantes, las curvas suavizadas de los incisos *a*) y *b*) se suponen como la curva de frecuencias porcentual y la ojiva porcentual para dicha población, respectivamente. Lo anterior es correcto solamente si la muestra se extrajo al azar, es decir, si todo estudiante tuvo la misma probabilidad que otro cualquiera de pertenecer a la muestra.

e) Puesto que los pesos entre 63 y 77 kilogramos registrados con aproximación de kilogramo realmente representan pesos entre 62.5 y 77.5 kilogramos, el porcentaje de estudiantes de la población que tiene estos pesos puede determinarse al dividir el área sombreada del polígono de frecuencias porcentual suavizado por el área total limitado por la curva suavizada y el eje *x*. Otra forma más simple de determinar el porcentaje de estudiantes de la población que tienen pesos entre 63 y 77 kilogramos es utilizando la ojiva porcentual menor que suavizada, en la que observamos:

Porcentaje de estudiantes con pesos menores que 77.5 kilogramos = 81.25%
Porcentaje de estudiantes con pesos menores que 62.5 kilogramos = 26.25%

De lo anterior, el porcentaje de estudiantes con pesos comprendidos entre 62.5 y 77.5 kilogramos es:

$$81.25\% - 26.25\% = 55\%$$

Por lo tanto, el número de estudiantes de la facultad que tienen peso entre 63 y 77 kilogramos es de:

$$55\% \text{ de } 12\,748 = 7\,011.4 \text{ estudiantes}$$

Otra manera de explicar lo anterior, es mencionar que la **probabilidad** de que una persona elegida al azar de entre los 12 748 estudiantes tenga un peso entre 63 y 77 kilogramos es de 55% o 0.55 o 55 de cada 100.

Por su relación con la probabilidad, las curvas de frecuencia relativa se conocen también con el nombre de **curvas de probabilidad o distribuciones de probabilidad**.

d) Se podría considerar la proporción encontrada del 55% aunque con menor garantía que el caso anterior, solamente si se tuviese la certeza de que la muestra de los 1 000 estudiantes del total de la población de la facultad en la República Mexicana es realmente una muestra al azar. Lo anterior es algo improbable debido a varias razones, por ejemplo:

a) Algunos estudiantes pueden no haber alcanzado su peso máximo, debido a su desarrollo biológico.
b) La generación actual puede tender a tener menor o mayor peso que el de sus padres (herencia biológica).

EJERCICIO 7

I. Realiza lo que se indica en cada caso.

1. Explica la finalidad de representar información por medio de gráficos.
2. ¿Qué y cuáles son las representaciones gráficas de datos estadísticos?
3. ¿Qué es un gráfico o diagrama lineal?
4. ¿Qué datos se agrupan en los gráficos o curvas de sucesiones cronológicas?
5. ¿Qué datos se agrupan en los gráficos o curvas de distribuciones de frecuencias?
6. Cita las recomendaciones que se requiera para trazar un gráfico lineal.
7. ¿Qué es un gráfico o diagrama de barras?
8. ¿Qué datos se agrupan en los gráficos o diagramas de barras?
9. Cita las recomendaciones que se requieren para trazar un gráfico o diagrama de barras.
10. ¿Qué es un pictograma o pictógrafo?
11. Cita algunas de las características de por qué se utilizan los pictogramas o pictógrafos en la estadística.
12. ¿Qué es un gráfico circular o diagrama de pastel?
13. Cita algunas características de por qué se utilizan los gráficos circulares o diagramas de pastel en la estadística.
14. ¿Cuál es el objetivo principal de graficar las distribuciones de frecuencia?
15. Cita los medios más comunes de mostrar gráficamente las distribuciones de frecuencias.
16. ¿Qué es un histograma?
17. Explica cómo se construye un histograma.
18. Conceptualmente cuáles son las diferencias entre un histograma y un diagrama de barras.

19. ¿Qué es un polígono de frecuencias?
20. Explica cómo se construye un polígono de frecuencias.
21. Explica cómo están relacionados un histograma y un polígono de frecuencias.
22. ¿Qué es una distribución de frecuencia relativa?
23. ¿Cómo se obtiene la frecuencia relativa de una clase?
24. Explica cómo están relacionados los gráficos de distribuciones de frecuencia relativa con los histogramas y polígonos de frecuencias.
25. ¿Qué es una distribución de frecuencia acumulada u ojiva?
26. ¿Cómo se obtiene la frecuencia acumulada?
27. ¿Cómo se le llama al diagrama que representa las frecuencias acumuladas **menor o mayor que**?
28. ¿Qué es una distribución de frecuencia relativa acumulada u ojiva porcentual?
29. ¿Qué es una curva de frecuencia u ojiva suavizada?

II. Con ayuda de tu profesor desarrolla los siguientes ejercicios.

Escribe los números correspondientes

Competencias genéricas

Competencias disciplinares

1. Las exportaciones de café mexicano durante la década de los 80's, fueron en millones de dólares: 1980: $1 689.4; 1981: $1 705.2; 1982: $1 753.6; 1983: 1 792.8; 1984: $1 838.5; 1985: $1 867.9; 1986: $1 942.3; 1987: $1 826.1; 1988: $1 749.7; 1989: $1 652.8; construye el gráfico lineal correspondiente.
2. El consumo de gasolina en la zona fronteriza del norte de Tamaulipas en el semestre enero-junio de 1991 en millones de litros es: enero, 1 274 400; febrero, 1 413 100; marzo, 1 362 80; abril, 1 562 700; mayo 1 422 500; junio 1 252 700; construye el gráfico lineal correspondiente.
3. La población de los Estados Unidos Mexicanos en millones, a partir de 1950 a 2010, de acuerdo con los censos nacionales de población y vivienda, son:

Año	1950	1960	1970	1980	1990	2000	2010
Población en millones	16.6	19.7	25.8	34.9	48.2	66.8	86.3

Construye el gráfico lineal correspondiente.

4. Dada la siguiente tabla que proporciona la oficina de censos nacionales de población y vivienda de la República Mexicana.

Año	Nacimientos	Defunciones (en millones)	Tasa de crecimiento (en millones)
1950	0.82	0.44	0.38
1960	0.87	0.46	0.41
1970	1.18	0.41	0.76
1980	1.60	0.40	1.21
1990	2.13	0.49	1.65
2000	2.43	0.43	1.99

Construye:

a) Un gráfico de barras para cada columna de información.
b) Un gráfico de barras compuesto para nacimientos y defunciones.
c) Un gráfico de barras compuesto para nacimientos y tasa de crecimiento.
d) Un gráfico de barras compuesto para defunciones y tasa de crecimiento.

5. La siguiente tabla registra la producción de autos en una planta al norte del pais en el período (1999-2006); construye el pictograma correspondiente.

Año	Producción
1999	105 715
2000	107 275
2001	109 829
2002	110 626
2003	112 844
2004	146 234
2005	154 139
2006	160 551

6. Construye el pictograma correspondiente a partir de la siguiente tabla que registra el número de los diferentes tipos de aviones utilizados por la fuerza aérea norteamericana en la guerra de Vietnam de 1960 a 1969:

Año	Número de aviones
1960	1 370
1961	1 185
1962	938
1963	812
1964	1 467
1965	1 791
1966	2 156
1967	2 379
1968	1 643
1969	1 224

7. Dada la siguiente tabla de distribución de frecuencias de los índices de accidentes semanales en 2008 de la *Exxon Oil Company Of Houston*, construye:

a) Su histograma.

b) Su polígono de frecuencias.

c) Su histograma y polígono de frecuencias en una misma representación gráfica.

Números de accidentes por 1000 horas-hombre (intervalos de clase)	Número de semanas Frecuencia de clase
1.5 - 1.9	1
2.0 - 2.4	3
2.5 - 2.9	8
3.0 - 3.4	16
3.5 - 3.9	14
4.0 - 4.4	7
4.5 - 4.9	2
5.0 - 5.4	1
	52

8. Traza el histograma y el polígono de frecuencias para la siguiente tabla de distribución de frecuencias.

Intervalos de clase	Frecuencia de clase
10 - 19	4
20 - 29	6
30 - 39	10
40 - 49	18
50 - 59	14
60 - 69	9
70 - 79	5
80 - 89	2
	68

9. La siguiente tabla registra la distribución de frecuencias de una muestra de 325 lámparas incandescentes de 60 watts sometidas a pruebas críticas de control de calidad; construye su histograma y polígono de frecuencias en una misma representación gráfica.

Duración en horas	Número de lámparas
820 - 919	22
920 - 1019	36
1020 - 1119	58
1120 - 1219	83
1220 - 1319	62
1320 - 1419	47
1420 - 1519	17
	325

10. La siguiente tabla de distribuciones de frecuencias registran los pesos de 40 estudiantes de la Universidad Autonoma de Tamaulipas en unidades de libras, construye:

a) Una distribución de frecuencias acumuladas **mayor que** y una porcentual **mayor que**

b) Una ojiva **mayor que** y una ojiva porcentual **mayor que.**

Intervalos de clase	Frecuencias de clase
116 - 124	3
125 - 133	5
134 - 142	9
143 - 131	12
152 - 160	6
161 - 169	3
170 - 178	2
	40

Escribe los números correspondientes

Competencias genéricas

Competencias disciplinares

11. Dada 1a siguiente tabla de distribución de frecuencias:

Intervalos de clase	Frecuencia de clase
50 - 52	2
53 - 55	6
56 - 58	17
59 - 61	32
62 - 64	25
65 - 67	18
68 - 70	9
	109

Construye:

a) Una distribución de frecuencias relativas o porcentual.

b) Un histograma de frecuencias relativas.

c) Un polígono de frecuencias relativas.

d) Una distribución de frecuencias acumuladas y distribución de frecuencias acumuladas porcentual.

e) Una ojiva y una ojiva porcentual.

12. Dada la siguiente tabla de distribuciones de frecuencias, construye:

a) Una distribución de frecuencias acumuladas mayor que y una porcentual mayor que.

b) Una ojiva mayor que y una ojiva porcentual mayor que.

Intervalos de clase	Frecuencia de clase
20 - 29	3
30 - 39	5
40 - 49	9
50 - 59	11
60 - 69	23
70 - 79	38
80 - 89	29
90 - 99	18
	136

13. La siguiente tabla de distribuciones de frecuencias, registran la duración de 600 tubos de televisión (cinescopios) de *Zenith Corporation*; los cuales realmente constituyen una muestra a de los 7 910 tubos de televisión de dicha empresa. Con la información anterior:

a) Construye un polígono de frecuencias porcentual suavizado (curva de frecuencias).

b) Construye una ojiva porcentual menor que suavizada.

c) Estima el número de tubos que tienen duración entre 6 500 y 8 000 horas; explica tu respuesta.

d) De acuerdo con los resultados, ¿pueden utilizarse para estimar la proporción de tubos en la población de 7 910 tengan duración entre 6 500 y 8 000 horas?

e) Estima el porcentaje de tubos que durará menos de 4 500 horas.

f) Estima el porcentaje de tubos que durará más de 9 000 horas.

Intervalos de clase	Frecuencia de clase
3 000 - 3 999	21
4 000 - 4 999	53
5 000 - 5 999	66
6 000 - 6 999	92
7 000 - 7 999	135
8 000 - 8 999	74
9 000 - 9 999	69
10 000 - 10 999	57
11 000 - 11 999	33
	600

14. La siguiente tabla de distribución registra el salario mensual de 100 varones censados en una avenida principal de Los Ángeles, California; contesta:

a) ¿Cuál es el tamaño o anchura del segundo intervalo de clase?, ¿y del quinto?

b) ¿Cuántos intervalos de clase diferentes hay?

c) ¿Cuántos intervalos de clase abiertos hay?

d) ¿Qué porcentaje de varones ganaron más de $8 000?, ¿y menos de $3 000?

e) ¿Qué porcentaje ganó entre $2 500 y $6 500?

Salarios (dólares)	Frecuencia (varones)
Menos de 800	31
800 - 999	24
1 000 - 1 999	18
2 000 - 2 999	11
3 000 - 5 999	8
6 000 - 8 999	5
9 000 y más	3
	100

15. La siguiente tabla de distribuciones de frecuencias registra las longitudes en centímetros que en una semana tienen 100 plantas de frijol:

Longitudes	Números de plantas
5.4 - 5.7	7
5.8 - 6.1	16
6.2 - 6.5	21
6.6 - 6.9	29
7.0 - 7.3	18
7.4 - 7.7	9
	100

Construye:

a) Una distribución de frecuencias relativas o porcentual.

b) Un histograma y polígono de frecuencias relativas.

c) Una distribución de frecuencias acumuladas mayor que y una porcentual mayor que.

d) Una ojiva porcentual menor que.

16. La siguiente tabla de distribución de frecuencias registra el número de canastos de manzanas que en una jornada (12 horas) levantan 200 laboristas.

Canastos	Número de laboristas
24 - 28	4
29 - 33	13
34 - 38	19
39 - 43	28
44 - 48	46
49 - 53	54
54 - 58	25
59 - 63	11
	200

Construye:

a) Una distribución de frecuencias relativas o porcentual.

b) Un histograma y polígono de frecuencias relativas.

c) Una distribución de frecuencias acumuladas mayor que y una porcentual mayor que.

d) Una ojiva porcentual menor que.

Verifica tus resultados en la sección de respuestas.

TEMA 2 Medidas de tendencia central

Propósito del tema

Que el estudiante:

- Aprenda e incorpore en situaciones reales las medidas de tendencia central: media, mediana y moda.
- Interprete y analice los resultados que obtiene en la resolución de cada uno de los problemas, y los aplique en situaciones reales.
- Ordene un conjunto de datos mediante una tabulación y su gráfica para un posterior análisis.
- Determine el promedio y variación de una serie de datos para su análisis en la toma de decisiones.

Competencias disciplinares

3. Explica e interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos o situaciones reales.
4. Argumenta la solución obtenida de un problema con métodos numéricos, gráficos, analíticos o variacionales, mediante lenguaje verbal, matemático y el uso de las tecnologías de la información y la comunicación.
5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.
8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

Contenidos que aborda el tema

Contenidos conceptuales

- Concepto de media, mediana y moda.
- Relación empírica entre media, mediana y moda.
- Concepto de las medias aritmética, geométrica y armónica.
- Relación entre las medias aritmética, geométrica y armónica.
- Raíz cuadrada del cuadrado de la media.
- Concepto de cuantil y percentil
- Cálculo e identificación de los cuantiles.

Contenidos procedimentales

- Expresa ideas y conceptos mediante representaciones lingüísticas, matemáticas o gráficas.
- Sigue procedimientos de manera reflexiva, comprendiendo como cada uno de sus pasos contribuye al alcance de un objetivo.
- Resolverá problemas utilizando elementos básicos de la estadística.
- Usará diversas estrategias de resolución de problemas.

Contenidos actitudinales

- Expresará ideas utilizando la terminología relativa a ecuaciones de primer grado.
- Aprenderá a valorar el trabajo de sus compañeros al resolver problemas.
- Expresará ideas utilizando modelos lineales.
- Colaborará en equipo y respetará a sus compañeros al resolver problemas.

Medidas de tendencia central

Un valor que es simbólico y representativo de un conjunto de datos se denomina **promedio**. Se hace notar que en un conjunto de datos ordenados de acuerdo con su magnitud, el promedio siempre tiende a situarse en el centro de dicho conjunto, razón por la cual los promedios se denominan también **medidas de centralización** o **de tendencia central**.

De los diferentes tipos de medidas de tendencia central tenemos que entre las más comunes se encuentran la media aritmética o media, la mediana, la moda, la media geométrica y la media armónica. Cada una de ellas se utiliza para describir y establecer comparaciones cuantitativas entre distribuciones de frecuencia.

Media aritmética

Es el promedio más utilizado y por lo general se denomina **media**. La media aritmética o media de un conjunto de elementos se define como la suma de los valores de estos elementos dividido entre el número total de ellos.

Si se tiene un conjunto $N(X_1, X_2, X_3, \ldots, X_N)$ de elementos, la media aritmética o media que se simboliza por $\overline{X}$ (léase $\overline{X}$ barra), matemáticamente se determina por la ecuación:

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_N}{N} = \frac{\sum_{d=1}^{N} X_N}{N} = \frac{\sum X}{N}$$

donde:

$\overline{X}$ = Media aritmética o media.
X = Valor de cada elemento.
N = Número total de elemento.

EJEMPLOS

Ejemplos

1 Determina la media aritmética de los números 18, 13, 15, 22 y 20.

Solución

Al sustituir los datos dados en la ecuación de la media aritmética o media resulta:

$$\overline{X} = \frac{\sum X}{N} = \frac{18+13+15+22+20}{5} = \frac{88}{5} = 17.6$$

La media aritmética o media de los números dados es 17.6.

2 Los precios del centenario de oro, en pesos, registrados en las seis primeras semanas del año 1992 fueron: 1 365 000, 1 370 000, 1 358 500, 1 330 000, 1 322 500, 1 320 000; determina la media aritmética de los precios.

Solución

Al sustituir los datos dados en la ecuación de la media aritmética o media resulta:

$$\overline{X} = \frac{\sum X}{N} = \frac{1\ 365\ 000 + 1\ 370\ 000 + 1\ 358\ 500 + 1\ 330\ 000 + 1\ 322\ 500 + 1\ 320\ 000}{6} = \frac{8\ 066\ 000}{6} = 1\ 344\ 333.3$$

La media aritmética o media de los precios es \$1 344 333.3.

3 El equipo de basquetbol de una preparatoria en la ciudad de Puebla tiene 10 jugadores (titulares y reservas) con las siguientes estaturas en metros: 1.89, 1.94, 1.81, 1.86, 1.76, 1.87, 1.91, 1.79, 1.84 y 1.82; determina la media aritmética de las estaturas.

Solución

Al sustituir los datos dados en la ecuación de la media aritmética o media resulta:

$$\overline{X} = \frac{\sum X}{N} = \frac{1.89+1.94+1.81+1.86+1.76+1.87+1.91+1.79+1.84+1.82}{10} \qquad \overline{X} = \frac{18.49}{10} = 1.849$$

La media aritmética o media de las estaturas es 1.849 metros.

Media aritmética con frecuencias

Si se tiene un conjunto de elementos ($X_1, X_2, X_3, \ldots, X_N$) que se presentan con ($f_1, f_2, f_3, \ldots, f_N$) veces de frecuencia, respectivamente, la media aritmética con frecuencia se define como la suma de los productos de cada frecuencia y con el respectivo valor del elemento, dividida entre el número total de casos; matemáticamente se determina por la ecuación:

$$\overline{X} = \frac{f_1 X_1 + f_2 X_2 + f_3 X_3 + \ldots + f_N}{X_N \quad f_1 + f_2 + f_3 + \ldots + f_N} = \frac{\sum_{j=1}^{N} f_1 X_j}{\sum_{j=1}^{N} f_j} = \frac{\sum f X}{\sum f} = \frac{\sum f N}{N}$$

donde:

$\overline{X}$ = Media aritmética con frecuencias.

$f\,X$ = Producto de cada frecuencia por el valor correspondiente de cada elemento, respectivamente.

$N = \sum f$ = Frecuencia total o el número total de casos.

EJEMPLOS

1 Determina la media aritmética de los números 12, 16, 18 y 15 que se presentan con frecuencias 5, 3, 2 y 4, respectivamente.

Solución

Después de sustituir los datos dados en la ecuación de la media aritmética con frecuencias resulta:

$$\overline{X} = \frac{\sum f X}{N} = \frac{(5)(12)+(3)(16)+(2)(18)+(4)(15)}{5+3+2+4} = \frac{60+48+36+60}{14} = \frac{204}{14} = 14.571$$

La media aritmética con frecuencias es 14.571.

2 De un total de 65 individuos, 16 registran una media de peso de 85 kg, 21, una de 77 kg, 12 registran una media de 92 kg y el resto, una de 64 kg; encuentra la media aritmética del peso medio de todos los individuos.

Solución

Al sustituir los datos dados en la ecuación de la media aritmética con frecuencias resulta:

$$\overline{X} = \frac{\sum f X}{N} = \frac{(16)(85)+(20)(77)+(12)(92)+(16)(64)}{16+21+12+16} = \frac{1\,360+1\,617+1\,104+1\,024}{65}$$

$$\overline{X} = \frac{5\,105}{65} = 78.538$$

La media aritmética con frecuencias es 78.538 kg.

Media aritmética ponderada

En algunas ocasiones a los elementos ($X_1, X_2, X_3, \ldots, X_N$) se les asocian ciertos **factores** o **pesos**, ($W_1, W_2, W_3, \ldots, W_N$) que destacan la significación e importancia de cada elemento; en tal situación se tiene una media aritmética ponderada, la cual se define como la suma de los productos del peso y el valor de cada elemento, respectivos, dividido entre el número total de pesos; matemáticamente se determina por la ecuación:

$$\overline{X} = \frac{W_1 X_1 + W_2 X_2 + W_3 X_3 + \ldots + W_N X_N}{W_1 + W_2 + W_3 + \ldots + W_N} = \frac{\sum_{j=1}^{N} W_j X_j}{\sum_{j=1}^{N} W_j} = \frac{\sum W X}{\sum W} = \frac{\sum W X}{N}$$

donde:

$\overline{X}$ = Media aritmética ponderada.
WX = Producto del peso y el valor de cada elemento, respectivamente.
$N = \sum W$ = Número total de pesos.

Debido a su gran similitud, la ecuación de la media aritmética con frecuencias se puede considerar como una ecuación de media aritmética ponderada en la cual los **factores** o **pesos** son precisamente las frecuencias de cada uno de los datos ($f_1, f_2, f_3, \ldots, f_N$).

EJEMPLOS

Ejemplos

1 Las puntuaciones finales de un estudiante en matemáticas, lectura y redacción, física e inglés son, respectivamente, 76, 87, 72 y 84. Si el número de créditos para cada una de estas asignaturas es 10, 6, 8 y 6, determina la media aritmética ponderada de las puntuaciones.

Solución

Al sustituir los datos dados en la ecuación de la media aritmética ponderada resulta:

$$\overline{X} = \frac{\sum W X}{N} = \frac{(10)(76) + (6)(87) + (8)(72) + (6)(84)}{10 + 6 + 8 + 6} = \frac{760 + 522 + 576 + 504}{30}$$

$$\overline{X} = \frac{2\,362}{30} = 78.73$$

La media aritmética ponderada de las puntuaciones es 78.73.

2 Si un examen final de curso se valora como tres veces los exámenes parciales y un estudiante tiene un resultado de examen final de 92 y calificaciones de exámenes parciales de 74 y 86, determina la media aritmética ponderada para la calificación final.

Solución

Después de sustituir los datos dados en la ecuación de la media aritmética ponderada resulta:

$$\overline{X} = \frac{\sum W X}{N} = \frac{(1)(74) + (1)(86) + (3)(92)}{1 + 1 + 3} = \frac{74 + 86 + 276}{5} = \frac{436}{5} = 87.2$$

La media aritmética ponderada para la calificación final es 87.2

Propiedades de la media aritmética

1. La suma algebraica de las desviaciones de un conjunto de elementos numéricos de su media aritmética es cero.

Comprobación

Las desviaciones de los números 7, 4, 9, 13 y 11, cuya media aritmética es 8.8, se determinan por la ecuación $d_j = X_j - \overline{X}$, es decir: $7 - 8.8 = -1.8$; $4 - 8.8 = -4.8$; $9 - 8.8 = 0.2$; $13 - 8.8 = 4.2$; $11 - 8.8 = 2.2$; demuestra que la suma algebraica de las desviaciones es igual a cero.

$$-1.8 - 4.8 + 0.2 + 4.2 + 2.2 = 0$$

2. La suma de los cuadrados de las desviaciones de un conjunto de elementos numéricos (X_j) de cualquier otro elemento numérico denotado por a es mínima solamente si $a = \overline{X}$.

Comprobación

Los números 6, 11, 13, 9 y 17 son el conjunto de elementos (X_j), cuya media aritmética es $\overline{X} = 11.2$; considerando que a es cualquier otro elemento numérico, se tiene que:

a) Si $a = \overline{X}$, la suma de los cuadrados de las desviaciones es:

$$\sum d_j^2 = \sum\left(X_j - \overline{X}\right)^2 = \sum\left(X_j - 11.2\right)^2 = (6-11.2)^2 + (11-11.2)^2 + (13-11.2)^2 + (9-11.2)^2 + (17-11.2)^2$$

$$\sum\left(X_j - 11.2\right)^2 = 27.04 + 0.04 + 3.24 + 4.84 + 33.64$$

$$\sum\left(X_j - 11.2\right)^2 = 68.8$$

b) Si $a > \overline{X}$, la suma de los cuadrados de las desviaciones respecto al valor $a = 12$ es:

$$\sum d_j^2 = \sum(X_j - a)^2 = \sum(X_j - 12)^2 = (6-12)^2 + (11-12)^2 + (13-12)^2 + (9-12)^2 + (17-12)^2$$

$$\sum(X_j - 12)^2 = 36 + 1 + 1 + 9 + 25$$

$$\sum(X_j - 12)^2 = 72$$

c) Si $a < X$, la suma de los cuadrados de las desviaciones respecto al valor $a = 11$ es:

$$\sum d_j^2 = \sum(X_j - a)^2 = \sum(X_j - 11)^2 = (6-11)^2 + (11-11)^2 + (13-11)^2 + (9-11)^2 + (17-11)^2$$

$$\sum(X_j - 11)^2 = 25 + 0 + 4 + 4 + 36$$

$$\sum(X_j - 11)^2 = 69$$

Por lo anterior se observa que cuando $a = \overline{X}$, la suma de los cuadrados de las desviaciones es mínima.

3. Si f_1 números tienen de media m_1, f_2 números tienen de media m_2, f_3 números tienen de media m_3, ..., f_N números tienen de media m_N entonces la media de todos los números se determina por la ecuación:

$$\overline{X} = \frac{f_1 m_1 + f_2 m_2 + f_3 m_3 + \ldots + f_N m_N}{f_1 + f_2 + f_3 + \ldots + f_N} = \frac{\sum fm}{\sum f}$$

La ecuación anterior, también se denomina **media aritmética ponderada** de todas las medias.

La comprobación de esta propiedad está dada en el problema 2 de los ejemplos para la media aritmética con frecuencias.

4. Si a es cualquier supuesta media aritmética (que puede ser cualquier número) y si $d_j = X_j - a$ son las desviaciones de X_j de a, entonces existen otras alternativas para calcular medidas de tendencia central:

a) La media aritmética con fórmula $\overline{X} = \frac{\sum_{j=1}^{N} X_j}{N} = \frac{\sum X}{N}$ también puede calcularse mediante la expresión

$$\overline{X} = a + \frac{\sum_{j=1}^{N} d_j}{N} = a + \frac{\sum d}{N}$$

b) La media aritmética con frecuencias $\overline{X} = \frac{\sum_{j=1}^{N} f_1 X_j}{\sum_{j=1}^{N} f_j} = \frac{\sum f X}{\sum f} = \frac{\sum f X}{N}$ también puede calcularse mediente la expresión

$$\overline{X} = a + \frac{\sum_{j=1}^{N} f_j d_j}{\sum_{j=1}^{N} f_j} = a + \frac{\sum f d}{N}$$

Comprobación

a) Las desviaciones de un conjunto de N elementos numéricos ($X_1, X_2, X_3, \ldots, X_N$) de un número cualquiera dado a son: $d_1 = X_1 - a$, $d_2 = X_2 - a$, $d_3 = X_3 - a$, ..., $d_N = X_N - a$, respectivamente, es decir, establecemos que $d_j = X_j - a$; si despejamos respecto a X_j, tenemos que: $X_j = a + d_j$.

Al sustituir la igualdad anterior en la ecuación de la media aritmética o media resulta:

$$\overline{X} = \frac{\sum_{j=1}^{N} X_j}{N} = \frac{\sum_{j=1}^{N} (a + d_j)}{N} = \frac{\sum_{j=1}^{N} a + \sum_{j=1}^{N} d_j}{N}$$

La sumatoria desde $j = 1$ hasta N de cualquier número constante es igual a N veces el valor de la constante, por tanto, se tiene que

$$\sum_{j=1}^{N} a = Na$$

Usando esta expresión en la fórmula que se obtuvo de media aritmética o media queda:

$$\overline{X} \frac{\sum_{j=1}^{N} a + \sum_{j=1}^{N} d_j}{N} = \frac{Na + \sum_{j=1}^{N} d_j}{N} = a + \frac{\sum_{j=1}^{N} d_j}{N} = a + \frac{\sum d}{N}$$

b) Si un conjunto de N elementos numéricos ($X_1, X_2, X_3, \ldots, X_N$) se presenta con frecuencias ($f_1, f_2, f_3, \ldots, f_N$), respectivamente, y si las desviaciones de los números se establecen por la ecuación $d_N = X_N - A$, de donde se tiene que $X_j = a + d_j$.

Al sustituir la igualdad anterior en la ecuación de la media aritmética con frecuencias resulta:

$$\overline{X} = \frac{\sum_{j=1}^{N} f_j x_j}{\sum_{j=1}^{N} f_j} = \frac{\sum_{j=1}^{N} f_j x_j}{N} = \frac{\sum_{j=1}^{N} f_j (a + d_j)}{N} = \frac{a \sum_{j=1}^{N} f_j + \sum_{j=1}^{N} f_j d_j}{N}$$

Como $\Sigma f_j = N$, por lo que la ecuación se simplifica como:

$$\overline{X} = \frac{aN + \sum_{j=1}^{N} f_j d_j}{N} = a + \frac{\sum_{j=1}^{N} f_j d_j}{N} = a + \frac{\sum fd}{N}$$

Otras propiedades de la media aritmética

5. La media aritmética siempre se puede determinar para cualquier conjunto numérico.
6. La media aritmética es única, es decir, existe una y sólo una media aritmética para un conjunto dado de elementos numéricos.
7. Si cambia algún valor del conjunto de elementos numéricos, entonces también cambia la media aritmética.
8. Si sumamos una constante de cada valor del conjunto de elementos dado, la media aritmética aumentará por la misma cantidad; si se multiplica o divide sucede lo mismo, es decir, la media aritmética resultará multiplicada o dividida por dicha constante.

Media aritmética para datos agrupados

Cuando los datos se presentan agrupados mediante una distribución de frecuencias, todos los valores caen dentro de los intervalos de clase dados que se consideran coincidentes con las marcas de clase o puntos medios de cada intervalo, es decir, las ponderaciones son las frecuencias y las marcas o puntos medios de clase son los valores que se ponderan.

Al determinar la media aritmética para datos agrupados, su resultado se aproximará mucho al resultado que se obtiene para datos no agrupados. El resultado de la media aritmética no será suficientemente aproximado si la distribución de frecuencias para datos agrupados es muy irregular, demasiado asimétrica o presenta imperfecciones.

Los métodos para calcular la media aritmética a partir de datos agrupados son los siguientes.

Método largo. En una distribución de frecuencias para datos agrupados, la media aritmética por el método largo se determina al multiplicar las distintas marcas o puntos medios de clase por sus respectivas frecuencias de clase; se suman los productos y el resultado se divide entre el número total de frecuencias. Matemáticamente se representa:

$$\overline{X} = \frac{\sum_{j=1}^{N} f_j x_j}{\sum_{j=1}^{N} f_j} = \frac{\sum f x}{\sum f} = \frac{\sum f x}{N}$$

donde:

$\overline{X}$ = Media aritmética.
X_j = Marcas o puntos medios de clases.
f_j = Frecuencia total de clase.
$N = \Sigma f$ = Número total de frecuencias.

Método corto. En una distribución de frecuencias para datos agrupados, la media aritmética por el método corto se determina al considerar una marca o punto medio de clase cualquiera, a la que se le aumenta la sumatoria del producto de las frecuencias de clase por las desviaciones de las marcas o puntos medios de clase de dicha marca o punto medio de clase seleccionado, dividido entre el número total de frecuencias. Matemáticamente se representa:

$$\overline{X} = a + \frac{\sum_{j=1}^{N} f_j d_j}{\sum_{j=1}^{N} f_j} = a + \frac{\sum f d}{\sum f} = \frac{\sum f d}{N}$$

donde:

$\overline{X}$ = Media aritmética.
A = Una marca o punto medio de clase cualquiera (media supuesta).
f_j = Frecuencias de clase.
$d_j = Xj - A$ = Las desviaciones de las marcas o puntos medios de clase de la marca o punto medio de clase seleccionado.
$N = \Sigma f$ = Número total de frecuencias.

Método clave. En una distribución de frecuencias para datos agrupados, la media aritmética por el método clave se determina cuando los intervalos de clase tienen igual tamaño C, las desviaciones $d_j = X_j - A$ pueden expresarse como Cu_j, donde u_j puede tomar el valor de un número entero positivo, negativo o cero, es decir, $u_j = 0, \pm 1, \pm 2, \pm 3, \ldots$, etcétera; entonces la ecuación de la media aritmética de frecuencias con desviaciones o fórmula del método corto se transforma en:

$$\overline{X} = A + \left(\frac{\sum_{j=1}^{N} f_j u_j}{N} \right) C = A + \left(\frac{\sum f u}{N} \right) C$$

La ecuación anterior es equivalente a la igualdad $\overline{X} = A + C\,\overline{u}$; se debe observar que en este método, los valores de la variable X se convierten en los valores de la variable u de acuerdo con la ecuación $X = A + C\,u$.

EJEMPLOS

Ejemplos

1 La siguiente tabla de distribución de frecuencias representa las edades de los 96 asistentes a un curso de informática en la ciudad de Querétaro.

Intervalos (años)	Marca de clase (X)	Frecuencia (f)
24 - 28	26	8
29 - 33	31	17
34 - 38	36	29
39 - 43	11	23
44 - 48	46	15
49 - 53	51	4
		$N = 96$

Determina la media aritmética de las edades por el método.

a) Largo.
b) Corto.
c) Clave.

Solución

a) La media aritmética por el método largo se determina al multiplicar las distintas marcas de clase X por sus respectivas frecuencias de clase f; se suman los productos y el resultado se divide entre el número total de frecuencias $\Sigma f = N$, es decir:

Intervalos (años)	Marca de clase (X)	Frecuencia (f)	fX
24 - 28	26	8	208
29 - 33	31	17	527
34 - 38	36	29	1 044
39 - 43	41	23	943
44 - 48	46	15	690
49 - 53	51	4	204
		$\Sigma f = 96 = N$	$\Sigma fX = 3\,616$

Al sustituir los datos anteriores en la ecuación para la media aritmética por el método largo resulta:

$$\overline{X} = \frac{\sum f X}{N} = \frac{3\,616}{96} = 37.666$$

La media aritmética de las edades por el método largo es 37.666 años.

b) La media aritmética por el método corto se determina al considerar una marca de clase cualquiera A, a la que se le aumenta la sumatoria del producto de las frecuencias de clases f por las desviaciones de las marcas de clase de dicha marca de clase seleccionada $d = X - A$, dividido entre el número total de frecuencias $\Sigma f = N$, es decir:

Marca de clase (X)	Desviaciones ($d = X - A$)	Frecuencias (f)	$f\,d$
26	−5	8	−40
$A \left(\text{Media supuesta}\right) \rightarrow$ 31	0	17	0
36	5	29	145
41	10	23	230
46	15	15	225
51	20	4	80
		$\Sigma f = 96 = N$	$\Sigma f d = 640$

Al sustituir los datos anteriores en la ecuación para la media aritmética por el método de corto, resulta:

$$\overline{X} = A + \frac{\sum f d}{N} = 31 + \frac{640}{96} = 31 + 6.666 = 37.666$$

La media aritmética de las
edades por el método corto es 37.666 años.

c) Para determinar la media aritmética por el método clave, primero debe observarse que los intervalos de clase tengan igual tamaño C; si se satisface la condición anterior, se selecciona una marca de clase cualquiera A, a la que se le aumenta la sumatoria del producto de las desviaciones $d = X - A$ que se expresan como Cu y donde los valores de u pueden ser 0, ±1, ±2, ±3, ..., etc.; el cero se ubicará en el renglón que corresponda a la media supuesta seleccionada y por debajo de ella decrecerán los valores negativamente (−1, −2, −3, etc.), por arriba de ella aumentará los valores positivamente (+1, +2, +3, etc.); por las frecuencias de clase f y dividido entre el número total de frecuencias $\Sigma f = N$. El resultado de la sumatoria se multiplica por el tamaño de la clase C, es decir:

Marca de clase (X)	Desviaciones (u)	Frecuencias (f)	$f\,u$
26	−4	8	−32
31	−3	17	−51
36	−2	29	−58
41	−1	23	−23
$A \left(\text{Media supuesta}\right) \rightarrow$ 46	0	15	0
51	1	4	4
		$\Sigma f = 96 = N$	$\Sigma f u = -160$

Sea C el tamaño del intervalo de clase igual a 5, y al sustituir los datos anteriores en la ecuación para la media aritmética por el método clave, resulta:

$$\overline{X} = A + \left(\frac{\sum f u}{N}\right)C = 46 + \left(\frac{-160}{96}\right)(5) = 46 - 8.334 = 37.666$$

La media aritmética de las edades por el método clave es 37.666 años.

2 La siguiente tabla de distribución de frecuencias registra la carga máxima en toneladas cortas (una tonelada corta equivale a 2 000 libras) que soportan ciertos cables de acero extraduro producidos por *Bego Steel Company*. Determina la media de la carga máxima por el método corto y comprueba su resultado por el método clave.

Intervalos de clase (máximo de carga en toneladas cortas)	Marca de clase (X)	Número de cables (frecuencias) (f)
10.4 - 11.1	10.75	8
11.2 - 11.9	11.55	13
12.0 - 12.7	12.35	21
12.8 - 13.5	13.15	30
13.6 - 14.3	13.95	23
14.4 - 15.1	14.75	17
15.2 - 15.9	15.55	10
16.0 - 16.7	16.35	6
16.8 - 17.5	17.15	2
		$f = 130$

Solución

Para aplicar el método corto se considera una marca de clase cualquiera A, a la cual le agregamos la sumatoria del producto de las frecuencias de clase f por las desviaciones $d = X - A$ y dividido entre el número total de frecuencias $\Sigma f = N$, es decir;

	Marca de clase (X)	Desviaciones ($d = X - A$)	Frecuencias (f)	$f\,d$
	10.75	−3.2	8	−25.6
	11.55	−2.4	13	−31.2
	12.35	−1.6	21	−33.6
	13.15	−0.8	30	−24.0
$A \left(\text{Media supuesta}\right) \rightarrow$	13.95	0	23	0
	14.75	0.8	17	13.6
	15.55	1.6	10	16.0
	16.35	2.4	6	14.4
	17.15	3.2	2	6.4
			$\Sigma f = 130 = N$	$\Sigma f d = -64$

Después de sustituir los datos anteriores en la ecuación para la media aritmética por el método corto resulta:

$$\overline{X} = A + \frac{\sum f d}{N} = 13.95 + \left(\frac{-64}{130}\right) = 13.95 - 0.4923 = 13.4577$$

La media aritmética de la carga máxima es 13.4577 toneladas cortas.

Para comprobar el resultado anterior con el método clave, se observa que el tamaño de los intervalos de clase son iguales ($C = 0.8$), seleccionando una marca de clase cualquiera A, a la cual se le agrega la sumatoria del producto de las desviaciones u por las frecuencias de clase f y dividido entre el número total de frecuencias $\Sigma f = N$. El resultado de la sumatoria se multiplica por el tamaño de la clase C, es decir:

	Marca de clase (X)	Desviaciones (u)	Frecuencias (f)	fu
	10.75	−2	8	−16
	11.55	−1	13	−13
A (Media supuesta) →	12.35	0	−21	0
	13.15	1	30	30
	13.95	2	23	46
	14.75	3	17	51
	15.55	4	10	40
	16.35	5	6	30
	17.15	6	2	12
			$\Sigma f = 130 = N$	$\Sigma fu = 180$

Al sustituir los datos anteriores en la ecuación para la media aritmética por el método clave resulta:

$$\overline{X} = A + \left(\frac{\sum f u}{N}\right)C = 12.35 + \left(\frac{180}{130}\right)(0.8) = 12.35 + 1.1077 = 13.4577$$

3 La siguiente tabla de distribución de frecuencias muestra los salarios semanales (cuarenta horas) en dólares para 120 técnicos de una empresa de electrónicos, de la división de televisores y pantallas. Determina la media aritmética de los salarios por el método largo y comprueba su resultado por el método corto.

Intervalos de clase (salario en dólares)	Marca de clase (X)	Número de técnicos Frecuencias (f)
380 - 459.99	420	4
360 - 379.99	370	13
350 - 359.99	335	22
340 - 349.99	345	31
330 - 339.99	335	25
320 - 329.99	325	18
310 - 319.99	315	7
		$N = 120$

Solución

El procedimiento para determinar la media aritmética por el método largo, consiste en multiplicar las marcas de clase X por sus respectivas frecuencias de clase f, se suman los productos y el resultado ΣfX se divide entre el número total de frecuencias $\Sigma f = N$, es decir:

Intervalos de clase	Marca de clase (X)	Frecuencias (f)	fX
380 - 459.99	420	4	1 680
360 - 379.99	370	13	4 810
350 - 359.99	355	22	7 810
340 - 349.99	345	31	10 695
330 - 339.99	335	25	8 375
320 - 329.99	325	18	5 850
310 - 319.99	315	7	2 205
		$\Sigma f = 120 = N$	$\Sigma fX = 41\,425$

Sustituyendo los datos anteriores en la ecuación de la media aritmética por el método largo, resulta:

$$\overline{X} = \frac{\sum f X}{N} = \frac{41\,425}{120} = 345.21$$

La media aritmética de los salarios es de 345.21 dólares.

Para comprobar el resultado anterior por el método corto, se considera una marca de clase cualquiera A, a la cual le agregamos la sumatoria del producto de las frecuencias de clases f por las desviaciones $d = X - A$ y dividido entre el número total de frecuencias $\Sigma f = N$, es decir:

	Marca de clase (X)	Desviaciones ($d = X - A$)	Frecuencias (f)	f d
	420	85	4	340
	370	35	13	455
	355	20	22	440
	345	10	31	310
A (Media supuesta) →	335	0	25	0
	325	−10	18	−180
	315	−20	7	−140
			$\Sigma f = 120 = N$	$\Sigma fd = 1\,225$

Sustituyendo los datos anteriores en la ecuación de la media aritmética por el método corto resulta:

$$\overline{X} = A + \frac{\sum f d}{N} = 335 + \frac{1\,225}{120} = 335 + 10.21 = 345.21$$

La media aritmética es de 345.21 dólares.

Media geométrica

Se define como la raíz enésima del producto de un conjunto de datos numéricos ($X_1, X_2, X_3, \ldots, X_N$). La media geométrica se representa por G y matemáticamente como:

$$G = \sqrt[N]{X_1 X_2 X_3 \ldots X_N}$$

Si se utilizan las propiedades de los exponentes y radicales, la media geométrica también se puede expresar como:

$$G = (X_1 X_2 X_3 \ldots X_N)^{\frac{1}{N}}$$

Esta operación se facilita con la aplicación de una calculadora científica que tenga:

$$X^{\frac{1}{N}}$$

EJEMPLOS

Ejemplos

1 Encuentra la media geométrica de los números 8, 12 y 18.

Solución

Del problema se observa que $N = 3$, ya que el conjunto de datos numéricos dado consta de tres valores; al sustituir los datos anteriores en la ecuación de la media geométrica resulta:

$$G = \sqrt[N]{X_1 X_2 X_3 \ldots X_N} = \sqrt[3]{(8)(12)(18)} = \sqrt[3]{1728} = 12$$

La media geométrica del conjunto de números dado es 12.

2 Determina la media geométrica de los números 2, 4, 6, 8, 10, 12, 14 y 16.

Solución

Del problema se observa que $N = 8$, ya que el conjunto de datos numéricos dado consta de ocho valores; después de sustituir los datos anteriores en la ecuación de la media geométrica resulta:

$$G = \sqrt[N]{(X_1 X_2 X_3 \ldots X_N)} = \sqrt[8]{(2)(4)(6)(8)(10)(12)(14)(16)} = \sqrt[8]{10\,321\,920} = 7.528$$

Al realizar la operación por medio de la calculadora científica resulta:

$$G = (X_1 X_2 X_3 \ldots X_N)^{\frac{1}{N}} = [(2)(4)(6)(8)(10)(12)(14)(16)]^{\frac{1}{8}} = (10321920)^{\frac{1}{8}}$$

$$G = 7.528$$

La media geométrica del conjunto de números dado es 7.528.

La media geométrica también se calcula por medio de logaritmos, es decir:

$$\log G = \frac{1}{N} \log(X_1 X_2 X_3 \ldots X_N)$$

EJEMPLOS

Ejemplos

1 Encuentra la media geométrica de los números 3, 5, 7 y 9, por medio de:

a) La ecuación radical.
b) La ecuación logarítmica.

Solución

a) Del problema se observa que $N = 4$, ya que el conjunto de datos numéricos dado consta de cuatro valores; al sustituir los datos anteriores en la ecuación radical de la media geométrica resulta:

$$G = \sqrt[N]{X_1 X_2 X_3 \ldots X_N} = \sqrt[4]{(3)(5)(7)(9)} = \sqrt[4]{945} = 5.544$$

b) Dado que $N = 4$, y sustituyendo los datos en la ecuación logarítmica de la media geométrica, tenemos:

$$\log\ G = \frac{1}{N} \log\,(X_1\, X_2\, X_3 \ldots\, X_N) = \frac{1}{4}\ \log = [(3(5)(7)(9)]$$

$$\log\ G = \frac{1}{4} \log\,(945) = \frac{1}{4}(2.975431) = 0.743857$$

$$G = \text{Antilog}(0.743857) = 5.544$$

La media geométrica del conjunto de números dado es 5.544.

2 Determina la media geométrica de los números 4, 5, 5, 7, 8, 8 y 9, por medio de:

a) La ecuación logarítmica.
b) La calculadora científica.

Solución

a) Del problema se observa que $N = 7$, ya que el conjunto de datos numéricos dado consta de siete valores; al sustituir los datos anteriores en la ecuación logarítmica de la media geométrica resulta:

$$\log\ G = \frac{1}{N} \log\,(X_1\, X_2\, X_3 \ldots X_N) = \frac{1}{7} \log = [(4)(5)(5)(7)(8)(8)(9)]$$

$$G = \text{antilog}(0.800788) = 6.321.$$

b) Al realizar la operación en la calculadora científica tenemos:

$$G = (X_1\, X_2\, X_3 \ldots X_N)^{\frac{1}{N}} = [(4)(5)(5)(7)(8)(8)(9)]^{\frac{1}{7}} = (403\ 200)^{\frac{1}{7}} = 6.321$$

La media geométrica del conjunto de números dado es 6.321.

La media geométrica se aplica en la solución de problemas de promedio de razones, cálculo de tasas de crecimiento, determinación de interés compuesto, etcétera.

EJEMPLOS

Ejemplos

1 Durante un año la razón entre los precios de un litro de aceite a un kilogramo de huevo fue de \$600, mientras que en el año siguiente la razón fue de \$800. Determina:

a) La media aritmética de las razones en los dos años.
b) La media aritmética de las razones del precio del huevo respecto al aceite en los dos años.
c) Si es o no conveniente el uso de la media aritmética en el promedio de razones.
d) Por qué es más conveniente el uso de la media geométrica en el promedio de razones.

Solución

a) La media aritmética de las razones del precio aceite-huevo, es:

$$\overline{X} = \frac{600 + 800}{2} = 700$$

b) La razón de los precios huevo-aceite en el primero y segundo año es, respectivamente:

$$\frac{1}{600} = 0.001666 \qquad \text{y} \qquad \frac{1}{800} = 0.00125$$

La media aritmética de las razones del precio huevo-aceite, es:

$$\overline{X} = \frac{0.001666 + 0.00125}{2} = 0.014583$$

c) Como era de esperarse, la media aritmética de las razones del precio aceite-huevo no es recíproca a la media aritmética de las razones del precio huevo-aceite, es decir:

$$\frac{1}{0.0014583} = 685.714 \neq 700$$

Por lo anterior se concluye que la media aritmética no es adecuada para determinar el promedio de razones.

d) La media geométrica de las razones del precio aceite-huevo, es:

$$G = \sqrt[N]{X_1 X_2 X_3 \ldots X_N} = \sqrt{(600)(800)} = 692.820$$

La media geométrica de las razones del precio huevo-aceite, es:

$$G = \sqrt[N]{X_1 X_2 X_3 \ldots X_N} = \sqrt{(0.001666)(0.00125)} = 0.001443$$

$$\frac{1}{0.001443} = 692.820$$

Dado que los promedios son recíprocos, se concluye que la media geométrica debe utilizarse para calcular los promedios de razones.

2 Un capital de \$800000 se invierte al 23% de interés anual. ¿Cuál será el capital total después de 5 años si el capital inicial no se ha retirado?

Solución

Si se aplica la fórmula de **interés compuesto** $A = P(1 + r)^n$, donde:

A = Capital total.
P = Capital inicial.
r = Porcentaje de inversión por unidad de tiempo.
n = Unidad de tiempo que dura la inversión.

Si el capital total después de un año es: $800000 + 800000\, r = 800000\,(1 + r)$
El capital total después de dos años es: $800000\,(1 + r) + 800000\,(1 + r)\, r = 800000\,(1 + r)^2$
El capital total después de tres años es: $800000\,(1 + r)^2 + 800000\,(1 + r)^2\, r = 800000\,(1 + r)^3$
De la misma manera, se concluye que el capital total después de cinco años es: $800000\,(1 + r)^5$

Después de resolver operaciones, tenemos que:

$$A = 800000\,(1 + 0.23)^5 = 800000\,(1.23)^5 = 800000\,(2.815305)$$

$$A = \$2\,252\,244.55$$

El capital total después de cinco años es \$2 252 244.55.

3 Los salarios profesionales en Argentina aumentaron en los últimos cuatro años en 29%, 22%, 26% y 28%; determina:

a) La media anual de crecimiento.
b) La media geométrica anual de crecimiento.

Solución

a) La media anual de crecimiento se obtiene aplicando la ecuación de la media aritmética, es decir:

$$\overline{X} = \frac{\sum X}{N} = \frac{20+22+26+28}{4} = 26.25$$

La media anual de crecimiento en los salarios profesionales de Argentina en los últimos cuatro años es del 26.25%.

b) La media geométrica anual de crecimiento se obtiene aplicando la ecuación radical de la media geométrica, es decir:

$$G = \sqrt[N]{X_1 X_2 X_3 \ldots X_N} = \sqrt[4]{(1.29)(1.22)(1.26)(1.28)}$$
$$G = 1.2622 \qquad \text{equivalente al } 26.22\%$$

Al comparar ambos resultados, se observa que la media geométrica siempre es menor o igual que la media aritmética.

Media geométrica para datos agrupados

Si consideramos los elementos $X_1, X_2, X_3, \ldots, X_N$ que se presentan con frecuencias respectivas $f_1, f_2, f_3, \ldots, f_N$, en donde $f_1 + f_2 + f_3 + \ldots + f_N = N$ representa la frecuencia total. La ecuación radical de la media geométrica se expresa por:

$$G = \sqrt[N]{\underbrace{X_1 X_1 \ldots X_1}_{f_1 veces}\ \underbrace{X_2 X_2 \ldots X_2}_{f_2 veces}\ \underbrace{X_3 X_3 \ldots X_3}_{f_3 veces}\ \ldots\ \underbrace{X_N X_N \ldots X_N}_{f_N veces}}$$

Al simplificar, resulta: $G = \sqrt[N]{X_1^{f_1}\ X_2^{f_2}\ X_3^{f_3} \ldots X_N^{f_N}}$

donde:

G = Media geométrica.
$N = \Sigma f$ = Número total de frecuencias.
$X_1, X_2, X_3, \ldots, X_N$ = Las marcas de clase de datos agrupados.
$f_1, f_2, f_3, \ldots, f_N$ = Las correspondientes frecuencias de clase.

La ecuación anterior se utiliza para determinar la media geométrica para datos agrupados y también se denomina **media geométrica ponderada**.

Tomando en cuenta las mismas consideraciones anteriores tenemos que la ecuación logarítmica de la media geométrica es:

$$\log G = \frac{1}{N} \log\left(X_1^{f1}\ X_2^{f2}\ X_3^{f3}\ X_N^{fN}\right)$$
$$\log G = \frac{1}{N}\left(f_1 \log X_1 + f_2 \log X_2 + f_3 \log X_3 + \ldots + f_N \log X_N\right)$$
$$\log G = \frac{1}{N}\sum_{j=1}^{N} f_1 \log X_j \qquad \log G = \frac{\sum f \log X}{N}$$

La ecuación anterior sólo es válida si todos los elementos son positivos, pues de otra forma no se satisface la operación logarítmica; también se hace notar que el logaritmo de la media geométrica de una serie de elementos positivos es la media aritmética de los logaritmos de dichos elementos.

EJEMPLO

Ejemplo 1 La siguiente tabla de distribuciones de frecuencias registra la duración en número de afeitadas de 40 navajas de las diferentes marcas más importantes del mundo y que fueron sometidas a esta prueba de control de calidad bajo condiciones normales de uso. Con la información anterior:

Intervalos de clase	Frecuencia de clase (f)
1 - 3	1
4 - 6	5
7 - 9	9
10 - 12	13
13 - 15	7
16 - 18	3
19 - 21	2
	40

a) Encuentra la media geométrica aplicando la ecuación radical.

b) Determina la media geométrica aplicando la ecuación logarítmica.

c) Comprueba que la media geométrica es menor o igual que la media aritmética.

Solución

a) Para aplicar la ecuación radical de la media geométrica es necesario construir la siguiente tabla de distribución de apoyo.

Intervalos de clase	Marca de clase (X)	Frecuencia de clase (f)	X^f
1 - 3	2	1	2
4 - 6	5	5	3125
7 - 9	8	9	134217728
10 - 12	11	13	3.452271214×10^{13}
13 - 15	14	7	105413504
16 - 18	17	3	4913
19 - 21	20	2	400
		$N = 40 = \Sigma f$	

Al sustituir los datos anteriores en la correspondiente ecuación tenemos:

$$G = \sqrt[N]{X_1^{f_1}\ X_2^{f_2}\ X_3^{f_3}\ \ldots\ X_N^{f_N}}$$

$$G = \sqrt[40]{(2)(3\,125)(134\,217\,728)(3.452271214\times10^{13})(105\,413\,504)(4\,913)(400)}$$

$$G = \sqrt[40]{5.999261773\times10^{39}} = 9.8731$$

La media geométrica calculada a partir de los datos agrupados es 9.8731.

b) Para aplicar la ecuación logarítmica de la media geométrica es necesario construir la siguiente tabla de distribución de apoyo.

Intervalos de clase	Marca de clase (*X*)	Frecuencia de clase (*f*)	*f* log *X*
1 - 3	2	1	0.301029
4 - 6	5	3	3.494830
7 - 9	8	9	6.127809
10 - 12	11	13	13.538104
13 - 15	14	7	8.022896
16 - 18	17	3	3.691346
19 - 21	20	2	2.602059
		$N = 40 = \Sigma f$	$\Sigma f \log X = 39.778093$

Después de sustituir los datos anteriores en la correspondiente ecuación tenemos:

$$\log G = \frac{\sum f \log X}{N} = \frac{39.778093}{40} = 0.994452$$

$$G = \text{antilog}\ 0.994452 = 9.8731$$

La media geométrica calculada a partir de los datos agrupados es 9.8731.

c) Para realizar la comprobación requerida, es necesario determinar la media aritmética y para ello se construye la siguiente tabla de distribución de apoyo.

Intervalos de clase	Marca de clase (*X*)	Frecuencia de clase (*f*)	*fX*
1 - 3	2	1	2
4 - 6	5	5	25
7 - 9	8	9	72
10 - 12	11	13	143
13 - 15	14	7	98
16 - 18	17	3	51
19 - 21	20	2	40
		$N = 40 = \Sigma f$	$\Sigma fX = 431$

Al sustituir los datos anteriores en la correspondiente ecuación tenemos:

$$\overline{X} = \frac{\sum fX}{N} = \frac{431}{40} = 10.775$$

$$\overline{X} > G, \text{ es decir, } 10.775 > 9.8731.$$

Media armónica

Se define como el recíproco de la media aritmética de los recíprocos de los datos numéricos en un determinado conjunto ($X_1, X_2, X_3, \ldots, X_N$) dado.

La media armónica se representa por H y matemáticamente como:

$$H = \frac{1}{\frac{1}{N}\sum_{j=1}^{N}\frac{1}{X_j}} = \frac{N}{\sum \frac{1}{X}}$$

Con el fin de facilitar su operatividad práctica, la ecuación anterior se expresa de la siguiente manera:

$$\frac{1}{H}=\frac{\sum\frac{1}{X}}{N}=\frac{1}{N}\sum\frac{1}{X}$$

EJEMPLOS

1 Encuentra la media armónica de los números 3, 5, 7 y 9.

Solución

El conjunto de datos numéricos dado consta de cuatro elementos, es decir, $N = 4$, y al sustituir en la ecuación de la media armónica, resulta:

$$H=\frac{N}{\sum\frac{1}{X}}=\frac{4}{\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{9}}=\frac{4}{\frac{105+63+45+35}{315}}=\frac{4}{\frac{248}{315}}$$

$$H=\frac{(4)(315)}{248}=\frac{1260}{248}=5.0806$$

La media armónica del conjunto de números dado es 5.0806.

Si se utiliza la ecuación de la media armónica en su forma práctica tenemos:

$$\frac{1}{H}=\frac{1}{N}\sum\frac{1}{X}=\frac{1}{4}\left(\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{9}\right)=\frac{1}{4}\left(\frac{105+63+45+35}{315}\right)$$

$$\frac{1}{H}=\frac{1}{4}\left(\frac{248}{315}\right)=\frac{248}{1260}=0.19682539$$

$$H=\frac{1}{0.19682539}=5.0806$$

2 Encuentra la media armónica de los números 4, 6, 10, 13, 15 y 19.

Solución

El conjunto de datos numéricos dado consta de seis elementos, es decir, $N = 6$, y al sustituir en la ecuación de la media armónica resulta:

$$H=\frac{N}{\sum\frac{1}{x}}=\frac{6}{\frac{1}{4}+\frac{1}{6}+\frac{1}{10}+\frac{1}{13}+\frac{1}{15}+\frac{1}{19}}$$

Se recomienda expresar las fracciones en forma decimal.

$$H=\frac{6}{0.25+0.1666+0.1+0.0769+0.0666+0.0526}=\frac{6}{0.7127}=8.4186$$

La media armónica del conjunto de números dado es 8.4186.

Para comprobar el resultado anterior, se utiliza la ecuación de la media armónica en su forma práctica, se obtiene:

$$\frac{1}{H}=\frac{1}{N}\sum\frac{1}{X}=\frac{1}{6}\left(\frac{1}{4}+\frac{1}{6}+\frac{1}{10}+\frac{1}{13}+\frac{1}{15}+\frac{1}{19}\right)$$

$$\frac{1}{H}=\frac{1}{6}(0.25+0.1666+0.1+0.0769+0.0666+0.0526)=\frac{1}{6}(0.7127)$$
$$\frac{1}{H}=\frac{0.7127}{6}=0.118783$$
$$H=\frac{1}{0.118783}=8.4186$$

La media armónica se aplica en el cálculo de rendimientos medios, promedios de costos, velocidades medias, tiempos promedios, etcétera.

EJEMPLOS

Ejemplos

1 Un agricultor puede arar un terreno en cuatro días empleando un tractor; su ayudante puede hacer el mismo trabajo con un tractor más pequeño en seis días. ¿Cuál es el rendimiento representativo de los dos tractores? ¿En cuántos días pueden arar el campo si trabajan conjuntamente?

Solución

Para calcular el rendimiento representativo se tiene que:

$$H=\frac{N}{\sum\frac{1}{X}}=\frac{2}{\frac{1}{4}+\frac{1}{6}}=\frac{2}{\frac{3+2}{12}}=\frac{2}{\frac{5}{12}}=\frac{(2)(12)}{5}=\frac{24}{5}=4\frac{4}{5}$$

El rendimiento representativo de los dos tractores es de $4\frac{4}{5}$ días.

Para determinar en cuantos días los tractores pueden arar el campo si trabajan conjuntamente, tenemos que el rendimiento representativo es de $4\frac{4}{5}$, por lo que al trabajar conjuntamente resulta:

$$\frac{4\frac{4}{5}}{2}=\frac{\frac{24}{5}}{2}=\frac{24}{10}=2\frac{2}{5}$$

Para arar el campo si los dos tractores trabajan conjuntamente, se requieren $2\frac{2}{5}$ días.

Comprobación

Los dos tractores pueden arar en un día $\frac{1}{4}+\frac{1}{6}=\frac{5}{12}$ del campo; en arar todo el campo tardarán $\frac{12}{5}=2\frac{2}{5}$ días.

2 En una ruta ferroviaria un tren se desplaza de la siguiente forma: los primeros 90 kilómetros a 70 km/h y en el segundo tramo de 70 kilómetros a 100 km/h; determina la velocidad promedio para todo el recorrido.

Solución

Dado que las distancias de los tramos no son iguales, se debe utilizar una ecuación representativa que denominamos **media armónica ponderada** para las velocidades donde los **pesos** son las distancias respectivas, es decir: $\frac{1}{H}=\frac{1}{N}\sum\frac{f}{X}$, donde, $N=\Sigma f\ \frac{N}{H}=\sum\frac{f}{X}$. Al sustituir los datos de velocidad y distancia resulta:

$$\frac{d_1+d_2+d_3+\ldots+d_N}{V}=\frac{d_1}{V_1}+\frac{d_2}{V_2}+\frac{d_3}{V_3}+\ldots+\frac{d_N}{V_N}$$

Al sustituir los datos en la ecuación de la media armónica ponderada para la velocidad media resulta:

$$\frac{90+70}{V}=\frac{90}{70}+\frac{70}{100}$$
$$\frac{160}{V}=1.2857+0.7$$
$$\frac{160}{V}=1.9857$$
$$V=\frac{160}{1.9857}=80.57$$

La velocidad promedio para todo el recorrido es 80.57 km/h.

Comprobación

Primero se determina el tiempo requerido para recorrer el primer tramo:

$$\text{Tiempo}(t)=\frac{\text{distancia}(d)}{\text{velocidad}(V)}=\frac{90\,\text{km}}{70\,\text{km/h}}=1.2857\text{ horas}$$

El tiempo requerido para recorrer el segundo tramo es:

$$t=\frac{d}{V}=\frac{70\,\text{km}}{100\,\text{km/h}}=0.7\text{ horas}$$

La velocidad media para todo el recorrido es:

$$\text{Velocidad media}=\frac{\text{distancia total}}{\text{tiempo total}}=\frac{90\text{ km}+70\text{ km}}{1.2857\text{ h}+0.7\text{ h}}=\frac{160\text{ km}}{1.9857\text{ h}}$$

$$\text{Velocidad media}=80.57\text{ km/h}$$

3 Una persona conduce su automóvil a una velocidad de 100 km/h al dirigirse a su trabajo; regresa por la misma ruta a una velocidad de 70 km/h. ¿Cuál es la velocidad promedio del viaje redondo?

Solución

Si se utiliza la ecuación de la media armónica, tenemos que la velocidad promedio del viaje redondo es:

$$H=\frac{N}{\sum\frac{1}{X}}=\frac{2}{\frac{1}{100}+\frac{1}{70}}=\frac{2}{0.01+0.014285}=\frac{2}{0.024285}=82.35$$

La velocidad promedio del viaje redondo es 82.35 km/h.

Comprobación

Si se supone que la distancia de ida es de 120 km (aunque puede ser cualquier distancia supuesta), entonces el tiempo requerido para la ida y para la vuelta es:

$$\text{Tiempo de ida}=\frac{\text{distancia de ida}}{\text{velocidad de ida}}=\frac{120\text{ km}}{100\text{ km/h}}=1.2\text{ horas}$$

$$\text{Tiempo de vuelta}=\frac{\text{distancia de vuelta}}{\text{velocidad de vuelta}}=\frac{120\text{ km}}{70\text{ km/h}}=1.7142\text{ horas}$$

La velocidad promedio del viaje redondo es:

$$\text{Velocidad media} = \frac{\text{distancia total}}{\text{tiempo total}} = \frac{120\text{ km} + 120\text{ km}}{1.2\text{ h} + 1.7142\text{ h}} = \frac{240\text{ km}}{2.9142\text{ h}}$$

$$\text{Velocidad media} = 82.35\text{ km/h}$$

Es importante mencionar que es fácil equivocarse al tomar la media aritmética de 100 y 70 km/h como 85 km/h, ya que el contenido del problema nos induce a ello.

Media armónica para datos agrupados

Si consideramos los elementos ($X_1, X_2, X_3, \ldots, X_N$) que se presentan con frecuencias ($f_1, f_2, f_3, \ldots, f_N$) donde ($f_1 + f_2 + f_3 + \ldots + f_N = N$) representa la frecuencia total, la ecuación de la media armónica para datos agrupados se expresa por:

$$H = \frac{f_1 + f_2 + f_3 + \ldots + f_N}{\frac{f_1}{X_1} + \frac{f_2}{X_2} + \frac{f_3}{X_3} + \ldots + \frac{f_N}{X_N}} = \frac{\sum f}{\sum \frac{f}{X}} = \frac{N}{\sum \frac{f}{X}}$$

También se puede expresar en su forma práctica por:

$$\frac{1}{H} = \frac{1}{f_1 + f_2 + f_3 + \ldots + f_N}\left(\frac{f_1}{X_1} + \frac{f_2}{X_2} + \frac{f_3}{X_3} + \ldots + \frac{f_N}{X_N}\right) = \frac{1}{\sum f}\left(\sum \frac{f}{X}\right) = \frac{1}{N}\sum \frac{f}{X}$$

Los elementos de ambas ecuaciones se denotan como:

H = Media armónica.
$N = \Sigma f$ = Número total de frecuencias.
($X_1, X_2, X_3, \ldots, X_N$) = Las marcas de clase de datos agrupados.
($f_1, f_2, f_3, \ldots, f_N$) = Las correspondientes frecuencias de clase.

Las ecuaciones anteriores se utilizan para determinar la media armónica para datos agrupados y también se denomina **media armónica ponderada**.

EJEMPLO

Ejemplo 1

La siguiente tabla de distribuciones de frecuencias registra las longitudes en centímetros que en una semana tienen 100 plantas de frijol. Con la información anterior:

a) Determina la media armónica.
b) Verifica que el resultado anterior sea el mismo que se obtiene al aplicar la ecuación de la media armónica en su forma práctica.

c) Comprueba que la media armónica es menor o igual que la media geométrica.

Intervalos (longitudes)	Frecuencias (f) (número de plantas)
5.4 - 5.7	7
5.8 - 6.1	16
6.2 - 6.5	21
6.6 - 6.9	29
7.0 - 7.3	18
7.4 - 7.7	9
	100

Solución

a) Para determinar la media armónica es necesario construir la siguiente tabla de distribución de apoyo.

Intervalos (longitudes)	Marca de clase (X)	Frecuencias (f) (número de plantas)	$\frac{f}{X}$
5.4 - 5.7	5.55	7	1.261261
5.8 - 6.1	5.95	16	2.689075
6.2 - 6.5	6.35	21	3.307086
6.6 - 6.9	6.75	29	4.296296
7.0 - 7.3	7.15	18	2.517482
7.4 - 7.7	7.55	9	1.192052
		$N = 100 = \Sigma f$	$\sum \frac{f}{X} = 15.263252$

Al sustituir los datos anteriores en la correspondiente ecuación tenemos:

$$H = \frac{N}{\sum \frac{f}{X}} = \frac{100}{15.263252} = 6.55$$

La media armónica calculada a partir de los datos agrupados es de 6.55 centímetros.

b) Para demostrar el resultado anterior, utilizaremos los datos que se tienen en la tabla de distribución de apoyo y los sustituimos en la ecuación de la media armónica en su forma práctica, de lo que resulta:

$$\frac{1}{H} = \frac{1}{N}\sum \frac{f}{X} = \frac{1}{100}(15.263252) = \frac{15.263252}{100} = 0.15263252$$

$$H = \frac{1}{0.15263252} = 6.55$$

c) Para realizar la comprobación requerida, es necesario determinar la media geométrica y para ello se construye la siguiente tabla de distribución de apoyo.

Intervalos (longitudes)	Marca de clase (X)	Frecuencias (f) (número de plantas)	X^f	$f \log X$
5.4 - 5.7	5.55	7	162200.01	5.210050
5.8 - 6.1	5.95	16	2.467581×10^{12}	12.392271
6.2 - 6.5	6.35	21	7.215197×10^{16}	16.858248
6.6 - 6.9	6.75	29	1.121526×10^{24}	24.049809
7.0 - 7.3	7.15	18	2.385111×10^{15}	15.377508
7.4 - 7.7	7.55	9	79711790.54	7.901522
		$N = 100 = \Sigma f$		$\Sigma f \log X = 81.789408$

Sustituyendo los datos anteriores en las correspondientes ecuaciones de la media geométrica (radical y logarítmica), tenemos:

$$G = \sqrt[N]{X_1^{f_1} X_2^{f_2} X_3^{f_3} \ldots X_N^{f_N}}$$

$$G = \sqrt[100]{(162\,200.0119)(2.467581\times10^{12})(7.215197\times10^{16})(1.121526\times10^{24})(2.385111\times10^{15})(79\,711\,790.5\text{-}}$$

$$G = \sqrt[100]{6.15759326\times10^{81}} = 6.57$$

$$\log G = \frac{\sum f \log X}{N} = \frac{81.789408}{100} = 0.81789408 \qquad G = \text{antilog } 0.81789408 = 6.57$$

Comprobamos que $G > H$, es decir, $6.57 > 6.55$.

Raíz cuadrada del cuadrado de la media (RMS)

Media cuadrática

La media cuadrática, también llamada valor cuadrático medio, se define como la raíz cuadrada de la media aritmética de los cuadrados de los datos de un conjunto de elementos numéricos ($X_1, X_2, X_3, \ldots, X_N$) dado. La media cuadrática, que también se denomina raíz cuadrática media o raíz cuadrada del cuadrado de la media, se simboliza por las siglas RMS (del inglés root mean square) o por la expresión $\left(\sqrt{\overline{X}^2}\right)$. Matemáticamente la media cuadrática se representa por:

$$\text{RMS} = \sqrt{\overline{X}^2} = \sqrt{\frac{\sum_{j=1}^{N} X_j^2}{N}} = \sqrt{\frac{\sum X^2}{N}}$$

La media cuadrática siempre es mayor o igual que las medias aritmética, geométrica y armónica; lo cual se simboliza como:

$$\text{RMS} \geq \overline{X} \geq G \geq H$$

EJEMPLOS

1 Determina la raíz cuadrática media para el siguiente conjunto de elementos numéricos (8, 10, 12, 14). Compara el resultado con las medias aritmética, geométrica y armónica.

Solución

Al sustituir los datos en la ecuación de la raíz cuadrática media resulta:

$$\text{RMS} = \sqrt{\frac{\sum X^2}{N}} = \sqrt{\frac{(8)^2 + (10)^2 + (12)^2 + (14)^2}{4}} = \sqrt{\frac{64 + 100 + 144 + 196}{4}}$$

$$\text{RMS} = \sqrt{\frac{504}{4}} = 11.2249$$

La raíz cuadrada del cuadrado de la media del conjunto de números dado es 11.2249.

La media aritmética del conjunto de elementos numéricos dado es:

$$\overline{X} = \frac{\sum X}{N} = \frac{8 + 10 + 12 + 14}{4} = \frac{44}{4} = 11$$

La media geométrica del conjunto de elementos numéricos dado es:

$$G = \sqrt[N]{X_1 X_2 X_3 \ldots X_N} = \sqrt[4]{(8)(10)(12)(14)} = \sqrt[4]{13440} = 10.7671$$

La media armónica del conjunto de elementos numéricos dado es:

$$H = \frac{N}{\sum \frac{1}{X}} = \frac{N}{\frac{1}{8} + \frac{1}{10} + \frac{1}{12} + \frac{1}{14}} = \frac{N}{\frac{105 + 84 + 70 + 60}{840}} = \frac{4}{\frac{319}{840}}$$

$$H = \frac{4(840)}{319} = \frac{3360}{319} = 10.5329$$

Al comparar los resultados, tenemos: $11.2249 > 11 > 10.7671 > 10.5329$

$$\text{RMS} > X > G > H$$

2 Determina la media cuadrática de los números 9, 9, 9, 9 y 9. Compara el resultado con las medias aritmética, geométrica y armónica.

Solución

Tras sustituir los datos dados en la ecuación de la media cuadrática resulta:

$$\text{RMS} = \sqrt{\frac{\sum X^2}{N}} = \sqrt{\frac{(9)^2 + (9)^2 + (9)^2 + (9)^2 + (9)^2}{5}} = \sqrt{\frac{81 + 81 + 81 + 81 + 81}{5}}$$

$$\text{RMS} = \sqrt{\frac{405}{5}} = 9$$

La media cuadrática del conjunto de números dado es 9.

La media aritmética del conjunto de elementos numéricos dado es:

$$\overline{X} = \frac{\sum X}{N} = \frac{9+9+9+9+9}{5} = \frac{45}{5} = 9$$

La media geométrica del conjunto de elementos numéricos dado, es:

$$G = \sqrt[N]{X_1 X_2 X_3 \ldots X_N} = \sqrt[5]{(9)(9)(9)(9)(9)} = \sqrt[5]{59\ 049} = 9$$

La media armónica del conjunto de elementos numéricos dado es:

$$H = \frac{N}{\sum \frac{1}{N}} = \frac{5}{\frac{1}{9}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}} = \frac{5}{\frac{1+1+1+1+1}{9}} = \frac{5}{\frac{5}{9}}$$

$$H = \frac{5(9)}{5} = \frac{45}{5} = 9$$

Al comparar los resultados, tenemos $9 = 9 = 9 = 9$

$$\text{RMS} = \overline{X} = G = H.$$

La raíz cuadrada del cuadrado de la media o media cuadrática se aplica generalmente a procesos físicos, por ejemplo, en el cálculo de la velocidad cuadrática media de las moléculas de un gas.

EJEMPLO

Ejemplo 1 Calcula la raíz cuadrada del cuadrado de la media para las moléculas de una mezcla de gases con oxígeno (O), dióxido de carbono (CO_2), flúor (F), helio (He) e hidrógeno (H), cuyas velocidades son 484, 354, 267, 329 y 508 m/s respectivamente; los gases se encuentran en un recipiente a 20 °C de temperatura y a una atmósfera de presión.

Solución

Al sustituir las velocidades en la ecuación de la media cuadrática tenemos:

$$\text{RMS} = \sqrt{\frac{\sum X^2}{N}} = \sqrt{\frac{(484)^2 + (354)^2 + (267)^2 + (329)^2 + (508)^2}{5}}$$

$$\text{RMS} = \sqrt{\frac{234\ 256 + 125\ 316 + 71\ 289 + 108\ 241 + 258\ 064}{5}} = \sqrt{\frac{797\ 166}{5}} = 399.29\,\text{m/s}$$

La velocidad cuadrática media de los gases es 399.29 m/s.

Media cuadrática para datos agrupados

Si consideramos los elementos ($X_1, X_2, X_3, \ldots, X_N$) que se presentan con frecuencias ($f_1, f_2, f_3, \ldots, f_N$) en una distribución de frecuencias, la ecuación de la media cuadrática o raíz cuadrada del cuadrado de la media para datos agrupados se expresa por:

$$\text{RMS} = \sqrt{\overline{X}^2} = \sqrt{\frac{\sum_{j=1}^{N} f_j X_j^2}{\sum_{j=1}^{N} f_j}} = \sqrt{\frac{\sum f X^2}{N}}$$

EJEMPLO

Ejemplo 1 Encuentra la media cuadrática o raíz cuadrada del cuadrado de la media para la tabla de distribución de frecuencias que muestra las ventas de una semana, en millones de pesos, de 60 vendedores de artículos electrónicos de una corporación de México, D.F. compara el resultado con las medias aritmética, geométrica y armónica.

Intervalos de clase (venta millones de pesos)	Frecuencia (f) (número de vendedores)
8 - 12	2
13 - 17	7
18 - 22	11
23 - 27	17
28 - 32	12
33 - 37	8
38 - 42	3
	$N = 60$

Solución

Se tiene que calcular la media cuadrática o raíz cuadrada del cuadrado de la media, la media aritmética, la media geométrica y la armónica, y utilizando las respectivas ecuaciones matemáticas, se elabora la siguiente tabla de apoyo.

Intervalos de clase	Marca de clase (X)	Frecuencia de clase (f)	fX	fX^2	X^f	$\frac{f}{X}$
8 - 12	10	2	20	200	100	0.200
13 - 17	15	7	105	1 575	170 859 375	0.467
18 - 22	20	11	220	4 400	2.048×10^{14}	0.550
23 - 27	25	17	425	10 625	5.820766×10^{23}	0.680
28 - 32	30	12	360	10 800	5.314409×10^{17}	0.400
33 - 37	35	8	280	9 800	2.251875×10^{12}	0.229
38 - 42	40	3	120	4 800	64 000	0.075
		$N = 60$	$\sum fX = 1530$	$\sum fX^2 = 42\,200$		$\sum\frac{f}{X} = 2.601$

Al sustituir los datos de la tabla en la ecuación de la media cuadrática o raíz cuadrada del cuadrado de la media resulta:

$$\text{RMS} = \sqrt{\frac{\sum fX^2}{N}} = \sqrt{\frac{42\,200}{60}} = 26.520$$

La media cuadrática de las ventas es de 26.520 millones de pesos.

Si sustituimos los datos de la tabla en la ecuación de la media aritmética resulta:

$$\overline{X} = \frac{\sum fX}{N} = \frac{1\,530}{60} = 25.5$$

La media aritmética de las ventas es de 25.5 millones de pesos.

Sustituyendo los datos de la tabla en la ecuación de la media geométrica, resulta:

$$G=\sqrt[N]{X_1^{f1}X_2^{f2}X_3^{f3}...X_N^{fN}}$$

$$G=\sqrt[60]{(100)(170\,859\,375)(2.048\times10^{14})(5.820766\times10^{23})(5.314409\times10^{17})(2.251875\times10^{12})(64\,000)}$$

$$G=\sqrt[60]{1.560012\times10^{83}}=24.352$$

La media geométrica de las ventas es de 24.352 millones de pesos.

Al sustituir los datos de la tabla en la ecuación de la media armónica resulta:

$$H=\frac{N}{\sum\frac{f}{X}}=\frac{60}{2.601}=23.068$$

La media armónica de las ventas es de 23.068 millones de pesos.

Si comparamos los resultados tenemos: $26.520 > 25.5 > 24.352 > 23.068$

$$\text{RMS} > X > G > H$$

EJERCICIO 8

I. Contesta las siguientes preguntas.

1. ¿Qué es la media cuadrática?
2. ¿Con qué otro nombre se conoce la media cuadrática?
3. ¿Cómo se simboliza la media cuadrática?
4. ¿Cuál es la ecuación matemática de la media cuadrática para datos no agrupados y para datos agrupados?
5. ¿Dónde se aplica la media cuadrática?

II. Resuelve los siguientes problemas.

1. Determina la media cuadrática para los siguientes conjuntos de elementos y compara el resultado con las medias aritmética, geométrica y armónica.

   a) 31, 35, 38.
   b) 15, 17, 19, 21.
   c) 345, 328, 337, 341, 334.
   d) 8.3, 12.7, 16.5, 9.8, 15.4, 11.6, 14.2.
   e) 18, 18, 18, 18.
   f) π, π, π, π.
   g) 2, 2, 2, 5, 5, 5, 5, 9, 9, 7.
   h) a, a, a, a, a, a.

2. Encuentra la media cuadrática para la siguiente tabla de distribución que muestra las calificaciones del examen final de edafología de la Escuela Superior de Agricultura; compara el resultado con las medias aritmética, geométrica y armónica.

Intervalos de clase (calificaciones)	Frecuencia de clase (*f*) (número de estudiantes)
30 - 39	2
40 - 49	3
50 - 59	6
60 - 69	13
70 - 79	21
80 - 89	16
90 - 100	4
	$N = 65$

3. Determina la media cuadrática para la siguiente tabla de distribución que muestra el tiempo de duración en horas de 25 baterías alcalinas seleccionadas para una prueba de control de calidad en la aplicación de uso para lámparas de mano; compara el resultado con las medias aritmética, geométrica y armónica.

Intervalos (horas)	Frecuencia (f) (baterías)
48.2 - 51.7	2
51.8 - 55.3	5
55.4 - 58.9	9
59.0 - 62.5	6
62.6 - 66.1	3
	$N = 25$

4. Calcula la media cuadrática para la siguiente tabla de distribución que muestra el número de boletos de rifa que vendieron 98 miembros de un club de servicio social; compara el resultado con las medias aritmética, geométrica y armónica.

Intervalos (boletos)	Frecuencia (f) (miembros)
10 - 24	34
25 - 39	21
40 - 54	16
55 - 69	11
70 - 84	8
85 - 99	5
100 - 114	3
	$N = 98$

5. Dadas las siguientes tablas de distribución de frecuencias determina la media cuadrática y compara el resultado con las medias aritmética, geométrica y armónica.

a)

Intervalos de clase	Frecuencia de clase (f)
16 - 22	3
23 - 29	5
30 - 36	9
37 - 43	12
44 - 50	7
51 - 57	4
	$N = 40$

b)

Intervalos de clase	Frecuencia de clase (f)
5.4 - 5.8	3
5.9 - 6.3	7
6.4 - 6.8	11
6.9 - 7.3	18
7.4 - 7.8	13
7.9 - 8.3	9
8.4 - 8.8	5
8.9 - 9.3	2
	$N = 68$

c)

Intervalos (segundos)	Frecuencia (atletas) (f)
10.8 - 11.2	7
11.3 - 11.7	12
11.8 - 12.2	17
12.3 - 12.7	14
12.8 - 13.2	6
13.3 - 13.7	4
	$N = 60$

d)

Intervalos (horas)	Frecuencia (alumnos) (*f*)
51 - 100	3
101 - 150	9
151 - 200	16
201 - 250	23
251 - 300	39
301 - 350	31
351 - 400	7
	$N = 128$

e)

Intervalos (horas)	Frecuencia (familias) (*f*)
30 - 39	4
40 - 49	9
50 - 59	13
60 - 69	19
70 - 79	12
80 - 89	3
	$N = 60$

f)

Intervalos de clase	Frecuencia de clase (*f*)
5.0 - 8.9	3
9.0 - 12.9	10
13.0 - 16.9	14
17.0 - 20.9	25
21.0 - 24.9	17
25.0 - 28.9	9
29.0 - 32.9	2
	$N = 80$

g)

Intervalos de clase	Frecuencia de clase (*f*)
0.583 - 0.588	4
0.589 - 0.594	8
0.595 - 0.600	19
0.601 - 0.606	24
0.607 - 0.612	39
0.613 - 0.618	17
0.619 - 0.624	9
0.625 - 0.630	5
	$N = 125$

h)

Intervalos de clase	Frecuencia de clase (*f*)
10 - 19	2
20 - 29	5
30 - 39	11
40 - 49	9
50 - 59	4
60 - 69	1
	$N = 32$

Verifica tus resultados en la sección de respuestas.

Mediana

Se define como el **valor medio** a aquel que divide a un conjunto de datos que se ordena de acuerdo con su magnitud (de forma ascendente o de forma descendente) en dos partes iguales, es decir, es aquel valor central que deja por debajo igual número de elementos que por arriba de él.

Si el conjunto de elementos ordenados de acuerdo con su magnitud, es **impar**, la mediana será el valor intermedio de dicha sucesión.

EJEMPLOS

Ejemplos

1 Determina la mediana de los números 25, 29, 24, 29, 27, 25, 29, 31 y 26.

Solución

Al ordenar ascendentemente el conjunto de elementos numéricos dado tenemos:

24, 25, 25, 26, 27, 29, 29, 29, 31

Por definición, la mediana de los números es el valor central, en este caso es 27.

Si el conjunto de elementos ordenados de acuerdo con su magnitud es par, la mediana será la media aritmética de los dos elementos medios.

2 Determina la mediana de los números 49, 38, 46, 40, 36, 42, 36, 43.

Solución

Al ordenar descendentemente el conjunto de elementos numéricos dado se tiene:

49, 46, 43, 42, 40, 38, 36, 36

La mediana es la media aritmética de los números 42 y 40, es decir:

$$\text{Mediana} = \frac{42+40}{2} = 41$$

La mediana del conjunto de elementos numéricos dado es 41.

Mediana para datos agrupados

Para determinar la mediana en datos agrupados se emplea el método de la **interpolación**, es decir, se fundamenta de la suposición de que los elementos en la **clase mediana** (es la clase en la cual se localiza el valor de la mediana) están distribuidos uniformemente en todo el intervalo.

El primer paso para determinar la mediana de una distribución de frecuencias consiste en hallar $\frac{N}{2}$, es decir, la suma total de todas las frecuencias se divide entre dos, para encontrar el elemento intermedio en la distribución; a continuación se fija la clase en la cual cae el elemento intermedio $\frac{N}{2}$, lo anterior se obtiene al sumar las frecuencias de todas las clases que están por debajo del elemento intermedio $\frac{N}{2}$. Una vez localizada la clase mediana, el siguiente paso es determinar la **mediana**, la cual se obtiene al restar de $\frac{N}{2}$ la frecuencia acumulada de la clase inferior siguiente $(\sum f)_1$; esta diferencia se divide entre la frecuencia de la clase mediana (f mediana) y su resultado se multiplica por la anchura del intervalo de la clase mediana C. Este valor así encontrado se aumenta al valor del límite inferior de la clase para dar lugar a la *mediana*.

El cálculo de la mediana de una distribución de frecuencias se determina matemáticamente por la ecuación:

$$\text{Mediana} = L_1 + \left[\frac{\frac{N}{2} - \left(\sum f\right)_1}{f \text{ mediana}}\right] C$$

donde:

L_1 = Límite real inferior de la clase mediana, es decir, la clase que contiene el valor de la mediana.

N = Número total de datos, es decir, la frecuencia total.

$(\Sigma f)_1$ = Suma de las frecuencias de todas las clases por debajo de la clase mediana.
f mediana = Frecuencia de la clase mediana, la cual nunca debe ser mayor que $\frac{N}{2}$.

EJEMPLO

C = Tamaño del intervalo de la clase mediana.

Ejemplo 1

Las siguientes son las puntuaciones ordenadas del primer parcial que obtuvieron 56 estudiantes en un examen de geometría analítica: 49, 53, 56, 58, 60, 62, 63, 64, 64, 65, 65, 67, 68, 68, 69, 69, 70, 70, 70, 72, 72, 72, 72, 73, 74, 74, 75, 75, 75, 76, 76, 76, 77, 77, 78, 78, 79, 79, 80, 81, 82, 83, 84, 85, 85, 85, 86, 88, 89, 90, 90, 91, 93, 95, 97, 100. La siguiente tabla de distribución de frecuencias agrupa los datos anteriores.

Intervalos de clase	Frecuencia de clase (f)
45 - 51	1
52 - 58	3
59 - 65	7
66 - 72	12
73 - 79	15
80 - 86	9
87 - 93	6
94 - 100	3
	$N = 56$

Calcula la mediana por medio de:

a) Los datos originales (no agrupados).
b) Interpolación (distribución de frecuencias).
c) Fórmula (distribución de frecuencias).

Solución

a) Puesto que hay 56 puntuaciones ordenadas y siendo éste un número par, existen dos valores numéricos medios que tendrán 27 valores numéricos por abajo de ellos y 27 valores numéricos por arriba. La mediana es la media aritmética de los dos valores numéricos que ocupen los lugares 28 y 29 y es igual a:

$$\text{Mediana} = \frac{75+75}{2} = 75$$

La mediana de las puntuaciones calculada a partir de datos no agrupados es 75.

b) Con base en la distribución de frecuencias dada, se supone que las puntuaciones se distribuyen en forma continua; en tal caso, la mediana es aquella puntuación que corresponda a la mitad de la frecuencia total $\frac{56}{2} = 28$, es decir, la mitad de las frecuencias queda por debajo y la otra mitad por arriba de ella. La suma de las frecuencias de las cuatro primeras clases es $1 + 3 + 7 + 12 = 23$; para llegar al valor deseado de 28 se necesitan cinco más de las quince puntuaciones correspondientes a la quinta clase. Puesto que el intervalo de la quinta clase es 73 - 79, realmente corresponde a las puntuaciones 72.5 - 79.5, por lo que la mediana estará a $\frac{5}{15} = \frac{1}{3}$ del camino entre 72.5 y 79.5, es decir:

$$\text{Mediana} = 72.5 + \frac{1}{3}(79.5 - 72.5) = 72.5 + \frac{1}{3}(7)$$
$$\text{Mediana} = 72.5 + 2.33 = 74.83 \approx 75$$

c) Considerando la tabla de distribución de frecuencias dada, se determina $\frac{N}{2}$, es decir, el número total de frecuencias se divide entre dos, con objeto de localizar la clase mediana:

$$\frac{N}{2} = \frac{56}{2} = 28$$

La suma de las frecuencias de las cuatro y cinco primeras clases es $1 + 3 + 7 + 12 = 23$ y $1 + 3 + 7 + 12 + 15 = 38$, respectivamente, por lo que es evidente que la mediana se encuentra en la quinta clase, que será, por lo tanto, la clase mediana.

	Intervalos de clase	Frecuencia de clase (f_1)	
		$(\Sigma f)_1$	
	45 - 51	1	
	52 - 58	3	
	59 - 65	7	
	66 - 72	12	
Clase mediana →	73 - 79	15	← (f mediana)
	80 - 86	9	
	87 - 93	6	
	94 - 100	3	
		$N = 56$	

Entonces observamos los siguientes datos:

L_1 = El límite real de la clase mediana se obtiene sumando el límite superior de un intervalo de clase y el límite inferior de intervalo de clase contiguo superior y dividiendo entre dos, es decir:

$$L_1 = \frac{72 + 73}{2} = 72.5$$

N = El número total de frecuencias es 56.
$(\Sigma f)_1$ = La suma de todas las frecuencias de clases inferiores a la clase mediana es 23.
f mediana = La frecuencia de la clase mediana es 15.
C = El tamaño del intervalo de la clase mediana es 7.

Al sustituir los datos anteriores en la ecuación de la mediana, resulta:

$$\text{Mediana} = L_1 + \left[\frac{\frac{N}{2} - \left(\sum f\right)_1}{f \text{ mediana}}\right]C = 72.5 + \left[\frac{\frac{56}{2} - (23)}{15}\right](7)$$
$$\text{Mediana} = 72.5 + \left[\frac{28 - 23}{15}\right](7) = 72.5 + \left(\frac{5}{15}\right)(7)$$
$$\text{Mediana} = 72.5 + 2.33 = 74.83 \approx 75$$

La mediana de las puntuaciones calculada a partir de datos agrupados es 75.

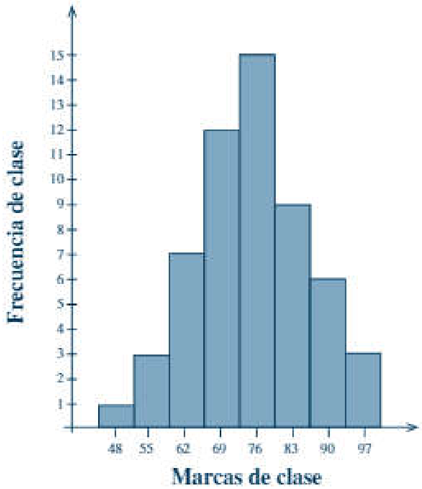
Cálculo de la mediana por métodos gráficos

Histograma. Geométricamente se hace notar que la mediana es el valor de la abscisa (x) que corresponde a la línea vertical que divide al histograma en dos partes de igual área.

La mediana puede determinarse en un histograma, el proceso consiste en contar las frecuencias que vienen siendo las áreas de los rectángulos cuando las bases son unitarias, localizar el rectángulo correspondiente a la clase que contiene la mediana y posteriormente dividir el área de este rectángulo mediante una línea perpendicular al eje x, de tal manera que el área total a la izquierda y a la derecha de dicha línea sean iguales.

EJEMPLO

Ejemplo 1 El gráfico siguiente es el histograma correspondiente a la tabla de distribución de frecuencias que agrupa las puntuaciones del primer examen parcial de cálculo diferencial de un grupo de estudiantes. Determina por el método gráfico su mediana.



Solución

Dado que, geométricamente, la mediana se define como el valor de abscisa que corresponde a la línea vertical que divide a un histograma en dos partes de igual área, tenemos que el área total del histograma dado es de 56 unidades de área, por lo que la mediana será la línea vertical que tenga $\frac{56}{2} = 28$ unidades de área a la izquierda y a la derecha en el gráfico.

Encontramos que en el quinto rectángulo del histograma se localiza la clase mediana, cuya marca de clase es 76 y área de 15 unidades; a la izquierda de dicho rectángulo hay $1 + 3 + 7 + 12 = 23$ unidades de área, por lo que se requieren 5 unidades del rectángulo de la clase mediana para completar 28 unidades de área, que es exactamente la mitad del área total del histograma; lo anterior se logra al dividir la base en tres partes iguales $\left(\text{cada parte es } \frac{5}{15} = \frac{1}{3}\right)$ y al trazar la perpendicular en el primer tercio de la base, se tiene la **mediana** como resultante gráfica.

Si los límites reales de la clase mediana (73 - 79) son:

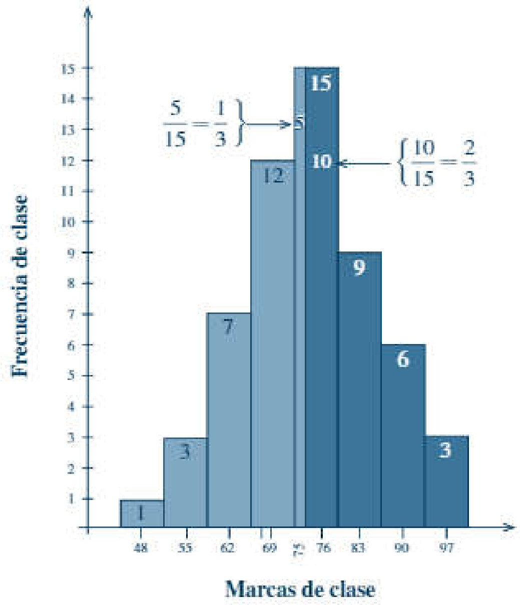
$$\text{Límite real inferior} = \frac{72 + 73}{2} = 72.5$$

$$\text{Límite real superior} = \frac{79 + 80}{2} = 79.5$$

El tamaño de la clase mediana es $79.5 - 72.5 = 7$. La tercera parte del tamaño de la clase mediana es $\frac{7}{3} = 2.33$, por lo que el valor de la mediana es:

$$\text{Mediana} = 72.5 + 2.33 = 74.83 \approx 75$$

Al elaborar la gráfica se tiene:

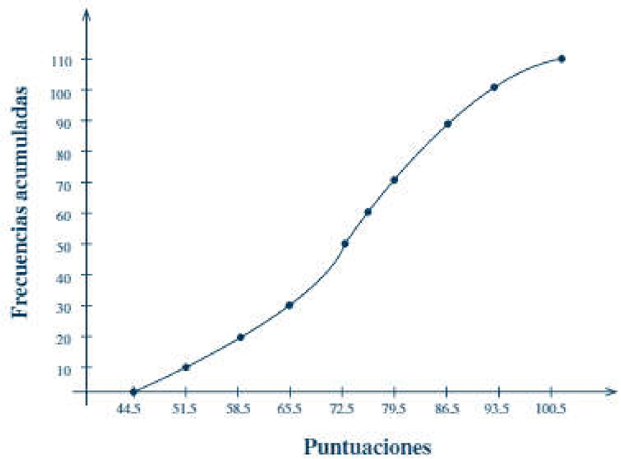


Ojiva porcentual. La mediana de una distribución de frecuencias relativas acumuladas, se determina por medio de una ojiva porcentual resultante de la distribución dada, donde la **mediana** es la abscisa x correspondiente a un punto sobre la ojiva, cuya ordenada es el 50%.

EJEMPLO

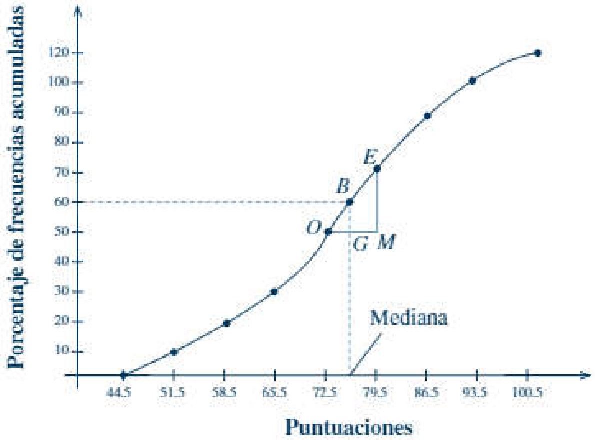
Ejemplo 1

El gráfico siguiente es una ojiva porcentual correspondiente a la tabla de distribución de frecuencias relativas acumulados que agrupan las puntuaciones del primer examen parcial de cálculo diferencial de un grupo de 56 estudiantes. Determina por el método gráfico su mediana.



Solución

La mediana es la abscisa del punto B sobre la ojiva porcentual, cuya ordenada es el 50%. Si se utiliza la relación de los triángulos semejantes BGO y OME, el valor de la mediana gráficamente es:



$$\frac{OG}{OM}=\frac{BG}{EM}=\frac{OG}{7}=\frac{50\%-41.07\%}{67.85\%-41.07\%}=\frac{8.93}{26.78}$$

$$OG=\frac{(8.93)(7)}{26.78}=\frac{62.51}{26.78}=2.33$$

Entonces:

Mediana $= 72.5 + OG = 72.5 + 2.33 = 74.83 \approx 75$.

EJERCICIO 9

Escribe los números correspondientes

Competencias genéricas

Competencias disciplinares

I. Realiza lo que se indica a continuación.

1. Define mediana.
2. Explica cómo se determina la mediana de un conjunto de elementos impar.
3. Explica cómo se determina la mediana de un conjunto de elementos par.
4. Explica cómo se determina la mediana para datos agrupados.
5. Escribe la ecuación matemática que se emplea para determinar la mediana de datos agrupados.
6. Explica cómo se calcula la mediana en un histograma.
7. Explica cómo se calcula la mediana en una ojiva porcentual.

II. Resuelve los siguientes problemas.

1. Determina la media aritmética, la mediana, la moda, la media geométrica y la media armónica, para los siguientes conjuntos de datos numéricos.

   a) 9, 4, 6, 8, 8, 12.

   b) 125, 116, 119, 121, 121, 124, 129.

   c) 8, 12, 16, 8, 10, 12, 10, 14, 10, 12, 10, 8, 10.

2. El costo de un litro de gasolina magna sin en la franja fronteriza norte de la República Mexicana, registrado en el primer semestre del año 2009 es: $820.00, $760.00, $760.00, $840.00, $890.00, $960.00; determina:

   a) La media aritmética de los costes.

   b) La mediana de los costes.

   c) La moda de los costes.

   d) La media geométrica de los costes.

   e) La media armónica de los costes.

3. Doce vendedores de autos usados vendieron en una semana, respectivamente, 59, 72, 86, 45, 64, 48, 69, 90, 43, 95, 64 y 81 automóviles. Determina:

   a) La media aritmética.

   b) La mediana.

   c) La moda.

   d) La media geométrica.

   e) La media armónica.

4. La siguiente tabla de distribución de frecuencias registra los salarios semanales en miles de pesos, de 80 jefes de familia. Determina:

Intervalos (salarios)	Frecuencia de clase (f)
96 - 125	4
126 - 155	9
156 - 185	17
186 - 215	25
216 - 245	18
246 - 275	7
	$N = 80$

   a) La media aritmética por el método corto.

   b) La mediana por interpolación.

   c) La moda.

   d) La media geométrica aplicando la ecuación logarítmica.

   e) Comprueba que la media geométrica es menor o igual que la media aritmética.

   f) La media armónica.

   g) Comprueba que la media armónica es menor o igual que la media geométrica.

5. La siguiente tabla de distribución de frecuencias registra los aciertos de un examen de admisión de 50 reactivos sobre la comprensión, razonamiento y aplicación de las matemáticas, aplicado a 582 alumnos de nuevo ingreso al nivel medio superior, en el CBTIS No. 125 de Ciudad Miguel Alemán, Tamaulipas. Determina:

Intervalos reactivos (aciertos)	Frecuencia de clase (f) (alumnos)
0 - 2	36
3 - 5	48
6 - 8	54
9 - 11	67
12 - 14	78
15 - 17	93
18 - 20	56
21 - 23	39
24 - 26	28
27 - 29	21
30 - 32	18
33 - 35	15
36 - 38	11
39 - 41	9
42 - 44	5
45 - 47	3
48 - 50	1
	$N = 582$

a) La media aritmética por el método clave.

b) Comprueba el resultado anterior por el método largo.

c) La mediana.

d) Construye el histograma y comprueba el valor de la mediana.

e) La moda.

f) La media geométrica.

g) La media armónica

6. Si el precio de un producto se triplica en un periodo de 5 años, ¿cuál es el promedio de incremento porcentual por año?

7. Las ventas de un almacén se incrementaron en los últimos 7 años en los siguientes porcentajes: 23%; 18%; 29%; 33%; 26%; 21%; 30%. Determina:

a) El porcentaje del incremento del último año con base en las ventas del primer año.

b) El promedio geométrico de incremento anual.

Verifica tus resultados en la sección de respuestas.

Moda

Se define como el valor que se presenta con la **mayor frecuencia**, es decir, es el **valor más común** de un conjunto de elementos numéricos dado.

Entre las características de la moda destacan que ésta puede no existir, incluso si existe puede no ser única.

Si un conjunto de valores dado presenta una sola moda se denomina **unimodal**. Si presenta dos modas se denomina **bimodal**; si presenta más de dos modas se denomina **multimodal**.

EJEMPLOS

1 Dado el siguiente conjunto de números 4, 6, 8, 10, 12, 14, 16, 18, 20; determina su moda.

Solución

Analizando el conjunto de números dado, se observa que todos ellos tienen igual frecuencia, por lo que se concluye que no presenta moda, es decir, no existe.

2 Dado el siguiente conjunto de números: 3, 3, 4, 6, 8, 8, 8, 11, 11, 13, 15; determina su moda.

Solución

Al analizar el conjunto de números dado, se observa que el número que se presenta con mayor frecuencia es el 8, por lo que se concluye que el valor de la moda es 8 y el conjunto se denomina unimodal.

3 Dado el siguiente conjunto de números: 10, 12, 12,12, 13, 14, 14, 14, 15, 15, 16, 17, 17, 17, 19; determina su moda.

Solución

Al analizar el conjunto de números dado, se observa que los números que se presentan con mayor frecuencia son el 12, 14 y 17, por lo que se concluye que los valores de la moda son 12, 14 y 17, en donde el conjunto se denomina multimodal.

Moda para datos agrupados

Para determinar la moda en datos agrupados se emplea el método de la interpolación, es decir, se fundamenta en encontrar primero la clase modal (es la clase en la cual se localiza el valor de la moda) que se caracteriza por tener la máxima frecuencia de la distribución.

El siguiente paso es determinar la diferencia absoluta entre la frecuencia de la clase modal y la frecuencia de la clase **premodal** o anterior; así mismo, la diferencia absoluta entre la frecuencia de la clase modal y la frecuencia de la clase **posmodal** o siguiente. Lo anterior es con el fin de observar hacia dónde se concentra el valor modal.

El cálculo de la moda de una distribución de frecuencias o de un histograma se determina matemáticamente por la ecuación:

$$\text{Moda} = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} C \right)$$

donde:

L_1 = Límite real inferior de la clase modal, es decir, la clase que contiene el valor de la moda.

Δ_1 = Es la diferencia absoluta entre la frecuencia de la clase modal y la frecuencia de la clase premodal o anterior.

Δ_2 = Es la diferencia absoluta entre la frecuencia de la clase modal y la frecuencia de la clase posmodal o siguiente.

C = Tamaño del intervalo de clase modal.

EJEMPLO

Ejemplo 1 La siguiente tabla de distribución de frecuencias agrupa los salarios semanales en miles de pesos de 130 empleados técnicos de la empresa *TRW*, división de cinturones de seguridad; determina la moda de dichos salarios.

Intervalo de clase	Frecuencia de clase (f)
654 - 677	3
678 - 701	7
702 - 725	13
726 - 749	19
750 - 773	26
774 - 797	21
798 - 821	17
822 - 845	12
846 - 869	8
870 - 893	4
	$N = 130$

Solución

Primero, se considera la tabla de distribución de frecuencias dada para localizar la clase modal que se caracteriza por tener la máxima frecuencia y el valor de la moda, que resulta ser (750 - 773).

El siguiente paso es determinar las diferencias absolutas, primero entre la frecuencia de la clase modal y la frecuencia de la clase premodal y enseguida respecto a la clase posmodal, es decir:

	Intervalo de clase	Frecuencia de clase (f)	
	654 - 677	3	
	678 - 701	7	
	702 - 725	13	
	726 - 749	19	$\Delta_1 = 26 - 19 = 7$
Clase modal →	750 - 773	26	
	774 - 797	21	$\Delta_2 = 26 - 21 = 5$
	798 - 821	17	
	822 - 845	12	
	846 - 869	8	
	870 - 893	4	
		$N = 130$	

También se observan los siguientes datos:

L_1 = El límite real de la clase modal se obtiene sumando el límite superior de un intervalo de clase y el límite inferior del intervalo de clase contiguo superior y dividiendo entre dos, es decir:

$$L_1 = \frac{749 + 750}{2} = 749.5$$

C = El tamaño del intervalo de la clase modal es 23.

Al sustituir los datos anteriores en la ecuación de la moda resulta:

$$\text{Moda} = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)C = 749.5 + \left(\frac{7}{7+5}\right)(23) = 749.5 + 13.42 = 762.92$$

La moda de los salarios calculada a partir de datos agrupados es 762.92 miles de pesos.

Cabe mencionar el valor resultante de la moda casi siempre se concentra aproximadamente hacia el valor de la marca de clase correspondiente a la clase modal, es decir, que el valor modal puede tender hacia la marca de clase de la clase premodal o hacia la marca de clase de la clase posmodal.

Ejercicio 10

Escribe los números correspondientes

Competencias genéricas	
Competencias disciplinares	

I. Realiza lo que se indica a continuación.

1. Define moda.
2. Cita algunas características que destacan a la moda.
3. Explica cómo se determina la moda para datos agrupados.
4. Escribe la ecuación matemática que se emplea para determinar la moda de datos agrupados.

II. Resuelve los siguientes problemas.

1. Veinticinco juegos de béisbol de la Liga Mexicana duraron, respectivamente 198, 202, 173, 186, 195, 202, 219, 217, 200, 217, 224, 181, 188, 202, 215, 199, 203, 218, 200, 202, 206, 183, 190 y 197 minutos. Determina.

   a) La media aritmética

   b) La mediana

   c) La moda

2. Al censar a 50 personas sobre cuál era su color favorito, los resultados fueron: azul, verde, verde, rojo, verde, morado, azul, verde, azul, azul, rojo, negro, verde, rojo, azul, verde, azul, rojo, rojo, amarillo, blanco, café, verde, verde, verde, azul, azul, morado, anaranjado, negro, rojo, verde, verde, negro, blanco, azul, verde, azul, café, rojo, café, verde, verde, anaranjado, azul, rojo, negro, amarillo, verde y azul. Determina:

   a) La media aritmética

   b) La mediana

   c) La moda

3. En un hospital, se le toma el pulso a 20 pacientes, registrando: 72, 76, 74, 82, 59, 96, 77, 65, 84, 86, 95, 78, 82, 89, 67, 82, 93, 93, 75, 82, 90. Determina:

   a) La media aritmética

   b) La mediana

   c) La moda

4. En 1970 y 1980 la población de la República Mexicana fue de 48 225 238 y 66 846 833 millones, respectivamente.
   a) ¿Cuál fue el promedio de incremento porcentual por año?
   b) ¿Cuál fue la población en 1976?
   c) Si el ritmo de incremento porcentual de la población por año desde 1970 a 1980 es el mismo que en a), ¿cuál será la población en 1990 y 2000?
5. Un capital de \$2 500 000 se invierte al 12.6% de interés anual, ¿cuál será el capital total después de 4 años si el capital inicial no se ha retirado?
6. Un capital de \$1 275 000 se invierte al 2.8% de interés trimestral, ¿cuál será el capital total después de 8 años si el capital inicial no se ha retirado?

III. Escribe en el paréntesis de la izquierda el número que corresponde a la respuesta correcta, tomándolo de la lista de la derecha.

( ) Valor simbólico y representación de un conjunto de datos.

Promedios que en un conjunto de datos ordenados de acuerdo con su magnitud, tienden a situarse al centro.

( ) Es la sumatoria de los valores de un conjunto de elementos dividido por el número total de ellos.

( ) Se determina por la ecuación: $\frac{\sum fX}{N}$

( ) Se representa por la ecuación matemática: $\frac{\sum WX}{N}$

( ) La suma algebraica de las desviaciones de un conjunto de elementos numéricos de su media aritmética es cero.

( ) Es el valor medio que divide a un conjunto de datos que se ordenan de acuerdo con su magnitud, en dos partes iguales.

( ) Es el valor más común de un conjunto de elementos numéricos dado.

( ) Es una forma matemática que se emplea para determinar esta medida.

( ) Se aplica en el cálculo de rendimientos medios, promedios de costos, velocidades medias, tiempos promedios, etcétera.

1. Media aritmética con frecuencias.
2. Media geométrica.
3. Propiedad de la media aritmética.
4. Mediana para datos agrupados.
5. Medidas de tendencia central.
6. Moda.
7. Media aritmética ponderada.
8. Promedio.
9. Media armónica.
10. Mediana.
11. Media aritmética.
12. $\log G = \frac{\sum f \log X}{N}$

Verifica tus resultados en la sección de respuestas.

Los cuantiles

Dado un conjunto de elementos que se ordenan de acuerdo con su magnitud, el valor medio que divide al conjunto de datos en dos partes iguales o la media aritmética de los dos valores medios de dicho conjunto, dan como resultado a la mediana.

Asociadas a la definición de mediana, tenemos otras medidas que se fundamentan en las divisiones proporcionales que pueden hacerse en datos agrupados o sin agrupar y que se denominan cuantiles. Existen diferentes tipos de cuantiles, a saber, **cuartiles**, **deciles** y **percentiles**.

Cuartiles

Se definen como los intervalos dentro de los cuales quedan proporcionalmente repartidos los datos sin agrupar o agrupados en una distribución formada por cuatro partes iguales.

Se tienen tres **cuartiles** que se simbolizan por Q_1 (primer cuartil), Q_2 (segundo cuartil) y Q_3 (tercer cuartil), en donde cada uno contendrá el mismo número de datos, es decir, el 25% del total.

El primer cuartil Q_1 es el valor que indica en el cual, o por debajo del cual, queda el 25% o un cuarto de todos los datos.

El segundo cuartil Q_2 es la medida igual a la mediana, es decir, es el valor que indica en el cual, o por debajo del cual, queda el 50% o la mitad de todos los datos.

El tercer cuartil Q_3 es el valor que indica en el cual, o por debajo del cual, queda el 75% o las tres cuartas partes de todos los datos.

El cálculo de los cuartiles de una distribución de frecuencias se determina matemáticamente por la ecuación general:

$$Q_k = L_1 + \left[\frac{\frac{KN}{4} - \left(\sum f\right)_1}{f_K}\right] C$$

donde:

K = K-ésimo cuartil = Se refiere al primero, segundo y tercer cuartil, según sea el que se tenga que determinar.
L_1 = Límite real inferior de la clase del K-ésimo cuartil, es decir, la clase que contiene el valor del K-ésimo cuartil.
N = Número total de datos, es decir, la frecuencia total.
$(\sum f)_1$ = Suma de las frecuencias de todas las clases por debajo de la clase de K-ésimo cuartil.
f_K = Frecuencia de la clase del K-ésimo cuartil, la cual nunca debe ser mayor que $\frac{KN}{10}$.
C = Tamaño del intervalo de la clase del K-ésimo cuartil.

Para determinar los **cuartiles** de un conjunto de datos sin agrupar, sólo es necesario aplicar los procedimientos empleados en el cálculo de la mediana.

Deciles

Se definen como los intervalos dentro de los cuales quedan proporcionalmente repartidos los datos sin agrupar o agrupados en una distribución formada por diez partes iguales. Se tienen nueve **deciles** que se simbolizan por D_1 (primer decil), D_2 (segundo decil), D_3 (tercer decil), …, D_9 (noveno decil), en donde cada uno contendrá el mismo número de datos, es decir, 10% del total.

El primer decil D_1 es el valor que indica en el cual, o por debajo del cual, queda el 10% o un décimo de todos los datos.

El segundo decil D_2 es el valor que indica en el cual, o por debajo del cual, queda el 20% o dos décimos de todos los datos.

El tercer decil D_3 es el valor que indica en el cual, o por debajo del cual, queda el 30% o tres décimos de todos los datos.

El cuarto decil D_4 es el valor que indica en el cual, o por debajo del cual, queda el 40% o cuatro décimos de todos los datos.

El quinto decil D_5 es la medida igual a la mediana, es decir, es el valor que indica en el cual o por debajo del cual queda el 50% o la mitad de todos los datos.

De manera similar, y en forma sucesiva tenemos que los deciles D_6, D_7, D_8 y D_9 son los valores que respectivamente indican en el cual, o por debajo del cual, quedan los porcentajes o décimas partes de todos los datos.

El cálculo de los deciles de una distribución de frecuencias se determinan matemáticamente por la ecuación general:

$$D_k = L_1 + \left[\frac{\frac{(KN)}{10} - \left(\sum f\right)_1}{f_K}\right] C$$

donde:

K = K-ésimo decil = Se refiere al primero, segundo, tercero, …, noveno decil, según sea el que se tenga que determinar.

L_1 = Límite real inferior de la clase del K-ésimo decil, es decir, la clase que contiene el valor del K-ésimo decil.

N = Número total de datos, es decir, la frecuencia total.

$(\sum f)_1$ = Suma de las frecuencias de todas las clases por debajo de la clase del K-ésimo cuartil.

f_K = Frecuencia de la clase del K-ésimo decil, la cual, nunca debe ser mayor que $\frac{KN}{10}$.

C = Tamaño del intervalo de la clase del K-ésimo decil.

Para determinar los deciles de un conjunto de datos sin agrupar sólo es necesario aplicar los procedimientos empleados en el cálculo de la mediana.

Percentiles

Se definen como los intervalos dentro de los cuales quedan proporcionalmente repartidos los datos sin agrupar o agrupados en una distribución formada por cien partes iguales.

Se tienen noventa y nueve **percentiles** que se simbolizan por P_1 (primer percentil), P_2 (segundo percentil), P_3 (tercer percentil), …, P_{99} (noventa y nueve percentil), en donde cada uno contendrá el mismo número de datos, es decir, el 1% del total.

El primer percentil P_1 es el valor que indica en el cual, o por debajo del cual, queda el 1% o un centésimo de todos los datos.

El segundo percentil P_2 es el valor que indica en el cual, o por debajo del cual, queda el 2 % o dos centésimos de todos los datos.

El tercer percentil P_3 es el valor que indica en el cual, o por debajo del cual, queda el 3% o tres centésimos de todos los datos.

De manera similar y en forma sucesiva tenemos que los percentiles $P_4, P_5, \ldots, P_{99}$ son los valores que respectivamente indican en el cual, o por debajo del cual, quedan los porcentajes o centésimas partes de todos los datos.

El vigésimo quinto percentil P_{25} es la medida igual al primer cuartil Q_1, es decir, es el valor que indica en el cual, o por debajo del cual, queda el 25% o un cuarto de todos los datos.

El quincuagésimo percentil P_{50} es la medida igual a la mediana, es decir, es el valor que indica en el cual, o por debajo del cual, queda el 50% o la mitad de todos los datos.

El septuagésimo quinto percentil P_{75} es la medida igual al tercer cuartil Q_3, es decir, es el valor que indica en el cual, o por debajo del cual, queda el 75% o tres cuartas partes de todos los datos.

Los percentiles P_{10}, P_{20}, P_{30}, P_{40}, P_{50}, P_{60}, P_{70}, P_{80} y P_{90}, son las medidas iguales a los deciles D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8 y D_9, respectivamente.

El cálculo de los **percentiles** de una distribución de frecuencias se determina matemáticamente por la ecuación general:

$$P_k = L_1 + \left[\frac{\frac{KN}{100} - \left(\sum f\right)_1}{f_K}\right] C$$

donde:

K = K-ésimo percentil. Se refiere al primero, segundo, tercero, …, nonogésimo noveno percentil, según sea el que se tenga que determinar.

L_1 = Límite real inferior de la clase del K-ésimo percentil, es decir, la clase que contiene el valor del K-ésimo percentil.
N = Número total de datos, es decir, la frecuencia total.
$(\sum f)_1$ = Suma de las frecuencias de todas las clases por debajo de la clase del K-ésimo percentil.
f_K = Frecuencia de la clase del K-ésimo percentil, la cual, nunca debe ser mayor que $\frac{KN}{10}$.
C = Tamaño del intervalo de la clase del K-ésimo percentil.

Para determinar los percentiles de un conjunto de datos sin agrupar, sólo es necesario aplicar los procedimientos empleados en el cálculo de la mediana.

EJEMPLOS

Ejemplos

1 En la garita fiscal del puente internacional Lic. Benito Juárez García de ciudad Reynosa, Tamaulipas, el número de vehículos que el semáforo fiscal indica aleatoriamente que deben revisarse por sus seis carriles de pase durante cada hora del día 1 de febrero de 2011, fueron: 27, 22, 19, 16, 28, 35, 33, 39, 54, 60, 53, 48, 65, 76, 68, 83, 89, 92, 103, 85, 57, 43, 46, 41. Determina los cuartiles Q_1, Q_2, Q_3, los deciles D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8 y D_9, los percentiles P_1, P_2, P_3, P_{10}, P_{15}, P_{35}, P_{45}, P_{55}, P_{65}, P_{85}, P_{95} y P_{99}.

Solución

Si ordenamos en forma ascendente los datos dados tenemos:

16, 19, 22, 27, 28, 33, 35, 39, 41, 43, 46, 48, 53, 54, 57, 60, 65, 68, 76, 83, 85, 89, 92, 103

Se observa que el conjunto tiene $N = 24$ datos.
Para determinar el valor de los cuartiles, aplicamos la expresión general $\frac{KN}{4}$ y las siguientes reglas:

a) Si KN es divisible exactamente entre 4, el valor del cuartil buscado será la media aritmética entre el dato ordenado cuya posición se obtiene de $\frac{KN}{4}$ y el siguiente dato de orden.

b) Si KN no es divisible exactamente entre 4, el valor del cuartil buscado será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado de $\frac{KN}{4}$.

Para el primer cuartil ($K = 1$), tenemos: $\frac{KN}{4} = \frac{1(24)}{4} = 6$

El valor del cuartil Q_1 será la media aritmética entre el 6º dato ordenado y el siguiente del conjunto, es decir:

$$Q_1 = \frac{33+35}{2} = 34$$ El valor del primer cuartil Q_1 es de 34.

Para el segundo cuartil ($K = 2$), tenemos: $\frac{KN}{4} = \frac{2(24)}{4} = 12$

El valor del cuartil Q_2 será la media aritmética entre el 12º dato ordenado y el siguiente del conjunto, es decir:

$$Q_2 = \frac{48+53}{2} = 50.5$$

El valor del segundo cuartil Q_2 es 50.5, equivalente a la **mediana**.

Para el tercer cuartil ($K = 3$), tenemos: $\frac{KN}{4} = \frac{3(24)}{4} = 18$

El valor del cuartil Q_3 será la media aritmética entre el 18º dato ordenado y el siguiente del conjunto, es decir:

$$Q_3 = \frac{68+76}{2} = 72$$

El valor del tercer cuartil Q_3 es 72.

Para determinar el valor de los deciles, aplicamos la expresión matemática $\frac{KN}{10}$ y adecuamos las reglas de los cuartiles.

Para el primer decil ($K = 1$), tenemos: $\frac{KN}{10} = \frac{1(24)}{10} = 2.4$

El valor del decil D_1 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{10}$, es decir, 2.4 redondeado al entero más próximo es 2.

El valor del primer decil D_1 es 19, ya que es el dato ordenado que ocupa la posición dos.

Para el segundo decil ($K = 2$), tenemos: $\frac{KN}{10} = \frac{2(24)}{10} = 4.8$

El valor del decil D_2 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{10}$, es decir, 4.8 redondeado al entero más próximo es 5.

El valor del segundo decil es 28, ya que es el dato ordenado que ocupa la posición cinco.

Para el tercer decil ($K = 3$), tenemos: $\frac{KN}{10} = \frac{3(24)}{10} = 7.2$

El valor del decil D_3 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{10}$, es decir, 7.2 redondeado al entero más próximo es 7.

El valor del tercer decil es 35, ya que es el dato ordenado que ocupa la posición siete.

Para el cuarto decil ($K = 4$), tenemos: $\frac{KN}{10} = \frac{4(24)}{10} = 9.6$

El valor del decil D_4 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{10}$, es decir, 9.6 redondeado al entero más próximo es 10.

El valor del cuarto decil es 43, ya que es el dato ordenado que ocupa la posición diez.

Para el quinto decil ($K = 5$), tenemos: $\frac{KN}{10} = \frac{5(24)}{10} = 12$

El valor del decil D_5 será el dato ordenado cuya posición es 12.

El valor del quinto decil es 48, ya que es el dato ordenado que ocupa la posición doce.

Para el sexto decil ($K = 6$), tenemos: $\frac{KN}{10} = \frac{6(24)}{10} = 14.4$

El valor del decil D_6 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{10}$, es decir, 14.4 redondeado al entero más próximo es 14.

El valor del sexto decil es 54, ya que es el dato ordenado que ocupa la posición catorce.

Para el séptimo decil ($K = 7$), tenemos: $\frac{KN}{10} = \frac{7(24)}{10} = 16.8$

El valor del decil D_7 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{10}$, es decir, 16.8 redondeado al entero más próximo es 17.

El valor del séptimo decil es 65, ya que es el dato ordenado que ocupa la posición diecisiete.

Para el octavo decil ($K = 8$), tenemos: $\frac{KN}{10} = \frac{8(24)}{10} = 19.2$

El valor del decil D_8 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{10}$, es decir, 19.2 redondeado al entero más próximo es 19.

El valor del octavo decil es 76, ya que es el dato ordenado que ocupa la posición diecinueve.

Para el noveno decil ($K = 9$), tenemos: $\frac{KN}{10} = \frac{9(24)}{10} = 21.6$

El valor del decil D_9 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{10}$, es decir, 21.6 redondeado al entero más próximo es 22.

El valor del noveno decil es 89, ya que es el dato ordenado que ocupa la posición veintidós.

Para determinar el valor de los percentiles, aplicamos la expresión matemática $\frac{KN}{100}$ y adecuamos las reglas de los cuartiles.

Para el primer percentil ($K = 1$), tenemos: $\frac{KN}{100} = \frac{1(24)}{100} = 0.24$

El valor del percentil P_1 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir, 0.24 redondeado al entero más próximo es 0.

El valor del primer percentil P_1 es 0, ya que no existe un dato ordenado que ocupe la posición cero en el conjunto.

Para el segundo percentil ($K = 2$), tenemos: $\frac{KN}{100} = \frac{2(24)}{100} = 0.48$

El valor del percentil P_2 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir, 0.48 redondeado al entero más próximo es 0.

El valor del segundo percentil P_2 es 0 ya que no existe un dato ordenado que ocupe la posición cero en el conjunto.

Para el tercer percentil ($K = 3$), tenemos: $\frac{KN}{100} = \frac{3(24)}{100} = 0.72$

El valor del percentil P_3 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir, 0.72 redondeado al entero más próximo es 1.

El valor del tercer percentil P_3 es16, ya que es el dato ordenado que ocupa la posición uno.

Para el décimo percentil ($K = 10$), tenemos: $\frac{KN}{100} = \frac{10(24)}{100} = 2.4$

El valor del percentil P_{10} será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir 2.4 redondeado al entero más próximo es 2.

El valor del décimo percentil es 19 ya que es el dato ordenado que ocupa la posición dos y es equivalente al valor del primer decil D_1.

Para el decimoquinto percentil ($K = 15$), tenemos: $\frac{KN}{100} = \frac{15(24)}{100} = 3.6$

El valor del percentil P_{15} será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir, 3.6 redondeado al entero más próximo es 4.

El valor del décimo percentil es 27, ya que es el dato ordenado que ocupa la posición cuatro.

Para el trigésimo quinto percentil ($K = 35$), tenemos: $\frac{KN}{100} = \frac{35(24)}{100} = 8.4$

El valor del percentil P_{35} será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir, 8.4 redondeado al entero más próximo es 8.

El valor del trigésimo quinto percentil es 39, ya que es el dato ordenado que ocupa la posición ocho.

Para el cuadragésimo quinto percentil ($K = 45$), tenemos: $\frac{KN}{100} = \frac{45(24)}{100} = 10.8$

El valor del percentil P_{45} será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir, 10.8 redondeado al entero más próximo es 11.

El valor del cuadragésimo quinto percentil es 46, ya que es el dato ordenado que ocupa la posición once.

Para el quincuagésimo quinto percentil ($K = 55$), tenemos: $\frac{KN}{100} = \frac{55(24)}{100} = 13.2$

El valor del percentil P_{55} será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir, 13.2 redondeado al entero más próximo es 13.

El valor del quincuagésimo quinto percentil es 53, ya que es el dato ordenado que ocupa la posición trece.

Para el sexagésimo quinto percentil ($K = 65$), tenemos: $\frac{KN}{100} = \frac{65(24)}{100} = 15.6$

El valor del percentil P_{65} será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir, 15.6 redondeado al entero más próximo es 16.

El valor del sexagésimo quinto percentil es 60, ya que es el dato ordenado que ocupa la posición dieciséis.

Para el octagésimo quinto percentil ($K = 85$), tenemos: $\frac{KN}{100} = \frac{85(24)}{100} = 20.4$

El valor del percentil P_{85} será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir, 20.4 redondeado al entero más próximo es 20.

El valor del octagésimo quinto percentil es 83, ya que es el dato ordenado que ocupa la posición veinte.

Para el nonagésimo quinto percentil ($K = 95$), tenemos: $\frac{KN}{100} = \frac{95(24)}{100} = 22.8$

El valor del percentil P_{95} será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir, 22.8 redondeado al entero más próximo es 23.

El valor del nonagésimo quinto percentil es 92, ya que es el dato ordenado que ocupa la posición veintitrés.

Para el nonagésimo noveno percentil ($K = 99$), tenemos: $\frac{KN}{100} = \frac{99(24)}{100} = 23.76$

El valor del percentil P_{99} será el dado ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir, 23.76 redondeado al entero más próximo es 24.

El valor del nonagésimo noveno percentil es 103, ya que es el dato ordenado que ocupa la posición veinticuatro.

2 Determina los cuartiles Q_1 y Q_3, los deciles D_3, D_6 y D_9 y los percentiles P_{40}, P_{50}, P_{70} y P_{80}, para el siguiente conjunto de datos ordenados: 76, 77, 81, 82, 83, 88, 92, 95, 99, 102, 105, 107, 108, 110, 112, 116, 119, 123, 125, 128, 131, 135, 137, 140, 142, 146, 149, 152, 155, 158, 161, 164, 166, 168 y 171.

Solución

Como el conjunto tiene $N = 35$ datos, se tiene:

Para el primer cuartil ($K = 1$), tenemos: $\frac{KN}{4} = \frac{1(35)}{4} = 8.75$

El valor del cuartil Q_1 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{4}$, es decir, 8.75 redondeado al entero más próximo es 9.

El valor del primer cuartil es 99, ya que es el dato ordenado que ocupa la posición nueve.

Para el tercer cuartil ($K = 3$), tenemos: $\frac{KN}{4} = \frac{3(35)}{4} = 26.25$

El valor del cuartil Q_3 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{4}$, es decir, 26.25 redondeado al entero más próximo es 26.

El valor del tercer cuartil es 146, ya que es el dato ordenado que ocupa la posición veintiséis.

Para el tercer decil ($K = 3$), tenemos: $\frac{KN}{10} = \frac{3(35)}{10} = 10.5$

El valor del decil D_3 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{10}$, es decir, 10.5 redondeado al entero más próximo es 11.

El valor del tercer decil es 105, ya que es el dato ordenado que ocupa la posición once.

Para el sexto decil ($K = 6$), tenemos: $\frac{KN}{10} = \frac{6(35)}{10} = 21$

El valor del decil D_6 será la media aritmética entre el 21° dato ordenado y el siguiente del conjunto, es decir,

$$D_6 = \frac{131 + 135}{2} = 133$$

El valor del sexto decil D_6 es 133.

Para el noveno decil ($K = 9$), tenemos: $\frac{KN}{10} = \frac{9(35)}{10} = 31.5$

El valor del decil D_9 será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{10}$, es decir, 31.5 redondeado al entero más próximo es 32.

El valor del noveno decil es 164, ya que es el dato ordenado que ocupa la posición treinta y dos.

Para el cuadragésimo percentil ($K = 40$), tenemos: $\frac{KN}{100} = \frac{40(35)}{100} = 14$

El valor del percentil P_{40} será la media aritmética entre el 14º dato ordenado y el siguiente del conjunto, es decir,

$$P_{40} = \frac{110 + 112}{2} = 111$$

El valor del cuadragésimo percentil P_{40} es 111.

Para el quincuagésimo percentil ($K = 50$), tenemos: $\frac{KN}{100} = \frac{50(35)}{100} = 17.5$

El valor del percentil P_{50} será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir, 17.5 redondeado al entero más próximo es 18.

El valor del quincuagésimo percentil es 123, ya que es el dato ordenado que ocupa la posición dieciocho.

Para el septuagésimo percentil ($K = 70$), tenemos: $\frac{KN}{100} = \frac{70(35)}{100} = 24.5$

El valor del percentil P_{70} será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100}$, es decir, 24.5 redondeado al entero más próximo es 25.

El valor del septuagésimo percentil es 142, ya que es el dato ordenado que ocupa la posición veinticinco.

Para el octogésimo percentil ($K = 80$), tenemos: $\frac{KN}{100} = \frac{80(35)}{100} = 28$

El valor del percentil P_{80} será la media aritmética entre el 28º dato ordenado y el siguiente del conjunto, es decir,

$$P_{80} = \frac{152 + 155}{2} = 153.5$$

El valor del octogésimo percentil P_{80} es 153.5.

EJEMPLOS

Ejemplos

1 La siguiente tabla de distribución de frecuencias registra los salarios semanales en miles de pesos de 75 empleados de la compañía *Festo Didactic*. Determina los cuartiles Q_1, Q_2 y Q_3, los deciles $D_1, D_2, D_3, \ldots, D_9$ y los percentiles $P_1, P_2, P_3, P_4, P_5, P_{15}, P_{25}, P_{35}, P_{40}, P_{50}, P_{65}, P_{75}, P_{90}$ y P_{95}.

Intervalos (salarios)	Frecuencia (empleados)
110 - 119	5
120 - 129	9
130 - 139	19
140 - 149	15
150 - 159	11
160 - 169	7
170 - 179	4
180 - 189	3
190 - 199	2
	$N = 75$

Solución

Para determinar el primer cuartil Q_1, usamos los siguientes datos.

Datos

$$\frac{KN}{4}=\frac{1(75)}{4}=18.75$$
$$L_1=\frac{129+130}{2}=129.5$$
$$(\Sigma f)_1=5+9=14$$
$$f_K=19$$
$$C=10$$

Fórmula

$$Q_1=L_1+\left[\frac{\frac{KN}{4}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$Q_1=129.5+\left[\frac{18.75-14}{19}\right](10)$$
$$Q_1=129.5+2.5=132$$

El valor del primer cuartil es \$132, es decir, que el 25% de los empleados gana \$132 o menos.

Para determinar el segundo cuartil Q_2, usamos los siguientes datos.

Datos

$$\frac{KN}{4}=\frac{2(75)}{4}=37.5$$
$$L_1=\frac{139+140}{2}=139.5$$
$$(\Sigma f)_1=5+9+19=33$$
$$f_K=15$$
$$C=10$$

Fórmula

$$Q_2=L_1+\left[\frac{\frac{KN}{4}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$Q_2=139.5+\left[\frac{37.5-33}{15}\right](10)$$
$$Q_2=139.5+3=142.5$$

El valor del segundo cuartil es equivalente a la mediana, es \$142.5, es decir, 50% de los empleados gana \$142.5 o menos.

Para determinar el tercer cuartil Q_3, usamos los siguientes datos.

Datos

$$\frac{KN}{4}=\frac{3(75)}{4}=56.25$$
$$L_1=\frac{149+150}{2}=149.5$$
$$(\Sigma f)_1=5+9+19+15=48$$
$$f_K=11$$
$$C=10$$

Fórmula

$$Q_3=L_1+\left[\frac{\frac{KN}{4}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$Q_3=149.5+\left[\frac{56.25-48}{11}\right](10)$$
$$Q_3=149.5+7.5=157$$

El valor del tercer cuartil es \$157, es decir, 75% de los empleados gana \$157 o menos.

Para determinar el primer decil D_1, se tiene:

Datos

$$\frac{KN}{10}=\frac{1(75)}{10}=7.5$$
$$L_1=\frac{119+120}{2}=119.5$$
$$(\Sigma f)_1=5$$
$$f_K=9$$
$$C=10$$

Fórmula

$$D_1=L_1+\left[\frac{\frac{KN}{10}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$D_1=119.5+\left[\frac{7.5-5}{9}\right](10)$$
$$D_1=119.5+2.78=122.28$$

El valor del primer decil es \$122.28, es decir, 10% de los empleados gana \$122.28 o menos.

Nota: los resultados que se presentan están en miles de pesos.

Para determinar el segundo decil D_2, se tiene:

Datos

$$\frac{KN}{10}=\frac{2(75)}{10}=15$$
$$L_1=\frac{129+130}{2}=129.5$$
$$(\Sigma f)_1=5+9=14$$
$$f_K=19$$
$$C=10$$

Fórmula

$$D_2=L_1+\left[\frac{\frac{KN}{10}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$D_2=129.5+\left[\frac{15-14}{19}\right](10)$$
$$D_2=129.5+0.53=130.03$$

El valor del segundo decil es \$130.03, es decir, 20% de los empleados gana \$130.03 o menos.

Para determinar el tercer decil D_3, se tiene:

Datos

$$\frac{KN}{10}=\frac{3(75)}{10}=22.5$$
$$L_1=\frac{129+130}{2}=129.5$$
$$(\Sigma f)_1=5+9=14$$
$$f_K=19$$
$$C=10$$

Fórmula

$$D_3=L_1+\left[\frac{\frac{KN}{10}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$D_3=129.5+\left[\frac{22.5-14}{19}\right](10)$$
$$D_2=129.5+4.47=133.97$$

El valor del tercer decil es \$133.97, es decir, 30% de los empleados gana \$133.97 o menos.

Para determinar el cuarto decil D_4, se tiene:

Datos

$$\frac{KN}{10}=\frac{4(75)}{10}=30$$
$$L_1=\frac{129+130}{2}=129.5$$
$$(\Sigma f)_1=5+9=14$$
$$f_K=19$$
$$C=10$$

Fórmula

$$D_4=L_1+\left[\frac{\frac{KN}{10}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$D_4=129.5+\left[\frac{30-14}{19}\right](10)$$
$$D_4=129.5+8.42=137.92$$

El valor del cuarto decil es \$137.92, es decir, 40% de los empleados gana \$137.92 o menos.

Para determinar el quinto decil D_5, se tiene:

Datos

$$\frac{KN}{10}=\frac{5(75)}{10}=37.5$$
$$L_1=\frac{139+140}{2}=139.5$$
$$(\Sigma f)_1=5+9+19=33$$
$$f_K=15$$
$$C=10$$

Fórmula

$$D_5=L_1+\left[\frac{\frac{KN}{10}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$D_5=139.5+\left[\frac{37.5-33}{15}\right](10)$$
$$D_5=139.5+3=142.5$$

El valor del quinto decil que es equivalente al segundo cuartil y a la mediana, es \$142.5, es decir, 50% de los empleados gana \$142.5 o menos.

Nota: los resultados que se presentan están en miles de pesos.

Para determinar el sexto decil D_6, se tiene:

Datos

$$\frac{KN}{10}=\frac{6(75)}{10}=45$$
$$L_1=\frac{139+140}{2}=139.5$$
$$\left(\sum f\right)_1=5+9+19=33$$
$$f_K=15$$
$$C=10$$

Fórmula

$$D_6=L_1+\left[\frac{\frac{KN}{10}-\left(\sum f\right)_1}{f_K}\right]C$$

Sustitución

$$D_6=139.5+\left[\frac{45-33}{15}\right](10)$$
$$D_6=139.5+8=147.5$$

El valor del sexto decil es \$147.5, es decir, 60% de los empleados gana \$147.5 o menos.

Para determinar el séptimo decil D_7, se tiene:

Datos

$$\frac{KN}{10}=\frac{7(75)}{10}=52.5$$
$$L_1=\frac{149+150}{2}=149.5$$
$$\left(\sum f\right)_1=5+9+19+15=48$$
$$f_K=11$$
$$C=10$$

Fórmula

$$D_7=L_1+\left[\frac{\frac{KN}{10}-\left(\sum f\right)_1}{f_K}\right]C$$

Sustitución

$$D_7=149.5+\left[\frac{52.5-48}{11}\right](10)$$
$$D_7=149.5+4.09=153.59$$

El valor del séptimo decil es \$153.59, es decir, 70% de los empleados gana \$153.59 o menos.

Para determinar el octavo decil D_8, se tiene:

Datos

$$\frac{KN}{10}=\frac{8(75)}{10}=60$$
$$L_1=\frac{159+160}{2}=159.5$$
$$\left(\sum f\right)_1=5+9+19+15+11=59$$
$$f_K=7$$
$$C=10$$

Fórmula

$$D_8=L_1+\left[\frac{\frac{KN}{10}-\left(\sum f\right)_1}{f_K}\right]C$$

Sustitución

$$D_8=159.5+\left[\frac{60-59}{7}\right](10)$$
$$D_8=159.5+1.43=160.93$$

El valor del octavo decil es \$160.93, es decir, 80% de los empleados gana \$160.93 o menos.

Para determinar el noveno decil D_9, se tiene:

Datos

$$\frac{KN}{10}=\frac{9(75)}{10}=67.5$$
$$L_1=\frac{169+170}{2}=169.5$$
$$\left(\sum f\right)_1=5+9+19+15+11+7=66$$
$$f_K=4$$
$$C=10$$

Fórmula

$$D_9=L_1+\left[\frac{\frac{KN}{10}-\left(\sum f\right)_1}{f_K}\right]C$$

Sustitución

$$D_9=169.5+\left[\frac{67.5-66}{4}\right](10)$$
$$D_9=169.5+3.75=173.25$$

El valor del noveno decil es \$173.25, es decir, 90% de los empleados gana \$173.25 o menos.

Nota: los resultados que se presentan están en miles de pesos.

Para determinar el primer percentil P_1, tenemos de la tabla:

Datos

$$\frac{KN}{100}=\frac{1(75)}{100}=0.75$$
$$L_1=\frac{109+110}{2}=109.5$$
$$(\Sigma f)_1=0$$
$$f_K=5$$
$$C=10$$

Fórmula

$$P_1=L_1+\left[\frac{\frac{KN}{100}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_1=109.5+\left[\frac{0.75-0}{5}\right](10)$$
$$P_1=109.5+1.5=111$$

El valor del primer percentil es \$111, es decir, 1% de los empleados gana \$111 o menos.

Para determinar el segundo percentil P_2, tenemos de la tabla:

Datos

$$\frac{KN}{100}=\frac{2(75)}{100}=1.5$$
$$L_1=\frac{109+110}{2}=109.5$$
$$(\Sigma f)_1=0$$
$$f_K=5$$
$$C=10$$

Fórmula

$$P_2=L_1+\left[\frac{\frac{KN}{100}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_2=109.5+\left[\frac{1.5-0}{5}\right](10)$$
$$P_2=109.5+3=112.5$$

El valor del segundo percentil es \$112.5, es decir, 2% de los empleados gana \$112.5 o menos.

Para determinar el tercer percentil P_3, tenemos de la tabla:

Datos

$$\frac{KN}{100}=\frac{3(75)}{100}=2.25$$
$$L_1=\frac{109+110}{2}=109.5$$
$$(\Sigma f)_1=0$$
$$f_K=5$$
$$C=10$$

Fórmula

$$P_3=L_1+\left[\frac{\frac{KN}{100}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_3=109.5+\left[\frac{2.25-0}{5}\right](10)$$
$$P_3=109.5+4.5=114$$

El valor del tercer percentil es \$114, es decir, 3% de los empleados gana \$114 o menos.

Para determinar el cuarto percentil P_4, tenemos de la tabla:

Datos

$$\frac{KN}{100}=\frac{4(75)}{100}=3$$
$$L_1=\frac{109+110}{2}=109.5$$
$$(\Sigma f)_1=0$$
$$f_K=5$$
$$C=10$$

Fórmula

$$P_4=L_1+\left[\frac{\frac{KN}{100}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_4=109.5+\left[\frac{3-0}{5}\right](10)$$
$$P_4=109.5+6=115.5$$

El valor del cuarto percentil es \$115.5, es decir, 4% de los empleados gana \$115.5 o menos.

Nota: los resultados que se presentan están en miles de pesos.

Para determinar el quinto percentil P_5, tenemos de la tabla:

Datos

$$\frac{KN}{100}=\frac{5(75)}{100}=3.75$$
$$L_1=\frac{109+110}{2}=109.5$$
$$(\Sigma f)_1=0$$
$$f_K=5$$
$$C=10$$

Fórmula

$$P_5=L_1+\left[\frac{\frac{KN}{100}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_5=109.5+\left[\frac{3.75-0}{5}\right](10)$$
$$P_5=109.5+7.5=117$$

El valor del quinto percentil es \$117, es decir, 5% de los empleados gana \$117 o menos.

Para determinar el decimoquinto percentil P_{15}, tenemos de la tabla:

Datos

$$\frac{KN}{100}=\frac{15(75)}{100}=11.25$$
$$L_1=\frac{119+120}{2}=119.5$$
$$(\Sigma f)_1=5$$
$$f_K=9$$
$$C=10$$

Fórmula

$$P_{15}=L_1+\left[\frac{\frac{KN}{100}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_{15}=119.5+\left[\frac{11.25-5}{9}\right](10)$$
$$P_{15}=119.5+6.94=126.44$$

El valor del decimoquinto percentil es \$126.44, es decir, 15% de los empleados gana \$126.44 o menos.

Para determinar el vigésimo quinto percentil P_{25}, tenemos de la tabla:

Datos

$$\frac{KN}{100}=\frac{25(75)}{100}=18.75$$
$$L_1=\frac{129+130}{2}=129.5$$
$$(\Sigma f)_1=5+9=14$$
$$f_K=19$$
$$C=10$$

Fórmula

$$P_{25}=L_1+\left[\frac{\frac{KN}{100}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_{25}=129.5+\left[\frac{18.75-14}{19}\right](10)$$
$$P_{25}=129.5+2.5=132$$

El valor del vigésimo quinto percentil que es equivalente al primer cuartil, es \$132, es decir, 25% de los empleados gana \$132 o menos.

Para determinar el trigésimo quinto percentil P_{35}, tenemos de la tabla:

Datos

$$\frac{KN}{100}=\frac{35(75)}{100}=26.25$$
$$L_1=\frac{129+130}{2}=129.5$$
$$(\Sigma f)_1=5+9=14$$
$$f_K=19$$
$$C=10$$

Fórmula

$$P_{35}=L_1+\left[\frac{\frac{KN}{100}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_{35}=129.5+\left[\frac{26.25-14}{19}\right](10)$$
$$P_{35}=129.5+6.45=135.95$$

El valor del trigésimo quinto percentil es \$135.95, es decir, 35% de los empleados gana \$135.95 o menos.

Nota: los resultados que se presentan están en miles de pesos.

Para determinar el cuadragésimo percentil P_{40}, tenemos de la tabla:

Datos

$$\frac{KN}{100} = \frac{40(75)}{100} = 30$$

$$L_1 = \frac{129+130}{2} = 129.5$$

$$\left(\sum f\right)_1 = 5+9 = 14$$

$$f_K = 19$$

$$C = 10$$

Fórmula

$$P_{40} = L_1 + \left[\frac{\frac{KN}{100} - \left(\sum f\right)_1}{f_K}\right]C$$

Sustitución

$$P_{40} = 129.5 + \left[\frac{30-14}{19}\right](10)$$

$$P_{40} = 129.5 + 8.42 = 137.92$$

El valor del cuadragésimo percentil que es equivalente al cuarto decil es $137.92, es decir, 40% de los empleados gana $137.92 o menos.

Para determinar el quincuagésimo percentil P_{50}, tenemos de la tabla:

Datos

$$\frac{KN}{100} = \frac{50(75)}{100} = 37.5$$

$$L_1 = \frac{139+140}{2} = 139.5$$

$$\left(\sum f\right)_1 = 5+9+19 = 33$$

$$f_K = 15$$

$$C = 10$$

Fórmula

$$P_{50} = L_1 + \left[\frac{\frac{KN}{100} - \left(\sum f\right)_1}{f_K}\right]C$$

Sustitución

$$P_{50} = 139.5 + \left[\frac{37.5-33}{15}\right](10)$$

$$P_{50} = 139.5 + 3 = 142.5$$

El valor del quincuagésimo percentil que es equivalente al segundo cuartil, al quinto decil y a la mediana, es $142.5, es decir, 50% de los empleados gana $142.5 o menos.

Para determinar el sexagésimo quinto percentil P_{65}, tenemos de la tabla:

Datos

$$\frac{KN}{100} = \frac{65(75)}{100} = 48.75$$

$$L_1 = \frac{149+150}{2} = 149.5$$

$$\left(\sum f\right)_1 = 5+9+19+15 = 48$$

$$f_K = 11$$

$$C = 10$$

Fórmula

$$P_{65} = L_1 + \left[\frac{\frac{KN}{100} - \left(\sum f\right)_1}{f_K}\right]C$$

Sustitución

$$P_{65} = 149.5 + \left[\frac{48.75-48}{11}\right](10)$$

$$P_{65} = 149.5 + 0.68 = 150.18$$

El valor del sexagésimo quinto percentil es $150.18, es decir, 65% de los empleados gana $150.18 o menos.

Para determinar el septuagésimo quinto percentil P_{75}, tenemos de la tabla:

Datos

$$\frac{KN}{100} = \frac{75(75)}{100} = 56.25$$

$$L_1 = \frac{149+150}{2} = 149.5$$

$$\left(\sum f\right)_1 = 5+9+19+15 = 48$$

$$f_K = 11$$

$$C = 10$$

Fórmula

$$P_{75} = L_1 + \left[\frac{\frac{KN}{100} - \left(\sum f\right)_1}{f_K}\right]C$$

Sustitución

$$P_{75} = 149.5 + \left[\frac{56.25-48}{11}\right](10)$$

$$P_{75} = 149.5 + 7.5 = 157$$

El valor del septuagésimo quinto percentil que es equivalente al tercer cuartil es $157, es decir, 75% de los empleados gana $157 o menos.

Nota: los resultados que se presentan están en miles de pesos.

Para determinar el nonagésimo percentil P_{90}, tenemos de la tabla:

Datos

$$\frac{KN}{100} = \frac{90(75)}{100} = 67.5$$

$$L_1 = \frac{169+170}{2} = 169.5$$

$$(\Sigma f)_1 = 5+9+19+15+11+7 = 66$$

$$f_K = 4$$

$$C = 10$$

Fórmula

$$P_{90} = L_1 + \left[\frac{\frac{KN}{100} - (\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_{90} = 169.5 + \left[\frac{67.5-66}{4}\right](10)$$

$$P_{90} = 169.5 + 3.75 = 173.25$$

El valor del nonagésimo percentil que es equivalente al noveno decil es $173.25, es decir, 90% de los empleados gana $173.25 o menos.

Para determinar el nonagésimo quinto percentil P_{95}, tenemos de la tabla:

Datos

$$\frac{KN}{100} = \frac{95(75)}{100} = 71.25$$

$$L_1 = \frac{179+180}{2} = 179.5$$

$$(\Sigma f)_1 = 5+9+19+15+11+7+4 = 70$$

$$f_K = 3$$

$$C = 10$$

Fórmula

$$P_{95} = L_1 + \left[\frac{\frac{KN}{100} - (\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_{95} = 179.5 + \left[\frac{71.25-70}{3}\right](10)$$

$$P_{95} = 179.5 + 4.17 = 183.67$$

El valor del nonagésimo quinto percentil es $183.67, es decir, 95% de los empleados gana $183.67 o menos.

Nota: los resultados que se presentan están en miles de pesos.

2 La siguiente tabla de distribución de frecuencias registra los pesos en libras de 42 atletas mexicanos que participaron en los *XXIX Juegos Olímpicos* en Beijing, en 2008. Encuentra:

a) Cuartiles Q_1 y Q_3.
b) Deciles D_3, D_6 y D_9.
c) Percentiles P_{35}, P_{50}, P_{65} y P_{80}.

Intervalos (pesos)	Frecuencia (atletas)
110 - 116	2
117 - 123	4
124 - 130	7
131 - 137	9
138 - 144	13
145 - 151	6
152 - 158	1
	$N = 42$

Solución

a) Para determinar el primer cuartil Q_1, usamos los siguientes datos:

Datos

$$\frac{KN}{4} = \frac{1(42)}{4} = 10.5$$

$$L_1 = \frac{123+124}{2} = 123.5$$

$$(\Sigma f)_1 = 2+4 = 6$$

$$f_K = 7$$

$$C = 7$$

Fórmula

$$Q_1 = L_1 + \left[\frac{\frac{KN}{4} - (\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$Q_1 = 123.5 + \left[\frac{10.5-6}{7}\right](7)$$

$$Q_1 = 123.5 + 4.5 = 128$$

El valor del primer cuartil es 128 libras, es decir, 25% de los atletas pesa 128 libras o menos.

Para determinar el tercer cuartil Q_3, usamos los siguientes datos:

Datos

$$\frac{KN}{4} = \frac{3(42)}{4} = 31.5$$
$$L_1 = \frac{137+138}{2} = 137.5$$
$$(\Sigma f)_1 = 2+4+7+9 = 22$$
$$f_K = 13$$
$$C = 7$$

Fórmula

$$Q_3 = L_1 + \left[\frac{\frac{KN}{4} - (\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$Q_3 = 137.5 + \left[\frac{31.5-22}{13}\right](7)$$
$$Q_3 = 137.5 + 5.12 = 142.62$$

El valor del tercer cuartil es 142.62 libras, es decir, 75% de los atletas pesa 142.62 libras o menos.

b) Para determinar el tercer decil D_3, se tiene:

Datos

$$\frac{KN}{10} = \frac{3(42)}{10} = 12.6$$
$$L_1 = \frac{123+124}{2} = 123.5$$
$$(\Sigma f)_1 = 2+4 = 6$$
$$f_K = 7$$
$$C = 7$$

Fórmula

$$D_3 = L_1 + \left[\frac{\frac{KN}{10} - (\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$D_3 = 123.5 + \left[\frac{12.6-6}{7}\right](7)$$
$$D_3 = 123.5 + 6.6 = 130.1$$

El valor del tercer decil es 130.1 libras, es decir, 30% de los atletas pesa 130.1 libras o menos.

Para determinar el sexto decil D_6, se tiene:

Datos

$$\frac{KN}{10} = \frac{6(42)}{10} = 25.2$$
$$L_1 = \frac{137+138}{2} = 137.5$$
$$(\Sigma f)_1 = 2+4+7+9 = 22$$
$$f_K = 13$$
$$C = 7$$

Fórmula

$$D_6 = L_1 + \left[\frac{\frac{KN}{10} - (\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$D_6 = 137.5 + \left[\frac{25.2-22}{13}\right](7)$$
$$D_6 = 137.5 + 1.72 = 139.22$$

El valor del sexto decil es 139.22 libras, es decir, 60% de los atletas pesa 139.22 libras o menos.

Para determinar el noveno decil D_9, se tiene:

Datos

$$\frac{KN}{10} = \frac{9(42)}{10} = 37.8$$
$$L_1 = \frac{144+145}{2} = 144.5$$
$$(\Sigma f)_1 = 2+4+7+9+13 = 35$$
$$f_K = 6$$
$$C = 7$$

Fórmula

$$D_9 = L_1 + \left[\frac{\frac{KN}{10} - (\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$D_9 = 144.5 + \left[\frac{37.8-35}{6}\right](7)$$
$$D_9 = 144.5 + 3.27 = 147.77$$

El valor del noveno decil es 147.77 libras, es decir, 90% de los atletas pesa 147.77 libras o menos.

c) Para determinar el trigésimo quinto percentil P_{35}, tenemos de la tabla:

Datos

$$\frac{KN}{100}=\frac{35(42)}{100}=14.7$$
$$L_1=\frac{130+131}{2}=130.5$$
$$(\Sigma f)_1=2+4+7=13$$
$$f_K=9$$
$$C=7$$

Fórmula

$$P_{35}=L_1+\left[\frac{\frac{KN}{100}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_{35}=130.5+\left[\frac{14.7-13}{9}\right](7)$$
$$P_{35}=130.5+1.32=131.82$$

El valor del trigésimo quinto percentil es 131.82 libras, es decir, 35% de los atletas pesa 131.82 libras o menos.

Para determinar el quincuagésimo percentil P_{50}, tenemos de la tabla:

Datos

$$\frac{KN}{100}=\frac{50(42)}{100}=21$$
$$L_1=\frac{130+131}{2}=130.5$$
$$(\Sigma f)_1=2+4+7=13$$
$$f_K=9$$
$$C=7$$

Fórmula

$$P_{50}=L_1+\left[\frac{\frac{KN}{100}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_{50}=130.5+\left[\frac{21-13}{9}\right](7)$$
$$P_{35}=130.5+6.22=136.72$$

El valor del quincuagésimo percentil que es equivalente a la mediana es 136.72 libras, es decir, 50% de los atletas pesa 136.72 libras o menos.

Para determinar el sexagésimo quinto percentil P_{65}, tenemos de la tabla:

Datos

$$\frac{KN}{100}=\frac{65(42)}{100}=27.3$$
$$L_1=\frac{137+138}{2}=137.5$$
$$(\Sigma f)_1=2+4+7+9=22$$
$$f_K=13$$
$$C=7$$

Fórmula

$$P_{65}=L_1+\left[\frac{\frac{KN}{100}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_{65}=137.5+\left[\frac{27.3-22}{13}\right](7)$$
$$P_{65}=137.5+2.85=140.35$$

El valor del sexagésimo quinto percentil es 140.35 libras, es decir, 65% de los atletas pesa 140.35 libras o menos.

Para determinar el octogésimo percentil P_{80}, tenemos de la tabla:

Datos

$$\frac{KN}{100}=\frac{80(42)}{100}=33.6$$
$$L_1=\frac{137+138}{2}=137.5$$
$$(\Sigma f)_1=2+4+7+9=22$$
$$f_K=13$$
$$C=7$$

Fórmula

$$P_{80}=L_1+\left[\frac{\frac{KN}{100}-(\Sigma f)_1}{f_K}\right]C$$

Sustitución

$$P_{80}=137.5+\left[\frac{33.6-22}{13}\right](7)$$
$$P_{80}=137.5+6.25=143.75$$

El valor del octogésimo percentil es 143.75 libras, es decir, 80% de los atletas pesa 143.75 libras o menos.

Cálculo de los cuantiles por métodos gráficos

Los cuantiles (cuartiles, deciles y percentiles) se determinan geométricamente en gráficos de histogramas de frecuencia porcentual, en polígonos de frecuencia porcentual y en ojivas porcentuales.

El cálculo puede hacerse de dos maneras:

a) Conocido el cuantil que se ubica sobre la ordenada (eje y), se determina su valor sobre la abscisa (eje x).

b) Conocido el valor del cuantil que se ubica sobre la abscisa (eje x), se determina su lugar correspondiente sobre la ordenada (eje y).

EJEMPLO

Ejemplo 1 A partir de los datos de la siguiente tabla de distribución de frecuencias registra las puntuaciones de una prueba de coordinación física dadas a 50 conductores de vehículos que habían consumido alcohol. Construye:

Intervalos (puntuaciones)	Frecuencia (f) (conductores)
50 - 59	8
60 - 69	18
70 - 79	12
80 - 89	7
90 - 99	5
	$N = 50$

a) El histograma porcentual indicando el valor de los cuantiles Q_1, D_6, P_{80}.

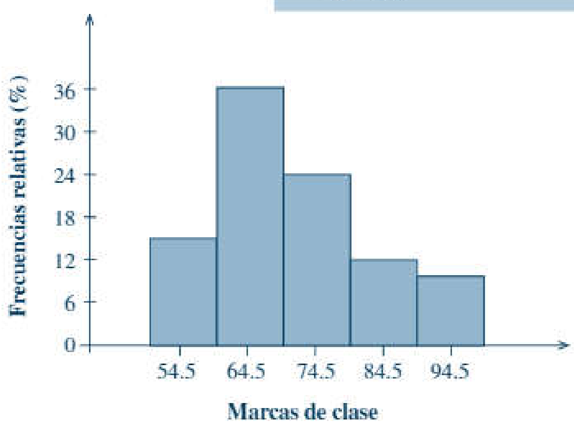
b) El polígono de frecuencias porcentual indicando el valor de los cuantiles Q_2, D_3, P_{70}.

c) La ojiva porcentual indicando el valor de los cuantiles Q_3, D_5, P_{90}.

Solución

a) Para construir la gráfica del histograma porcentual, se requiere la siguiente tabla de apoyo:

Intervalos de clase	Marca de clase (X)	Frecuencia de clase (f)	Frecuencia relativa de clase (FRC %)
50 - 59	54.5	8	16%
60 - 69	64.5	18	36%
70 - 79	74.5	12	24%
80 - 89	84.5	7	14%
90 - 99	94.5	5	10%
		$N = 50$	100%



Dado que geométricamente los cuartiles Q_K se definen como el valor de abscisa que corresponde a la línea vertical que divide a un histograma en cuatro partes de igual área.

Si el área total del histograma es de 50 unidades de área, el cuartil Q_1 será la línea vertical que tenga $\frac{KN}{4}$, es decir, $\frac{1(50)}{4} = 12.5$ unidades de área a la izquierda en el histograma.

Encontramos que en el segundo rectángulo, se localiza la clase del primer cuartil Q_1, cuya marca de clase es 64.5 y área de 18 unidades; a la izquierda de dicho rectángulo hay 8 unidades de área, por lo que se requieren 4.5 unidades del rectángulo de la clase del primer cuartil Q_1 para completar 12.5 unidades de área que es exactamente la cuarta parte del área total del histograma. Lo anterior se obtiene al dividir la base en cuatro partes iguales $\left(\text{cada parte} \frac{4.5}{18} = \frac{1}{4}\right)$ y trazando la perpendicular en el primer cuarto de la base, se tiene el valor del primer cuartil como resultado gráfico.

Si los límites reales de la clase del primer cuartil (60 - 69) son:

$$\text{Límite real inferior} = \frac{59+60}{4} = 59.5 \qquad \text{Límite real superior} = \frac{69+70}{2} = 69.5$$

El tamaño de la clase del primer cuartil es $69.5 - 59.5 = 10$. La cuarta parte del tamaño de la clase del primer cuartil es $\frac{10}{4} = 2.5$; por lo que el valor del primer cuartil Q_1 es:

$$Q_1 = 59.5 + 2.5 = 62$$

Los deciles D_X son el valor de abscisa que corresponde a la línea vertical que divide a un histograma en diez partes de igual área.

El decil (D_6) será la línea vertical que tenga $\frac{KN}{10}$, es decir, $\frac{6(50)}{10} = 30$ unidades de área a la izquierda en el histograma.

Encontramos que en el tercer rectángulo se localiza la clase del sexto decil D_6 cuya marca de clase es 74.5 y área de 12 unidades; a la izquierda de dicho rectángulo hay $8 + 18 = 26$ unidades de área, por lo que se requieren 4 unidades del rectángulo de la clase del sexto decil D_6 para completar 30 unidades de área que es exactamente $\frac{6}{10} = \frac{3}{5}$, las tres quintas partes del área total de histograma. Lo anterior se obtiene al dividir la base en tres partes iguales $\left(\text{cada parte} \frac{4}{12} = \frac{1}{3}\right)$ y trazando la perpendicular en el primer tercio de la base, se tiene el valor del sexto decil como resultado gráfico.

Si los límites reales de la clase del sexto decil (70 - 79) son:

$$\text{Límite real inferior} = \frac{69+70}{2} = 69.5 \qquad \text{Límite real superior} = \frac{79+80}{2} = 79.5$$

El tamaño de la clase del sexto decil es $79.5 - 69.5 = 10$. La tercera parte del tamaño de la clase del sexto decil es $\frac{10}{3} = 3.33$; por lo que el valor del sexto decil es: $D_6 = 69.5 + 3.33 = 72.83$

Los percentiles P_X son el valor de abscisa que corresponde a la línea vertical que divide a un histograma en cien partes de igual área.

El percentil P_{80} será la línea vertical que tenga $\frac{KN}{100}$, es decir, $\frac{80(50)}{100} = 40$ unidades de área a la izquierda en el histograma.

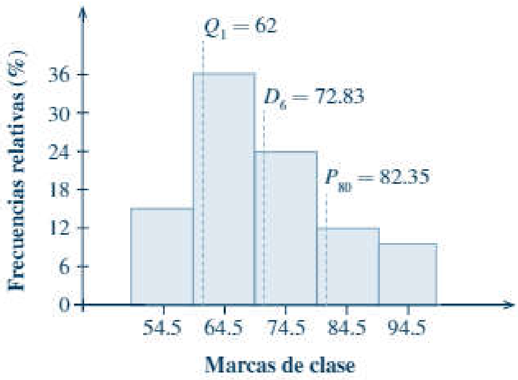
Encontramos que en el cuarto rectángulo, se localiza la clase del octogésimo percentil P_{80} cuya marca de clase es 84.5 y área de 7 unidades; a la izquierda de dicho rectángulo hay $8 + 18 + 12 = 38$ unidades del área, por lo que se requieren 2 unidades del rectángulo de la clase del octogésimo percentil (P_{80}) para completar 40 unidades de área que es exactamente $\frac{80}{100} = \frac{4}{5}$, las cuatro quintas partes del área total de histograma. Lo anterior se obtiene al dividir la base en dos séptimas parte iguales y trazando la perpendicular en el primer dos séptimos de la base, se tiene el valor del octogésimo percentil P_{80} como resultado gráfico.

Si los límites reales de la clase del octogésimo percentil (80 - 89) son:

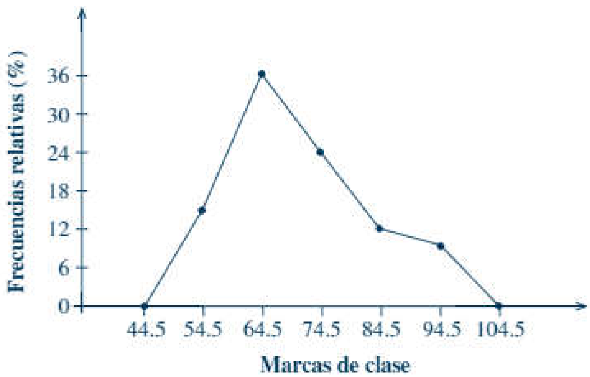
$$\text{Límite real inferior} = \frac{79+80}{4} = 79.5 \qquad \text{Límite real superior} = \frac{89+90}{2} = 89.5$$

El tamaño de la clase del octogésimo percentil es $89.5 - 79.5 = 10$. Las dos séptimas partes del tamaño de la clase del octogésimo percentil es $\frac{10}{7/2} = 2.85$; por lo que el valor del octogésimo percentil P_{80} es $P_{80} = 79.5 + 2.85 = 82.35$.

Gráficamente se tiene:



b) Para construir la gráfica del polígono de frecuencias relativas, nos apoyamos con los datos de la tabla del inciso anterior, es decir:



Dado que geométricamente los cuartiles Q_K se definen como el valor de abscisa que corresponde a la línea vertical que divide a un polígono de frecuencias en cuatro partes de igual área.

Si el área total del polígono es de 50 unidades de frecuencia, el segundo cuartil Q_2 será la línea vertical que tenga $\frac{KN}{4}$, es decir, $\frac{2(50)}{4} = 25$ unidades de frecuencia a la izquierda y a la derecha en el polígono.

A continuación se fija la clase que contiene al segundo cuartil Q_2, lo anterior se obtiene al sumar las frecuencias de todas las clases que se ubiquen por debajo de $\frac{KN}{4} = 25$ unidades de frecuencia, resultando ser $(\Sigma f)_1 = 8$. La clase que se localizó es (60 - 69) y su frecuencia es 18; determinando los límites reales de dicha clase, tenemos:

$$\text{Límite real inferior} = \frac{59+60}{2} = 59.5 \qquad \text{Límite real superior} = \frac{69+70}{2} = 69.5$$

El tamaño de la clase es $69.5 - 59.5 = 10$.

La diferencia entre $\frac{KN}{4} = 25$ y $\left(\sum f\right)_1 = 8$ se divide entre la frecuencia de la clase que contiene al cuartil, es decir, $f_{Q_2} = 18$; su resultado se multiplica por el tamaño de dicha clase $C = 10$. Este valor se suma al valor del límite inferior, es decir:

$$Q_2 = 59.5 + \left[\frac{25.8}{18}\right](10) = 59.5 + 9.44 = 68.94$$

El valor del segundo cuartil $Q_2 = 68.94$ es equivalente a la mediana.

Dado que geométricamente los deciles D_K se definen como el valor de abscisa que corresponde a la línea vertical que divide a un polígono de frecuencias en diez partes de igual área.

El decil D_3 será la línea vertical que tenga $\frac{KN}{10}$, es decir, $\frac{3(50)}{10} = 15$ unidades de frecuencia a la izquierda en el polígono.

A continuación se fija la clase que contiene al tercer decil (D_3), lo anterior se obtiene al sumar las frecuencias de todas las clases que estén por debajo de $\frac{KN}{10} = 15$ unidades de frecuencia, así $(\sum f)_1 = 8$.

La clase que se localizó es (60 - 69) y su frecuencia es 18; determinando los límites reales de dicha clase, tenemos:

$$\text{Límite real inferior} = \frac{59+60}{2} = 59.5 \qquad \text{Límite real superior} = \frac{69+70}{2} = 69.5$$

El tamaño de la clase es $69.5 - 59.5 = 10$.

La diferencia entre $\frac{KN}{10} = 15$ y $\left(\sum f\right)_1 = 8$ se divide entre la frecuencia de la clase que contiene al decil, es decir, $f_{D3} = 18$; su resultado se multiplica por el tamaño de dicha clase $C = 10$. Este valor así encontrado se suma al valor del límite inferior, es decir:

$$D_3 = 59.5 + \left[\frac{15-8}{18}\right](10) = 59.5 + 3.88 = 63.38$$

Los percentiles P_K son el valor de abscisa que corresponde a la línea vertical que divide al polígono de frecuencias en cien partes de igual área.

El percentil P_{70} será la línea vertical que tenga $\frac{KN}{100}$, es decir, $\frac{70(50)}{100} = 35$ unidades de frecuencias a la izquierda en el polígono.

A continuación se fija la clase que contiene al septuagésimo percentil P_{70}, lo anterior se obtiene al sumar las frecuencias de todas las clases que estén por debajo de $\frac{KN}{100} = 35$ unidades de frecuencia, resultando ser $(\Sigma f)_1 = 8 + 18 = 26$. La clase que se localizó es (70 - 79) y su frecuencia es 12; determinando los límites reales de dicha clase, tenemos:

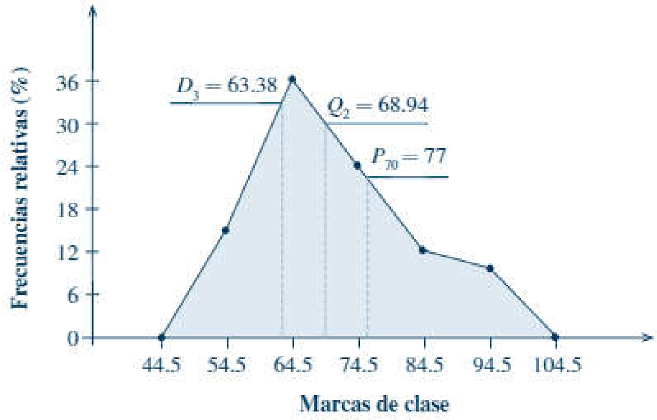
$$\text{Límite real inferior} = \frac{69+70}{2} = 69.5 \qquad \text{Límite real superior} = \frac{79+80}{2} = 79.5$$

El tamaño de la clase es $79.5 - 69.5 = 10$.

La diferencia entre $\frac{KN}{100} = 35$ y $\left(\sum f\right)_1 = 26$ se divide entre la frecuencia de la clase que contiene al percentil, es decir, $f_{P70} = 12$; su resultado se multiplica por el tamaño de dicha clase $C = 10$. Este valor se suma al valor del límite inferior, es decir:

$$P_{70} = 69.5 + \left[\frac{35-26}{12}\right](10) = 69.5 + 7.5 = 77$$

Al elaborar la gráfica correspondiente se tiene:

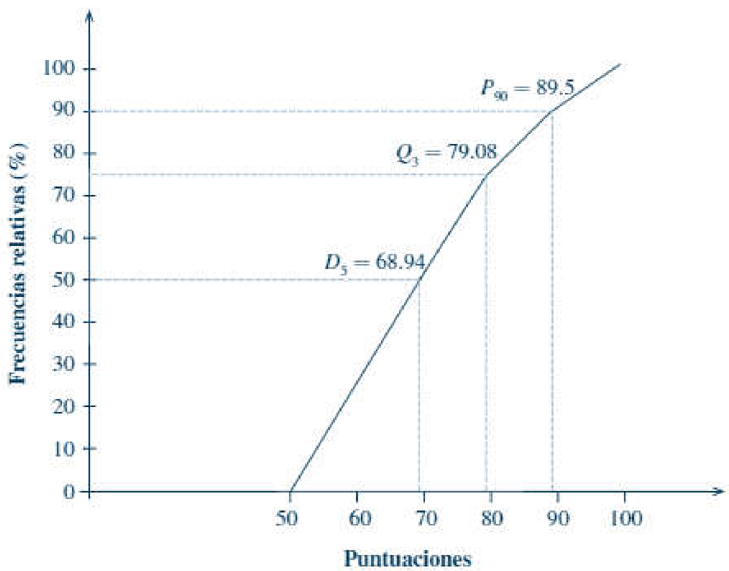


c) Para construir la ojiva porcentual, se requiere la siguiente tabla de apoyo:

Intervalos de clase	Marca de clase (*X*)	Frecuencia de clase (*f*)	Frecuencias acumuladas	Frecuencias acumuladas porcentuales
50 - 59	54.5	8	8	16%
60 - 69	64.5	18	26	52%
70 - 79	74.5	12	38	76%
80 - 89	84.5	7	45	90%
90 - 99	94.5	5	50	100%
		$N = 50$		

La ojiva porcentual se grafica al marcar sobre la escala vertical izquierda las frecuencias acumuladas porcentuales y sobre la escala horizontal las puntuaciones de clase; también suele construirse como ojiva porcentual **menor que** o también como ojiva porcentual o **mayor que**.

La intersección sobre la ojiva porcentual de la abscisa (puntuaciones de clase) y la ordenada (frecuencias relativas acumuladas en porcentaje) dan lugar a los cuartiles Q_3, deciles D_5 y percentiles P_{90}, es decir:



Los valores respectivos de los cuantiles Q_3, D_5, P_{90} se pueden determinar por la lectura aproximada sobre el eje horizontal o se calculan por los métodos descritos anteriormente en el tema.

Los resultados son $Q_3 = 79.08$, $D_5 = 68.94$ y $P_{90} = 89.5$.

EJERCICIO 11

Escribe los números correspondientes

Competencias genéricas

Competencias disciplinares

I. Realiza lo que se indica en cada caso.

1. ¿Qué son los cuantiles?
2. Define cuartil.
3. Explica las características propias de los cuartiles.
4. Escribe la ecuación general para calcular los cuartiles de una distribución de frecuencias.
5. Define decil.
6. Explica las características propias de los deciles.
7. Escribe la ecuación general para calcular los deciles de una distribución de frecuencias.
8. Define percentil.
9. Explica las características propias de los percentiles.
10. Escribe la ecuación general para calcular los percentiles de una distribución de frecuencias.
11. ¿En qué tipo de gráficos es posible el cálculo de los cuantiles?
12. Explica las maneras en que pueden calcularse los cuantiles por métodos gráficos.

II. Resuelve los siguientes problemas.

1. Treinta fallas de energía eléctrica duraron 23, 130, 49, 100, 36, 31, 85, 54, 123, 67, 44, 38, 93, 17, 97, 75, 42, 81, 62, 128, 19, 39, 96, 26, 80, 53, 85, 77, 29 y 86 minutos. Determina:
   a) Los cuartiles Q_1, Q_2 y Q_3.
   b) Los deciles D_1, D_3, D_5, D_7 y D_9.
   c) Los percentiles P_5, P_{15}, P_{45}, P_{65}, P_{85} y P_{95}.
2. Quince traductores de inglés-español lograron traducir 92, 107, 123, 90, 78, 81, 76, 94, 105, 88, 109, 121, 95, 101 y 89 páginas de un mismo libro de control de calidad en una hora. Determina:
   a) Los cuartiles Q_1 y Q_3.
   b) Los deciles D_2, D_4, D_6 y D_8.
   c) Los percentiles P_{10}, P_{30}, P_{50}, P_{70}, P_{90} y P_{99}.
3. Encuentra los cuartiles Q_2 y Q_3; los deciles D_1, D_3, D_6 y D_9, y los percentiles P_{20}, P_{40}, P_{70}, P_{80} y P_{95}, para los siguientes conjuntos de datos numéricos.
   a) 27, 30, 32, 34, 31, 23, 28, 37, 33, 41, 39, 18, 22, 29.
   b) 238, 242, 213, 226, 235, 242, 259, 257, 240, 258, 221, 228, 243, 264, 255, 239, 248, 257, 240, 218, 245, 256, 223, 230, 237.
   c) 52, 84, 40, 57, 61, 65, 77, 64, 62, 35, 82, 58, 50, 78, 103, 71, 76, 42, 55, 67, 60, 95, 58, 49, 89, 48, 63, 75, 56, 93, 84, 87, 59.

4. La siguiente tabla de distribución de frecuencias registra los salarios anuales en miles de dólares de cincuenta familias que trabajan en el departamento de agricultura de California; determina los cuartiles Q_1, Q_2, Q_3; los deciles D_1, D_2, D_3, ..., D_9, y los percentiles P_5, P_{15}, P_{25}, P_{35}, P_{45}, P_{65}, P_{80}, P_{95}.

Intervalos (salarios)	Frecuencia (familias)
9500 - 9999	3
10 000 - 10 499	8
10 500 - 10 999	13
11 000 - 11 499	10
11 500 - 11 999	7
12 000 - 12 499	5
12 500 - 12 999	3
13 000 - 13 499	1
	$N = 50$

5. La siguiente tabla de distribución de frecuencias registra los tiempos en segundos que 35 atletas varones tardan en recorrer los 100 metros planos. Determina los cuartiles Q_1 y Q_2; los deciles D_3, D_6 y D_9, y los percentiles P_{35}, P_{50}, P_{65} y P_{80}.

Intervalos (segundos)	Frecuencia (atletas)
8.8 - 9.2	2
9.3 - 9.7	5
9.8 - 10.2	13
10.3 - 10.7	9
10.8 - 11.2	6
	$N = 35$

6. La siguiente tabla de distribución de frecuencias registra las calificaciones de un segundo examen parcial de Física II de un grupo de la Escuela Superior de Ingeniería Mecánica y Eléctrica; determina el tercer cuartil, el cuarto y séptimo decil y los percentiles P_{25}, P_{50} y P_{90}. Indica también:

a) La calificación más baja lograda por el 20% superior de la clase.

b) La calificación más alta lograda por el 25% inferior de la clase.

c) Cuál es la calificación del 50% de los alumnos.

Intervalos (calificaciones)	Frecuencia (estudiantes)
30 - 39	1
40 - 49	2
50 - 59	4
60 - 69	7
70 - 79	11
80 - 89	8
90 - 100	3
	$N = 36$

7. La siguiente tabla de distribución de frecuencias registra los diámetros en pulgadas de las cabezas de los remaches fabricados por una compañía. Determina:

a) Los cuartiles Q_1, Q_2 y Q_3.

b) Los deciles D_1, D_3, D_5, D_7 y D_9.

c) Los percentiles P_5, P_{15}, P_{20}, P_{35}, P_{40}, P_{45}, P_{60}, P_{65}, P_{85}, P_{95} y P_{99}.

Intervalos (pulgadas)	Frecuencia (remaches)
0.247 - 0.249	6
0.250 - 0.252	8
0.253 - 0.255	15
0.256 - 0.258	43
0.259 - 0.261	65
0.262 - 0.264	51
0.265 - 0.267	29
0.268 - 0.270	17
0.271 - 0.273	11
0.274 - 0.276	5
	$N = 250$

8. Con los datos de los problemas 4, 5, 6 y 7, construye:
   a) El histograma porcentual indicando el valor de los cuantiles Q_1, D_6 y P_{80}.
   b) El polígono de frecuencias porcentual indicando el valor de los cuantiles Q_1, D_3 y P_{70}.
   c) La ojiva porcentual indicando el valor de los cuantiles Q_3, D_5 y P_{90}.

Verifica tus resultados en la sección de respuestas.

TEMA 3 Medidas de dispersión

Propósito del tema

Que el estudiante:

- Aprenda e incorpore en situaciones reales las medidas de dispersión.
- Interprete y analice los resultados que obtiene en la resolución de problemas.
- Exprese razones para fundamentar una respuesta y obtenga conclusiones pertinentes a partir de datos estadísticos.
- Determine la desviación y variación de una serie de datos para su análisis en la toma de decisiones.

Competencias disciplinares

1. Construye e interpreta modelos matemáticos mediante la aplicación de procedimientos aritméticos, algebraicos, geométricos y variacionales, para la comprensión y análisis de situaciones reales, hipotéticas o formales.

3. Explica e interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos o situaciones reales.

5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.

8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

Contenidos que aborda el tema

Contenidos conceptuales

- Definición de rango.
- Definición de desviación media.
- Definición de desviación típica.
- Definición de varianza.
- Definición de dispersión absoluta y relativa.
- Definición de referencias tipificadas.

Contenidos procedimentales

- Sintetizará evidencias obtenidas mediante la experimentación para producir conclusiones y formulará nuevas preguntas.
- Aprenderá a identificar qué medida de dispersión utilizar para resolver distintos problemas de aplicación.
- Sigue procedimientos de manera reflexiva, comprendiendo como cada uno de sus pasos contribuye al alcance de un objetivo.
- Resolverá problemas aplicando la desviación media y desviación típica.
- Usará diversas estrategias de resolución de problemas y los aplicara a situaciones reales.

Contenidos actitudinales

- Expresará sus ideas mediante los elementos básicos de las medidas de dispersión.
- Trabajará en equipo y respetará a sus compañeros al resolver problemas.
- Aprenderá a valorar el trabajo de sus compañeros al resolver problemas.
- Estructurará ideas y argumentos de manera clara, coherente y sintética.

Medidas de dispersión

Introducción

Para describir el comportamiento de una serie de datos que se concentran en una distribución, se emplean las **medidas de tendencia central**; para explicar el grado de dispersión o variación de los datos que se expanden alrededor de una medida central, se utilizan las **medidas de dispersión**.

Entre los diferentes tipos de medidas de dispersión, las más comunes son: el rango, la desviación media, el rango intercuartítico, el rango semiintercuartítico, el rango entre percentiles 10 - 90, la varianza, la desviación típica, la dispersión absoluta y la dispersión relativa.

Rango o amplitud

Es la medida de dispersión más simple que se define como la diferencia que existe entre el mayor y el menor de los datos de un conjunto, es decir:

$$\text{Rango} = D_{\text{mayor}} - D_{\text{menor}}$$

EJEMPLO

Ejemplo 1 Determina el rango de los números 4, 5, 5, 8, 8, 8, 11, 13 y 15.

Solución

$$\text{Rango} = D_{\text{mayor}} - D_{\text{menor}} = 15 - 4 = 11$$

El rango o amplitud del conjunto de números dado es 11.

Para datos agrupados, el rango se define de dos formas:

a) Es la diferencia entre el límite superior de la clase más alta (LSC_{Alta}) y el límite inferior de la clase más baja (LIC_{Baja}), es decir:

$$\text{Rango} = LSC_{\text{Alta}} - LIC_{\text{Baja}}$$

b) Es la diferencia entre la marca de clase de la clase superior (MCC_{Superior}) y la marca de clase de la clase inferior (MCC_{Inferior}), es decir:

$$\text{Rango} = MCC_{\text{Superior}} - MCC_{\text{Inferior}}$$

EJEMPLO

Ejemplo 1 Encuentra el rango para la siguiente tabla de distribución de frecuencias, a partir del método de los límites y el de las marcas de clase.

Intervalos de clase	Marcas de clase (X)	Frecuencia (familias)
22 - 25	23.5	3
26 - 29	27.5	7
30 - 33	31.5	11
34 - 37	35.5	18
38 - 41	39.5	13
42 - 45	43.5	8
		$N = 60$

Solución

Si utilizamos la ecuación de los límites, tenemos:

$$\text{Rango} = LSC_{\text{Alta}} - LIC_{\text{Baja}} = 45 - 22 = 23$$

Si empleamos la ecuación de las marcas de clase, tenemos:

$$\text{Rango} = MCC_{\text{Superior}} - MCC_{\text{Inferior}} = 43.5 - 23.5 = 20$$

La diferencia entre los valores encontrados se debe a que los límites de clase se extienden siempre más que las marcas de clase, que tienden a eliminar los casos más extremos. También se piensa que cuanto mayor es el valor del rango, mayor es la dispersión o variación de los datos.

El rango o recorrido es muy útil para situaciones que se concretan a inquirir el tamaño de las variaciones extremas, por ejemplo: la temperatura máxima o mínima en un día, los costos máximo y mínimo de una serie de acciones en el mercado bursátil, las velocidades máximas y mínimas que se exigen en una carretera expresa, etcétera.

El rango o amplitud también presenta desventajas ya que depende solamente de los valores mayor y menor, por lo que es notoria su delicadeza hacia los valores desproporcionados que puedan presentarse en un conjunto.

EJEMPLO

Ejemplo 1 Determina el rango para cada uno de los siguientes conjuntos de datos numéricos:

a) 16, 10, 11, 7, 19, 14, 22, 9.

b) 22, 13, 12, 13, 12, 12, 7, 13.

Solución

a) El resultado de ordenar en forma ascendente los elementos del conjunto es: 7, 9, 10, 11, 14, 16, 19, 22; su rango es:

$$\text{Rango} = D_{\text{mayor}} - D_{\text{menor}} = 22 - 7 = 15$$

b) El resultado de ordenar en forma ascendente los elementos del conjunto es: 7, 12, 12, 12, 13, 13, 13, 22; su rango es:

$$\text{Rango} = D_{\text{mayor}} - D_{\text{menor}} = 22 - 7 = 15$$

Ambos conjuntos tienen el mismo rango; sin embargo, de acuerdo con el ordenamiento, de sus datos se observa que existe más dispersión o variación en el conjunto (*a*) que en el conjunto (*b*), ya que éste se compone de datos más compactos entre sí (principalmente 12 y 13). Se puede obtener una mejor información si en ambos conjuntos se eliminan los datos extremos 7 y 22.

EJERCICIO **12**

I. Contesta las siguientes preguntas.

1. ¿Para qué se utilizan las medidas de dispersión?
2. ¿Cuáles son las medidas de dispersión más comunes?
3. ¿Qué es el rango o amplitud?
4. ¿Cuál es la definición del rango o amplitud en sus dos formas para datos agrupados?

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

5. ¿Cuál es la causa de la diferencia de valores del rango calculado a partir de los límites y de las marcas de clase, respectivamente?
6. ¿En que situaciones el rango o recorrido resulta ser muy útil?
7. ¿Cuáles son las desventajas de aplicación que presenta el rango o amplitud?

II. Resuelve los siguientes problemas.

1. Encuentra el rango o amplitud de los siguientes conjuntos numéricos:
   a) 3.95, 4.09, 3.91, 3.98, 4.03, 3.89, 4.05, 3.90, 4.13, 4.01, 3.85, 3.97.
   b) 85, 92, 73, 69, 89, 78, 69, 96, 90.
   c) 11, 5, 6, 2, 14, 9, 17, 4.
   d) 12, 6, 11, 11, 12, 11, 12, 19, 13, 5.
2. Determina el rango de las siguientes tablas de distribución de frecuencias, por medio del método de los límites y el de las marcas de clase.

a)

Intervalos de clase	Frecuencia de clase (f)
80 - 82	7
83 - 85	13
86 - 88	21
89 - 91	34
92 - 94	42
95 - 97	35
98 - 100	28
101 - 103	19
104 - 106	11
107 - 109	5
	$N = 215$

b)

Intervalos de clase	Frecuencia de clase (f)
100 - 199	12
200 - 299	26
300 - 399	38
400 - 499	47
500 - 599	59
600 - 699	42
700 - 799	33
800 - 899	21
	$N = 278$

Verifica tus resultados en la sección de respuestas.

Desviación media o promedio de desviación

Se define como la media aritmética de los valores absolutos de las desviaciones de las variables respecto de la media aritmética; matemáticamente se expresa por la ecuación:

$$\text{Desviación media} = DM = \frac{\sum_{j=1}^{N} |X_j - \overline{X}|}{N} = \frac{\sum |X - \overline{X}|}{N} = \overline{|X - \overline{X}|}$$

donde:

N = Número total de elementos del conjunto.
$\overline{X}$ = Media aritmética del conjunto de números.
$|X - \overline{X}|$ = Valor absoluto de las desviaciones de los diferentes elementos numéricos de su media aritmética.

El valor absoluto de un número X denotado por $|X|$, indica la magnitud del número X sin considerar su signo; por ejemplo: $|3| = 3, |-7| = 7, |-2.5| = 2.5, |38| = 38$.

EJEMPLO

Ejemplo 1 Deteremina la desviación media del conjunto de elementos numéricos 5, 8, 11, 13, 17, 21 y 24.

Solución

Primero se calcúla la media aritmética del conjunto de elementos numéricos:

$$\overline{X} = \frac{\sum X}{N} = \frac{5+8+11+13+17+21+24}{7} = \frac{99}{7} = 14.142$$

Ahora, la desviación media: $DM = \dfrac{\sum|X-\overline{X}|}{N}$

$$DM = \frac{|5-14.142|+|8-14.142|+|11-14.142|+|13-14.142|+|17-14.142|+|21-14.142|+|24-14.142|}{7}$$

$$DM = \frac{9.142+6.142+3.142+1.142+2.858+6.858+9.858}{7} = \frac{39.142}{7} = 5.591 \approx 5.6$$

La desviación media del conjunto de elementos numéricos es 5.6.

Desviación media o promedio de desviación para datos agrupados

Si se tiene un conjunto de elementos $X_1, X_2, X_3, ..., X_N$ que se presentan con $f_1, f_2, f_3, ..., f_N$ veces de frecuencia, respectivamente, la desviación media o promedio de desviación puede determinarse por:

$$\text{Desviación media} = DM = \frac{\sum_{j=1}^{N} f_j|X_j-\overline{X}|}{N} = \frac{\sum f|X-\overline{X}|}{N} = |\overline{X-\overline{X}}|$$

donde:

$(X_1, X_2, X_3, ..., X_N)$ = Representan las marcas de clase.

$(f_1, f_2, f_3, ..., f_N)$ = Las frecuencias totales de clase.

$\sum f = N$ = Numero total de datos.

$\overline{X}$ = Media aritmética con frecuencias $\left(\dfrac{\sum f X}{N}\right)$.

$|X-\overline{X}|$ = Valor absoluto de las desviaciones de las diferentes marcas de clase de su media aritmética.

La desviación media o promedio de desviación presenta desventajas en cuanto a la media de tendencia central que se utiliza para su cálculo. El problema más grave se presenta cuando los signos de las desviaciones no se toman en cuenta (por ser valores absolutos), lo cual no muestra si la desviación está por arriba o por abajo de la media aritmética. Si se respeta el signo, la suma de las desviaciones respecto a la media es igual a cero y se aproxima a cero cuando las desviaciones son respecto a la mediana; por lo anterior, la desviación media o promedio de desviación se le debe denominar **desviación media absoluta**.

Si la distribución de frecuencias presenta forma simétrica y gráficamente tiene forma de **campana** o **normal**, donde se hace notar que el 57.5% de las observaciones **caen** en el intervalo $\overline{X} \pm DM$, es decir, que más de la mitad de las observaciones se concentran dentro de un intervalo de una unidad de la desviación media o promedio de desviación a uno y otro lado de la media aritmética; incluso si es moderadamente asimétrica.

EJEMPLOS

Ejemplos

1 Determina la desviación media de los salarios semanales en miles de pesos de 80 jefes de familia.

Intervalos (salarios)	Frecuencia de clase (f)
96 - 125	4
126 - 155	9
156 - 185	17
186 - 215	25
216 - 245	18
246 - 275	7
	$N = 80$

Solución

Primero se determina la media aritmética, para ello se utiliza la siguiente tabla de distribución y la ecuación del método largo (puede ser por cualquier otro método).

Intervalos (salarios)	Marca de clase (X)	Frecuencia de clase (f)	fX
96 - 125	110.5	4	442.0
126 - 155	140.5	9	1 264.5
156 - 185	170.5	17	2 898.5
186 - 215	200.5	25	5 012.5
216 - 245	230.5	18	4 149.0
246 - 275	260.5	7	1 823.5
		$N = 80$	$\Sigma fX = 15\,590$

$$\bar{X} = \frac{\sum f X}{N} = \frac{15\,590}{80} = 194.875$$

Para determinar la desviación media o promedio de desviación, se requiere complementar la tabla anterior con las siguientes columnas de apoyo.

$\lvert X - \bar{X}\rvert$	$f\lvert X - \bar{X}\rvert$
84.375	337.500
54.375	489.375
24.375	414.375
5.625	140.625
35.625	641.250
65.625	459.375
	$\Sigma f\lvert X - \bar{X}\rvert = 2\,482.5$

$$DM = \frac{\sum f\lvert X - \bar{X}\rvert}{N} = \frac{2\,482.5}{80} = 31.03125$$

La desviación media de los salarios es $31.03125 miles de pesos.

2 La siguiente tabla de distribución de frecuencias registra el tiempo de duración en horas de 25 baterías alcalinas seleccionadas para una prueba de control de calidad en la aplicación de uso para lámparas de mano. Determina la desviación media respecto al promedio.

a) Media aritmética.

b) Mediana.

Intervalos (horas)	Frecuencia de clase (f)
48.2 - 51.7	2
51.8 - 55.3	5
55.4 - 58.9	9
59.0 - 62.5	6
62.6 - 66.1	3
	$N = 25$

Solución

a) Para determinar la media aritmética, utilizaremos la siguiente tabla de distribución y la ecuación del método largo.

Intervalos (horas)	Marca de clase (X)	Frecuencia de clase (f)	fX
48.2 - 51.7	49.95	2	99.90
51.8 - 55.3	53.55	5	267.75
55.4 - 58.9	57.15	9	514.35
59.0 - 62.5	60.75	6	364.50
62.6 - 66.1	64.35	3	193.05
		$N = 25$	$\sum fX = 1\,439.55$

$$\overline{X} = \frac{\sum fX}{N} = \frac{1\,439.55}{25} = 57.582$$

Para determinar la desviación media o promedio de desviación, se requiere complementar la tabla anterior con las siguientes columnas de apoyo.

$\lvert X - \overline{X}\rvert$	$f\lvert X - \overline{X}\rvert$
7.632	15.264
4.032	20.160
0.432	3.888
3.168	19.008
6.768	20.304
	$\sum f\lvert X - X\rvert = 78.624$

$$DM = \frac{\sum f\lvert X - \overline{X}\rvert}{N} = \frac{78.624}{25} = 3.14496$$

La desviación media respecto a la media es 3.14496.

b) Para determinar la mediana, se calcula $\frac{N}{2}$, con objeto de localizar la clase mediana.

$$\frac{N}{2} = \frac{25}{2} = 12.5$$

La suma de las frecuencias de las segunda y tercera primeras clases es $2 + 5 = 7$ y $2 + 5 + 9 = 16$, respectivamente. Al comparar, se observa que la mediana se encuentra en la tercera clase; por lo tanto, la clase mediana también.

	Intervalos (horas)	Frecuencia de clase (f)	
		$(\sum f)_2$	
	48.2 - 51.7	2	
	51.8 - 55.3	5	
Clase mediana $\} \longrightarrow$	55.4 - 58.9	9	$\longleftarrow \{$ f mediana
	59.0 - 62.5	6	
	62.6 - 66.1	3	
		$N = 25$	

De la tabla anterior se obtienen los siguientes datos:

$L_1 = \frac{55.3 + 55.4}{2} = 55.35$

$N = 25$

$(\sum f)_1 = 7$

$f \text{ mediana} = 9$

$C = 3.6$

Al sustituir los datos anteriores en la ecuación de la mediana resulta:

$$\text{Mediana} = L_1 + \left[\frac{\frac{N}{2} - (\sum f)_1}{f \text{ mediana}}\right] C = 55.35 + \left[\frac{\frac{25}{2} - (7)}{9}\right](3.6)$$

$$\text{Mediana} = 55.35 + \left[\frac{12.5 - 7}{9}\right](3.6) = 55.35 + \left(\frac{5.5}{9}\right)(3.6) = 55.3 + 2.2 = 57.55$$

Para determinar la desviación media o promedio de desviación, se emplean los datos de la siguiente tabla.

Intervalos (horas)	Marca de clase (X)	Frecuencia de clase (f)	$\|X - \text{Mediana}\|$	$f\|X - \text{Mediana}\|$
48.2 - 51.7	49.95	2	7.6	15.2
51.8 - 55.3	53.55	5	4.0	20.0
55.4 - 58.9	57.15	9	0.4	3.6
59.0 - 62.5	60.75	6	3.2	19.2
62.6 - 66.1	64.35	3	6.8	20.4
		$N = 25$		78.4

$$DM = \frac{\sum f|X - \text{Mediana}|}{N} = \frac{78.4}{25} = 3.136$$

La desviación media respecto a la mediana es 3.136.

En la solución de este problema se observa, como un principio, que la desviación media respecto a la media aritmética siempre es mayor o igual que la desviación media respecto a la mediana, es decir:

$$\frac{\sum f|X - \overline{X}|}{N} \geq \frac{\sum f|X - \text{Mediana}|}{N}$$

3 ●●· Determina la desviación media del número de horas a la semana que 60 familias miran programas televisivos; encuentra el porcentaje de televidentes cuyas horas caen dentro de los siguientes intervalos:

a) $\overline{X} \pm DM$ *b)* $\overline{X} \pm 2DM$ *c)* $\overline{X} \pm 3DM$

Construye la gráfica representativa de dichos porcentajes.

Intervalos (horas)	Frecuencia (familias) (f)
30 - 39	4
40 - 49	9
50 - 59	13
60 - 69	19
70 - 79	12
80 - 89	3
	$N = 60$

Solución

Primero, calculamos la media aritmética.

Intervalos (horas)	Marca de clase (*X*)	Frecuencia de clase (*f*)	*f X*
30 - 39	34.5	4	138.0
40 - 49	44.5	9	400.5
50 - 59	54.5	13	708.5
60 - 69	64.5	19	1 225.5
70 - 79	74.5	12	894.0
80 - 89	84.5	3	253.5
		$N = 60$	$\Sigma fX = 3\,620$

$$\overline{X} = \frac{\sum fX}{N} = \frac{3\,620}{60} = 60.34$$

Ahora, utilizaremos la siguiente tabla de apoyo para determinar la desviación o promedio de desviación.

$\lvert X - \overline{X}\rvert$	$f\lvert X - \overline{X}\rvert$
25.84	103.36
15.84	142.56
5.84	75.92
4.16	79.04
14.16	169.92
24.16	72.48
	$\Sigma f\lvert X - \overline{X}\rvert = 643.28$

$$DM = \frac{\sum\lvert X - \overline{X}\rvert}{N} = \frac{643.28}{60} = 10.72$$

La desviación media de las horas es 10.72.

a) Para determinar el porcentaje $\overline{X} \pm DM$, tenemos:

Sea $\overline{X} \pm DM = 60.34 \pm 10.72$, por tanto, el intervalo 49.62 a 71.06 horas.

Si se considera el valor de 49.62, se tiene que el límite real superior de la segunda clase es 49.5; al tomar el valor de 71.06, se tiene que el límite real inferior de la quinta clase es 69.5. El tamaño de los intervalos de clase es de 10 horas.

El intervalo de 49.62 a 71.06 está integrado por todas las familias de la tercera clase (13), menos $\frac{1}{10}(49.62 - 49.5)$ de las familias de la tercera clase (13), más todas las familias de la cuarta clase (19), más $\frac{1}{10}(71.06 - 69.5)$ de las familias de la quinta clase (12), es decir:

$$13 - \frac{1}{10}(49.62 - 49.5)(13) + 19 + \frac{1}{10}(71.06 - 69.5)(12)$$
$$13 - 0.156 + 19 + 1.872 = 33.716$$

Después de resolver la siguiente regla de tres, resulta:

$$\begin{array}{rcl} 60 \text{ familias} & \longrightarrow & 100\% \\ 33.716 \text{ familias} & \longrightarrow & X \end{array} \qquad X = 56.19\% \approx 56\%$$

El porcentaje de $\overline{X} \pm DM$ es el 56% del total.

b) Para determinar el porcentaje $\overline{X} \pm 2\ DM$, tenemos:

Sea $\overline{X} \pm 2\ DM = 60.34 \pm 2\,(10.72) = 60.34 \pm 21.44$; por tanto, el intervalo resultante es de 38.9 a 81.78 horas.

Si consideramos el valor de 38.9, se tiene que el límite real superior de la primera clase es 39.5; al tomar el valor 81.73, se tiene que el límite real inferior de la sexta clase es 79.5. El tamaño de los intervalos de clase es de 10 horas.

El intervalo de 38.9 a 81.78 se integra por todas las familias de la primera clase (4), menos $\frac{1}{10}(39.5-38.9)$ de las familias de la primera clase (4), más todas las familias de la segunda, tercera, cuarta y quinta clase $(9+13+19+12)$, más $\frac{1}{10}(81.78-79.5)$ de las familias de la sexta clase (3), es decir:

$$4-\frac{1}{10}(39.5-38.9)(4)+9+13+19+12+\frac{1}{10}(81.78-79.5)(3)$$
$$4-0.24+9+13+19+12+0.684=57.444$$

Al resolver la siguiente regla de tres, resulta:

$$\begin{array}{rcl} 60 \text{ familias} & \longrightarrow & 100\% \\ 57.444 \text{ familias} & \longrightarrow & X \end{array} \qquad X = 95.74\% \approx 96\%$$

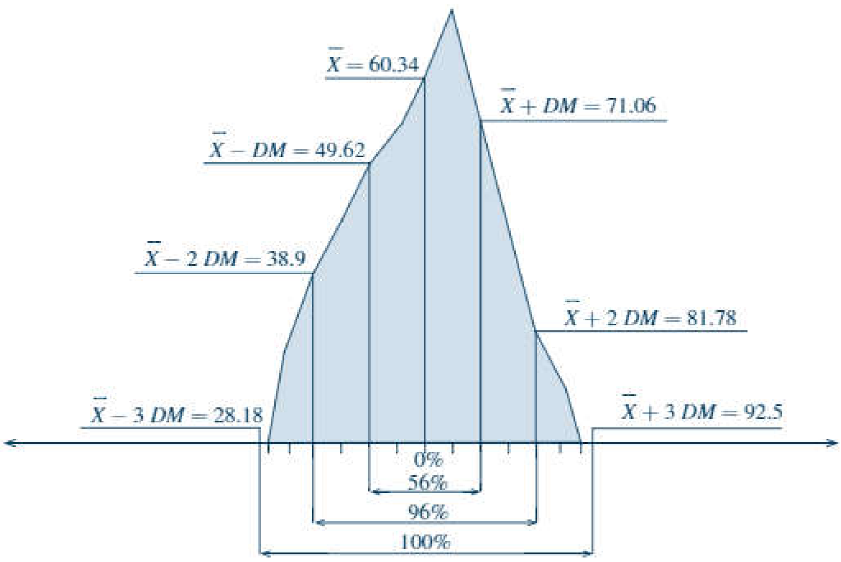
El porcentaje de $\overline{X} \pm 2\ DM$ es el 96% del total.

c) Para determinar el porcentaje $\overline{X} \pm 3\ DM$, tenemos:

Sea $\overline{X} \pm 3\ DM = 60.34 \pm 3\,(10.72) = 60.34 \pm 32.16$, resultando; por lo tanto, el intervalo de 28.18 a 92.5 horas.

Se observa que el intervalo 28.18 a 92.5 queda fuera de los límites reales de la primera y de la última clase, es decir, que el porcentaje $\overline{X} \pm 3\ DM$ es el 100% del total.

La representación gráfica de los porcentajes $\overline{X} \pm DM$, $\overline{X} \pm 2\ DM$ y $\overline{X} \pm 3\ DM$ resultantes es:



Para comprobar la notación anterior, tenemos:

Si tomamos el valor de 28.18, se tiene que el límite real superior de la clase que se ubicará antes de la primera clase es 29.5; si consideramos el valor de 92.5, se tiene que el límite real inferior de la clase que se ubicará después de la última clase es 89.5. El tamaño de los intervalos de clase es de 10 horas.

El intervalo de 28.18 a 92.5 está integrado por todas las familias de la clase que se ubicará antes de la primera clase (0), menos $\frac{1}{10}(29.5-28.18)$ de las familias de la clase que se ubicará antes de la primera clase (0), más todas las familias de la primera, segunda, tercera, cuarta, quinta y sexta clase $(4+9+13+19+12+3)$, más $\frac{1}{10}(92.5-89.5)$ de las familias de clase que se ubicará después de la última clase (0), es decir:

$$0-\frac{1}{10}(29.5-28.18)(0)+4+9+13+19+12+3+\frac{1}{10}(92.5-89.5)(0)$$
$$4+9+13+19+12+3=60$$

Como 60 familias representa el 100%, tenemos que:

El porcentaje de $\overline{X} \pm 3\ DM$ es el 100% del total.

Rango intercuartílico

Dado que el **rango** es una medida de dispersión muy influenciable por los valores extremos de un conjunto de datos ordenados o de una distribución de frecuencias, con el fin de eliminar la influencia de dichos valores extremos se utiliza el **rango intercuartílico**, el cual se define como la diferencia entre el tercer cuartil Q_3 y el primer cuartil Q_1. Matemáticamente se expresa como:

$$\text{Rango intercuartílico} = Q = Q_3 - Q_1$$

Rango semiintercuartílico o desviación cuartílica

Para un conjunto de datos numéricos o para una distribución de frecuencias, el **rango semiintercuartílico** o **desviación cuartílica** se define como la mitad del **rango intercuartílico**.

Matemáticamente se representa como:

$$\text{Rango semiintercuartílico} = Q_{SI} = \frac{Q_3 - Q_1}{2} = \frac{Q}{2}$$

El rango intercuartílico y el rango semiintercuartílico son las medidas de dispersión más apropiadas para utilizarse en lugar del **rango**; por ello, es necesario conocer también las siguientes desventajas de aplicación que presenta, es decir:

a) No toma en cuenta a todos los valores de una distribución de frecuencias y puede ser que los valores por abajo de Q_1 o por arriba de Q_3 estén demasiado compactos o demasiado dispersos, por lo que el valor del rango intercuartílico es el mismo.

b) Cuando sólo se conoce el valor del rango intercuartílico, no es posible precisar la ubicación de cualquiera de las observaciones que están dentro de la distribución.

c) Ninguna de las medidas tiene propiedades que les permitan participar en las relaciones matemáticas que se emplean en la estadística.

EJEMPLOS

Ejemplos

1 Determina el rango intercuartílico y la desviación cuartílica del siguiente conjunto de datos numéricos ordenados: 17, 19, 21, 24, 27, 29, 31, 33, 36, 38, 40, 43, 47, 50, 53, 58, 61, 63, 65, 69.

Solución

El conjunto consta de $N = 20$ datos, por lo que primero se determinan los valores correspondientes a los cuartiles Q_1 y Q_3.

Para el primer cuartil Q_1, el valor de $\frac{KN}{4}$ es:

$$\frac{KN}{4} = \frac{1(20)}{4} = 5$$

El valor del cuartil Q_1 será la media aritmética entre el 5º dato ordenado y el siguiente del conjunto, de lo que resulta:

$$Q_1 = \frac{27+29}{2} = 28$$

El valor del primer cuartil Q_1 es 28.

Para el tercer cuartil Q_3, el valor de $\frac{KN}{4}$ es:

$$\frac{KN}{4} = \frac{3(20)}{4} = 15$$

El valor del cuartil Q_3 será la media aritmética entre el 15º dato ordenado y el siguiente del conjunto, con lo que se obtiene:

$$Q_3 = \frac{53+58}{2} = 55.5$$

El valor del tercer cuartil Q_3 es 55.5.

Al aplicar la ecuación matemática del rango intercuartílico, tenemos:

$$Q = Q_3 - Q_1 = 55.5 - 28 = 27.5$$

El rango intercuartílico del conjunto de datos es 27.5.

Si empleamos la ecuación matemática del rango semiintercuartílico, tenemos:

$$Q_{S.I.} = \frac{Q_3 - Q_1}{2} = \frac{Q}{2} = \frac{27.5}{2} = 13.75$$

El rango semiintercuartílico o desviación cuartílica del conjunto de datos es 13.75.

Se observa que el 50% de los datos se ubican entre Q_1 y Q_3, es decir, que 10 de los 20 datos que forman el conjunto, se encuentran entre los valores de 28 y 55.5.

Si la medida de centralización es:

$$\frac{Q_1 + Q_2}{2} = \frac{28+55.5}{2} = 41.75$$

Por lo anterior, el 50% de los datos se encuentran en el intervalo de 41.75 ± 13.75, es decir, que el 50% de los datos tienen una desviación de 13.75 respecto a la medida central de 41.75.

2 La siguiente tabla de distribución de frecuencias registra el número de horas de prestación del servicio social de 130 alumnos del cuarto semestre de la especialidad de trabajo social de un bachillerato en la ciudad de Mérida, Yucatán. Determina el rango intercuartílico y la desviación cuartílica y representa gráficamente los resultados.

Intervalos (horas)	Frecuencia (alumnos)
51 - 100	5
101 - 150	9
151 - 200	16
201 - 250	23
251 - 300	39
301 - 350	31
351 - 400	7
	$N = 130$

Solución

Para determinar el primer cuartil Q_1, tenemos de la tabla:

Datos

$$\frac{KN}{4} = \frac{1(130)}{4} = 32.5$$
$$L_1 = \frac{200 + 201}{2} = 200.5$$
$$\left(\sum f\right)_1 = 5 + 9 + 16 = 30$$
$$f_K = 23$$
$$C = 50$$

Fórmula

$$Q_1 = L_1 + \left[\frac{\frac{KN}{4} - (\sum f)_1}{f_K}\right]C$$

Sustitución

$$Q_1 = 200.5 + \left[\frac{32.5 - 30}{23}\right](50)$$
$$Q_1 = 200.5 + 5.43 = 205.93$$

El valor del primer cuartil es 205.93 horas, es decir, 25% de los alumnos ha prestado 205.93 o menos horas de servicio social.

Para determinar el tercer cuartil Q_3, tenemos de la tabla:

Datos

$$\frac{KN}{4} = \frac{3(130)}{4} = 97.5$$
$$L_1 = \frac{300 + 301}{2} = 300.5$$
$$\left(\sum f\right)_1 = 5 + 9 + 16 + 23 + 39 = 92$$
$$f_K = 31$$
$$C = 50$$

Fórmula

$$Q_3 = L_1 + \left[\frac{\frac{KN}{4} - (\sum f)_1}{f_K}\right]C$$

Sustitución

$$Q_3 = 300.5 + \left[\frac{97.5 - 92}{31}\right](50)$$
$$Q_1 = 300.5 + 8.87 = 309.37$$

El valor del tercer cuartil es 309.37 horas, es decir, 75% de los alumnos ha prestado 309.37 o menos horas de servicio social.

Al aplicar la ecuación matemática del rango intercuartílico, tenemos:

$$Q = Q_3 - Q_1 = 309.37 - 205.93 = 103.44$$

El rango intercuartílico de la tabla de distribución es 103.44 horas.

Si se emplea la ecuación matemática del rango semi-intercuartílico, tenemos:

$$Q_{SI} = \frac{Q_3 - Q_1}{2} = \frac{Q}{2} = \frac{103.44}{2} = 51.72$$

El rango semiintercuartílico o desviación cuartílica de la tabla de distribución es 51.72 horas.

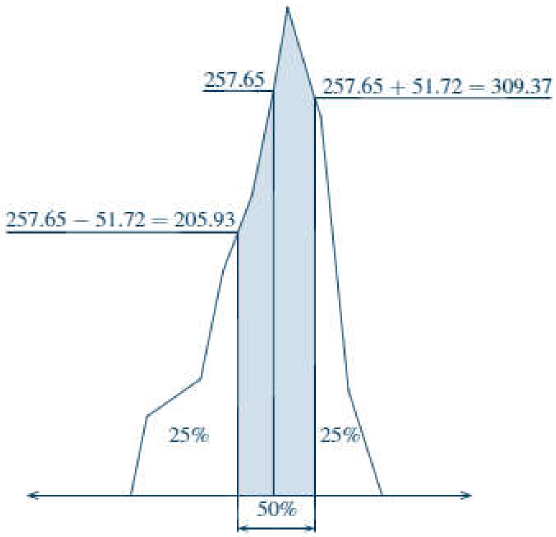
Se observa que el 50% de los alumnos se ubican entre Q_1 y Q_3, es decir, que 65 de los 130 alumnos que forman la tabla de distribución, se encuentran entre los valores de 205.93 y 309.37 horas.

Se considera como medida de centralización:

$$\frac{Q_1 - Q_3}{2} = \frac{205.93 + 309.37}{2} = 257.65$$

Por lo anterior, el 50% de los alumnos se encuentran en el intervalo de 257.65 ± 51.72, es decir, que el 50% de los alumnos tienen una desviación de 51.72 horas respecto a la medida central de 257.65 horas.

La representación gráfica del rango intercuartílico y la desviación cuartílica, es:



Rango entre percentiles 10-90

Para un conjunto de datos ordenados o de una distribución de frecuencias, el **rango entre percentiles 10 - 90** se define como la diferencia entre el nonagésimo percentil P_{90} y el décimo percentil P_{10}; matemáticamente se expresa por la ecuación:

$$\text{Rango entre percentiles 10-90} = P = P_{90} - P_{10}$$

Rango semipercentílico o desviación percentílica

Para un conjunto de datos numéricos o para una distribución de frecuencias, el **rango semipercentílico** o **desviación percentílica** se define como la mitad del rango entre percentiles 10-90; matemáticamente se representa por la expresión:

$$\text{Rango semipercentílico} = P_{SI} = \frac{P_{90} - P_{10}}{2} = \frac{P}{2}$$

EJEMPLOS

Ejemplos

1 Encuentra el rango entre percentiles 10-90 y la desviación percentílica para el siguiente conjunto de datos numéricos ordenados: 3, 5, 6, 7, 8, 9, 10, 12, 15, 18, 21, 23, 27, 30, 32, 33, 35, 36, 38, 39, 41, 42, 44, 45, 47, 48, 51, 53, 54, 55, 57, 58, 59, 60.

Solución

El conjunto consta de $N = 35$ datos, por lo que primero determinamos los valores correspondientes a los percentiles P_{10} y P_{90}.

Para el décimo percentil P_{10} tenemos que:

$$\frac{KN}{100} = \frac{(10)(35)}{100} = 3.5$$

El valor del percentil P_{10} será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100} = 3.5$, al redondear el valor de 3.5 obtenemos el número 4. Por tanto, el percentil es el dato ordenado que se encuentra en la cuarta posición.

El valor del décimo percentil es 7 ya que es el dato ordenado que ocupa la posición cuatro.

Para el nonagésimo percentil P_{90} tenemos que:

$$\frac{KN}{100} = \frac{(90)(35)}{100} = 31.5$$

El valor del percentil P_{90} será el dato ordenado cuya posición se obtiene de redondear al entero más próximo el resultado $\frac{KN}{100} = 31.5$, , al redondear el valor de 31.5 obtenemos el número 32. Por tanto, el nonagésimo percentil es el dato ordenado que se encuentra en la trigésimo segunda posición.

El valor del nonagésimo percentil es 57 ya que es el dato ordenado que ocupa la posición treinta y dos.

La ecuación matemática del rango entre percentiles 10 - 90 arroja:

$$P = P_{90} - P_{10} = 57 - 7 = 50$$

El rango entre percentiles 10 - 90 del conjunto de datos es 50.

Si empleamos la ecuación matemática del rango semipercentílico, tenemos:

$$P_{SI} = \frac{P_{90} - P_{10}}{2} = \frac{P}{2} = \frac{50}{2} = 25$$

El rango semipercentílico o desviación percentílica del conjunto de datos es 25.

Se observa que el 30% de los datos se ubican entre P_{10} y P_{90}, es decir, que 28 de los 35 datos que forman el conjunto, se encuentran entre los valores de 7 y 57.

Considerando como medida de centralización a: $\frac{P_{10} + P_{90}}{2} = \frac{7 + 57}{2} = 32$

Por lo anterior, el 80% de los datos se encuentran en el intervalo de 32 ± 25, es decir, que el 80% de los datos tienen una desviación de 25 respecto a la medida central de 32.

2 La siguiente tabla de distribución de frecuencias registra las calificaciones del examen final de cálculo vectorial de la Escuela Superior de Matemáticas. Determina el rango entre percentiles 10 - 90 y la desviación percentílica y representa gráficamente los resultados.

Intervalos de clase (calificaciones)	Frecuencia de clase (f) (número de estudiantes)
30 - 39	2
40 - 49	3
50 - 59	6
60 - 69	13
70 - 79	21
80 - 89	16
90 - 99	4
	$N = 65$

Solución

Para determinar el décimo percentil P_{10}, tenemos de la tabla:

Datos

$$\frac{KN}{100} = \frac{10(65)}{100} = 6.5$$

$$L_1 = \frac{40+50}{2} = 49.5$$

$$(\sum f)_1 = 2+3 = 5$$

$$f_K = 6$$

$$C = 10$$

Fórmula

$$P_{10} = L_1 + \left[\frac{\frac{KN}{100} - (\sum f)_1}{f_K}\right] C$$

Sustitución

$$P_{10} = 49.5 + \left[\frac{6.5-5}{6}\right](10)$$

$$P_{10} = 49.5 + 2.5 = 52$$

El valor del décimo percentil es 52 puntos, es decir, que el 10% de los alumnos ha registrado calificación de 52 puntos o menos.

Para determinar el nonagésimo percentil P_{90}, tenemos de la tabla:

Datos

$$\frac{KN}{100} = \frac{90(65)}{100} = 58.5$$

$$L_1 = \frac{79+80}{2} = 79.5$$

$$(\sum f)_1 = 2+3+6+13+21 = 45$$

$$f_K = 16$$

$$C = 10$$

Fórmula

$$P_{90} = L_1 + \left[\frac{\frac{KN}{100} - (\sum f)_1}{f_K}\right] C$$

Sustitución

$$P_{90} = 79.5 + \left[\frac{58.5-45}{16}\right](10)$$

$$P_{90} = 79.5 + 8.4375 = 87.9375 \approx 88$$

El valor del nonagésimo percentil es 88 puntos, es decir, que el 90% de los alumnos ha registrado calificación de 88 puntos o menos.

Al aplicar la ecuación matemática del rango entre percentiles 10 - 90, tenemos:

$$P = P_{90} - P_{10} = 88 - 52 = 36$$

El rango entre percentiles 10 - 90 de la tabla de distribución es 36 puntos.

Si se emplea la ecuación matemática del rango semipercentílico, tenemos:

$$P_{SI} = \frac{P_{90} - P_{10}}{2} = \frac{P}{2} = \frac{36}{2} = 18$$

El rango semipercentílico o desviación percentílica de la tabla de distribución es 18 puntos.

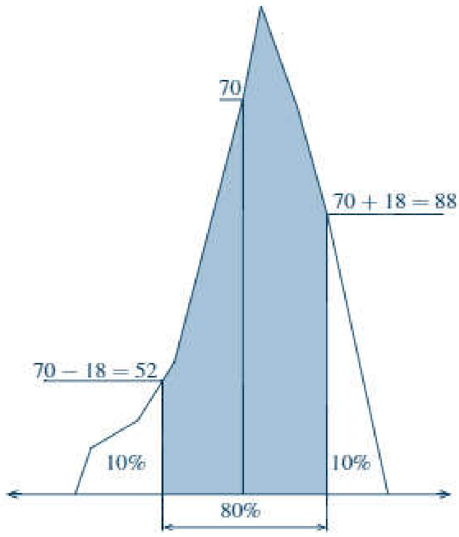
Se observa que el 80% de los alumnos se ubican entre P_{10} y P_{90}, es decir, que 52 de los 56 alumnos que forman la tabla de distribución, se encuentran entre los valores de 52 y 88 puntos.

Considerando como medida de centralización a:

$$\frac{P_{10}+P_{90}}{2}=\frac{52+88}{2}=70$$

Por lo anterior, el 80% de los alumnos se encuentran en el intervalo de 70 ± 18, es decir, que el 80% de los alumnos tiene una desviación de 18 puntos respecto a la medida central de 70 puntos.

La representación gráfica del rango entre percentiles 10 - 90 y la desviación percentílica, es:



EJERCICIO 13

I. Realiza lo que se indica en cada caso.

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

1. Define la desviación media o promedio de desviación.
2. ¿Cuáles son las desventajas de la desviación media en cuanto a la medida de tendencia central que se utiliza para su cálculo?
3. ¿Qué significado tiene el intervalo $\overline{X} \pm DM$?
4. ¿Por qué razón se utiliza el rango intercuartílico?
5. Define el rango intercuartílico
6. Define el rango semiintercuartílico.
7. Cita las desventajas de aplicación que presentan el rango intercuartílico y el rango semiintercuartílico.
8. Define el rango entre percentiles 10 - 90.
9. Define el rango semipercentílico.

II. Resuelve en equipo los siguientes problemas y discute los resultados en plenaria.

1. Determina la desviación media de los conjuntos de elementos numéricos indicados en el inciso *a*.
2. Determina la desviación media de las siguientes tablas de distribución de frecuencias y las indicadas en el inciso *b*.

a)

Intervalos de clase	Frecuencia de clase (f)
127 - 135	6
136 - 144	10
145 - 153	13
154 - 162	9
163 - 171	5
	$N = 43$

b)

Intervalos de clase	Frecuencia de clase (f)
2.3 - 2.8	3
2.9 - 3.4	7
3.5 - 4.0	13
4.1 - 4.6	20
4.7 - 5.2	15
5.3 - 5.8	8
5.9 - 6.4	4
	$N = 70$

3. La siguiente tabla de distribución de frecuencias registra los salarios diarios en dólares de 72 profesionales de la industria petrolera. Determina:

Intervalos (salarios)	Frecuencia (f) (profesionales)
30 - 39	7
40 - 49	12
50 - 59	19
60 - 69	16
70 - 79	10
80 - 89	6
90 - 99	2
	$N = 72$

a) La desviación media respecto a la media aritmética.

b) La desviación media respecto a la mediana.

c) Los porcentajes de los siguientes intervalos: $\overline{X} \pm DM$ $\overline{X} \pm 2\ DM$ $\overline{X} \pm 3\ DM$

d) Construye la gráfica respectiva de los porcentajes.

4. La siguiente tabla de distribución de frecuencias registra la masa en libras de 50 estudiantes de un bachillerato en el estado de Hidalgo. Determina:

Intervalos (pesos)	Frecuencia (f) (estudiantes)
115 - 127	3
128 - l40	15
141 - 153	16
154 - 166	11
167 - l79	4
180 - 192	1
	$N = 50$

a) La desviación media.

b) Los porcentajes para los intervalos: $\overline{X} \pm DM$ $\overline{X} \pm 2\ DM$ y $\overline{X} \pm 3\ DM$

c) Construye la gráfica respectiva de los porcentajes.

5. Determina el rango intercuartílico y la desviación cuartílica para los siguientes conjuntos de datos numéricos:
   a) 12, 8, 7, 18, 9, 2, 5, 14, 16, 22, 6, 27.
   b) 165, 169, 173, 177, 182, 186, 190, 193, 196, 198, 201, 205, 209, 213, 217, 220, 222, 226, 229, 231, 234, 238, 242, 244, 247, 250, 253, 255, 259, 261, 262, 268, 270, 273, 277, 280, 282, 285.
6. Determina el rango intercuartílico y el rango semiintercuartílico y representa gráficamente los resultados para la siguiente tabla de distribución de frecuencias que registra los diámetros interiores en milésimas de pulgada de 40 inyectores para motores Diesel producidos por una compañía.

Intervalos (diámetros)	Frecuencia (inyectores)
0.423 - 0.425	1
0.426 - 0.428	2
0.429 - 0.431	5
0.432 - 0.434	7
0.435 - 0.437	11
0.438 - 0.440	8
0.441 - 0.443	4
0.444 - 0.446	2
	$N = 40$

Verifica tus resultados en la sección de respuestas.

Varianza

Con base en la definición de la **desviación media**, en donde se toman los valores absolutos de las desviaciones respecto a su media aritmética, tenemos que al elevar al cuadrado las desviaciones, obtenemos valores positivos. Al sumar los cuadrados de las desviaciones y dividir entre el número total de elementos N, tenemos como resultado la **varianza**.

La varianza de un conjunto de elementos $X_1, X_2, X_3, ..., X_N$ se define como la media aritmética de los cuadrados de las desviaciones respecto a la media aritmética; se representa por el símbolo S^2 y matemáticamente se expresa por la ecuación:

$$S^2 = \frac{\sum_{j=1}^{N}(X_j - \overline{X})^2}{N} = \frac{\sum(X - \overline{X})^2}{N} = \frac{\sum \overline{X}^2}{N} = (X - \overline{X})^2 \quad \Bigg\} \text{ Fórmula para datos no agrupados}$$

$$S^2 = \frac{\sum_{j=1}^{N} f_j(X_j - \overline{X})^2}{N} = \frac{\sum f(X - \overline{X})^2}{N} = \frac{\sum f X^2}{N} = (X - \overline{X})^2 \quad \Bigg\} \text{ Fórmula para datos agrupados}$$

Es necesario hacer una distinción entre lo que es la **varianza de la población** que se simboliza por σ^2 letra griega minúscula llamada **sigma** y la **varianza de la muestra** extraída de la misma población y que se simboliza por S^2.

Si en las ecuaciones de la varianza para datos no agrupados y agrupados, se dividen entre $N - 1$ en lugar de N, el valor resultante representa un **estimador** de una población de la que se ha tomado una muestra, es decir, se utiliza la varianza de la muestra para determinar la varianza de la población de la que proviene. Cuando $N > 30$, prácticamente no existe diferencia entre los dos denominadores, es decir, se puede utilizar N o $N - 1$.

Desviación típica

Para un conjunto de elementos numéricos $X_1, X_2, X_3, \ldots, X_N$, la desviación típica se define como la raíz cuadrada del cuadrado de la media de los cuadrados de las desviaciones respecto a la media aritmética, es decir, es la raíz cuadrada de la varianza; se representa por el símbolo S y matemáticamente se expresa por la ecuación:

$$S = \sqrt{\frac{\sum_{j=1}^{N}(X_j - \bar{X})^2}{N}} = \sqrt{\frac{\sum(X - \bar{X})^2}{N}} = \sqrt{\frac{\sum X^2}{N}} = \sqrt{(X - \bar{X})^2}$$

La desviación típica también se denomina **desviación estándar** o **raíz cuadrado medio de las desviaciones.**

Desviación típica para datos agrupados

Si se tiene un conjunto $X_1, X_2, X_3, \ldots, X_N$ que se presenta con $f_1, f_2, f_3, \ldots, f_N$ veces de frecuencia, respectivamente, la desviación típica o estándar se determina matemáticamente por la ecuación:

$$S = \sqrt{\frac{\sum_{j=1}^{N} f_j (X_j - \bar{X})^2}{N}} = \sqrt{\frac{\sum f(X - \bar{X})^2}{N}} = \sqrt{\frac{\sum fX^2}{N}} = \sqrt{(X - \bar{X})^2}$$

Al igual que en la varianza, tenemos que la **desviación típica de la población** se simboliza por la letra griega σ y la **desviación típica de la muestra** que proviene de la misma población se simboliza por S.

La desviación típica o estándar es la medida de dispersión más importante y que también presenta la característica de que sus ecuaciones para datos no agrupados y agrupados pueden dividirse entre $N - 1$ en lugar de N.

EJEMPLOS

Ejemplos

1 Encuentra la varianza y la desviación típica para el conjunto de datos numéricos 8, 15, 11, 5, 10, 12, 8 y 13.

Solución

Primero determinamos la media aritmética del conjunto de datos.

$$\bar{X} = \frac{\sum X}{N} = \frac{8+15+11+5+10+12+8+13}{8} = \frac{82}{8} = 10.25$$

Al sustituir en la ecuación de la varianza, tenemos: $S^2 = \dfrac{\sum(X - \bar{X})^2}{N}$

$$S^2 = \frac{(8-10.25)^2 + (15-10.25)^2 + (11-10.25)^2 + (5-10.25)^2 + (10-10.25)^2 + (12-10.25)^2 + (8-10.25)^2 + (13-10.25)^2}{8}$$

$$S^2 = \frac{5.0625 + 22.5625 + 0.5625 + 27.5625 + 0.0625 + 3.0625 + 5.0625 + 7.5625}{8} = \frac{71.5}{8} = 8.9375$$

La varianza del conjunto de datos es 8.9375.

Si sustituimos en la ecuación de la desviación típica o estándar, tenemos: $S = \sqrt{\dfrac{\sum(X - \bar{X})^2}{N}}$

$$S=\sqrt{\frac{(8-10.25)^2+(15-10.25)^2+(11-10.25)^2+(5-10.25)^2+(10-10.25)^2+(12-10.25)^2+(8-10.25)^2+(13-10.25)^2}{8}}$$

$$S=\sqrt{\frac{5.0625+22.5625+0.5625+27.5625+0.0625+3.0625+5.0625+7.5625}{8}}=\sqrt{\frac{71.5}{8}}$$

$S=\sqrt{8.9375}=2.9895$ La desviación típica o estándar es 2.9895.

La desviación típica o estándar se obtiene directamente de la extracción de la raíz cuadrada de la varianza, es decir:

$$S=\sqrt{S^2}=\sqrt{8.9375}=2.9895$$

2 ••• La siguiente tabla de distribución de frecuencias registra el número de horas al mes que 108 miembros de un club altruista dedican a labores sociales en su comunidad. Determinar su varianza y su desviación típica.

Intervalos (horas)	Frecuencia (f) (miembros)
10 - 14	6
15 - 19	14
20 - 24	23
25 - 29	28
30 - 34	19
35 - 39	11
40 - 44	7
	$N = 108$

Solución

Se determina la media aritmética de la tabla de distribución de frecuencias, a partir de la siguiente información:

Intervalos (horas)	Marca de clase (X)	Frecuencia (f)(miembros)	fX
10 - 14	12	6	72
15 - 19	17	14	238
20 - 24	22	23	506
25 - 29	27	28	756
30 - 34	32	19	608
35 - 39	37	11	407
40 - 44	42	7	294
		$N = 108$	$\sum fX = 2\,881$

Al sustituir en la ecuación de la media aritmética, resulta:

$$\overline{X}=\frac{\Sigma fX}{N}=\frac{2\,881}{108}=26.68$$

Para determinar la varianza y la desviación típica, se requiere la siguiente tabla de apoyo:

Marca de clase (X)	Frecuencia (f) (miembros)	$(X-\overline{X})$	$(X-\overline{X})^2$	$f(X-\overline{X})^2$
12	6	−14.68	215.5024	1 293.0144
17	14	−9.68	93.7024	1 311.8336
22	23	−4.68	21.9024	503.7552
27	28	0.32	0.1024	2.8672
32	19	5.32	28.3024	537.7456
37	11	10.32	106.5024	1 171.5264
42	7	15.32	234.7024	1 642.9168
	$N = 108$			$\sum f(X-\overline{X})^2 = 6\,463.6592$

Después de sustituir en la ecuación de la varianza, tenemos:

$$S^2 = \frac{\sum f(X-\bar{X})^2}{N} = \frac{6463.6592}{108} = 59.85$$

La varianza de la tabla de distribución es 59.85.

Si sustituimos en la ecuación de la desviación típica, tenemos:

$$S = \sqrt{\frac{\sum f(X-\bar{X})^2}{N}} = \sqrt{\frac{6463.6592}{108}} = \sqrt{59.85} = 7.736$$

La desviación típica o estándar de la tabla de distribución es 7.736.

Métodos para calcular la varianza y la desviación típica

Métodos largos. Las ecuaciones correspondientes a la **varianza** y la **desviación típica** para datos no agrupados y agrupados, pueden expresarse mediante una forma equivalente que es más sencilla.

1. Para la ecuación de la varianza para datos no agrupados, tenemos:

$$S = \frac{\sum f\left(X-\bar{X}\right)^2}{N} = \frac{\sum(X^2-2\bar{X}+\bar{X}^2)}{N} = \frac{\sum X^2 - 2\bar{X}\sum X + \bar{X}^2}{N} = \frac{\sum X^2}{N} - 2\bar{X}\frac{\sum X}{N} + \frac{\sum \overline{X}^2}{N}$$

$$S = \frac{\sum X^2}{N} - 2\bar{X}(\bar{X}) + \frac{N\bar{X}^2}{N} = \frac{\sum X^2}{N} - 2\bar{X}^2 + \bar{X}^2 = \frac{\sum X^2}{N} - \bar{X}^2$$

$$S^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2 = \bar{X}^2 - (\bar{X})^2$$

2. Para la ecuación de la desviación típica para datos no agrupados, tenemos:

$$S^2 = \sqrt{\frac{\sum (X-\bar{X})^2}{N}} - \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2} = \sqrt{\overline{X}^2 - (\bar{X})^2}$$

3. Para la ecuación de la varianza para datos agrupados, tenemos:

$$S^2 = \frac{\sum f\left(X-\bar{X}\right)^2}{N} = \frac{\sum f\left(X^2 - 2\bar{X}X + \bar{X}^2\right)}{N} = \frac{\sum f X^2 - 2\bar{X}\sum f X + \sum f X^2}{N}$$

$$S^2 = \frac{\sum f X^2}{N} - 2\bar{X}\frac{\sum fXN}{N} + \frac{\sum f X^2}{N} = \frac{\sum f X^2}{N} - 2\bar{X}(\bar{X}) + \frac{N\bar{X}^2}{N}$$

$$S^2 = \frac{\sum f X^2}{N} - 2\bar{X}^2 + \bar{X}^2 = \frac{\sum f X^2 \sum f X^2}{N} - \bar{X}^2 = \frac{\sum f X^2}{N} - \left(\frac{\sum f X^2}{N}\right)^2 = \bar{X}^2 - (\bar{X})^2$$

4. Para la ecuación de la desviación típica para datos agrupados, tenemos:

$$S = \sqrt{\frac{\sum f\,(X-\bar{X})^2}{N}} = \sqrt{\frac{\sum f\,X^2}{N} - \left(\frac{\sum f\,X}{N}\right)^2} = \sqrt{\bar{X}^2 - (\bar{X})^2}$$

En todas las ecuaciones anteriores identificamos a $\overline{X}^2$ como la media de los cuadrados de los distintos valores de X; mientras que $(\bar{X})^2$ representa el cuadrado de la media de los diferentes valores de X.

Métodos cortos. Si la ecuación $d = X - A$, representa las desviaciones de los distintos valores de X respecto a un valor arbitrario y constante A; al asociar lo anterior con las ecuaciones correspondientes de la **varianza** y la **desviación típica** para datos no agrupados y agrupados, resulta:

1. Ecuación de la varianza para datos no agrupados:

$$\text{Si } d = X - A, \text{ donde } X = A + d \text{ y } \overline{X} = A + \overline{d}; \text{ entonces tenemos que:}$$
$$X - \overline{X} = (A + d) - (A + \overline{d}) = A + d - A - \overline{d} = d - \overline{d}$$

Al sustituir en la ecuación de la varianza, tenemos:

$$S^2 = \frac{\sum(X-\overline{X})^2}{N} = \frac{\sum(d-\overline{d})^2}{N} = \frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2 = \overline{d^2} - (\overline{d})^2$$

2. Para la ecuación de la desviación típica para datos no agrupados, tenemos:

$$S = \sqrt{\frac{\sum(X-\overline{X})^2}{N}} = \sqrt{\frac{\sum(d-\overline{d})^2}{N}} = \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2} = \sqrt{\overline{d^2}(\overline{d})^2}$$

3. Para la ecuación de la varianza para datos agrupados, tenemos:

$$S = \frac{\sum f(X-\overline{X})^2}{N} = \frac{\sum F(d-\overline{d})^2}{N} = \frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2 = \overline{d}^2 - (\overline{d})^2$$

4. Para la ecuación de la desviación típica para datos agrupados, tenemos:

$$S = \sqrt{\frac{\sum f(X-\overline{X})^2}{N}} = \sqrt{\frac{\sum f(d-\overline{d})^2}{N}} = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} = \sqrt{\overline{d}^2 - (\overline{d})^2}$$

Métodos claves. Las ecuaciones correspondientes a la varianza y la desviación típica para datos agrupados en donde los intervalos de clase deben tener igual tamaño C y desviaciones $d = X - A$ se expresan como $d = Cu$ o $X = A + Cu$, donde $u = \dfrac{X-A}{C}$.

1. Ecuación de la varianza para datos agrupados:

$$\text{Si } X = A + Cu \text{ y } \overline{X} = A + C\overline{u}$$
$$X - \overline{X} = (A + Cu) - (A + C\overline{u}) = A + Cu - A - C\overline{u} = Cu - C\overline{u} = C(u - \overline{u})$$

Al sustituir en la ecuación de la varianza, tenemos:

$$S^2 = \frac{\sum f(X-\overline{X})^2}{N} = \frac{\sum f(Cu-C\overline{u})^2}{N} = \frac{\sum f(Cu)^2}{N} - \left[\frac{\sum f(Cu)}{N}\right]^2 = C^2\frac{\sum fu^2}{N} - C^2\left(\frac{\sum fu}{N}\right)^2$$

$$S^2 = \left[\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2\right]C^2 = \left[\overline{u^2} - (\overline{u})^2\right]C^2$$

2. Para la ecuación de la desviación típica para datos agrupados, tenemos:

$$S = \sqrt{\frac{\sum f(X-\overline{X})^2}{N}} = \sqrt{\frac{\sum f(Cu-C\overline{u})^2}{N}} = \sqrt{\frac{\sum f(Cu)^2}{N} - \left[\frac{\sum f(Cu)}{N}\right]^2}$$

$$S^2 = \sqrt{C^2\frac{\sum fu^2}{N} - C^2\left(\frac{\sum fu}{N}\right)^2} = C\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2} = C\sqrt{\overline{u^2} - (\overline{u})^2}$$

EJEMPLOS

Ejemplos

1 Dado el conjunto de elementos numéricos 11, 3, 5, 2, 14, 9, 7, 15, 4 y 6, determina la varianza y la desviación típica mediante:

a) Método largo. *b*) Método corto.

Solución

a) Primero se calcula:

$$\overline{X}^2=\frac{\sum X^2}{N}=\frac{(11)^2+(3)^2+(5)^2+(2)^2+(14)^2+(9)^2+(7)^2+(15)^2+(4)^2+(6)^2}{10}$$

$$\overline{X}^2=\frac{\sum X^2}{N}=\frac{121+9+25+4+196+81+49+225+16+36}{10}=\frac{762}{10}=76.2$$

Ahora determinamos:

$$\overline{X}=\frac{\sum X}{N}=\frac{11+3+5+2+14+9+7+15+4+6}{10}=\frac{76}{10}=7.6$$

Al sustituir en la ecuación de la varianza, tenemos:

$$S^2=\frac{\sum X^2}{N}-\left(\frac{\sum X}{N}\right)^2=76.2-(7.6)^2=76.2-57.76=18.44$$

La varianza del conjunto de datos por el método largo es 18.44.

Al sustituir en la ecuación de la desviación típica, tenemos:

$$S=\sqrt{\frac{\sum X^2}{N}-\left(\frac{\sum X}{N}\right)^2}=\sqrt{76.2-(7.6)^2}=\sqrt{76.2-57.76}=\sqrt{18.44}\approx 4.3$$

La desviación típica del conjunto de datos por el método largo es 4.3.

b) Al tomar como valor arbitrario y constante de $A=9$ y aplicando la ecuación $d=X-A$, tenemos que:

$$\overline{d}^2=\frac{\sum d^2}{N}=\frac{(11-9)^2+(3-9)^2+(5-9)^2+(2-9)^2+(14-9)^2+(9-9)^2+(7-9)^2(15-9)^2+(4-9)^2+(6-9)^2}{10}$$

$$\overline{d}^2=\frac{\sum d^2}{N}=\frac{4+36+16+49+25+0+4+36+25+9}{10}=\frac{204}{10}=20.4$$

Ahora determinamos:

$$\overline{d}=\frac{\sum d}{N}=\frac{(11-9)+(3-9)+(5-9)+(2-9)+(14-9)+(9-9)+(7-9)+(15-9)+(4-9)+(6-9)}{10}$$

$$\overline{d}=\frac{\sum d}{N}=\frac{2+(-6)+(-4)+(-7)+5+0+(-2)+6+(-5)+(-3)}{10}=\frac{-14}{10}=-1.4$$

Si se sustituye en la ecuación de la varianza, tenemos:

$$\overline{S}^2=\frac{\sum d^2}{N}-\left(\frac{\sum d}{N}\right)^2=20.4-(-1.4)^2=20.4-1.96=18.44$$

La varianza del conjunto de datos por el método corto es 18.44.

Al sustituir en la ecuación de la desviación típica, tenemos:

$$S=\sqrt{\frac{\sum d^2}{N}-\left(\frac{\sum d^2}{N}\right)^2}=\sqrt{20.4-(-1.4)^2}=\sqrt{20.4-1.96}=\sqrt{18.44}\approx 4.3$$

La desviación típica del conjunto de datos por el método corto es 4.3.

Comprobamos por ambos métodos que obtenemos los mismos resultados.

2 La siguiente tabla de distribución de frecuencias registra la masa en kilogramos de 126 niños que nacieron en los distintos hospitales y sanatorios, el 13 de febrero de 2007; determina la varianza y la desviación típica por el:

a) Método largo.

b) Método corto.

c) Método clave.

Intervalos (masa)	Frecuencia (niños) (f)
2.3 - 2.6	1
2.7 - 3.0	7
3.1 - 3.4	13
3.5 - 3.8	25
3.9 - 4.2	32
4.3 - 4.6	24
4.7 - 5.0	16
5.1 - 5.4	8
	$N = 126$

Solución

a) Para determinar la varianza y la desviación típica por el método largo, nos apoyamos en la siguiente tabla de distribución:

Intervalos (masa)	Marca de clase (X)	X^2	Frecuencia (niños) (f)	fX	fX^2
2.3 - 2.6	2.45	6.0025	1	2.45	6.0025
2.7 - 3.0	2.85	8.1225	7	19.95	56.8575
3.1 - 3.4	3.25	10.5625	13	42.25	137.3125
3.5 - 3.8	3.65	13.3225	25	91.25	333.0625
3.9 - 4.2	4.05	16.4025	32	129.60	524.8800
4.3 - 4.6	4.45	19.8025	24	106.80	475.2600
4.7 - 5.0	4.85	23.5225	16	77.60	376.3600
5.1 - 5.4	5.25	27.5625	8	42.00	220.5000
			$N = 126$	$\Sigma fX = 511.9$	$\Sigma fX^2 = 2\,130.235$

Al sustituir en la ecuación de la varianza, tenemos:

$$S^2 = \frac{\sum f X^2}{N} - \left(\frac{\sum f X}{N}\right)^2 = \frac{2130.235}{126} - \left(\frac{511.9}{126}\right)^2 = 16.9066 - 16.5055 = 0.4011$$

La desviación típica de la tabla de distribución por el método largo es 0.4011.

Al sustituir en la ecuación de la desviación típica, tenemos:

$$S = \sqrt{\frac{\sum f X^2}{N} - \left(\frac{\sum f X}{N}\right)^2} = \sqrt{\frac{2130.235}{126} - \left(\frac{511.9}{126}\right)^2} = \sqrt{16.9066 - 16.5055}$$

$$S = \sqrt{0.4011} = 0.6333$$

La desviación típica de la tabla de distribución por el método largo es 0.6333.

b) Para determinar la varianza y la desviación típica por el método corto, seleccionamos como marca de clase constante y arbitraria al valor de $A = 3.25$ (puede ser cualquier otro), también nos apoyamos en la siguiente tabla de distribución:

Marca de clase (X)	$d = X - A$	d^2	Frecuencia (niños) (*f*)	fd	fd^2
2.45	−0.8	0.64	1	−0.8	0.64
2.85	−0.4	0.16	7	−2.8	1.12
3.25	0	0	13	0	0
3.65	0.4	0.16	25	10.0	4.00
4.05	0.8	0.64	32	25.6	20.48
4.45	1.2	1.44	24	28.8	34.56
4.85	1.6	2.56	16	25.6	40.96
5.25	2.0	4.00	8	16.0	32.00
			$N = 126$	$\Sigma fd = 102.4$	$\Sigma fd^2 = 133.76$

Al sustituir en la ecuación de la varianza, tenemos:

$$S^2 = \frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2 = \frac{133.76}{126} - \left(\frac{102.4}{126}\right)^2 = 1.0615 - 0.6604 = 0.4011$$

La varianza de la tabla de distribución por el método corto es 0.4011.

Al sustituir en la ecuación de la desviación típica, tenemos:

$$S^2 = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} = \sqrt{\frac{133.76}{126} - \left(\frac{102.4}{126}\right)^2} = \sqrt{1.0615 - 0.6604}$$

$$S = \sqrt{0.4011} = 0.6333$$

La desviación típica de la tabla de distribución por el método corto es 0.6333.

c) Para determinar la varianza y la desviación típica por el método clave, seleccionamos el valor de la constante arbitraria $A = 4.45$ (puede ser cualquier otro de los valores que se encuentran en la marca de clase), el tamaño de los intervalos de clase en $C = 0.4$ y nos apoyamos con la siguiente tabla de distribución:

Marca de clase (X)	$u = \frac{X - A}{C}$	u^2	Frecuencia (niños) (*f*)	fu	fu^2
2.45	−5	25	1	−5	25
2 .85	−4	16	7	−28	112
3.25	−3	9	13	−39	117
3.65	−2	4	25	−50	100
4.05	−1	1	32	−32	32
4.45	0	0	24	0	0
4.85	1	1	16	16	16
5.25	2	4	8	16	32
			$N = 126$	$\Sigma fu = -122$	$\Sigma fu^2 = 434$

Si sustituimos en la ecuación de la varianza, tenemos:

$$S^2 = \left[\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2\right]C^2 = \left[\frac{434}{126} - \left(\frac{-122}{126}\right)^2\right](0.4)^2 = (3.4444 - 0.9375)(0.16) = 0.4011$$

La varianza de la tabla de distribución por el método clave es 0.4011.

Después sustituir en la ecuación de la desviación típica, tenemos:

$$S = C\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum f\,u}{N}\right)^2} = (0.4)\sqrt{\frac{434}{126} - \left(\frac{-122}{126}\right)^2} = (0.4)\sqrt{3.4444 - 0.9375}$$

$$S = (0.4)\sqrt{25069} = (0.4)(1.5833) = 0.6333$$

La desviación típica de la tabla de distribución por el método clave es 0.6333.

Observamos que por cualquiera de los métodos usados, obtenemos los mismos resultados para la varianza y la desviación típica o estándar.

Propiedades de la desviación típica o estándar

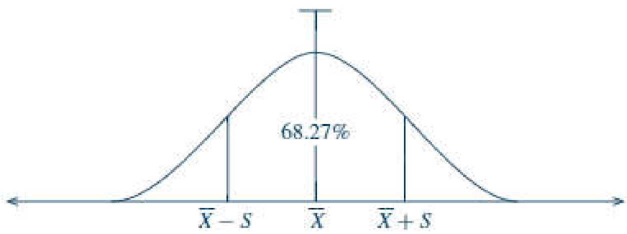
1. Si a es cualquier promedio diferente de la media aritmética, la desviación típica o estándar se puede definir por la expresión matemática:

$$S = \sqrt{\frac{\sum_{j=1}^{N}(X_j - a)^2}{N}} = \sqrt{\frac{\sum(X-a)^2}{N}}$$

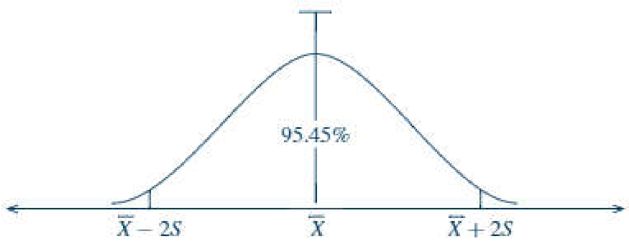
De todas las desviaciones típicas que resulten para cualquier elección del valor de a, la desviación típica o estántar será mínima cuando $a = X$.

2. Si se tienen distribuciones normales, se observa que:

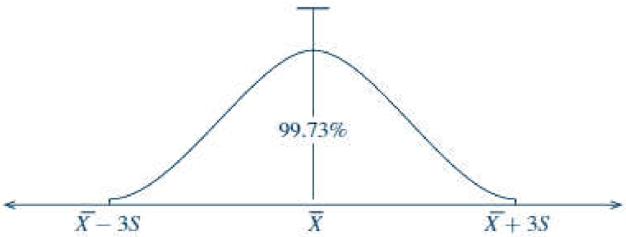
   a) El 68.27% de las observaciones quedan comprendidas entre $\overline{X} \pm S$; es decir, es el valor de la desviación típica o estándar a ambos lados de la media aritmética. Gráficamente se representa como:



   b) El 95.45% de las observaciones quedan comprendidas entre $\overline{X} \pm 2S$; es decir, es el doble de valor de la desviación típica o estándar a ambos lados de la media aritmética. Gráficamente se representa como:



c) El 99.73% de las observaciones quedan comprendidas entre $\overline{X} \pm 3S$; es decir, es el triple de valor de la desviación típica o estándar a ambos lados de la media aritmética. Gráficamente se representa por:



Cuando las distribuciones son moderadamente asimétricas o sesgadas, los porcentajes señalados anteriormente suelen mantenerse aproximados.

3. Si suponemos dos conjuntos de elementos numéricos de N_1 y N_2 datos o dos distribuciones de frecuencias totales, con variaciones respectivas de S_1^2 y S_2^2 y que además presentan la misma media aritmética ($\overline{X}$); en tal caso la **varianza combinada** para ambos conjuntos o ambas distribuciones, se expresa por la siguiente ecuación matemática:

$$S^2 = \frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2}$$

La expresión anterior es la media aritmética ponderada de las varianzas de dichos conjuntos de datos o distribuciones de frecuencias.

La ecuación matemática considerada se puede generalizar a más de 3 conjuntos de datos o distribuciones de frecuencia, siempre y cuando tengan la misma media aritmética.

EJEMPLOS

Ejemplos

1 Dado el conjunto de elementos numéricos (5, 9, 9, 11, 13, 16 y 19), verifica la primera propiedad de la desviación típica o estándar.

Solución

Determinamos primero las medidas de centralización (media aritmética, mediana y moda), resultando:

$$\overline{X} = \frac{\sum X}{N} = \frac{5+9+9+11+13+16+19}{7} = \frac{82}{7} = 11.71$$

Por definición la mediana es 11 y también, por definición, la moda es 9.

Si $a = \overline{X}$, la desviación típica del conjunto de elementos es: $S = \sqrt{\frac{\sum(X-a)^2}{N}}$

Si $a < \overline{X}$, es decir, que a representa el valor de la mediana (11), la desviación típica del conjunto de elementos es:

$$S = \sqrt{\frac{\sum(X-a)^2}{N}} = \sqrt{\frac{(5-11)^2+(9-11)^2+(9-11)^2+(11-11)^2+(13-11)^2+(16-11)^2+(19-11)^2}{7}}$$

$$S = \sqrt{\frac{36+4+4+0+4+25+64}{7}} = \sqrt{\frac{137}{7}} = \sqrt{19.57142857} = 4.4239$$

Si $a < \overline{X}$, es decir, que a representa el valor de la moda (9), la desviación típica del conjunto de elementos es:

$$S = \sqrt{\frac{\sum(X-a)^2}{N}} = \sqrt{\frac{(5-9)^2+(9-9)^2+(9-9)^2+(11-9)^2+(13-9)^2+(16-9)^2+(19-9)^2}{7}}$$

$$S = \sqrt{\frac{16+0+0+4+16+49+100}{7}} = \sqrt{\frac{185}{7}} = \sqrt{26.42857143} = 5.1408$$

Comparando los resultados anteriores, se hace notar que cuando $a = \overline{X}$, el valor de la desviación típica es mínimo.

2 La siguiente tabla de distribución de frecuencias registra las tasas porcentuales de interés que pagó un banco sobre 525 cuentas de retiro individual de tasa variable a 12 meses en 2010.

Comprueba la segunda propiedad de la desviación típica o estándar y construye una gráfica representativa.

Intervalos (porcentajes)	Frecuencia (cuentas)
14.21 - 14.40	7
14.41 - 14.60	24
14.61 - 14.80	52
14.81 - 15.00	79
15.01 - 15.20	93
15.21 - 15.40	81
15.41 - 15.60	75
15.61 - 15.80	68
15.81 - 16.00	36
16.01 - 16.20	10
	$N = 525$

Solución

Para calcular la media aritmética y la desviación típica o estándar, se requiere la siguiente tabla de apoyo:

Intervalos (porcentajes)	Marca de clase (X)	X^2	Frecuencia cuentas (f)	fX	fX^2
14.21 - 14.40	14.305	204.633	7	100.135	1 432.431
14.41 - 14.60	14.505	210.395	24	348.120	5 049.480
14.61 - 14.80	14.705	216.237	52	764.660	11 244.325
14.81 - 15.00	14.905	222.159	79	1 177.495	17 550.563
15.01 - 15.20	15.105	228.161	93	1 404.765	21 218.975
15.21 - 15.40	15.305	234.243	81	1 239.705	18 973.685
15.41 - 15.60	15.505	240.405	75	1 162.875	18 030.376
15.61 - 15.80	15.705	246.647	68	1 067.940	16 771.997
15.81 - 16.00	15.905	252.969	36	572.580	9 106.884
16.01 - 16.20	16.105	259.371	10	161.050	2 593.710
			$N = 525$	$\sum fX = 7\,999.325$	$\sum fX^2 = 1\,211\,972.4307$

Al sustituir en la ecuación de la media aritmética, resulta:

$$\overline{X}=\frac{\sum f X}{N}=\frac{7999.325}{525}=15.2368$$

Al sustituir en la ecuación de la desviación típica, resulta:

$$S=\sqrt{\frac{\sum f X^2}{N}-\left(\frac{\sum f X}{N}\right)^2}=\sqrt{\frac{1211972.4307}{525}-\left(\frac{7999.325}{252}\right)^2}=\sqrt{232.3284394-232.1603644}$$

$$S=\sqrt{0.1680750}=0.41$$

1. Para determinar el porcentaje $\overline{X}\pm S$, tenemos:

Sea $\overline{X}\pm S = 15.2368 \pm 0.41$, el intervalo resultante es de $14.8268 \approx 14.83$ a $15.6468 \approx 15.65$ porcentaje de interés.

Si se considera el valor 14.83, se tiene que el límite real superior de la tercera clase es 14.805; considerando el valor de 15.65, se tiene que el límite real inferior de la séptima clase es 15.605. El tamaño de los intervalos de clase es de 0.2 porcentaje de interés.

El intervalo de 14.83 a 15.65, está integrado por todas las cuentas de la cuarta clase (79), menos $\frac{1}{0.2}(14.83-14.805)$, de las cuentas de la cuarta clase (79), más todas las cuentas de la quinta, sexta y séptima clase (93 + 81 + 75), más $\frac{1}{0.2}(15.65-15.605)$ de las cuentas de la octava clase (68), es decir:

$$79-\frac{1}{0.2}(14.83-14.805)(79)+93+81+75+\frac{1}{0.2}(15.65-15.605)(68)$$

$$79-9.875+93+81+75+15.3=333.425$$

Al resolver la siguiente regla de tres, resulta:

$$\begin{array}{rcl} 525 \text{ cuentas} & \longrightarrow & 100\% \\ 333.425 \text{ cuentas} & \longrightarrow & X \end{array} \qquad X = 63.51\%$$

El porcentaje de $\overline{X}\pm S$ es el 63.51% del total.

2. Para determinar el porcentaje $\overline{X}\pm 2S$, tenemos:

Sea $\overline{X}\pm 2S = 15.2368 \pm 2\,(0.41)$, el intervalo resultante es de $14.4168 \approx 14.42$ a $16.0568 \approx 16.06$ porcentaje de interés.

Si consideramos el valor de 14.42, se tiene que el límite real superior de la primera clase es 14.405; considerando el valor de 16.06, se tiene que el límite real inferior de la novena clase es 16.005; el tamaño de los intervalos de clase es de 0.2 porcentaje de interés.

El intervalo 14.42 a 16.06, se integra por todas las cuentas de la segunda clase (24); menos $\frac{1}{0.2}$ (14.42 – 14.405) de las cuentas de la segunda clase (24); más todas las cuentas de la tercera, cuarta, quinta, sexta, séptima, octava y novena clase (52 + 79 + 93 + 81 + 75 + 68 + 36); más $\frac{1}{0.2}(16.06-16.005)$ de las cuentas de la décima clase (10), es decir:

$$24-\frac{1}{0.2}(14.42-14.405)(24)+52+79+93+81+75+68+36+\frac{1}{0.2}(16.06-16.005)(10)$$

$$24-1.8+52+79+93+81+75+68+36+2.75=508.95$$

Al resolver la siguiente regla de tres, resulta:

$$\begin{array}{rcl} 525 \text{ cuentas} & \longrightarrow & 100\% \\ 508.95 \text{ cuentas} & \longrightarrow & X \end{array} \qquad X = 96.94\%$$

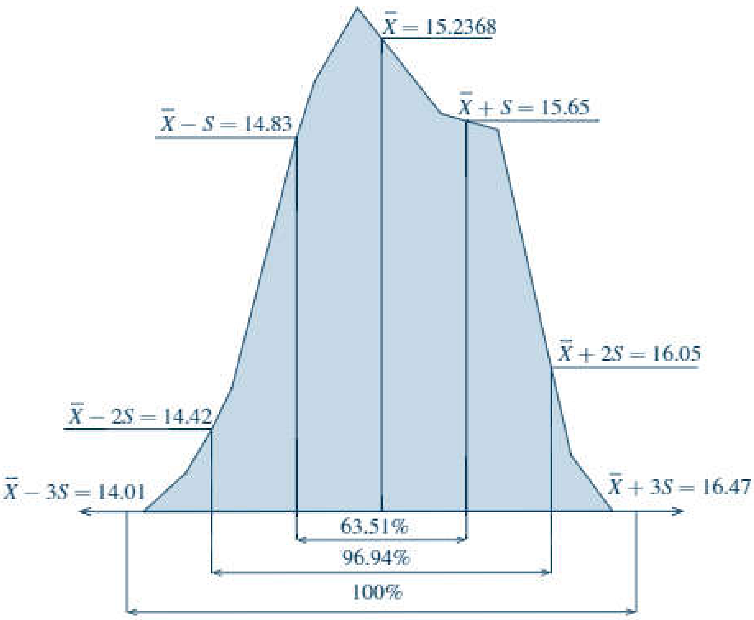
El porcentaje de $\overline{X}\pm 2S$ es el 96.94% del total.

3. Para determinar el porcentaje $\overline{X} \pm 3S$, tenemos:

Sea $\overline{X} \pm 3S = 15.2368 \pm 3\ (0.41)$, el intervalo resultante es de $14.0068 \approx 14.01$ a $16.4668 \approx 16.47$ porcentaje de interés.

Se observa que el intervalo de 14.01 a 16.47 queda fuera de los límites reales de la primera y última clase, es decir, que el porcentaje $\overline{X} \pm 3S$ es el 100% del total.

La representación gráfica de los porcentajes $\overline{X} \pm S$, $\overline{X} \pm 2S$, $\overline{X} \pm 3S$ es:



3 Dados los siguientes conjuntos de elementos numéricos (2, 6, 9, 13, 16, 20) y (4, 8, 14, 18), determina:

a) La media aritmética de cada conjunto.
b) La varianza de cada conjunto.
c) La media aritmética de los dos conjuntos combinados.
d) La varianza de los dos conjuntos combinados (demostración de la propiedad tres de la desviación típica o estándar).

Solución

a) La media aritmética del primer conjunto es:

$$\overline{X}_1 = \frac{\sum X}{N_1} = \frac{2+6+9+13+16+20}{6} = \frac{66}{6} = 11$$

La media aritmética del segundo conjunto es: $\overline{X}_2 = \dfrac{\sum X}{N_2} = \dfrac{4+8+14+18}{4} = \dfrac{44}{4} = 11$

La media aritmética de cada conjunto es 11.

b) La varianza del primer conjunto es:

$$S_1^2 = \frac{\sum(X-\overline{X})^2}{N_1} = \frac{(2-11)^2+(6-11)^2+(9-11)^2+(13-11)^2+(16-11)^2+(20-11)^2}{6}$$

$$S_1^2 = \frac{81+25+4+4+25+81}{6} = \frac{220}{6} = 36.667$$

La varianza del segundo conjunto es:

$$S_2^2 = \frac{\sum (X-\overline{X})^2}{N_2} = \frac{(4-11)^2+(8-11)^2+(14-11)^2+(18-11)^2}{4} = \frac{49+9+9+49}{4} = \frac{116}{4} = 29$$

La varianza del primer conjunto es 36.667 y la del segundo conjunto es 29.

c) La media aritmética de los dos conjuntos combinados es:

$$\overline{X} = \frac{\sum X_{N1} + \sum X_{N2}}{N_1 + N_2} = \frac{2+6+9+13+16+20+4+8+14+18}{6+4} + \frac{110}{10} = 11$$

La media aritmética combinada de los dos conjuntos es 11.

d) La varianza de los dos conjuntos combinados es:

$$\overline{X} = \frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2} = \frac{(6)(36.667)+(4)(29)}{6+4} = \frac{220+116}{10} = \frac{336}{10} = 33.6$$

La varianza combinada de los dos conjuntos es 33.6.

Comprobación Charlier

En los cálculos de medias aritméticas y desviaciones típicas o estándar por el método clave, se emplea la comprobación **Charlier** para demostrar los resultados obtenidos mediante las siguientes igualdades:

a) Para la media aritmética, tenemos:

$$\sum f(u+1) = \sum fu + \sum f = \sum fu + N$$

b) Para la desviación típica o estándar, tenemos:

$$\sum f(u+1)^2 = \sum f(u^2+2u+1) = \sum fu^2 + 2\sum fu + \sum f = \sum fu^2 + 2\sum fu + N$$

EJEMPLO

Ejemplo 1 La siguiente tabla de distribución de frecuencias registra los diámetros en pulgadas de las cabezas de los remaches fabricados por una compañía. Mediante la comprobación Charlier, verifica los cálculos de la media aritmética y la desviación típica o estándar.

Intervalos (diámetros)	Frecuencia (remaches)
0.247 - 0.249	6
0.250 - 0.252	8
0.253 - 0.255	15
0.256 - 0.258	43
0.259 - 0.261	65
0.262 - 0.264	51
0.265 - 0.267	29
0.268 - 0.270	17
0.271 - 0.273	11
0.274 - 0.276	5
	$N = 250$

Solución

Para determinar la media aritmética por el método clave, se requiere la siguiente tabla de apoyo:

	Marca de clase (X)	Desviaciones $u=\frac{X-A}{C}$	Frecuencias (remaches) (f)	fu
	0.248	−3	6	−18
	0.251	−2	8	−16
	0.254	−1	15	−15
$A\left(\text{Media supuesta}\right)\rightarrow$	0.257	0	43	0
	0.260	1	65	65
	0.263	2	51	102
	0.266	3	29	87
	0.269	4	17	68
	0.272	5	11	55
	0.275	6	5	30
			$N = 250$	$\sum fu = 358$

Sea C el tamaño del intervalo de clase igual a (0.003) y sustituyendo en la ecuación respectiva, resulta:

$$\overline{X} = A + \left(\frac{\sum f u}{N}\right)C = 0.257 + \left(\frac{358}{250}\right)(0.003) = 0.257 + 0.004296 = 0.261296$$

La media aritmética de los diámetros por el método clave es 0.261296 pulg.

Para realizar la comprobación Charlier, se requieren los siguientes datos que se obtienen de la tabla de apoyo anterior:

u	$u+1$	Frecuencia (remaches) (f)	$f(u+1)$
−3	−2	6	−12
−2	−1	8	−8
−l	0	15	0
0	1	43	43
1	2	65	130
2	3	51	153
3	4	29	116
4	5	17	85
5	6	11	66
6	7	5	35
		$N = 250$	$\sum f(u+1) = 608$

Al sustituir en la identidad, resulta: $\sum f(u+1) = \sum f u + N$

$$608 = 358 + 250$$

$$608 = 608$$

Para determinar la desviación típica o estándar por el método clave, se requieren los siguientes datos que se obtienen de la tabla de apoyo:

u	u^2	Frecuencia (remaches) (f)	$f\,u$	$f\,u^2$
−3	9	6	−18	54
−2	4	8	−16	32
−1	1	15	−15	15
0	0	43	0	0
1	1	65	65	65
2	4	51	102	204
3	9	29	87	261
4	16	17	68	272
5	25	11	55	275
6	36	5	30	180
		$N = 250$	$\sum f\,u = 358$	$\sum f\,u^2 = 1\,358$

Sea ***C*** el tamaño del intervalo de clase igual a (0.003) y sustituyendo en la ecuación respectiva, resulta:

$$S = C\sqrt{\frac{\sum f\,u^2}{N} - \left(\frac{\sum f\,u}{N}\right)^2} = (0.003)\sqrt{\frac{1\,358}{250} - \left(\frac{358}{250}\right)^2} = (0.003)\sqrt{5.432 - 2.050624}$$

$$S = (0.003)\sqrt{3.381376} = 0.00551655$$

La desviación típica de los diámetros por el método clave es 0.00551655 pulgadas.

Para realizar la comprobación Charlier, se requieren los siguientes datos que se obtienen de las tablas de apoyo anteriores:

u	u^2	$(u+1)$	$(u+1)^2$	Frecuencias (remaches) (f)	$f\,u$	$f\,u^2$	w
−3	9	−2	4	6	−18	54	24
−2	1	−1	1	8	−16	32	8
−1	1	0	0	15	−15	15	0
0	0	1	1	43	0	0	43
1	1	2	4	65	65	65	260
2	4	3	9	51	102	204	459
3	9	4	16	29	87	261	464
4	16	5	25	17	68	272	425
5	25	6	36	11	55	275	396
6	36	7	49	5	30	180	245
				$N = 250$	$\sum f\,u = 358$	$\sum f\,u^2 = 1\,358$	$\sum f(u+1)^2 = 2\,324$

Sustituyendo en la identidad, resulta: $\sum f\,(u+1)^2 = \sum f\,u^2 + 2\sum f\,u + N$

$$2\,324 = 1\,358 + 2\,(358) + 250$$

$$2\,324 = 2\,324$$

Corrección Sheppard para la varianza

Al determinar la desviación típica o estándar en datos agrupados, se obtiene algo de error; generalmente se debe al error de agrupar los datos en clases. Para ser consistentes con la realidad de la información y eliminar el error, se emplea la **varianza corregida** que matemáticamente se expresa por la ecuación:

$$\text{Varianza corregida} = (S^2) - \frac{C^2}{12}$$

El tamaño de los intervalos de clase se representa por C y la corrección **Sheppard** por $\left(\frac{C^2}{12}\right)$.

La corrección Sheppard para la varianza se aplica en distribuciones continuas donde los **sesgos** o **colas** tienden en forma gradual a cero en ambos extremos.

EJEMPLO

Ejemplo 1 La siguiente tabla de distribución de frecuencias registra la masa de 40 estudiantes de una Universdad en Toluca, Estado de México en unidades de libras. Determina:

a) La varianza y la desviación típica.
b) La desviación típica introduciendo la corrección Sheppard.

Intervalos (libras)	Frecuencia (f) (estudiantes)
116 - 124	3
125 - 133	5
134 - 142	9
143 - 151	12
152 - 160	6
161 - 169	3
170 - 178	2
	$N = 40$

Solución

a) Primero, se determina la media aritmética, apoyándonos con la siguiente información.

Intervalos (libras)	Marca de clase (X)	Frecuencia (f) (estudiantes)	fX
116 - 124	120	3	360
125 - 133	129	5	645
134 - 142	138	9	1 242
143 - 151	147	12	1 764
152 - 160	156	6	936
161 - 169	165	3	495
170 - 178	174	2	348
		$N = 40$	$\sum fX = 5\,790$

Al sustituir en la ecuación de la media aritmética, resulta $\overline{X} = \frac{\sum f X}{N} = \frac{5\,790}{40} = 144.75$.

Para determinar la varianza y la desviación típica, se requiere la siguiente tabla de apoyo.

Marca de clase (X)	Frecuencia (f) (estudiantes)	$(X - \overline{X})$	$(X - \overline{X})^2$	$f(X - \overline{X})^2$
120	3	−24.75	612.56	1 837.687
129	5	−15.75	248.06	1 240.312
138	9	−6.75	45.56	410.062
147	12	2.25	5.06	60.750
156	6	11.25	126.56	759.375
165	3	20.25	410.06	1 230.187
174	2	29.25	855.56	1 711.125
	$N = 40$			$\sum f(X - X)^2 = 7\,249.5$

Después de sustituir en la ecuación de la varianza, tenemos $S^2 = \frac{\sum f(X - \overline{X})^2}{N} = \frac{7\,249.5}{40} = 181.2375$

La varianza de la tabla de distribución es 181.2375.

Al sustituir en la ecuación de la desviación típica, tenemos:

$$S = \sqrt{\frac{\sum f(X - \overline{X})^2}{N}} = \sqrt{\frac{7\,249.5}{40}} = \sqrt{181.2375} = 13.4624$$

La desviación típica o estándar de la tabla de distribución es 13.4624.

b) El tamaño de los intervalos de clase es $C = 9$, y sustituyendo en la ecuación de la varianza corregida, tenemos:

$$\text{Varianza corregida} = S^2 - \left(\frac{C^2}{12}\right) = 181.2375 - \left(\frac{9^2}{12}\right) = 181.2375 - 6.75 = 174.4875$$

La desviación típica corregida se obtiene por:

$$\text{Desviación típica corregida} = \sqrt{\text{varianza corregida}} = \sqrt{174.4875} = 13.2093$$

La desviación típica corregida por la introducción de la corrección Sheppard es 13.2093.

Se observa la diferencia en los resultados de la desviación típica calculada por la ecuación respectiva (13.4624) y la calculada por la introducción Sheppard (13.2093).

EJERCICIO 14

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

I. Contesta las siguientes preguntas.

1. ¿Qué es la varianza?
2. ¿Qué es la distribución de varianza de población y la varianza de la muestra?
3. ¿Qué es la desviación típica?
4. ¿Cuál es la distribución entre desviación típica de la población y la desviación típica de la muestra?
5. ¿Cuáles son las propiedades de la desviación típica o estándar?
6. ¿Para qué se usa la comprobación Charlier?
7. ¿Qué es la varianza corregida?
8. ¿En qué consiste la corrección Sheppard?
9. ¿Cuál es la aplicación de la corrección Sheppard?

II. Resuelve los siguientes problemas y en plenaria discute tus resultados.

1. Encuentra la varianza y la desviación típica para los siguientes conjuntos numéricos.
   a) 3.95, 4.09, 3.91 3.98, 4.03, 3.89, 4.05, 3.90, 4.13, 4.01, 3.85, 3.97.
   b) 85, 92, 73, 69, 89, 78, 69, 96, 90.
   c) 11, 5, 6, 2, 14, 9, 17, 4.
   d) 12, 6,11, 11, 12, 11, 12, 19, 13, 5.
2. Determina la varianza y la desviación típica para las tablas de distribución de frecuencias indicadas en los problemas 4, 6 y 11.
3. Determina la varianza y la desviación típica para los conjuntos numéricos indicados en el problema 1, aplicando el método corto.
4. La siguiente tabla de distribución de frecuencias registra la duración en horas de 325 lámparas incandescentes de 60 watts sometidas a pruebas críticas de control de calidad. Determina la varianza y la desviación típica por el:
   a) Método largo. *b*) Método corto. *c*) Método clave.

Intervalos (duración)	Frecuencia (lámparas)
820 - 919	22
920 - 1 019	36
1 020 - 1 119	58
1 120 - 1 219	83
1 220 - 1 319	62
1 320 - 1 419	47
1 420 - 1 519	17
	$N = 325$

5. Para los siguientes conjuntos de elementos numéricos, verifica la primera propiedad de la desviación típica o estándar.
   a) 3, 4, 6, 8, 8, 8, 1 l, 13, 15, 17.
   b) 10, 12, 13, 14, 14, 15, 16, 17, 19, 22.
   c) 12, 6, 8, 13, 9, 17, 17.
   d) 125, 116, 119, 121, 121, 124, 129.

6. La siguiente tabla de distribución de frecuencias registra los salarios semanales en miles de pesos de 80 jefes de familias. Verifica la segunda propiedad de la desviación típica y construye la gráfica representativa.

Intervalos (salarios)	Frecuencia (jefes) (f)
96 - 125	4
126 - 155	9
156 - 185	17
186 - 215	25
216 - 245	18
246 - 275	7
	$N = 80$

7. Verifica la segunda propiedad de la desviación típica y construye la gráfica representativa para las tablas de distribución de frecuencias indicadas en los problemas 2, 4, 5, 6, 10 y 16.
8. Dados los siguientes conjuntos de elementos numéricos (3, 5, 7, 9, 11, 13) y (2, 4, 6, 8, 10, 12, 14) determina:
   a) La media aritmética de cada conjunto.
   b) La varianza de cada conjunto.
   c) La media aritmética de los dos conjuntos combinados.
   d) La varianza de dos conjuntos combinados.
9. Mediante la comprobación Charlier, verifica los cálculos de la media aritmética y la desviación típica para las tablas de distribución de frecuencias indicadas en los problemas 2, 4, 5, 6, 10, 16 y 18.
10. Determina la desviación típica introduciendo la corrección Sheppard para las tablas de distribución de frecuencias indicadas en los problemas 2, 4, 5, 6, 10, 16 y 18.
11. Dada la siguiente tabla de distribución de frecuencias, calcula:

Intervalos de clase	Frecuencia (f)
47.5 - 52.4	2
52.5 - 57.4	4
57.5 - 62.4	10
62.5 - 67.4	15
67.5 - 72.4	30
72.5 - 77.4	20
77.5 - 82.4	9
82.5 - 87.4	6
87.5 - 92.4	3
92.5 - 97.4	1
	$N = 100$

   a) La varianza y la desviación típica o estándar.
   b) Verifica el cálculo de la desviación típica o estándar mediante la comprobación Charlier.
   c) La desviación típica o estándar introduciendo la corrección Sheppard.

Verifica tus resultados en la sección de respuestas.

Relaciones empíricas entre las medidas de dispersión

Introducción

En distribuciones normales se tiene que la desviación media y el rango semiintercuartílico presentan, respectivamente, valores iguales a $0.7979 \approx \frac{4}{5}$ y $0.6745 \approx \frac{2}{3}$ veces la desviación típica.

Como una consecuencia de lo anterior y aplicado en distribuciones que son moderadamente sesgadas o asimétricas, tenemos las siguientes ecuaciones matemáticas que relacionan empíricamente entre sí a dichas medidas de dispersión, es decir:

$$\text{Desviación media}\,(DM) = \left(\frac{4}{5}\right) \text{Desviación típica}\,(S)$$

$$DM = \left(\frac{4}{5}\right) S$$

$$\text{Rango semi-intercuartílico}\,(Q_{SI}) = \left(\frac{2}{3}\right) \text{Desviación típica}\,(S)$$

$$Q_{SI} = \left(\frac{2}{3}\right) S$$

EJEMPLO

Ejemplo 1 La siguiente tabla de distribución de frecuencias registra la fabricación de camisas de un taller de costura durante un mes. Discute la validez de las ecuaciones empíricas que relacionan entre sí a las medidas de dispersión.

Intervalos (camisas)	Frecuencia (días) (f)
60 - 65	1
66 - 71	3
72 - 77	7
78 - 83	11
84 - 89	6
90 - 95	2
	$N = 30$

Solución

Primero, se calcula la media aritmética, para ello se utiliza la siguiente tabla de distribución:

Intervalos (camisas)	Marca de clase (X)	Frecuencia (días) (f)	fX
60 - 65	62.5	1	62.5
66 - 71	68.5	3	205.5
72 - 77	74.5	7	521.5
78 - 83	80.5	11	885.5
84 - 89	86.5	6	519.0
90 - 95	92.5	2	185.0
		$N = 30$	$\sum fX = 2\,379$

Al sustituir en la ecuación de la media aritmética, resulta $\overline{X} = \frac{\sum fX}{N} = \frac{2\,379}{30} = 79.3$.

Para determinar la desviación media, se requiere complementar la tabla anterior con las siguientes columnas de apoyo:

$\lvert X-\overline{X}\rvert$	$f\lvert X-\overline{X}\rvert$
16.8	16.8
10.8	32.4
4.8	33.6
1.2	13.2
7.2	43.2
13.2	26.4
	$\sum f\lvert X-\overline{X}\rvert = 165.6$

$$MD=\frac{\sum f\lvert X-\overline{X}\rvert}{N}=\frac{165.6}{30}=5.52$$

La desviación media es 5.52.

Para determinar el rango semiintercuartílico, se requiere conocer los valores de los cuartiles Q_1 y Q_3. Para ello tenemos de la tabla de distribución los siguientes datos:

Datos

$$\frac{KN}{4}=\frac{1(30)}{4}=7.5$$
$$L_1=\frac{71+72}{2}=71.5$$
$$\left(\sum f\right)_1=1+3=4$$
$$f_K=7$$
$$C=6$$

Fórmula

$$Q_1=L_1+\left[\frac{\frac{KN}{4}-\left(\sum f\right)_1}{f_K}\right]C$$

Sustitución

$$Q_1=71.5+\left[\frac{7.5-4}{7}\right](6)$$
$$Q_1=71.5+3=74.5$$

El valor del primer cuartil es 74.5.

Datos

$$\frac{KN}{4}=\frac{3(30)}{4}=22.5$$
$$L_1=\frac{83+84}{2}=83.5$$
$$\left(\sum f\right)_1=1+3+7+11=22$$
$$f_K=6$$
$$C=6$$

Fórmula

$$Q_3=L_1+\left[\frac{\frac{KN}{4}-\left(\sum f\right)_1}{f_K}\right]C$$

Sustitución

$$Q_3=83.5+\left[\frac{22.5-22}{6}\right](6)$$
$$Q_3=83.5+0.5=84$$

El valor del tercer cuartil es 84.

Al aplicar la ecuación matemática del rango semiintercuartílico, tenemos:

$$Q_{SI.}=\frac{Q_3-Q_1}{2}=\frac{84-74.5}{2}=\frac{9.5}{2}=4.75$$

El rango semiintercuartílico es 4.75.

Para determinar la desviación típica, se requiere de la siguiente tabla de apoyo que se complementa con las tablas anteriores.

Marca de clase (X)	Frecuencia (días) (f)	$(X-\overline{X})$	$(X-\overline{X})^2$	$f(X-\overline{X})^2$
62.5	1	−16.8	282.24	282.24
68.5	3	−10.8	116.64	349.92
74.5	7	−4.8	23.04	161.28
80.5	11	1.2	1.44	15.84
86.5	6	7.2	51.84	314.04
92.5	2	13.2	174.24	346.48
	$N=30$			$\sum f(X-\overline{X})^2 = 1\,468.8$

Al sustituir en la ecuación de la desviación típica, resulta:

$$S=\sqrt{\frac{\sum f(X-\overline{X})^2}{N}}=\sqrt{\frac{1\,468.8}{30}}=\sqrt{48.96}=6.9971$$

La desviación típica es 6.9971.

Para discutir la validez de las ecuaciones empíricas, tenemos:

$$DM=\left(\frac{4}{5}\right)S \text{ o } \frac{DM}{S}=\frac{4}{5}, \text{ al sustituir los datos resulta:}$$

$$\frac{DM}{S}=\frac{5.52}{6.9971}=0.7888, \text{ lo cual se aproxima al valor de } \frac{4}{5}$$

$$\text{Si } Q_{SI}=\left(\frac{2}{3}\right)S \text{ o } \frac{Q_{SI}}{S}=\frac{2}{3}, \text{ al sustituir los datos resulta:}$$

$$\frac{Q_{SI}}{S}=\frac{4.75}{6.9971}=0.6788, \text{ lo cual se aproxima al valor de } \frac{2}{3}$$

Lo anterior permite comprobar la validez de las ecuaciones empíricas.

Es necesario hacer notar que en la desviación típica no debe introducirse la corrección Sheppard, ya que no se ha hecho ninguna corrección en la desviación media ni en el rango semiintercuartílico.

Dispersión absoluta y relativa

Todas las medidas de dispersión estudiadas anteriormente se denominan **medidas de dispersión absoluta** ya que presentan una característica en común, es decir, se expresan en las mismas unidades con las que se miden sus elementos variables.

Cuando las unidades de los conjuntos son diferentes, por ejemplo, si una dispersión o variación de 100 g en la medida de una pesada de 10 000 kg tiene un efecto totalmente diferente al que resulta de tener la misma variación de 100 g en una pesada de 50 kg; otro ejemplo, es cuando se comparan un conjunto de elementos que se expresan en dólares y otro conjunto cuyos elementos se expresan en pesos.

Para realizar comparaciones entre conjuntos de elementos se utiliza la **dispersión relativa** que se define como el cociente entre la dispersión absoluta y la medida de centralización o promedio, matemáticamente se expresa como:

$$\text{Dispersión relativa} = \frac{\text{Dispersión absoluta}}{\text{Medida de centralización}}$$

Coeficiente de variación

Si la dispersión absoluta se representa por la desviación típica o estándar S y la medida de centralización o promedio se representa por la media aritmética $\overline{X}$, entonces la dispersión relativa se denomina **coeficiente de variación** o **coeficiente de dispersión** y su ecuación matemática queda expresada por:

$$\text{Coeficiente de variación } (V) = \frac{S}{\overline{X}}$$

Las unidades resultantes del coeficiente de variación se expresan en porcentaje, ya que son unidades independientes de las medidas que intervienen en la ecuación, es decir, son **adimensionales**. Una desventaja del coeficiente de variación se presenta cuando el valor de la media aritmética ($\overline{X}$) está próximo a cero.

EJEMPLOS

Ejemplos

1 Un fabricante de llantas para automóvil tiene dos tipos de llantas, las convencionales (A) y las radiales (B). Las llantas tienen sus duraciones medias respectivas de $X_A = 35\,000$ km y $X_B = 50\,000$ km, las desviaciones típicas de $S_A = 6\,650$ km y $S_B = 8\,500$ km. ¿Cuál tipo de llantas presenta la mayor dispersión absoluta y cuál la mayor dispersión relativa?

Solución

La dispersión absoluta del tipo de llanta convencional ($S_A = 6\,650$ km) y la dispersión absoluta del tipo de llanta radial ($S_B = 8\,500$ km).

Al compararse se observa que el tipo (B) de llanta presenta la mayor dispersión absoluta.

Al sustituir los datos en la ecuación del coeficiente de variación, tenemos que para el tipo de llanta convencional (A).

$$V_A = \frac{S_A}{\overline{X}_A} = \frac{6\ 650}{35\ 000} = 19\%$$

y para el tipo de llantas radial (B), se tiene: $V_B = \dfrac{S_B}{\overline{X}_B} = \dfrac{8\ 500}{50\ 000} = 17\%$

Al comparar los coeficientes de variación se observa que el tipo (A) de llanta presenta la mayor dispersión relativa.

2 La siguiente tabla de distribución de frecuencia registra los salarios semanales en miles de pesos de 80 jefes de familia. Determina el coeficiente de variación, a partir de la desviación típica corregida y no corregida.

Intervalos (salarios)	Frecuencia de clase (f)
96 - 125	4
126 - 155	9
156 - 185	17
186 - 215	25
216 - 245	18
246 - 275	7
	$N = 80$

Solución

Primero, se determina la media aritmética, apoyándonos con la siguiente información.

Intervalos (salarios)	Marca de clase (*X*)	Frecuencia de clase (*f*)	*fX*
96 - 125	110.5	4	442
126 - 155	140.5	9	1 264.5
156 - 185	170.5	17	2 898.5
186 - 215	200.5	25	5 012.5
216 - 245	230.5	18	4 149
246 - 275	260.5	7	1 823.5
		$N = 80$	$\sum fX = 15\ 590$

Al sustituir en la ecuación de la media aritmética, resulta $\overline{X} = \frac{\sum fX}{N} = \frac{15\ 590}{80} = 194.875$

Para determinar la varianza y la desviación típica, se requiere la siguiente tabla de apoyo.

Marca de clase (*X*)	Frecuencia de clase (*X*)	$(X - \overline{X})$	$(X - \overline{X})^2$	$f(X - \overline{X})^2$
110.5	4	−84.375	7 119.1406	28 476.562500
140.5	9	−54.375	2 956.6406	26 609.765630
170.5	17	−24.375	594.1406	10 100.390630
200.5	25	5.625	31.6406	791.015625
230.5	18	35.625	1 269.1406	22 844.531250
260.5	7	65.625	4 306.6406	30 146.484380
	$N = 80$			$\sum f(X - \overline{X})^2 = 118\ 968.75$

Al sustituir en la ecuación de la varianza, tenemos:

$$S^2 = \frac{\sum f\left(X - \overline{X}\right)^2}{N} = \frac{118\ 968.75}{80} = 1\ 487.109$$

La varianza no corregida de la distribución es 1 487.109.

Si sustituimos en la ecuación de la desviación típica, tenemos:

$$S = \sqrt{\frac{\sum f\left(X - \overline{X}\right)^2}{N}} = \sqrt{\frac{118\ 968.75}{80}} = \sqrt{1\ 487.109375} = 38.563$$

La desviación típica no corregida de la distribución es 38.563.

Como el tamaño de los intervalos de clase es $C = 30$, al sustituir en la ecuación de la varianza corregida, resulta:

$$\text{Varianza corregida} = S^2 - \left(\frac{C^2}{12}\right) = 1\ 487.109 - \left[\frac{(30)^2}{12}\right] = 1\ 487.109 - 75 = 1\ 412.109$$

Varianza corregida de la distribución es 1 412.109.

La desviación típica corregida, se obtiene por:

$$\text{Desviación típica corregida} = \sqrt{\text{Varianza corregida}} = \sqrt{1\ 412.109} = 37.578$$

La desviación típica corregida de la distribución es 37.578.

Al sustituir en la ecuación del coeficiente de dispersión o variación, resulta:

$$V = \frac{S(\text{no corregida})}{\overline{X}} = \frac{38.563}{194.875} = 19.78\% \approx 20\%$$

El coeficiente de variación no corregido de la distribución es 20%.

Para los datos corregidos, tenemos:

$$V = \frac{S(\text{corregida})}{\overline{X}} = \frac{37.578}{194.875} = 19.28\% \approx 19\%$$

El coeficiente de variación corregido de la distribución es 19%.

3 Deduce una ecuación para una medida de dispersión relativa que pueda aplicarse en un conjunto de datos en cuyos cuartiles fueran conocidos. Usa la fórmula resultante para la siguiente tabla de distribución de frecuencias.

Intervalos de clase	Frecuencia de clase (f)
127 - 135	6
136 - 144	10
145 - 153	13
154 - 162	9
163 - 171	5
	$N = 43$

Solución

Si los cuartiles Q_1 y Q_3 son conocidos para un conjunto de datos, la ecuación $\frac{Q_1 + Q_3}{2}$ representa una medida de centralización o promedio, mientras que la expresión

$$Q_{SI} = \frac{Q_3 - Q_1}{2} = \frac{Q}{2}$$

representa el rango semiintercuartílico, que es una medida de dispersión.

Por lo anterior, es posible establecer una ecuación que defina a una medida de dispersión relativa que se denomina **coeficiente de variación cuartílica** o **coeficiente dispersión relativa cuartílica**, es decir:

$$V_Q = \frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_1 + Q_3}{2}} = \frac{Q_3 - Q_1}{Q_1 + Q_3}$$

Para aplicar la ecuación del coeficiente de variación cuartílica, primero se determina el cuartil (Q_1), por lo que de la tabla de distribución dada, tenemos:

Datos

$$\frac{KN}{4} = \frac{1(43)}{4} = 10.75$$

$$L_1 = \frac{135 + 136}{2} = 135.5$$

$$(\Sigma f)_1 = 6$$

$$f_K = 10$$

$$C = 9$$

Fórmula

$$Q_1 = L_1 + \left[\frac{\frac{KN}{4} - (\Sigma f)_1}{f_K}\right] C$$

Sustitución

$$Q_1 = 135.5 + \left[\frac{10.75 - 6}{10}\right](9)$$

$$Q_1 = 135.5 + 4.275 = 139.775$$

Para determinar el tercer cuartil (Q_3), tenemos de la tabla:

Datos

$$\frac{KN}{4}=\frac{3(43)}{4}=32.25$$

$$L_1=\frac{153+154}{2}=153.5$$

$$\left(\sum f\right)_1=6+10+13=29$$

$$f_K=9$$

$$C=9$$

Fórmula

$$Q_3=L_1+\left[\frac{\frac{KN}{4}-\left(\sum f\right)_1}{f_K}\right]C$$

Sustitución

$$Q_3=153.5+\left[\frac{32.25-29}{9}\right](9)$$

$$Q_3=153.5+3.25=156.75$$

Al sustituir en la ecuación del coeficiente de variación cuartílica, resulta:

$$V_Q=\frac{Q_3-Q_1}{Q_1+Q_3}=\frac{156.75-139.775}{139.775+156.75}=\frac{16.975}{296.525}=5.72\%$$

El coeficiente de variación cuartílica de la distribución es 5.72%.

Variable normalizada

Se define como el cociente entre la desviación de una **medición** respecto a la media aritmética y la desviación estándar; matemáticamente se expresa por la ecuación:

$$Z=\frac{X-\overline{X}}{S}$$

La variable **Z** que mide la desviación de la media aritmética en unidades de desviación típica o estándar se denomina **variable normalizada** o **calificación estándar** y sus medidas son **adimensionales**.

Referencias tipificadas

Cuando las desviaciones de la media aritmética están expresadas en unidades de desviación típica o estándar, se establece que están expresadas en **unidades tipificadas** o **referencias tipificadas**. Lo anterior indica que dichas **unidades** son útiles para comparar datos individuales de distribuciones que presentan distintas unidades de medición, así como distintas medidas y desviaciones típicas o estándares.

EJEMPLOS

Ejemplos

1 Un estudiante obtuvo una puntuación en el examen final de Química de 85, con una puntuación media en la asignatura de 79 y una desviación típica de 9. En el examen final de Física, él obtuvo una puntuación de 88 con una media general de 81 y una desviación de 8.2. ¿En qué asignatura su puntuación relativa fue superior?

Solución

Al aplicar la ecuación de la variable normalizada o calificación estándar, tenemos para:

$$\text{Química: } Z=\frac{X-\overline{X}}{S}=\frac{85-79}{9}=0.67$$

$$\text{Física: } Z=\frac{X-\overline{X}}{S}=\frac{88-81}{8.2}=0.85$$

El estudiante tiene una puntuación de 0.67 desviación típica por encima de la media aritmética de la asignatura de Química; para la asignatura de Física presenta una puntuación de 0.85 de desviación típica por arriba de su media aritmética.

La puntuación relativa superior se presenta en la asignatura de Física.

2 Dado el conjunto de elementos numéricos 8, 6, 11, 13, 10, 7 y 9, conviértelos en referencias tipificadas.

Solución

Primero, se determina la media aritmética del conjunto:

$$\overline{X}=\frac{\sum X}{N}=\frac{8+6+11+13+10+7+9}{7}=\frac{64}{7}=9.143$$

Ahora se determina la desviación típica o estándar:

$$S=\sqrt{\frac{\sum(X-\overline{X})^2}{N}}=\sqrt{\frac{(8-9.143)^2+(6-9.143)^2+(11-9.143)^2+(13-9.143)^2+(10-9.143)^2+(7-9.143)^2+(9-9.143)^2}{7}}$$

$$S=\sqrt{\frac{1.306+9.878+3.448+14.876+0.734+4.592+0.020}{7}}=\sqrt{\frac{37.854}{7}}=2.231$$

Al realizar la conversión del conjunto de elementos numéricos a referencias tipificadas, tenemos:

X	8	6	11	13	10	7	9
$Z=\frac{X-\overline{X}}{S}$	−0.512	−1.409	0.832	1.729	0.384	−0.961	−0.064

3 La siguiente tabla de distribución de frecuencias registra los coeficientes intelectuales de 265 alumnos de la Facultad de Ciencias Químicas en una universidad; conviértelos en referencias tipificadas y construye una gráfica de frecuencias relativas de la variable normalizada.

Intervalos de clase	Frecuencia de clase (f)
60 - 68	0
69 - 77	3
78 - 86	11
87 - 95	27
96 - 104	44
105 - 113	75
114 - 122	58
123 - 131	31
132 - 140	16
	$N = 265$

Solución

Primero se determina la media aritmética, por lo que se requiere la siguiente información:

Intervalos de clase	Marca de clase (X)	Frecuencia de clase (f)	fX
60 - 68	64	0	0
69 - 77	73	3	219
78 - 86	82	11	902
87 - 95	91	27	2 457
96 - 104	100	44	4 400
105 - 115	109	75	8 175
114 - 122	118	58	6 844
123 - 131	127	31	3 937
132 - 140	136	16	2 176
		$N = 265$	$\sum fX = 29\ 110$

Al sustituir en la ecuación de la media aritmética, resulta: $\overline{X} = \frac{\sum fX}{N} = \frac{29\ 110}{265} = 109.85$

Para determinar la desviación típica, se requiere la siguiente tabla de apoyo:

Marca de clase (X)	Frecuencia de clase (f)	$(X - \overline{X})$	$(X - \overline{X})^2$	$f(X - \overline{X})^2$
64	0	−45.85	2 102.2225	0
73	3	−36.85	1 357.9225	4 073.7675
82	11	−27.85	775.6225	8 531.8475
91	27	−18.85	355.3225	9 593.7075
100	44	−9.85	97.0225	4 268.9900
109	75	−0.85	0.7225	54.1875
118	58	8.15	66.4225	3 852.5050
127	31	17.15	294.1225	9 117.7975
136	16	26.15	683.8225	10 941.1600
	$N = 265$			$\sum f(X - \overline{X})^2 = 50\ 433.9625$

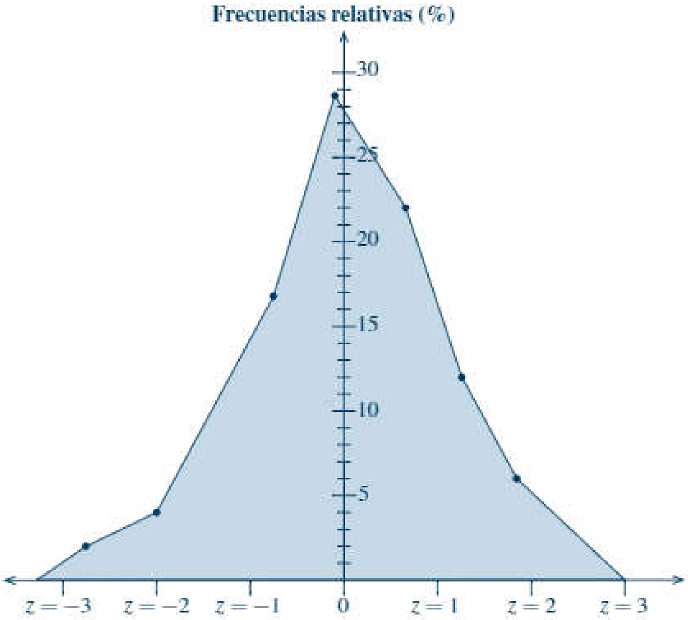
Si sustituimos en la ecuación de la desviación típica, tenemos:

$$S = \sqrt{\frac{\sum f(X - \overline{X})^2}{N}} = \sqrt{\frac{50\ 433.9625}{265}} = \sqrt{190.3168} = 13.8$$

Ahora se construye la tabla que contiene las referencias tipificadas con sus respectivas frecuencias relativas, de decir:

Marca de clase (X)	Frecuencia de clase (f)	$(X - \overline{X})$	$z = \frac{X - \overline{X}}{S}$	Frecuencias relativas (%)
64	0	−45.85	−3.32	0
73	3	−36.85	−2.67	1.13
82	11	−27.85	−2.02	4.15
91	27	−18.85	−1.37	10.19
100	44	−9.85	−0.71	16.60
109	75	−0.85	−0.06	28.30
118	58	8.15	0.60	21.89
127	31	17.15	1.24	11.70
136	16	26.15	1.90	6.04
	$N = 265$			100%

Gráficamente, se tiene:



Se observa que la distribución es moderadamente asimétrica y ligeramente sesgada a la izquierda.

EJERCICIO 15

I. Contesta las siguientes preguntas.

1. ¿Cuáles son los valores equivalentes entre la desviación media y la desviación típica, así como el rango semiintercuartílico y la desviación típica?
2. ¿Cuáles son las ecuaciones que relacionan empíricamente entre sí a dichas medidas de dispersión?
3. ¿Qué característica presentan las medidas de dispersión absoluta?
4. ¿Qué es la dispersión relativa?
5. ¿Cómo se expresa el coeficiente de variación o de dispersión?
6. ¿Qué son las unidades adimensionales?
7. ¿Cuál es la desventaja que presenta en su aplicación el coeficiente de variación?
8. ¿Qué es a la variable normalizada?

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

II. Resuelve los siguientes problemas.

1. La siguiente tabla de distribución de frecuencias registra el número de llamadas de auxilio que durante 24 horas de un día son recibidas en el Departamento de Policía y Tránsito en la ciudad de Cuernavaca, Morelos. Determina:

   a) La validez de las ecuaciones empíricas que relacionan entre sí a las medidas de dispersión.

   b) El coeficiente de variación, utilizando la desviación típica corregida y no corregida.

   c) Las referencias tipificadas y construye un gráfico de frecuencias relativas de la variable normalizada.

Intervalos (horas)	Frecuencia de clase (f)
0 - 4	21
5 - 9	14
10 - 14	9
15 - 19	11
20 - 24	17
	$N = 72$

2. La siguiente tabla de distribución de frecuencias muestra los salarios semanales (40 horas) en dólares para 120 técnicos de la división de televisores de una empresa. Determina:

   a) La validez de las ecuaciones empíricas que relacionan entre sí a las medidas de dispersión.

   b) El coeficiente de variación, utilizando la desviación típica corregida y no corregida.

   c) Las referencias tipificadas y construye un gráfico de frecuencias relativas de la variable normalizada.

Intervalos de clase (salarios en dólares)	Frecuencia (f) (número de técnicos)
310 - 319.99	7
320 - 329.99	18
330 - 339.99	25
340 - 349.99	31
350 - 359.99	22
360 - 369.99	13
370 - 379.99	4
	$N = 120$

3. En un examen final de Cálculo Integral, la puntuación media de un grupo de 187 estudiantes fue de 82 y la desviación típica fue de 10.6; en cálculo avanzado, la media final del grupo fue de 76 y la desviación típica de 8.8 ¿En qué asignatura hubo mayor dispersión absoluta y en cuál mayor dispersión relativa?

4. Una compañía fabricante de lámparas de halógeno tiene dos tipos de lámpara, X y Y. Las lámparas tienen una duración media de 1 950 horas y 2 125 horas, respectivamente, las desviaciones típicas de 236 horas y 345 horas, respectivamente. ¿Qué lámpara tiene la mayor dispersión absoluta y cuál mayor dispersión relativa?

5. Deduce una ecuación para una medida de dispersión relativa que se pueda aplicar en un conjunto de datos en los que los cuartiles sean conocidos. Usa la fórmula resultante para las siguientes tablas de distribución de frecuencias.

a)

Intervalos de clase	Frecuencia de clase (f)
40 - 49	4
50 - 59	12
60 - 69	15
70 - 79	24
80 - 89	21
90 - 99	17
100 - 109	8
110 - 119	3
	$N = 104$

b)

Intervalos de clase	Frecuencia de clase (f)
150 - 154	5
155 - 159	13
160 - 164	21
165 - 169	30
170 - 174	42
175 - 179	33
180 - 184	25
185 - 189	18
190 - 194	9
	$N = 196$

6. Un alumno obtuvo una puntuación en el examen final de inglés de 74, con una puntuación media de 76 y una desviación típica de 7. En el examen final de francés con una media general de 80 y una desviación típica de 13, obtuvo una puntuación de 88. ¿En qué asignatura fue su puntuación relativa superior?

7. La ficha antropométrica de un grupo de delincuentes registra para las estaturas una media aritmética de 164.2 cm y una desviación típica de 9.8 cm; para los pesos una media aritmética de 59.75 kg y una desviación típica de 10.3 kg. Para un delincuente cuya estatura es 169 cm y que pesa 65 kg, determina:

   a) La variable normalizada para cada medida.

   b) ¿Cuál de las dos medidas tiene mayor desviación?

8. Dados los siguientes conjuntos de elementos numéricos, conviértelos en referencias tipificadas.

   a) 7, 2, 4, 6, 8, 11.

   b) 6, 5, 11, 9, 16, 13.

   c) 59, 65, 71, 54, 62, 68, 73.

   d) 3.25, 3.64, 3.91, 4.08, 4.17.

9. Para las tablas de distribución de frecuencias del problema 5, determina:

   a) La validez de las ecuaciones empíricas que relacionan entre sí a las medidas de dispersión.

   b) El coeficiente de variación, utilizando la desviación típica corregida y no corregida.

   c) Las referencias tipificadas y construye un gráfico de frecuencias relativas de la variable normalizada.

Verifica tus resultados en la sección de respuestas.

TEMA 4 Medidas de forma

Propósito del tema

Que el estudiante:

- Haga uso del sesgo y coeficiente de Fisher como herramienta para determinar la simetría de una distribución.
- Comprenda las ventajas del uso de las medidas de forma para la solución de problemas que involucren datos estadísticos.
- Interprete los resultados que obtiene al identificar los elementos de las medidas de forma en la solución de problemas.

Competencias disciplinares

3. Explica e interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos o situaciones reales.
4. Argumenta la solución obtenida de un problema con métodos numéricos, gráficos, analíticos o variacionales, mediante lenguaje verbal, matemático y el uso de las tecnologías de la información y la comunicación.
5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.
8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

Contenidos que aborda el tema

Contenidos conceptuales

- Concepto de sesgo.
- Concepto del coeficiente de Fisher.
- Concepto de curtosis.

Contenidos procedimentales

- Usará procedimientos de manera reflexiva, comprendiendo cómo cada uno de sus pasos contribuye al alcance de un objetivo.
- Resolverá problemas utilizando las mediadas de forma como herramienta para determinar la simetría de una distribución.
- Usará diversas estrategias de resolución de problemas y a los aplicará a situaciones reales.

Contenidos actitudinales

- Expresará sus ideas mediante los elementos básicos de las medidas de forma.
- Trabajará en equipo y respetará a sus compañeros al resolver problemas.
- Aprenderá a valorar el trabajo de sus compañeros al resolver problemas.
- Estructurará ideas y argumentos de manera clara, coherente y sintética.

Medidas de forma

Introducción

Las medidas de forma son indicadores que ofrecen información acerca de la manera en que los datos de encuentran contenidos dentro de una distribución; se clasifican en dos grupos: las **medidas de sesgo** o **asimetría** y las **medidas de apuntamiento** o **curtosis**.

Sesgo

El sesgo es una medida que indica qué tan simétrica o asimétrica es una distribución. Se dice que una distribución es **simétrica** si los datos que se encuentran contenidos en la misma se ubican repartidos de forma semejante en ambos lados de la media **f**. En cambio, una distribución **asimétrica** es aquella en la que los datos con las frecuencias más bajas se ubican a la derecha o a la izquierda de la media; por tanto, existen dos tipos de asimetría: la asimetría a la izquierda y la asimetría a la derecha.

Asimetría a la izquierda. También llamada **asimetría negativa**, es aquella en la que los datos con las frecuencias más bajas se ubican a la izquierda de la media, mientras que los datos con mayores frecuencias tienden a aglomerar se a la derecha.

Asimetría a la derecha. También llamada **asimetría positiva**, es aquella en la que los datos con las menores frecuencias están ubicados a la derecha de la media y los datos con mayores frecuenciasse encuentran localizados a la izquierda.

EJEMPLOS

Ejemplos

1 A continuación se enlistan las calificaciones de que obtuvieron 45 alumnos de primer semestre de licenciatura en un examen de cálculo diferencial.

6.4, 6.4, 6.3, 6.4, 6.4, 6.4, 6.4, 6.5, 6.5, 6.6, 6.6, 6.7, 6.7, 6.7, 6.8, 6.8, 6.9, 6.9, 7.0, 7.1, 7.2, 7.2, 7.2, 7.2, 7.2, 7.2, 7.2, 7.3, 7.3, 7.4, 7.4, 7.5, 7.5, 7.5, 7.6, 7.6, 7.6, 7.7, 7.7, 8.0, 8.1, 8.1, 8.2, 8.4, 8.5.

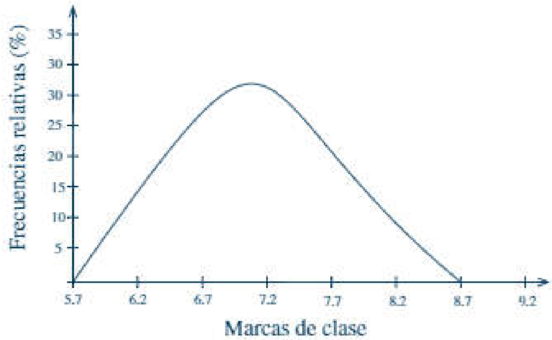
Con la información anterior.

a) Construye un polígono de frecuencias.
b) Calcula la media aritmética.
c) Determina cualitativamente la simetría o simetría de la distribución.
d) Calcula la moda.
e) Calcula la mediana.

Solución

a) Para construir el polígono de frecuencias, primero se elabora una tabla de distribución de frecuencias.

Intervalo de clase	Marca de clase	Frecuencia de clase	Frecuencia relativa de clase (FRC %)
6.0 - 6.4	6.2	7	16%
6.5 - 6.9	6.7	11	24%
7.0 - 7.4	7.2	13	29%
7.5 - 7.9	7.7	8	18%
8.0 - 8.4	8.2	5	11%
8.5 - 8.9	8.7	1	2%



b) Al sustituir los datos proporcionados en la ecuación de media aritmética con frecuencias resulta:

$$\overline{X} = \frac{\sum f X}{N} = \frac{(7)(6.2)+(11)(6.7)+(13)(7.2)+(8)(7.7)+(5)(8.2)+(1)(8.7)}{9+11+13+9+5+1} = \frac{43.4+73.7+93.6+61.6+41+8.7}{45}$$

$$\overline{X} = \frac{322}{45} = 7.15 \approx 7.2$$

La media aritmética es 7.2.

c) A partir del polígono de frecuencias se puede observar que las frecuencias están ubicadas de forma semejante en ambos lados de la cúspide de la distribución; por esta razón es una distribución simétrica.

La distribución es simétrica.

d) Como los datos se encuentran ordenados, es posible ver que el dato con mayor frecuencia es 7.2.

La moda es 7.2.

e) Puesto que hay 45 calificaciones ordenadas de forma ascendente, existe un valor numérico ubicado en la posición 23. Ese dato corresponde a la calificación de 7.2.

La mediana es 7.2.

Del ejercicio anterior se puede concluir que una distribución de frecuencias es simétrica si y sólo si:

Media aritmética = Mediana = Moda

2 En una evaluación de la carrera de ingeniería en una universidad, se encuestó a 550 alumnos; se les preguntó cuántas horas a la semana en promedio dedicaban al estudio. Se obtuvo la siguiente tabla de frecuencias.

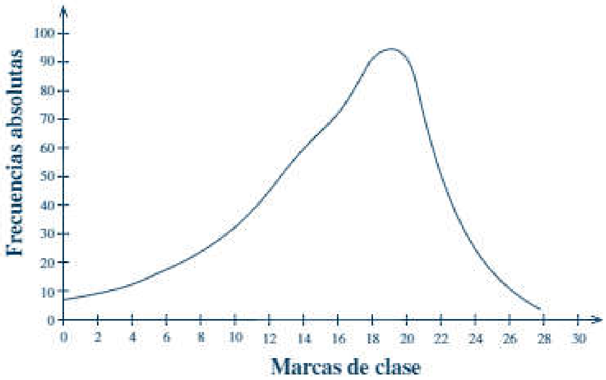
Número de horas de estudio a la semana	Frecuencias absolutas
0	7
2	9
4	12
6	17
8	23
10	32
12	44
14	59
16	71
18	90
20	92
22	54
24	25
26	12
28	3

Con la información anterior:

a) Construye un polígono de frecuencias.
b) Calcula la media aritmética.
c) Determina cualitativamente la presencia de una asimetría o un sesgo.
d) Calcula la moda.
e) Calcula la mediana.

Solución

a) Es posible construir un polígono a partir de las frecuencias absolutas, para tener así lo siguiente:



b) Al utilizar la fórmula de media aritmética con frecuencias resulta:

$$\overline{X}=\frac{\sum fX}{N}=\frac{(7)(0)+(9)(2)+(12)(4)+(17)(6)+(23)(8)+(32)(10)+(44)(12)+(59)(14)}{7+9+12+17+23+32+44+59}$$

$$\frac{+(71)(16)+(90)(18)+(92)(20)+(54)(22)+(25)(24)+(12)(26)+(3)(28)}{+71+90+92+54+25+12+3}$$

$$=\frac{0+18+48+102+184+320+528+826+1136+1620+1840+1188+600+312+84}{550}$$

$$=\frac{8806}{550}=16.01\approx 16$$

La media aritmética es 16.

c) Es posible observar que los valores con menores frecuencias se encuentran generalmente a la izquierda de la distribución, justo por debajo del valor de la media aritmética; entonces, la distribución es asimétrica a la izquierda o tiene un sesgo negativo.

La distribución tiene un sesgo negativo.

d) A partir de la información de la tabla, se determina que el dato con mayor frecuencia absoluta es 20.

La moda es 20.

e) Para encontrar la mediana, primero es necesario determinar la clase mediana. Para ello se calcula:

$$\frac{N}{2}=\frac{550}{2}=275$$

La clase mediana es la que tiene una marca de clase de 18. Como el número total de datos es par, la mediana será el promedio entre el dato que ocupa la posición número 75 y el dato que ocupa la posición número 76. Como ambos datos son iguales a 18, se tiene que:

$$\text{Mediana}=\frac{18+18}{2}=18$$

La mediana es 18.

La distribución del ejemplo anterior es asimétrica a la izquierda; además se verificó que:

$$\text{Media}=16<\text{Mediana}=18<\text{Moda}=20$$

De lo anterior se puede concluir que una distribución de frecuencias tiene sesgo negativo o es asimétrica hacia la izquierda cuando:

$$\text{Media}<\text{Mediana}\leq\text{Moda}$$

En ocasiones, en una distribución con sesgo negativo, la mediana puede ser igual que la moda. De modo semejante, una distribución de frecuencias tiene sesgo positivo si:

$$\text{Moda}<\text{Mediana}\leq\text{Media}$$

Coeficiente de Fisher

El coeficiente de Fisher es una herramienta que sirve para determinar la simetría o la asimetría de una distribución. Su fórmula es:

$$S_f = \frac{\left(\frac{1}{N}\right)\sum_{i=1}^{k}(X_i - \overline{X})^3}{S^3}$$

donde:

N = Número total de datos.
$\overline{X}$ = La media aritmética.
S = Desviación estándar.

El coeficiente de Fisher cumple con las siguientes características.

1. Si $S_f = 0$ significa que la distribución es simétrica.
2. Si $S_f > 0$ significa que la distribución tiene sesgo positivo.
3. Si $S_f < 0$ significa que la distribución tiene sesgo negativo.

EJEMPLO 1

Un censo realizado en la planta docente de una facultad arrojó los siguientes datos sobre las edades de sus académicos, que suman un total de 720.

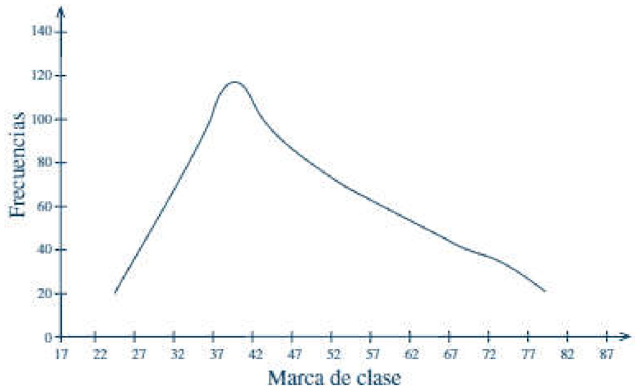
Edad	Marca de clase	Frecuencia
20 - 24	22	21
25 - 29	27	53
30 - 34	32	84
35 - 39	37	117
40 - 44	42	95
45 - 49	47	80
50 - 54	52	68
55 - 59	57	59
60 - 64	62	49
65 - 69	67	40
70 - 74	72	33
75 - 79	77	21

Con la información anterior:

a) Construye el polígono de frecuencias respectivo.
b) Calcula la media.
c) Calcula la moda.
d) Calcula la mediana.
e) Determina la asimetría de la distribución, evaluando medidas de tendencia central.
f) Determina la asimetría de la distribución mediante el coeficiente de Fisher.

Solución

a) A partir de los datos de la tabla de distribución de frecuencias, se realiza la gráfica del polígono de frecuencias.



b) Al utilizar la fórmula de media aritmética con frecuencias se tiene que:

$$\overline{X}=\frac{\sum fX}{N}=\frac{(21)(22)+(53)(27)+(84)(32)+(117)(37)+(95)(42)+(80)(47)+(}{21+53+84+117+95+80+(}$$

$$\frac{68)(52)+(59)(57)+(49)(62)+(40)(67)+(33)(72)+(21)(77)}{58+59+49+40+33+21}$$

$$=\frac{462+1\,431+2\,688+4\,329+3\,990+3\,760+3\,536+3\,363+3\,038+2\,680+2\,376+1}{720}$$

$$\frac{617}{\quad}=\frac{33\,270}{720}46$$

La media aritmética es 46.208.

c) La moda es el dato que posee la mayor frecuencia. La marca de clase con mayor frecuencia es 37 .

La moda es 37.

d) Para calcular la mediana es necesario encontrar la clase mediana. Entonces:

$$\frac{N}{2}=\frac{720}{2}=360$$

La mediana está en la clase con marca de clase 42. Entonces, con la fórmula de mediana resulta:

$$\text{Mediana}=L_1+\left|\frac{\frac{N}{2}-\left(\sum f\right)_1}{f_{\text{mediana}}}\right|C=40+\left|\frac{360-275}{95}\right|5=40+4.47=44.7$$

La mediana es 44.7.

e) Al comparar las tres medidas de tendencia central calculadas previamente, se observa que:

$$\text{Moda}<\text{Mediana}\leq\text{Media}$$

La distribución tiene sesgo positivo o sesgo a la derecha.

f) Es necesario calcular la varianza y la desviación estándar para obtener el coeficiente de Fisher; para ello, se completa la tabla de distribución de frecuencias de la siguiente manera.

Intervalo	Marca de clase	Frecuencia absoluta	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})^3$	$f(x_i - \overline{x})^3$
20 - 24	22	21	–24.208	586.043	–14 187.134	–297 929.814
25 - 29	27	53	–19.208	368.960	–7 087.108	–375 616.722
30 - 34	32	84	–14.208	201.877	–2 868.332	–240 939.883
35 - 39	37	117	–9.208	84.793	–780.806	–91 354.291
40 - 44	42	95	–4.208	17.710	–74.530	–7 080.338
45 - 49	47	80	0.792	0.627	0.496	39.693
50 - 54	52	68	5.792	33.543	194.272	13 210.510
55 - 59	57	59	10.792	116.460	1 256.798	74 151.097
60 - 64	62	49	15.792	249.377	3 938.074	192 965.641
65 - 69	67	40	20.792	432.293	8 988.100	359 524.015
70 - 74	72	33	25.792	665.210	17 156.876	566 176.923
75 - 79	77	21	30.792	948.127	29 194.403	613 082.453
Suma				3 705.021		806 229.285

La varianza se calcula mediante la fórmula respectiva:

$$S^2 = \frac{\sum (X - \overline{X})^2}{N} = \frac{3705.021}{720} = 5.1458$$

La varianza es de 195.832.

La desviación estándar es:

$$S^2 = \sqrt{\frac{\sum (X - \overline{X})^2}{N}} = \sqrt{5.11458} = 2.2684$$

Ahora, con la fórmula para el coeficiente de Fisher se tiene:

$$S_f = \frac{\left(\frac{1}{N}\right)\sum_{i=1}^{k} f\left(X_i - \overline{X}\right)^3}{S^3} = \frac{\frac{1}{720}(806229.285)}{(2.2684)^3} = 95.9328$$

Como $S_f > 0$, entonces la distribución tiene sesgo positivo.

La distribución tiene sesgo positivo.

Apuntamiento o curtosis

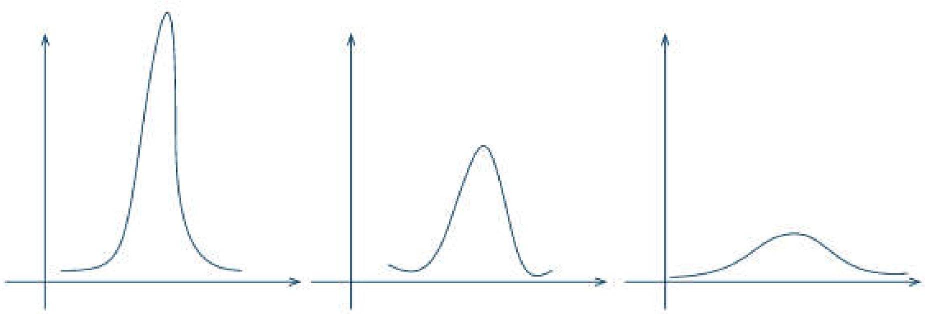
El apuntamiento o curtosis mide qué tan achatada o tan puntiaguda es una distribución. Así, la curtosis indica cuan conglomerados se encuentran los datos en la zona central de la distribución.

De acuerdo con este análisis, las distribuciones se clasifican en tres tipos:

Distribución leptocúrtica. Es una distribución que contiene una gran concentración de datos en la zona central. La forma de su polígono de frecuencias es muy puntiaguda.

Distribución mesocúrtica. Es una distribución que contiene una concentración de datos mediana en la zona central. La forma de su polígono de frecuencias es más achatada que la de una distribución leptocúrtica.

Distribución platicúrtica. Es una distribución que contiene una baja conglomeración de datos en su región central y por lo tanto, presenta un polígono de frecuencias muy achatado.



Para determinar el grado de achatamiento de una distribución se utiliza el coeficiente de apuntamiento de Fisher.

$$A_f = \frac{\left(\frac{1}{n}\right)\sum_{i=1}^{k} f(X-\overline{X})^4}{S^4}$$

Las características de este coeficiente son las siguientes.

1. Si $A_f = 3$ significa que la distribución es mesocúrtica
2. Si $A_f < 3$ significa que la distribución es platicúrtica
3. Si $A_f > 3$ significa que la distribución es leptocúrtica.

EJEMPLOS

Ejemplos

1 Determina si la distribución del ejemplo anterior es mesocúrtica, platicúrtica o leptocúrtica.

Solución

Para determinar el coeficiente de apuntamiento de Fisher se evalúa la siguiente tabla de distribución de frecuencias.

Intervalo	Marca de clase	Frecuencia absoluta	$(x_i-\bar{x})^4$	$f(x_i-\bar{x})^4$
20 - 24	22	21	343 446.868	7 212 384.229
25 - 29	27	53	136 131.532	7 214 971.191
30 - 34	32	84	40 754.216	3 423 354.161
35 - 39	37	117	7 189.921	841 220.7629
40 - 44	42	95	313.647	29 796.42223
45 - 49	47	80	0.393	31.42385753
50 - 54	52	68	1 125.160	76 510.87292
55 - 59	57	59	13 562.948	800 213.9286
60 - 64	62	49	62 188.757	3 047 249.095
65 - 69	67	40	186 877.587	7 475 103.491
70 - 74	72	33	442 504.439	14 602 646.48
75 - 79	77	21	898 944.312	18 877 830.54
Suma				63 601 312.6

El coeficiente de Fisher se calcula con la fórmula:

$$A_f = \frac{\left(\frac{1}{N}\right)\sum_{i=1}^{k} f(X-\overline{X})^4}{S^4}$$

$$A_f = \frac{\left(\frac{1}{N}\right)\sum_{i=1}^{k} f(X-\overline{X})^4}{S^4} = \frac{\left(\frac{1}{720}\right)(63\,601\,312.6)}{(2.2684)^4} = 3\,336.22$$

Por lo tanto, la distribución es leptocúrtica.

2 Determina el apuntamiento de la siguiente distribución de frecuencias.

Intervalo	Marca de clase	Frecuencia
18 - 20	19	15
20 - 22	21	33
22 - 24	23	78
24 - 26	25	36
26 - 28	37	21

La distribución de frecuencias se completa de la siguiente manera:

Intervalo	Marca de clase	Frecuencia	$(x_i-\overline{x})^2$	$(x_i-\overline{x})^4$	$f(x_i-\overline{x})^4$
18 - 20	19	15	28.212	795.90318	11938.5477
20 - 22	21	33	10.966	120.250087	3968.25286
22 - 24	23	78	1.720	2.95827634	230.745555
24 - 26	25	36	0.474	0.22474052	8.09065867
26 - 28	37	21	160.999	25920.5788	544332.156
		Suma	202.370		560477.792

Cálculo de la desviación estándar.

$$S=\sqrt{\frac{\sum(X-\overline{X})^2}{N}}=\sqrt{\frac{202.370}{183}}=1.05$$

Cálculo del coeficiente de apuntamiento de Fisher.

$$A_f=\frac{\left(\frac{1}{N}\right)\sum_{i=1}^{k}f(X-\overline{X})^4}{S^4}=\frac{\left(\frac{1}{183}\right)(560\ 477.792)}{(1.05)^4}=2\ 519.707$$

Como dicho valor es mayor que 3 se tiene una distribución leptocúrtica.

EJERCICIO 16

I. Realiza lo que se indica.

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

1. Explica con tus propias palabras qué son las medidas de forma.
2. Define sesgo.
3. Explica la diferencia que existe entre asimetría positiva y asimetría negativa.
4. Describe qué es el coeficiente de Fisher.
5. Indica cómo se clasifican las distribuciones de acuerdo con el valor que adquiere el coeficiente de Fischer.
6. Define apuntamiento o curtosis.
7. Define distribución leptocúrtica.
8. Define distribución mesocúrtica.
9. Define distribución platicúrtica.
10. Explica qué es el coeficiente de apuntamiento de Fisher.

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

II. Resuelve los siguientes problemas.

1. Para el siguiente conjunto de datos:

   a) Construye el polígono de frecuencias.

   b) Calcula la media aritmética.

   c) Calcula la moda.

   d) Calcula la mediana.

   e) Determina la simetría o la asimetría de la distribución.

24.7	24.8	25	25.4	25.5	25.8	25.9	26.5
24.7	24.8	25.1	25.5	25.5	25.8	26	26.7
24.7	24.9	25.1	25.5	25.6	25.8	26	26.8
24.7	24.9	25.2	25.5	25.6	25.8	26.3	
24.7	25	25.2	25.5	25.7	25.9	26.4	
24.7	25	25.3	25.5	25.7	25.9	26.5	

2. Una encuesta hecha a los empleados de una empresa muestra la cantidad de dinero en pesos que cada persona invierte en medicamentos cada quincena. Los resultados obtenidos se muestran en la siguiente tabla.

Dinero invertido (pesos)	Número de personas
100 - 140	120
150 - 190	170
200 - 240	220
250 - 290	270
300 - 340	320
350 - 390	370
400 - 440	420
450 - 490	470
500 - 540	520
550 - 590	570
600 - 640	620
650 - 690	670
700 - 740	720
750 - 790	770

   Para los datos anteriores encuentra.

   a) El polígono de frecuencias de la distribución.

   b) La media.

   c) La moda.

d) La mediana.

e) La simetría o la asimetría de la función.

f) El coeficiente de Fisher.

3. A partir de la siguiente distribución de frecuencias, encuentra:

a) El polígono de frecuencias

b) La media.

c) La moda.

d) La mediana.

e) El coeficiente de Fisher.

f) La simetría o la asimetría de la función.

g) El coeficiente de apuntamiento de Fisher.

h) El apuntamiento de la distribución.

Marca de clase	Frecuencia
50	25
55	34
60	55
65	82
70	124
75	76
80	45
85	26
90	11

Verifica tus resultados en la sección de respuestas.

TEMA 5 Medidas de correlación

Propósito del tema

Que el estudiante:

- Aprenda e incorpore en situaciones reales las medidas de correlación.
- Identifique y analice la relación que existe entre diversas cantidades para generar una conclusión.
- Incorpore los conocimientos adquiridos para determinar el mejor procedimiento de resolución de problemas.

Competencias disciplinares

1. Construye e interpreta modelos matemáticos mediante la aplicación de procedimientos aritméticos, algebraicos, geométricos y variacionales, para la comprensión y análisis de situaciones reales, hipotéticas o formales.
3. Explica e interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos o situaciones reales.
5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.
8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

Contenidos que aborda el tema

Contenidos conceptuales

- Concepto de coeficiente de correlación.
- Concepto de coeficiente de Pearson
- Concepto de recta de regresión.
- Error estándar de estimación.

Contenidos procedimentales

- Expresará ideas y conceptos mediante representaciones lingüísticas, matemáticas o gráficas.
- Aprenderá y aplicará como herramienta la regresión lineal.
- Resolverá problemas utilizando las medidas de correlación en situaciones reales.
- Integrará los conocimientos adquiridos en la resolución de problemas.

Contenidos actitudinales

- Expresará sus ideas mediante los elementos básicos de las medidas de correlación.
- Trabajará en equipo y respetará a sus compañeros al resolver problemas.
- Aprenderá a valorar el trabajo de sus compañeros al resolver problemas.
- Estructurará ideas y argumentos de manera clara, coherente y sintética.

Medidas de correlación

En muchas disciplinas suele ser de gran interés conocer la relación que existe entre diversas cantidades. Situaciones como investigar la relación entre el precio de un producto y el consumo del mismo, o bien, la manera en que una población de células incrementa su número respecto al tiempo son tan sólo ejemplos de la manera como se inquiere a través de un método que permite determinar, con un buen grado de certeza, si existe una relación específica entre dos magnitudes diferentes. Para analizar la posible relación entre dos variables se construyen pares de coordenadas, en donde por cada coordenada se coloca una variable. Así, al realizar la gráfica de dichas coordenadas como puntos en un plano se obtiene una dispersión de puntos, también llamada **nube de puntos**, o bien, **gráfica de dispersión**. A partir de dicha gráfica es posible inferir la posible relación que existe entre dos variables.

EJEMPLO

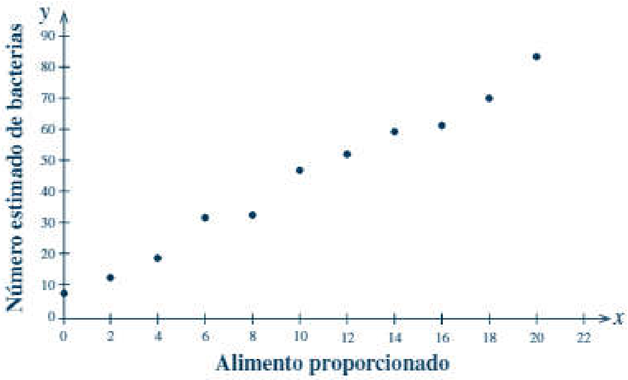
Ejemplo 1 En una universidad se analizó el crecimiento de una población de bacterias con relación a la cantidad de nutrientes que se le proporcionaba. Los estudios se llevaban a cabo en lapsos de tiempo de tres días y al final de cada periodo, se registraba el número estimado de bacterias observadas contra los microgramos de nutrientes suministrados. La tabla de datos que se obtuvo es la siguiente:

Tabla de valores del número estimado de bacterias contra la cantidad de alimento proporcionado en microgramos	
Cantidad de alimento	**Número estimado de bacterias**
0	8
2	13
4	19
6	32
8	33
10	47
12	52
14	59
16	61
18	70
20	83

Elabora la dispersión de puntos correspondiente.

Solución

Si las coordenadas x están asociadas con la cantidad de alimento y las coordenadas y, con el número estimado de bacterias, cada punto se puede trazar en un plano cartesiano para obtener una dispersión como la siguiente:



Como los puntos están *prácticamente* alineados, se puede establecer que la relación entre el número estimado de bacterias y el alimento proporcionado es lineal.

La relación que existe entre dos variables puede ser de varios tipos: lineal, cuadrática, cúbica, exponencial, etcétera. Si dos variables mantienen alguna u otra de estas relaciones, entonces, se dice que existe una correlación entre ellas.

Coeficiente de correlación

El coeficiente de correlación es una cantidad que permite conocer qué tan bien se ajusta un conjunto de pares ordenados a un comportamiento determinado (únicamente se trabajará con el comportamiento lineal). Así, se dice que dos magnitudes diferentes están correlacionadas linealmente si los puntos pertenecientes a una dispersión se pueden ubicar —con menor o mayor grado de precisión— en una línea recta.

Coeficiente de correlación de Pearson

El coeficiente de correlación de Pearson se calcula mediante la siguiente expresión:

$$r = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{\sqrt{\left(\sum (X-\overline{X})^2\right)\left(\sum (Y-\overline{Y})^2\right)}}$$

donde X es la variable a la que se le designó el eje de las abscisas, Y es la coordenada a la que le corresponde el eje de las ordenadas y, $\overline{X}$ y $\overline{Y}$ son las medias aritméticas respectivas.

Cuando $r = 0$, no existe correlación lineal alguna entre las variables que se estudian; en cambio, si $r = 1$, o bien, $r = -1$, entonces se dice que existe una correlación lineal perfecta entre las variables. Los valores ubicados entre 1 y -1 indican una correlación que no es exacta. Por lo tanto, mientras el valor de r esté más cerca de 1 o de -1, la correlación entre las variables se semejará más al comportamiento de una línea recta.

EJEMPLO 1

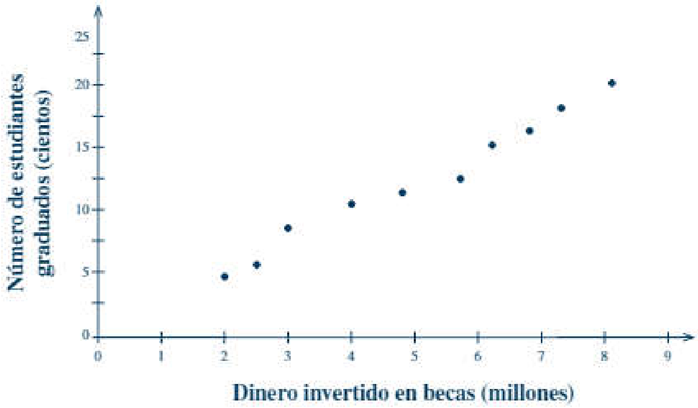
Al hacer un estudio de la población estudiantil en una universidad pública, se decidió comparar el número de estudiantes graduados contra la cantidad de dinero destinada a ofrecer becas. Se sabe que el número de becas ha incrementado año con año, pero se quiere saber si existe alguna relación entre el número de estudiantes graduados contra el presupuesto destinado a becas. La información obtenida se presenta en la siguiente tabla:

Presupuesto dedicado a becas (millones de pesos)	Número de alumnos graduados (cientos)
2	5
2.5	6
3	9
4	11
4.8	12
5.7	13
6.2	16
6.8	17
7.3	19
8.1	21

a) Obtén la gráfica de dispersión de los datos anteriores.
b) Determina el coeficiente de correlación de Pearson.

Solución

a) Al trazar los puntos de la tabla en un plano cartesiano se obtiene una gráfica como la siguiente:



b) Para calcular el coeficiente de correlación, primero se colocan todos los datos en una tabla:

2	5	−3.04	−7.9	24.016	9.2416	62.41
2.5	6	−2.54	−6.9	17.526	6.4516	47.61
3	9	−2.04	−3.9	7.956	4.1616	15.21
4	11	−1.04	−1.9	1.976	1.0816	3.61
4.8	12	−0.24	−0.9	0.216	0.0576	0.81
5.7	13	0.66	0.1	0.066	0.4356	0.01
6.2	16	1.16	3.1	3.596	1.3456	9.61
6.8	17	1.76	4.1	7.216	3.0976	16.81
7.3	19	2.26	6.1	13.786	5.1076	37.21
8.1	21	3.06	8.1	24.786	9.3636	65.61
Media = 5.04	Media = 12.9			101.14	40.344	258.9

De acuerdo con la fórmula para el coeficiente de correlación de Pearson se tiene que:

$$r = \frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\left(\sum(X-\overline{X})^2\right)\left(\sum(Y-\overline{Y})^2\right)}} = \frac{101.14}{\sqrt{(40.344)(258.9)}} = 0.9896$$

Como el coeficiente de correlación es muy cercano a 1, se puede decir que existe una buena correlación lineal entre las variables estudiadas

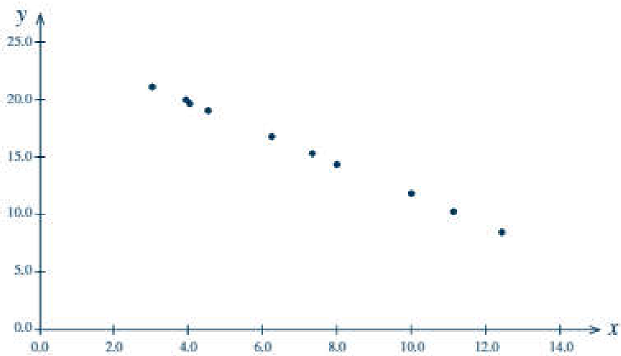
Cuando el coeficiente de correlación es positivo, la recta a la que se aproximan los datos de la dispersión tiene pendiente positiva y por tanto, se dice que la correlación entre las variables es positiva. En cambio, cuando el coeficiente de correlación es negativo, la recta a la que se aproxima la dispersión de datos tiene pendiente negativa y entonces se dice que la correlación lineal que existe entre las variables es negativa.

EJEMPLO

Ejemplo 1 Con la siguiente colección de datos, elabora la gráfica de dispersión correspondiente y determina si existe una correlación lineal positiva o negativa.

x	y
3.0	21.0
3.9	19.8
4.0	19.7
4.5	19.0
6.2	16.7
7.3	15.3
8.0	14.3
9.9	11.8
11.1	10.2
12.4	8.5

Se obtiene la gráfica de dispersión al trazar cada uno de los puntos anteriores en un plano cartesiano:



Para calcular el coeficiente de correlación de Pearson se elabora una tabla con todos los datos necesarios:

X	Y	$(X-\overline{X})$	$(Y-\overline{Y})$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
3.0	21.0	−4.0000	5.4	−21.60	16.00	29.16
3.9	19.8	−3.1000	4.2	−13.02	9.61	17.64
4.0	19.7	−3.0000	4.1	−12.20	9.00	16.53
4.5	19.0	−2.5000	3.4	−8.50	6.25	11.56
6.2	16.7	−0.8000	1.1	−0.90	0.64	1.28
7.3	15.3	0.3000	−0.3	−0.10	0.09	0.11
8.0	14.3	1.0000	−1.3	−1.26	1.00	1.60
9.9	11.8	2.9000	−3.8	−11.02	8.41	14.44
11.1	10.2	4.1000	−5.4	−22.14	16.81	29.16
12.4	8.5	5.4000	−7.1	−38.52	29.16	50.88
Media = 7.0	Media = 15.6			−129.27	96.97	172.38

Con los datos de la tabla se puede calcular el coeficiente de correlación de Pearson:

$$r=\frac{\sum\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)}{\sqrt{\left(\sum\left(X-\overline{X}\right)^2\right)\left(\sum\left(Y-\overline{Y}\right)^2\right)}}=\frac{-129.27}{\sqrt{(96.97)(172.38)}}=-1$$

Como el coeficiente de correlación es igual a -1, entonces se tiene una correlación perfecta negativa.

Recta de regresión

Cuando los puntos presentes en una dispersión tienen una correlación lineal significa que dichos puntos pueden ajustarse adecuadamente mediante una recta. Sin embargo, ¿cuál es la mejor recta que se ajusta a los puntos de la dispersión?

El procedimiento mediante el cual se obtiene la ecuación (y, por lo tanto, la gráfica) de la recta que mejor se ajusta a la dispersión de puntos en un plano cartesiano se conoce como **regresión lineal**, y la recta a la que se ajustan los valores se denomina **recta de regresión**.

La ecuación de la recta de regresión de la variable y sobre la variable x tiene la forma:

$$y=mx+b$$

donde m es la pendiente y b la ordenada al origen de la recta. Cada uno de estos valores se calcula a partir de las medias aritméticas mediante las siguientes expresiones:

$$m=\frac{\sum\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)}{\sum\left(X-\overline{X}\right)^2}$$

$$b=\overline{Y}-m\overline{X}$$

donde $\overline{X}$ y $\overline{Y}$ son las medias aritméticas respectivas de las cantidades presentes en la dispersión.

Si, en cambio, se desea hacer la recta de regresión de x sobre y, entonces la ecuación toma la forma:

$$x=my+b$$

donde:

$$m=\frac{\sum\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)}{\sum\left(Y-\overline{Y}\right)^2}$$

$$b=\overline{X}-m\overline{Y}$$

EJEMPLOS

Ejemplos

1 En una práctica de laboratorio un pequeño carro de juguete se echó a andar desde una línea de salida. Los estudiantes midieron la distancia que el carro recorría desde la línea de salida en intervalos de tiempo determinados. Los valores de distancia y tiempo obtenidos por los estudiantes se muestran a continuación:

Tiempo (segundos)	Distancia (metros)
2	7.5
4	14.1
6	18.0
8	23.7

(continúa)

(*continuación*)

Tiempo (segundos)	Distancia (metros)
10	27.0
12	34.0
14	37.1
16	43.0
18	48.5
20	52.7

a) Encuentra el coeficiente de correlación para los datos obtenidos.
b) Encuentra la pendiente y la ordenada al origen de la recta de regresión de distancia sobre tiempo.
c) Elabora la gráfica de dispersión y sobre ella, dibuja la recta de regresión obtenida.

Solución

a) Para encontrar el coeficiente de correlación, primero se elabora una tabla con los datos necesarios;

Tiempo (s)	Distancia (*m*)	$(X-\overline{X})$	$(Y-\overline{Y})$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
2	7.5	−8	−20.5545	164.4364	64	422.4893
4	14.1	−6	−13.9545	83.7273	36	194.7293
6	18.0	−4	−10.0545	40.2182	16	101.0939
8	23.7	−2	−4.3545	8.7091	4	18.9621
10	27.0	0	−1.0545	0.0000	0	1.1121
12	34.0	2	5.9455	11.8909	4	35.3484
14	37.1	4	9.0455	36.1818	16	81.8202
16	43.0	6	14.9455	89.6727	36	223.3666
18	48.5	8	20.4455	163.5636	64	418.0166
20	52.7	10	24.6455	246.4545	100	607.3984
Media = 10	Media = 28.0545			1095.4	440	2732.0673

Con los datos anteriores se tiene que:

$$r=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\left(\sum(X-\overline{X})^2\right)\left(\sum(Y-\overline{Y})^2\right)}}=\frac{1\,095.4}{\sqrt{(440)(2732.0673)}}=0.9991$$

El coeficiente de correlación es 0.9991.

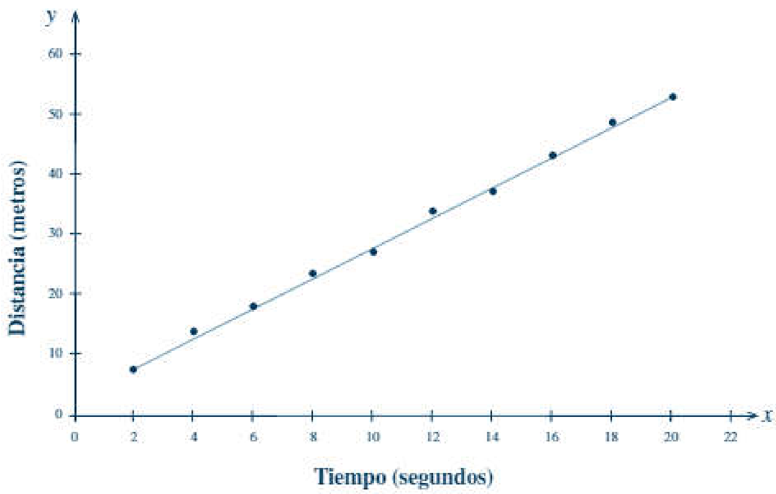
b) Para la recta de regresión se tiene que:

$$m=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sum(X-\overline{X})^2}=\frac{1\,095.4}{440}=2.4895\approx 2.5$$

$$b=\overline{Y}-m\overline{X}=28.0545-(2.5)(10)=3.1591\approx 3.1$$

La ecuación de la recta de regresión es $y=2.5x+3.1$.

c) A partir de la tabla anterior, al representar cada punto se obtiene la dispersión deseada y al trazar la recta de regresión resulta:



2 Un estudio de bioenergética se encargó de analizar el número de Calorías (kcal) que una persona pierde en una hora cuando se encuentra en reposo. Se busca relacionar la cantidad de energía perdida por hora en función de la edad de la persona. Los datos obtenidos se muestran en la siguiente tabla.

Edad (años)	Energía perdida (kcal/h)
10	49.5
15	48.3
20	47.9
25	48.5
30	47.3
35	45.8
40	44.4
45	43.1
50	41
55	40.2
60	37.7
65	37.1
70	36.8

a) Calcula el coeficiente de correlación para los datos anteriores.
b) Encuentra la recta de regresión respectiva.
c) Elabora un diagrama de dispersión de los puntos anteriores junto con la correspondiente recta de regresión.

Solución

a) Para calcular el coeficiente de correlación se construye una tabla con la información necesaria:

Edad (años)	Calorías (kcal/h)	$(X-\overline{X})$	$(Y-\overline{Y})$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
10	49.5	−30	5.8385	−175.1538	900	34.0876
15	48.3	−25	4.6385	−115.9615	625	21.5153
20	47.9	−20	4.2385	−84.7692	400	17.9646
25	48.5	−15	4.8385	−72.5769	225	23.4107
30	47.3	−10	3.6385	−36.3846	100	13.2384
35	45.8	−5	2.1385	−10.6923	25	4.5730
40	44.4	0	0.7385	0.0000	0	0.5453
45	43.1	5	−0.5615	−2.8077	25	0.3153
50	41	10	−2.6615	−26.6154	100	7.0838
55	40.2	15	−3.4615	−51.9231	225	11.9822
60	37.7	20	−5.9615	−119.2308	400	35.5399
65	37.1	25	−6.5615	−164.0385	625	43.0538
70	36.8	30	−6.8615	−205.8462	900	47.0807
Media = 40	Media = 43.66			−1 066	4 550	260.3908

Con la información de la tabla se tiene que;

$$r=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\left(\sum(X-\overline{X})^2\right)\left(\sum(Y-\overline{Y})^2\right)}}=\frac{-1\,066}{\sqrt{(4\,550)(260.3908)}}=-0.9793$$

El coeficiente de correlación es −0.9793.

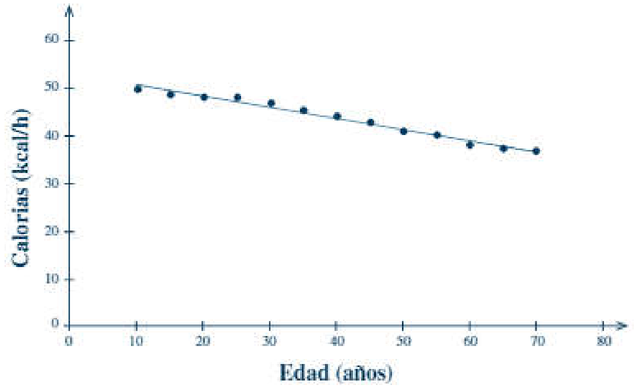
b) Para encontrar la recta de regresión, primero se necesita determinar la pendiente y la ordenada al origen, esto es:

$$m=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sum(X-\overline{X})^2}=\frac{-1\,066}{4\,550}=-0.2342\approx-0.23$$

$$b=\overline{Y}-m\overline{X}=43.6615-(-0.23)(40)=53.0330\approx53.0$$

La ecuación de la recta de regresión es $y=-0.23x+53.0$.

c) Como existe una correlación negativa entre la edad y las Calorías consumidas por hora, entonces:



Error estándar de estimación

El error estándar de estimación es una cantidad que indica qué tan confiable es la recta de regresión elaborada para un conjunto de datos en una dispersión. Gráficamente, es posible tener una idea de dicha confiabilidad al observar el comportamiento de los puntos en la dispersión: si los puntos están muy próximos a la recta de regresión, entonces ésta será una recta confiable. Por el contrario, si los puntos se ubican en posiciones lejanas al lugar donde se encuentra la recta de regresión, entonces la recta no será confiable. Puesto que dicho análisis es meramente cualitativo, el error estándar de la estimación, también llamado error estándar de regresión, permite cuantificar la cercanía o lejanía de los puntos de la dispersión a la recta de regresión.

El error estándar de la estimación se calcula mediante la siguiente fórmula:

$$S = \sqrt{\frac{\sum (Y - Y')^2}{N - 2}}$$

donde Y es cada uno de los valores de la dispersión ubicados en la coordenada y, Y' es el valor calculado para estos valores de acuerdo con la recta de regresión y N es el número total de datos.

EJEMPLO

Ejemplo 1 Las medidas de la altura y la masa de una muestra tomada de un grupo de alumnos de cuarto semestre de bachillerato se presentan en la siguiente tabla:

Altura (cm)	Masa (kg)
156	77.5
160	78.3
163	78.5
165	80.1
172	83.1
178	83.5
184	87.7
186	87
189	88.7
191	90.5

a) Determina la recta de regresión.
b) Calcula el error estándar de la estimación y el intervalo de confianza de la misma.

Solución

a) Para determinar la recta de regresión se construye una tabla con la información siguiente:

Altura (cm)	Masa (kg)	$(X-\overline{X})$	$(Y-\overline{Y})$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
156	77.5	−18.4	−5.99	110.216	338.56	35.8801
160	78.3	−14.4	−5.19	74.736	207.36	26.9361
163	78.5	−11.4	−4.99	56.886	129.96	24.9001
165	80.1	−9.4	−3.39	31.866	88.36	11.4921
172	83.1	−2.4	−0.39	0.936	5.76	0.1521
178	83.5	3.6	0.01	0.036	12.96	0.0001
184	87.7	9.6	4.21	40.416	92.16	17.7241
186	87	11.6	3.51	40.716	134.56	12.3201
189	88.7	14.6	5.21	76.066	213.16	27.1441
191	90.5			431.874	1222.84	156.5489

La pendiente y la ordenada al origen de la recta de regresión está dada por:

$$m=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sum(X-\overline{X})^2}=\frac{431.874}{1\,222.84}=0.3517\approx 0.35$$

$$b=\overline{Y}-m\overline{X}=89.49-(0.35)(174.4)=21.8967\approx 21.90$$

La ecuación de la recta de regresión es $y=0.35x+21.90$.

b) Para determinar el error estándar de la desviación se elabora otra tabla auxiliar con los datos de la variable *Y*, obtenidos a partir de la ecuación de la recta de regresión.

X	Y	Y'	Y − Y'
156	77.5	76.5	1
160	78.3	77.9	0.4
163	78.5	78.95	−0.45
165	80.1	79.65	0.45
172	83.1	82.1	1
178	83.5	84.2	−0.7
184	87.7	86.3	1.4
186	87	87	0
189	88.7	88.05	0.65
191	90.5	88.75	1.75
			5.5

Como el total de valores en la tabla es $N = 10$, entonces:

$$S = \sqrt{\frac{\sum(Y - Y')^2}{N - 2}} = \sqrt{\frac{5.5}{8}} = 0.8291 \approx 0.83$$

El error estándar de estimación se utiliza para calcular los intervalos de confianza.

Si la ecuación de la recta de regresión es:

$$y = 0.35x + 21.90$$

entonces, el intervalo de confianza está definido mediante:

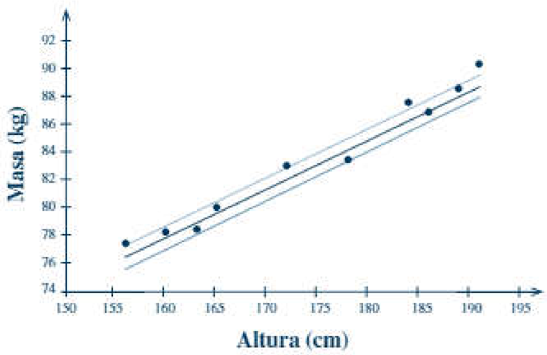
$$y = 0.35x + 21.90 \pm 0.83$$

Así, el intervalo de confianza queda definido por las rectas:

$$y = 0.35x + 22.73$$

$$y = 0.35x + 21.07$$

Gráficamente, el intervalo se muestra en la siguiente figura, que contiene la dispersión de puntos.



EJERCICIO 17

I. Contesta las siguientes preguntas.

1. ¿Qué es una gráfica de dispersión?
2. ¿Para qué sirve una nube de puntos?
3. ¿Qué es el coeficiente de correlación?
4. ¿Cuál es la fórmula del coeficiente de correlación de Pearson?
5. ¿Qué valores puede adquirir el coeficiente de correlación de Pearson y qué es lo que indica cada uno de ellos?
6. ¿Qué es una regresión lineal?
7. ¿Qué es una recta de regresión?
8. ¿Qué es un error estándar de estimación?
9. ¿Para qué sirve el error estándar de estimación?

II. Resuelve los siguientes problemas.

1. Los siguientes datos corresponden a las distancias que recorre un móvil por cada unidad de tiempo.

Tiempo (segundos)	Distancia (metros)
1	−4.9
1.5	−4.9
2.3	−3.5
2.8	−2.4
3.3	−2.6
4.2	−0.4
4.7	0
5.4	5.1
6.8	3.7
7.5	4
8.2	5.6

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

Encuentra:

a) El diagrama de dispersión.

b) El coeficiente de correlación.

c) La recta de regresión respectiva.

d) La gráfica de la recta de regresión previamente calculada.

2. Para la siguiente colección de datos encuentra:

a) El diagrama de dispersión.

b) El coeficiente de correlación.

c) La recta de regresión.

d) La gráfica de la recta de regresión previamente calculada.

e) El error estándar de estimación de la recta.

f) El intervalo de confianza de la recta de regresión.

Variable X	Variable Y
0.5	17.3
0.9	16.4
1.4	15.1
2.2	13.5
2.7	12.2
3.5	9.9
3.9	8.8
4.2	7.8
5	6.2
5.3	4.9
6	3.7
6.8	1.1

3. En la materia de farmacología, un grupo de estudiantes compararon el número de bacterias presentes en un cultivo contra el tiempo transcurrido para evaluar el funcionamiento de un antibiótico. Los resultados obtenidos fueron los siguientes:

Tiempo (minutos)	Número de bacterias
1.1	68
3.7	60
4.9	53
6.2	50
7.8	42
8.8	39
9.9	35
12.2	27
13.5	22
15.1	14
16.4	9
17.3	5

Para los datos anteriores, encuentra:

a) El diagrama de dispersión.

b) El coeficiente de correlación.

c) La recta de regresión.

d) La gráfica de la recta de regresión previamente calculada.

e) El error estándar de estimación de la recta.

f) El intervalo de confianza de la recta de regresión.

Verifica tus resultados en la sección de respuestas.

Autoevaluación

1. Los 45 alumnos de una clase han obtenido las siguientes puntuaciones, sobre 50, en un examen de Química.

   3, 15, 24, 28, 33, 35, 38, 42, 23, 38, 36, 34, 29, 25, 17, 7, 34, 36, 39, 44, 31, 26, 20, 11, 13, 22, 27, 47, 39, 37, 34, 32, 35, 28, 38, 41, 48, 15, 32, 13, 45, 32, 48,24, 49.

   a) Construye la **tabla de frecuencias**.

   b) Dibuja el **histograma** y el **polígono de frecuencias**.

2. Un dentista observa el número de caries en cada uno de los 100 niños de cierto colegio. La información obtenida aparece resumida en la siguiente **tabla**:

Nº de caries	f_i	n_i
0	25	0.25
1	20	0.2
2	x	z
3	15	0.15
4	y	0.05

   Completar la **tabla** obteniendo los valores x, y, z.

   a) Elabora un **diagrama de sectores**.

   b) Calcula el número medio de caries.

3. Dadas las series estadísticas:

   3, 5, 2, 7, 6, 4, 9 y 3, 5, 2, 7, 6, 4, 9, 1.

   Calcula:

   a) La moda, la mediana y la media.

   b) La desviación media, la varianza y la desviación típica.

   c) Los cuartiles 1º y 3º

   d) Los deciles 2º y 7º

   e) Los percentiles 32 y 85

4. Considérense los siguientes datos: 3, 8, 4, 10, 6, 2. Se pide:

   a) Calcula su media y su varianza.

   b) Si los todos los datos anteriores los multiplicamos por 3, ¿cúal será la nueva media y varianza?

5. El resultado de lanzar dos dados 120 veces viene dado por la **tabla**:

Sumas	2	3	4	5	6	7	8	9	10	11	12
Veces	3	8	9	11	20	19	16	13	11	6	4

   a) Calcula la **media** y la **desviación típica**.

   b) Determina el porcentaje de valores comprendidos en el intervalo $(x - \sigma, x + \sigma)$.

UNIDAD 2

Probabilidad

Evaluación diagnóstica

1. Determina la probabilidad de que al lanzar al aire dos monedas, salgan:

   a) Dos caras

   b) Dos cruces

   c) Una cara y una cruz

2. Determina la probabilidad de que al levantar unas fichas de dominó se obtenga un número de puntos mayor que 9 o que sea múltiplo de 4.

3. Un dado está trucado, de forma que las probabilidades de obtener las distintas caras son proporcionales a los números de estas. Determina:

   a) La probabilidad de obtener el 6 en un lanzamiento

   b) La probabilidad de conseguir un número impar en un lanzamiento

4. Se lanzan dos dados al aire y se anota la suma de los puntos obtenidos. Determina:

   a) La probabilidad de que salga el 7

   b) La probabilidad de que el número obtenido sea par

   c) La probabilidad de que el número obtenido sea múltiplo de tres

5. Se lanzan tres dados. Encontrar la probabilidad de que:

   a) Salga 6 en todos

   b) Los puntos obtenidos sumen 7

TEMA 6 Teoría de conjuntos

Propósito del tema

Que el estudiante:

- Hago uso del lenguaje de conjuntos para establecer procedimientos de solución de problemas.
- Exprese razones para fundamentar una respuesta y obtenga conclusiones pertinentes a partir de datos probabilistícos.
- Sea capaz de resolver problemas a partir de las operaciones entre conjuntos.

Competencias disciplinares

1. Construye e interpreta modelos matemáticos mediante la aplicación de procedimientos aritméticos, algebraicos, geométricos y variacionales, para la comprensión y análisis de situaciones reales, hipotéticas o formales.
3. Interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos o situaciones reales.
5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.
7. Elige un enfoque determinista o uno aleatorio para el estudio de un proceso o fenómeno, y argumenta su pertinencia.
8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

Contenidos que aborda el tema

Contenidos conceptuales

- Definición de los elementos básicos de conjuntos.
- Concepto de conjunto.
- Definición de las propiedades de conjunto
- Operación con conjuntos.
- Definición de diagrama de Venn.

Contenidos procedimentales

- Sintetizará evidencias obtenidas mediante la experimentación para producir conclusiones y formulará nuevas preguntas.
- Utilizará el diagrama de Venn para entender de manera gráfica la teoría de conjuntos.
- Seguirá procedimientos de manera reflexiva, comprendiendo como cada uno de sus pasos contribuye al alcance de un objetivo.
- Usará diversas estrategias de resolución de problemas.

Contenidos actitudinales

- Expresará sus ideas mediante los elementos básicos de la teoría de conjuntos.
- Trabajará en equipo y respetará a sus compañeros al resolver problemas.
- Aprenderá a valorar el trabajo de sus compañeros al resolver problemas.
- Estructurará ideas y argumentos de manera clara, coherente y sintética.

Los conjuntos y sus operaciones

El lenguaje de los conjuntos

Existe un buen número de palabras que denotan conjuntos, como: **recua** de mulas, **hatos** de vacas, **clases** de conscriptos, **equipos** de béisbol, **manada** de cerdos, **racimos** de uvas, **bandadas** de golondrinas y **grupos** étnicos.

Un conjunto consta de **elementos** o **miembros**, por ejemplo, una colección de monedas antiguas es un conjunto y las monedas son los elementos de la colección.

Ser miembro de un conjunto no necesariamente implica la existencia de una característica común y evidente o algún tipo de colectividad natural. Arbitrariamente se puede designar como conjunto a cualquier colección de objetos o imágenes dadas, por ejemplo: \$, B, 93, Ω son miembros de un conjunto y por tal hecho, sus elementos están relacionados.

Métodos para describir un conjunto

Existen tres métodos que permiten describir un conjunto.

Descripción por comprensión. Es la manera más sencilla de dar a conocer el contenido de un conjunto, y consiste en una descripción verbal del mismo.

Ejemplo

El conjunto de los números superiores a 25, el conjunto de los números mayores que 5 y menores que 100, el conjunto de los días de la semana, el conjunto de las estaciones del año, el conjunto de los billetes actuales de los Estados Unidos Mexicanos, etcétera.

Descripción por extensión. Forma que permite denotar el contenido de un conjunto mediante la presentación de una lista de sus elementos separados por comas y encerrados entre llaves.

Ejemplo

El conjunto de los números inferiores a 12, puede representarse como: $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$.

El conjunto de las vocales del alfabeto castellano puede representarse como: {a, e, i, o, u}.

Si este método se aplica para describir todos los números inferiores de 1000, se tendrían que escribir 999 números; sin embargo, este conjunto puede representarse como: $\{1, 2, 3, ..., 997, 998, 999\}$.

Ahora bien, si se quiere describir el conjunto de **todos los números superiores a 5**, sería imposible escribir todos los elementos del conjunto; sin embargo, éste se podría representar como: $\{6, 7, 8, 9, 10, ...\}$ donde los puntos suspensivos indican la continuidad indefinida de los elementos.

Notación de constitución de conjuntos. La simbología (|) se denomina **notación de constitución de conjuntos** y describe un conjunto con base en las condiciones de un elemento arbitrario del grupo, es decir, establece las condiciones bajo las cuales, un elemento cualquiera puede o no pertenecer al conjunto.

Ejemplo

El conjunto de los números impares, mayores que 4 y menores que 14, puede expresarse como $\{X|X$ es un número impar, mayor que 4 y menor que 14$\}$.

Las llaves { } indican la notación del conjunto; la línea vertical | se lee como **tal que**; la letra X es un elemento arbitrario del conjunto y a su vez, una variable; el lado izquierdo de la línea vertical se lee como **el conjunto de las** X y al lado derecho de la línea vertical se tienen las **condiciones necesarias** para pertenecer al conjunto.

Conjunto bien definido

Un conjunto está **bien definido** cuando su descripción determina claramente y sin ambigüedad cuáles elementos pertenecen y cuáles no pertenecen a él; por ejemplo: los número impares que van de 5 a 15, es decir: {5, 7, 9, 11, 13, 15} es un conjunto bien definido.

Pertenencia de un elemento a un conjunto

Para indicar **pertenencia** se usa el símbolo $\in$.

Ejemplo

Si $\boldsymbol{a}$ es un objeto que se encuentra dentro del conjunto A se escribe $a \in A$, que se lee *"$\boldsymbol{a}$ es un elemento que pertenece al conjunto A"*. En caso contrario, la expresión $\boldsymbol{a} \notin \boldsymbol{A}$ indica que el elemento $\boldsymbol{a}$ no se encuentra dentro del conjunto A y se lee *"$\boldsymbol{a}$ no pertenece al conjunto A"*.

Notación matemática de conjunto

Un conjunto se puede designar por medio de una letra mayúscula, y es necesario citar esa letra, en lugar de escribir el conjunto completo cada vez que se requiera.

Ejemplo

Sea $\boldsymbol{G}$ el conjunto de los meses del año que comienzan con la letra A, es decir:

$$G = \{\text{abril, agosto}\}$$

Como se utiliza un número limitado de literales constantemente, el conjunto al que se hace referencia con determinada letra m no puede ser el mismo en otros casos; todo depende de cómo se establezcan las descripciones. Algunas literales llegan a asociarse con conjuntos particulares, por ejemplo, el conjunto universal.

Conjunto universal

El conjunto que contiene a los elementos de los conjuntos que se toman en cuenta en un análisis cualquiera se denomina **conjunto universal** o también **universo** y se representa por la letra mayúscula $\boldsymbol{U}$.

Cabe mencionar que un conjunto puede variar de un contexto a otro, o bien, cambiar dentro de un mismo análisis.

Igualdad

Dos conjuntos A y B son iguales cuando A contiene exactamente los mismos elementos que B, o viceversa.

Ejemplo

Sean los conjuntos $A = \{b, g, o\}$ y $B = \{g, o, b\}$, se establece que son iguales, ya que contienen exactamente los mismos elementos, es decir:

$$A = B \qquad \text{o} \qquad \{b, g, o\} = \{g, o, \text{b}\}$$

Se observa que el orden en que se enuncian los elementos del conjunto carece de importancia.

Si el conjunto A no es igual al conjunto B, es decir, cuando no tienen exactamente los mismos elementos, se utiliza la expresión $\boldsymbol{A \neq B}$, que se lee "**A diferente de B**".

Correspondencia uno a uno

Dos conjuntos $\boldsymbol{M}$ y N tienen **correspondencia uno a uno** (o también llamada **correspondencia biunívoca**), si cada elemento de $\boldsymbol{M}$ puede aparearse exactamente con uno de N y cada elemento de N puede aparearse exactamente con uno de $\boldsymbol{M}$.

Ejemplo

Sean los conjuntos $M = \{2, 4, 6\}$ y $N = \{\propto, \theta, \varnothing\}$; el apareamiento puede ser:

$$\begin{matrix} 2 & 4 & 6 \\ \updownarrow & \updownarrow & \updownarrow \\ \propto & \theta & \varnothing \end{matrix} \qquad \begin{matrix} 4 & 6 & 2 \\ \updownarrow & \updownarrow & \updownarrow \\ \propto & \theta & \varnothing \end{matrix} \qquad \begin{matrix} 6 & 2 & 4 \\ \updownarrow & \updownarrow & \updownarrow \\ \propto & \theta & \varnothing \end{matrix}$$

¿Cuántos apareamientos adicionales se pueden establecer?

Conjuntos equivalentes

Dos conjuntos X y Y son **equivalentes** cuando presentan correspondencia uno a uno.

Ejemplo

Sean los conjuntos $X = \{\text{febrero, junio, diciembre}\}$ y $Y = \{\text{Garza, Gómez, Méndez}\}$, se establece que son equivalentes porque pueden aparearse uno a uno, es decir:

$$\begin{matrix} X = \{ & \text{febrero,} & \text{junio,} & \text{diciembre} & \} \\ & \updownarrow & \updownarrow & \updownarrow & \\ Y = \{ & \text{Garza,} & \text{Gómez,} & \text{Méndez} & \} \end{matrix}$$

Subconjuntos

Cuando todos los elementos de un conjunto cualquiera A son también miembros de otro conjunto B, se dice que el conjunto A es **subconjunto** del conjunto B; el símbolo empleado para indicar esta relación es $\subset$.

Ejemplo

Sean los conjuntos $K = \{a, b, c, d, e\}$ y $L = \{a, c, e\}$, se establece que el conjunto L es subconjunto del conjunto K, lo que simbólicamente se representa como: $L \subset K$.

¿Cuántos subconjuntos posibles de L se pueden establecer?

Todos los subconjuntos posibles de L serán:

$$\begin{array}{llll} A = \{\ \} & A \subset L & E = \{a, c\} & E \subset L \\ B = \{a\} & B \subset L & F = \{a, e\} & F \subset L \\ C = \{c\} & C \subset L & G = \{c, e\} & G \subset L \\ D = \{e\} & D \subset L & H = \{a, c, e\} & H \subset L \text{ o } H = L \end{array}$$

Se hace notar que el subconjunto de L es el conjunto H, que es exactamente igual al conjunto L; por lo anterior, se establece que **cualquier conjunto es un subconjunto de sí mismo**.

Conjunto vacío

Cuando un conjunto carece de elementos se denomina **conjunto vacío** o **nulo**. El conjunto vacío se expresa simbólicamente por $\varnothing$ o $\{\ \}$. También se establece que el conjunto vacío o nulo **es un subconjunto de todos los conjuntos**.

Conjunto potencia

Si se pregunta cuántos subconjuntos se pueden obtener de un conjunto dado, la respuesta es que el número de subconjuntos resultantes de cualquier conjunto dado se obtiene con la expresión $T = 2^n$, en donde T es el número de subconjuntos y n representa el número de elementos del conjunto dado.

Ejemplo

¿Cuántos subconjuntos se obtienen del conjunto $F = \{a\}$?

Se observa que el conjunto dado tiene un solo elemento, es decir, $n = 1$ y al sustituir los datos en la ecuación $T = 2^n$ se tiene:

$$T = 2^1 = 2, \text{ es decir: } \qquad G = \{\} \text{ y } H = \{a\}$$

Dado el conjunto $A = \{2, x\}$, ¿cuántos subconjuntos sse pueden obtener?

Se observa que el conjunto dado tiene dos elementos, es decir, $\boldsymbol{n = 2}$ al sustituir los datos en la ecuación $T = 2^n$ se tiene:

$$T = 2^2 = 4, \text{ es decir: } B = \{\ \}, C = \{2\}, D = \{x\} \text{ y } E = \{2, x\}$$

Subconjunto propio

Se establece que el conjunto $\boldsymbol{A}$ es un **subconjunto propio** del conjunto $\boldsymbol{B}$, si todos los elementos de $\boldsymbol{A}$ están contenidos en $\boldsymbol{B}$ y si en $\boldsymbol{B}$ se encuentra por lo menos un elemento no contenido en $\boldsymbol{A}$.

Ejemplo

Sean los conjuntos $A = \{$lunes, miércoles, viernes$\}$ y $B = \{$lunes, miércoles, viernes, domingo$\}$, entonces **A es un subconjunto propio de B.**

Conjunto infinito

Se tiene un **conjunto infinito** cuando no es posible indicar el número de elementos que están contenidos en él.

Ejemplo

El conjunto de todos los números naturales: $N = \{1, 2, 3, 4, 5, \ldots\}$

Conjunto finito

Se tiene un **conjunto finito** cuando es posible indicar el número de elementos que están contenidos en él. Simbólicamente el número de elementos de un conjunto finito se expresa por $\boldsymbol{n}$.

Ejemplo

Se tiene el conjunto de los días de la semana, es decir, $K = \{$lunes, martes, miércoles, jueves, viernes, sábado, domingo$\}$, $\boldsymbol{K}$ es un conjunto finito ya que consta de 7 elementos, es decir: $n(K) = 7$.

Es necesario aclarar que un conjunto puede ser finito, aunque físicamente sea muy difícil o esté fuera de la capacidad humana, determinar cuántos elementos están contenidos.

Ejemplo

El conjunto de estrellas en el firmamento se considera un conjunto finito, aunque, ¿quién podrá contarlas?

Operaciones con conjuntos

Una operación con conjuntos es un roceso que conduce a la formación de conjuntos a partir de otros conjuntos. Las principales operaciones con conjuntos son la unión, la intersección, la diferencia, el complemento y el producto cartesiano.

Unión

A partir de dos conjuntos A y B se obtiene otro conjunto cuyos elementos son todos los contenidos en $\boldsymbol{A}$ junto con todos los contenidos en $\boldsymbol{B}$. A este proceso se le denomina **unión de conjuntos** y se representa con el símbolo $\cup$.

Ejemplo

Sean los conjuntos $A = \{a, c, e, g\}$ y $B = \{a, b, d, f\}$, la unión de estos conjuntos da como resultado:

$$A \cup B = \{a, b, c, d, e, f, g\}$$

Intersección

A partir de dos conjuntos A y B se obtiene otro conjunto cuyos elementos deben ser comunes en ambos conjuntos. A este proceso se le denomina **intersección de conjuntos** y se representa con el símbolo $\cap$.

Ejemplo

Sean los conjuntos $M = \{$oro, plata, níquel, cobre$\}$ y $N = \{$plata, aluminio, paladio, cobre$\}$, la **intersección** de estos conjuntos da como resultado:

$$M \cap N = \{\text{plata, cobre}\}$$

Dados los conjuntos $P = \{a, e, i, o, u\}$ y $Q = \{w, x, y, z\}$, como no tienen ningún elemento en común su **intersección** es el **conjunto vacío**, es decir:

$$P \cap Q = \varnothing$$

En tal situación, se establece que los conjuntos $\boldsymbol{P}$ y $\boldsymbol{Q}$ son **disjuntos**.

Diferencia

A partir de dos conjuntos A y B se obtiene otro conjunto cuyos elementos son aquellos que pertenecen al conjunto A pero no están contenidos en el conjunto $\boldsymbol{B}$. A esté proceso se le denomina **diferencia de conjuntos** o **complemento relativo de B respecto de A** y se representa con el signo "$-$".

Ejemplo

Sean los conjuntos $A = \{4, 6, 7, 8, 10, 12\}$ y $B = \{10, 11, 12, 13, 14, 15\}$, la **diferencia** de estos conjuntos da como resultado:

$$A - B = \{4, 6, 7, 8\}$$

Complemento

A partir de un conjunto A y un conjunto universal $\boldsymbol{U}$ se obtiene otro conjunto cuyos elementos deben ser todos los que estén contenidos en el conjunto $\boldsymbol{U}$ que no pertenecen el conjunto A. Este proceso se denomina **complemento de un conjunto cualquiera en relación con un conjunto universal dado**. Simbólicamente se representa con una **comilla** que se ubica en la parte superior derecha de la literal que define al conjunto cualquiera.

Ejemplo

Dados el conjunto universal $U = \{1, 3, 5, 7, 9, 11, 13\}$ y el conjunto $K = \{5, 7, 9\}$ se puede formar otro conjunto que contenga todos los elementos de conjunto $\boldsymbol{U}$ que no sean miembros del conjunto $\boldsymbol{K}$, esto es:

$$K' = \{1, 3, 11, 13\}$$

El conjunto K', que se lee **K prima**, se denomina complemento del conjunto K.

Producto cartesiano

A partir de dos conjuntos A y B se obtiene otro conjunto cuyos elementos de denominan **pares ordenados**, que se escriben entre paréntesis. El orden significativo de dichos pares ordenados se indica de acuerdo con la posición de los elementos, es decir, el primer componente pertenece al conjunto A y el segundo componente pertenece al conjunto $\boldsymbol{B}$. Este proceso se denomina **producto cartesiano** y simbólicamente se representa como $\times$.

Ejemplo

Sean los conjuntos $A = \{a, b, c\}$ y $B = \{c, b, s\}$, se puede formar otro conjunto que contenga todos los pares ordenados que resulten de la ejecución del producto cartesiano $A \times B$, es decir:

$$A \times B = \{(a,c), (a,b), (a,s), (b,c), (b,b), (b,s), (c,c), (c,b), (c,s)\}$$

Los conjuntos que se emplean en la construcción de un **producto cartesiano** no necesariamente tienen que ser diferentes.

Ejemplo

Sea el conjunto $G = \{2, 4, 6\}$, realiza la operación $G \times G$.

$$G \times G = \{(2, 2), (2, 4), (2, 6), (4, 2), (4, 4), (4, 6), (6, 2), (6, 4), (6, 6)\}$$

Propiedades de los conjuntos

a) Conmutativa $A \cup B = B \cup A$

$A \cap B = B \cap A$

$A - B \neq B - A$

b) Asociativa $A \cup (B \cup C) = (A \cup B) \cup C$

$A \cap (B \cap C) = (A \cap B) \cap C$

c) Distributiva $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

d) Leyes de Morgan $(A \cup B)' = A' \cap B'$

$(A \cap B)' = A' \cup B'$

Otras propiedades de los conjuntos

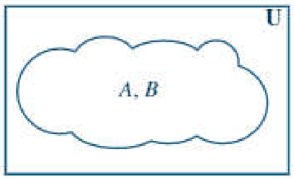
$U' = \varnothing$	$A \cup A' = U$	$U \cap \varnothing = \varnothing$	$A \cap A' = \varnothing$	$A \cap A = A$
$U \cup \varnothing = U$	$A \cup U = U$	$A \cup \varnothing = A$	$A \cup A = A$	$(A')' = A$
	$\varnothing' = U$	$A \cap U = A$	$A \cap \varnothing = \varnothing$	

Diagramas de Venn

La manera más fácil de comprender las ideas de la teoría de conjuntos es por medio de los **diagramas de Venn**. Estos gráficos ayudan a relacionar la igualdad y las operaciones de unión, intersección, diferencia y complemento entre conjuntos.

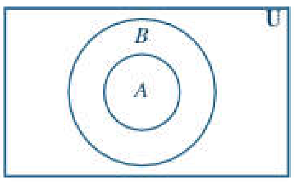
En los diagramas de Venn, los conjuntos se representan mediante óvalos, círculos o nubes y el punto de referencia es el conjunto universal U que se representa por un rectángulo.

Si A y B son conjuntos contenidos en el conjunto universal U, por medio de diagramas de Venn se pueden determinar las siguientes relaciones:

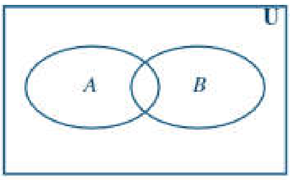


En este diagrama se representa la igualdad entre A y B; también se establece que el conjunto A es un subconjunto del conjunto B o viceversa.

$$A = B \quad \text{si} \quad A \subset B \text{ o } B \subset A$$

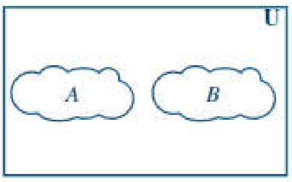


En este diagrama el conjunto A representa un subconjunto propio del conjunto B, es decir, todos los elementos de A están contenidos en B, mientras que B tiene por lo menos un elemento no contenido en A.



En este diagrama los conjuntos A y B tienen en común algunos pero no todos los elementos, es decir, representan la intersección entre A y B.

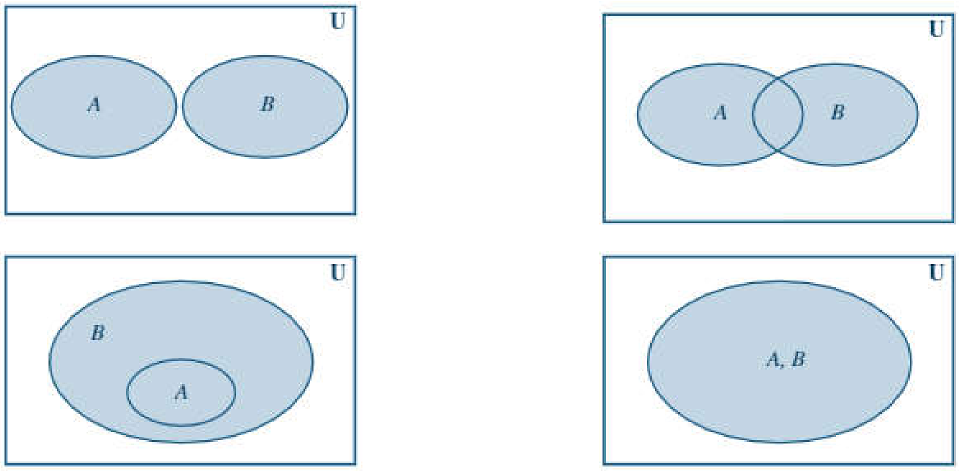
$$A \cap B$$



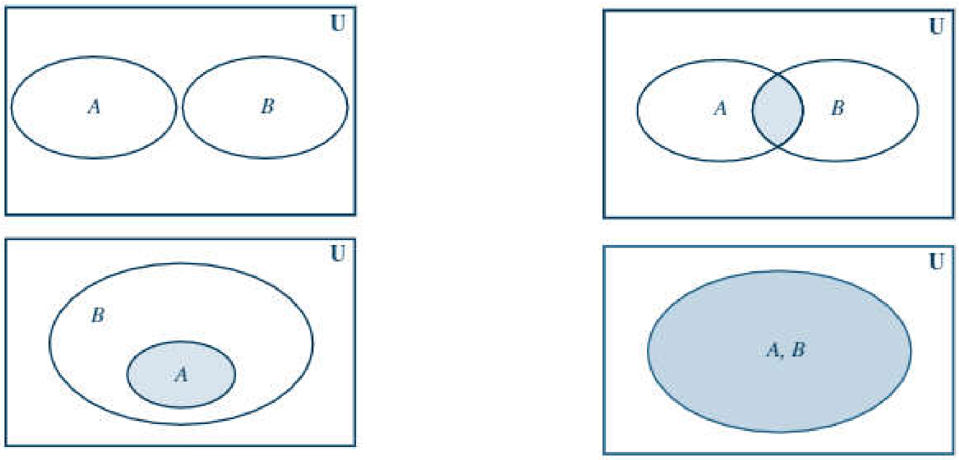
En este diagrama los conjuntos A y B no tienen ningún elemento en común, es decir, representan dos conjuntos disjuntos cuya intersección es el conjunto vacío.

$$A \cap B = \varnothing$$

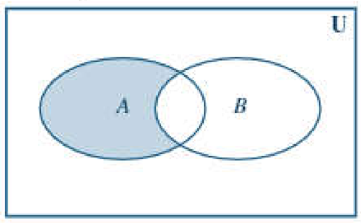
Las principales operaciones entre dos conjuntos se representan por medio de los siguientes diagramas de Venn. Las superficies sombreadas en las siguientes figuras ilustran la unión del conjunto A con el conjunto B.



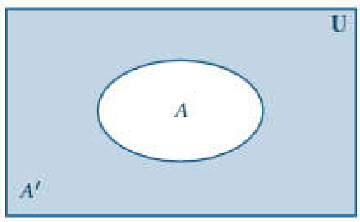
Las superficies sombreadas en las siguientes figuras ilustran la intersección del conjunto A con el conjunto B.



La superficie sombreada en la siguiente figura ilustra la diferencia del conjunto A con el conjunto B.



El complemento A' del conjunto A se obtiene sombreando la superficie del conjunto universal U no contenida en A, es decir:



EJEMPLOS

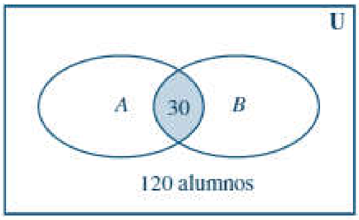
Ejemplos

1 En una escuela de idiomas hay 120 alumnos, de los cuales 65 estudian alemán, 55 ejercitan su inglés y 30 estudian ambos idiomas. Utiliza un diagrama de Venn para determinar:

a) Los alumnos que sólo estudian alemán.
b) Los alumnos que sólo estudian inglés.
c) El número de alumnos que estudian alemán o inglés.
d) El número de alumnos que no estudian ninguno de estos idiomas.

Solución

Se utiliza un diagrama de Venn para representar el conjunto universal U de 120 alumnos, el conjunto A de alumnos, que estudian alemán y el conjunto B de alumnos que estudian inglés. En la intersección de ambos conjuntos se tienen a los 30 alumnos que estudian alemán e inglés simultáneamente, es decir:



a) Los alumnos que sólo estudian alemán son:

$n(A) - n(A \cap B) = 65 - 30 = 35$ 35 alumnos solamente estudian alemán.

b) Los alumnos que sólo estudian inglés son:

$n(B) - n(A \cap B) = 55 - 30 = 25$ 25 alumnos solamente estudian inglés.

c) El número de alumnos que estudian alemán o inglés son:

$$35 + 25 + n(A \cap B) = 35 + 25 + 30 = 90$$

d) El número de alumnos que no estudian ninguno de estos idiomas es:

$$U - [35 + 25 + n(A \cap B)] = 120 - 90 = 30$$

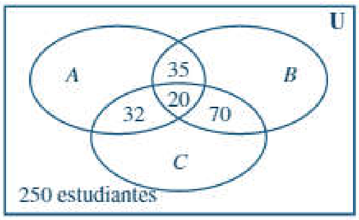
2 Una encuesta aplicada a 250 estudiantes de nivel medio superior dio lugar a la siguiente información acerca de su ingreso a los cursos de química, física y matemáticas:

- 101 estudian química
- 163 estudian física
- 163 estudian matemáticas
- 35 estudian química y física
- 32 estudian química y matemáticas
- 70 estudian física y matemáticas
- 20 estudian química, física y matemáticas

a) ¿Cuántos estudiantes llevan química como único curso?
b) ¿Cuántos no siguen ninguno de los tres cursos?
c) ¿Cuántos estudian química y matemáticas, pero no física?
d) ¿Cuántos alumnos no toman cursos de química ni física?

Solución

Se utiliza un diagrama de Venn para representar el conjunto universal U de 250 estudiantes, el conjunto A de estudiantes que cursan química, el conjunto B de estudiantes que cursan física y el conjunto C de estudiantes que cursan matemáticas. En la intersección de los tres conjuntos se tienen a los 20 estudiantes que cursan química, física y matemáticas simultáneamente es decir:



a) Los estudiantes que sólo cursan química son:

$$n(A) - n(A \cap B) - n(A \cap C) - n[A \cap (B \cap C)] = 101 - 35 - 32 - 20 = 14$$

14 estudiantes cursan sólo química.

b) Los estudiantes que no llevan ninguno de los tres cursos son:
Primero se determina cuántos estudiantes cursan solamente física, es decir:

$$n(B) - n(A \cap B) - n(B \cap C) - n[A \cap (B \cap C)] = 163 - 35 - 70 - 20 = 38$$

También se determina cuántos estudiantes cursan solamente matemáticas, es decir:

$$n(C) - n(A \cap C) - n(B \cap C) - n[A \cap (B \cap C)] = 163 - 32 - 70 - 20 = 41$$

Ahora se determina cuántos estudiantes no llevan ninguno de los tres cursos, es decir:

$$U = \{14 + 38 + 41 + n(A \cap B) + n(A \cap C) + n(B \cap C) + n[A \cap (B \cap C)]\}$$

$$U = 250 - 14 - 38 - 41 - 35 - 32 - 70 - 20 = 0$$

Todos los estudiantes siguen por lo menos uno de los tres cursos.

c) Los estudiantes que cursan química y matemáticas pero no física son:

$$14 + 41 + n(A \cap C) = 14 + 41 + 32 = 87$$

87 estudiantes cursan química y matemáticas pero no física.

d) Los estudiantes que no cursan química ni física, son:

$$14 + 38 + n(A \cap B) = 14 + 38 + 35 = 87$$

87 estudiantes no cursan química ni física.

EJERCICIO 18

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

I. Resuelve lo que se indica en cada caso.

1. ¿Qué es un conjunto?
2. Menciona cinco términos descriptivos que se utilizan para nombrar ciertos conjuntos.
3. Escribe los elementos de cada uno de los siguientes conjuntos.
   a) El conjunto de los países de Centroamérica.
   b) El conjunto de asignaturas que estás cursando este semestre.
   c) El conjunto de los meses del año.
   d) El conjunto de los días de la semana que comienzan con la letra M.
   e) El conjunto de los números pares menores que 25.
4. Escribe los elementos de cada uno de los siguientes conjuntos descritos por la notación de constitución:
   a) $B = \{x|x$ es el cuadrado de un número mayor que 2 y menor que $11\}$.
   b) $G = \{x|x$ es un día de la semana$\}$.
   c) $O = \{x|x$ es un triciclo de dos ruedas$\}$.
   d) $A = \{x|x$ es un presidente de los Estados Unidos Mexicanos de 1934 a $1970\}$.
   e) $M = \{x|x$ es una figura de los naipes americanos$\}$.
5. Describe cada uno de los siguientes conjuntos por medio de la notación de constitución.
   a) $K = \{$octubre, noviembre, diciembre$\}$.
   b) $P = \{a, b, c, d, e, f\}$.
   c) $H = \{$violeta, azul turquí, azul, verde, amarillo, anaranjado, rojo$\}$.
   d) $Q = \{$aritmética, álgebra, geometría plana, trigonometría, geometría analítica, cálculo integral, estadística, probabilidad$\}$.
   e) $C = \{$John, Paul, George, Ringo$\}$.

6. ¿Cuándo se tiene un conjunto bien definido?
7. Indica si los siguientes conjuntos están definidos o enumerados:
   a) {automóvil, bicicleta, tren, avión, barco}. ______________
   b) {1, 4, 9, 16, 25, …}. ______________
   c) El conjunto de todos los alumnos de la clase de filosofía cuyo apelativo comience con la letra *G*. ______________
   d) {rey, caballo, sota,…, as}. ______________
   e) El conjunto de los profesores de matemáticas de tu escuela. ______________
8. Indica si los siguientes conjuntos están o no bien definidos:
   a) El conjunto de astronautas norteamericanos que han alunizado. ______________
   b) El conjunto de bebidas sabrosas. ______________
   c) El conjunto de los políticos mundialmente famosos. ______________
   d) El conjunto de los números divisibles entre 5. ______________
   e) El conjunto de directores de todas las escuelas del estado de Tabasco. ______________
9. ¿Cuál es el símbolo que indica pertenencia de un elemento a un conjunto?
10. Menciona tres ejemplos de la pertenencia de elementos a un conjunto.
11. ¿Qué es el conjuntos universal?
12. Explica la igualdad entre dos conjuntos.
13. Indica si los siguientes pares de conjuntos son iguales:
    a) $\{a, e, i, o, u\}$ y $\{u, a, o, e, i\}$. ______________
    b) {letras del alfabeto español} y {28}. ______________
    c) {letras de la palabra amor}, $\{a, m, o, r\}$. ______________
    d) {números pares menores que 15} y {2, 4, 6, 8, 11, 14}. ______________
    e) {número de los colores del arcoiris} y {7}. ______________
14. ¿Qué es una correspondencia biunívoca?
15. Indica cuáles de los siguientes conjuntos pueden colocarse en correspondencia de uno a uno.
    a) El conjunto de todos los números naturales y el conjunto de granos de arena de la playa de Cancún. ______________
    b) {3, 6, 9, 12, 15,…} y {5, 10, 15, 20, 25,…}. ______________
    c) {lunes, miércoles, viernes} y {rojo, azul, verde}. ______________
    d) El conjunto de automóviles de tu ciudad y el conjunto de habitantes de la misma. ______________
    e) {Alberto, Manuel, Álvaro, Carlos} y {Olivia, Ana, María}. ______________
16. ¿Cuáles de los siguientes pares de conjuntos son iguales? ¿Cuáles son equivalentes?
    a) $M = \{1, 2, 3\}$ y $N = \{2, 4, 6\}$. ______________
    b) $K = \{I, b, g, e\}$ y $L = \{g, c, I, b\}$. ______________
    c) $A = \{\ \}$ y $B = \varnothing$. ______________
    d) $X = \left\{\frac{1}{2}, \frac{1}{4}, \frac{1}{5}, \frac{1}{8}\right\}$ y $Y = \{0.125, 0.2, 0.25, 0.5\}$. ______________
    e) $P = \{a, e, i, o, u\}$ y $Q = \{1, 2, 3, 4, 5\}$. ______________
17. ¿Qué entiendes por subconjunto?

18. ¿Cuántos subconjuntos posibles se pueden establecer para los siguientes conjuntos?
    a) $\{k, g, b, t, r\}$.
    b) {John, Pau1, George, Ringo}.
    c) $\{2, 4, 6, 8, 10, 12\}$.
19. Si el conjunto $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, selecciona:
    a) Un subconjunto cuyos elementos sean impares.
    b) Un subconjunto cuyos elementos sean divisibles entre 2.
    c) Un subconjunto cuyos elementos sean números primos.
    d) Un subconjunto cuyos elementos sean múltiplos de 3.
20. ¿Qué es el conjunto vacío?
21. ¿Qué es un conjunto potencia?
22. Explica la diferencia entre un conjunto infinito y uno finito.
23. Indica si los siguientes conjuntos son finitos o infinitos:
    a) $\{b\}$. _______________
    b) $\{3, 6, 9, \ldots 45\}$. _______________
    c) $\{2, 4, 6, 8, \ldots\}$. _______________
    d) $\{x|x$ es el cuadrado de un número mayor que $1\}$. _______________
    e) $\{x|x$ es la existencia del tiempo$\}$. _______________
    f) $\{x|x$ es el número de cabellos en la cabeza de una persona$\}$. _______________
24. ¿Cómo son las operaciones con conjuntos?

II. Resuelve las siguientes operaciones de conjuntos.

1. Para los siguientes pares de conjuntos dados, determina $A \cup B$, $A \cap B$ y $A - B$.
    a) $A = \{1, 2, 3, 4, 5, 6, 7\}$ y $B = \{1, 3, 5, 7, 9\}$.
    b) $A =$ {Brasil, México, Colombia, Venezuela, Canadá} y
    $B =$ {Perú, Argentina, México, Ecuador, Venezuela}.
    e) $A = \{a, b, c, d, e, f\}$ y $B = \{1, 3, 5, 7, 9\}$.
    d) $A =$ {rojo, verde, azul, café} y $B =$ {amarillo, rojo, café, gris}
    e) $A =$ {Pedro, Héctor, David, Álvaro, Manuel} y $B =$ {Héctor, Alberto, Samuel, Álvaro, Rubén}.
2. Dado el conjunto universal $U =$ {Mercurio, Venus, Tierra, Marte, Júpiter, Saturno, Urano, Neptuno} y el conjunto $R =$ {Tierra, Marte, Júpiter}, determina el complemento del conjunto R.
3. Efectúa el producto cartesiano entre los pares de conjuntos dados en el problema I.
4. ¿Cuál es el resultado de aplicar el producto cartesiano a los conjuntos $P = \{13\}$ y $Q = \{\}$?
5. Describe el conjunto de todos los resultados posibles que pueden obtenerse al lanzar simultáneamente dos monedas al aire. Relaciona dicho conjunto con el producto cartesiano de $\{a, s\} \times \{a, s\}$, donde ***a*** representa al águila y *s* representa al sol.
6. Describe los elementos pertenecientes al conjunto de resultados posibles, al lanzar una moneda al aire tres veces consecutivas.
7. Define la unión de dos conjuntos M y N por medio de la notación de constitución.
8. Define la intersección de dos conjuntos K y L por medio de la notación constitución.

9. Define la diferencia de dos conjuntos P y Q por medio de la notación de constitución.
10. Define el producto cartesiano de dos conjuntos X y Y por medio de la notación de constitución.

III. Resuelve las siguientes operaciones de conjuntos.

1. Represента los siguientes conjuntos por medio de diagramas de Venn.
   a) $M \cup N'$ *b*) $K' \cap L$ *c*) $(P' \cup Q)'$
2. Representa las siguientes igualdades de conjuntos por medio de diagramas de Venn.
   a) $(P \cap Q) \cap R = P \cap (Q \cap R)$
   b) $(M \cup N)' = M' \cap N'$
   c) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
3. Sea el conjunto universal $U = \{1, 2, 3, \ldots, 13\}$ y los conjuntos $A = \{3, 5, 6, 10, 11, 13\}$, $B = \{2, 4, 6, 9, 11, 13\}$ y $C = \{1, 3, 5, 9, 11, 13\}$, determina los elementos de los siguientes conjuntos y represéntalos por medio de diagramas de Venn.
   a) $(A \cup B\} \cap C'$ *c*) $(A \cup B)' \cap C'$ *e*) $(A' \cap B') \cap C'$
   b) $A' \cap (B' \cap C')$ *d*) $A \cup (B \cap C')$ *f*) $[(A \cup B) \cup C]'$
4. Utiliza la regla: $n(A \cup B = n(A) + n(B) - n(A \cap B)$ para determinar $n(A \cup B)$ de los datos del problema 3.
5. En un grupo de 55 personas hay solamente dos tipos de individuos, los especialistas en economía y los estadistas. Si 35 son especialistas en economía y 31 son estadistas, ¿cuántos son a la vez estadistas y especialistas en economía? Represéntalos por medio de diagramas de Venn.
6. En una reunión de 400 delegados, 180 pertenecen al partido republicano, 160 pertenecen al partido demócrata y 160 pertenecen al partido independiente; 30 delegados a su vez pertenecen al partido republicano, demócrata e independiente; 60 delegados pertenecen al partido republicano y demócrata, 40 delegados pertenecen al partido demócrata e independiente y 50 delegados pertenecen al partido republicano e independiente. Determina por medio de un diagrama de Venn:
   a) ¿Cuántos delegados pertenecen sólo al partido republicano?
   b) ¿Cuántos delegados pertenecen sólo al partido demócrata?
   c) ¿Cuántos delegados pertenecen sólo al partido independiente?
   d) ¿Cuántos delegados no pertenecen a ningún partido?
7. De 120 alumnos 60 estudian francés, 50 español y 20 estudian francés y español. Determina por medio de un diagrama de Venn:
   a) ¿Cuántos alumnos sólo estudian francés?
   b) ¿Cuántos alumnos sólo estudian español?
   c) El número de alumnos que estudian francés o español.
   d) El número de alumnos que no estudian ninguno de estos idiomas.
8. Entre los 200 profesores del departamento de matemáticas hay 150 titulados, 60 del total dedican parte de su tiempo a trabajos de álgebra lineal y 40 de los 150 titulados dedican parte de su tiempo a trabajos de álgebra lineal. Por medio de un diagrama de Venn determina:
   a) ¿Cuántos profesores titulados no dedican su tiempo a trabajos de álgebra lineal?
   b) ¿Cuántos profesores no titulados se dedican a trabajos de álgebra lineal?
   c) ¿Cuántos profesores del total no se dedican a trabajos de álgebra lineal?

Verifica tus resultados en la sección de respuestas.

TEMA 7 Técnicas de conteo

Propósito del tema

Que el estudiante:

- Utilice las técnicas de conteo para describir y obtener información de bases de datos.
- Reconozca la diferencia entre permutación y combinación.
- Aprenda el principio de la suma y multiplicación.
- Haga uso de los elementos básicos de las técnicas de conteo para dar argumentos a sus soluciones.

Competencias disciplinares

1. Construye e interpreta modelos matemáticos mediante la aplicación de procedimientos aritméticos, algebraicos, geométricos y variacionales, para la comprensión y análisis de situaciones reales, hipotéticas o formales.
3. Interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos o situaciones reales.
5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.
7. Elige un enfoque determinista o uno aleatorio para el estudio de un proceso o fenómeno, y argumenta su pertinencia.
8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

Contenidos que aborda el tema

Contenidos conceptuales

- Definición de los elementos básicos de las técnicas de conteo.
- Definición del diagrama de árbol.
- Definición del principio de la suma y de la multiplicación.
- Concepto de permutación y combinación.

Contenidos procedimentales

- Sintetizará evidencias obtenidas mediante la experimentación para producir conclusiones y formulará nuevas preguntas.
- Integrará los conocimientos adquiridos hasta este momento.
- Seguirá procedimientos de manera reflexiva, comprendiendo como cada uno de sus pasos contribuye al alcance de un objetivo.
- Resolverá problemas aplicando la notación factorial.
- Aprenderá a identificar las permutaciones y combinaciones.
- Usará diversas estrategias de resolución de problemas y los aplicara a situaciones reales.

Contenidos actitudinales

- Expresará sus ideas mediante los elementos básicos de las técnicas de conteo.
- Trabajará en equipo y respetará a sus compañeros al resolver problemas.
- Aprenderá a valorar el trabajo de sus compañeros al resolver problemas.
- Estructurará ideas y argumentos de manera clara, coherente y sintética.

Técnicas de conteo

Elementos básicos

Para determinar sin describir directamente el número de resultados posibles de un experimento en particular o el número de elementos de un conjunto en especial, se requieren algunos principios básicos que faciliten el proceso; donde destacan: el diagrama de árbol, el principio fundamental del conteo, la notación factorial, el teorema del binomio, el triángulo de Pascal, las permutaciones, las combinaciones y las particiones.

Diagrama de árbol

Es un gráfico que ilustra cómo enumerar todos los resultados posibles de una serie de experimentos, donde cada experimento puede suceder de un número finito de maneras.

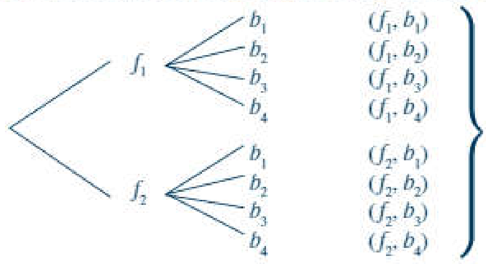
El diagrama de árbol se construye generalmente de izquierda a derecha y el número de ramas en cada punto es el número de resultados posibles del experimento.

EJEMPLOS

1 Si una dama tiene dos faldas $\{f_1, f_2\}$ y cuatro blusas $\{b_1, b_2, b_3, b_4\}$, ¿de cuántas maneras puede seleccionar una falda y luego una blusa?

Solución

Al trazar el diagrama de árbol correspondiente, cada falda se representa con una letra, respectivamente, es decir, por una falda es posible seleccionar cuatro blusas se tiene que:



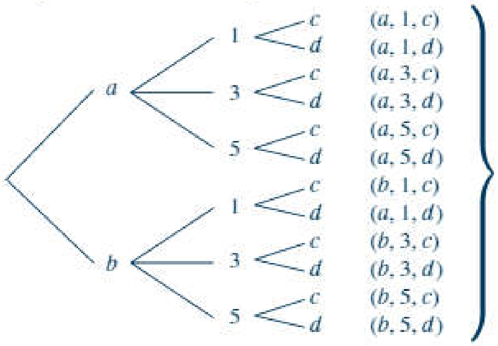
Puede seleccionar una falda y luego una blusa de 8 maneras.

2 Determina el conjunto resultante del producto de $P \times Q \times R$ donde:

$$P = \{a, b\},\ Q = \{1, 3, 5\} \text{ y } R = \{c, d\}.$$

Solución

Para trazar el diagrama de árbol correspondiente se considera que por cada letra del conjunto P es posible elegir un mismo número del conjunto Q y, al mismo tiempo, por cada número perteneciente a Q existe la posibilidad de escoger una letra perteneciente al conjunto R. Esto es:

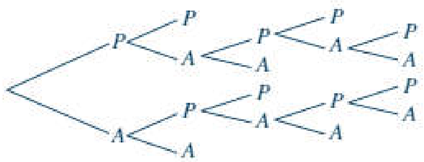


El conjunto resultante consta de 12 elementos.

3 Patricia y Aurora participarán en un torneo de boliche, la primera que gane dos juegos seguidos o que complete tres, ganará el torneo. Representa gráficamente los posibles resultados del torneo.

Solución

Para aplicar la técnica de diagrama de árbol, se utiliza una P en caso de que Patricia gane el juego o una A en el caso que Aurora lo gane; sin embargo, se debe tomar en cuenta que dos juegos consecutivos ganados por la misma persona interrumpen la secuencia de juegos. Entonces se tiene:



Observa que se tienen 10 ramas que indican los resultados posibles del torneo, es decir:

$$\{PP, PAPP, PAPAP, PAPAA, PAA, APP, APAPP, APAPA, APAA, AA\}$$

Si se sigue el recorrido desde el punto de partida del diagrama de árbol hacia las puntas finales de las ramas, se observa quién ganaría cada juego del torneo.

Principio fundamental del conteo

Si una operación puede efectuarse independientemente de n_1 maneras diferentes y continuando el procedimiento, una segunda operación puede efectuarse independientemente de n_2 maneras diferentes y si, después de realizadas, una tercera operación puede efectuarse independientemente de n_3 maneras diferentes y así sucesivamente, para un número finito arbitrario de operaciones, entonces el número total de maneras en las que pueden efectuarse todas las operaciones en el orden indicado es el producto $n_1 n_2 n_3 \ldots n_t$.

EJEMPLOS

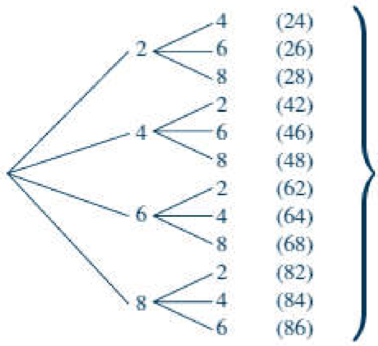
Ejemplos

1 Determina cuántos números de dos cifras distintas pueden formarse con los enteros 2, 4, 6 y 8 sin repetir ninguno.

Solución

Se elige cualquiera de los enteros como la cifra de las decenas; una vez elegido no será posible volver a seleccionar el mismo entero, por lo que quedarán tres enteros entre los que se podrá elegir la cifra de las unidades. De este modo, por cada una de las cuatro alternativas iniciales existen tres más, lo que matemáticamente se puede expresar como: $(4)(3) = 12$.

La ilustración de este ejemplo mediante un diagrama de árbol es:



Se pueden formar 12 números de dos cifras con los cuatro enteros dados.

Se comprueba que se pueden formar 12 número de dos cifras con los cuatro enteros dados.

¿Cuántos números de dos cifras podrán formarse con los cuatro enteros dados, si ahora sí es posible repetir alguno de ellos? Si se aplica el principio fundamental del conteo, se obtiene: $(4)(4) = 16$.

Se pueden formar 16 números de dos cifras con los cuatro enteros dados, permitiendo que se repitan.

2 ¿Cuántas placas distintas de circulación para vehículos pueden construirse usando dos letras distintas del alfabeto seguidas de tres dígitos? Con la condición de que los dígitos sí pueden repetirse pero el primer dígito no puede ser cero.

Solución

Si se considera el alfabeto de 26 letras, la primera letra puede seleccionarse de 26 maneras distintas, mientras que la segunda letra puede escogerse de 25 maneras diferentes. Por otra parte, el primer dígito puede seleccionarse de nueve maneras distintas, ya que no debe usarse el cero, y para cada uno de los otros dos dígitos requeridos existen 10 maneras diferentes para elegirlos. Al aplicar el principio fundamental del conteo se obtiene: $(26)(25)(9)(10)(10) = 585\,000$

Se pueden construir 585 000 placas diferentes para vehículos.

3 Si un hombre tiene 15 camisas, 11 corbatas y 9 pantalones, ¿de cuántas maneras puede elegir una camisa, una corbata y luego un pantalón?

Solución

El proceso contempla tres pasos a efectuarse en forma sucesiva, es decir, para escoger una camisa se tienen 15 maneras distintas, para seleccionar una corbata se tienen 11 maneras diferentes y para elegir un pantalón se tienen 9 maneras distintas. Al aplicar el principio fundamental del conteo, se obtiene:

$$(15)(11)(9) = 1\,485.$$

Puede escoger una camisa, una corbata y luego un pantalón de 1 485 maneras.

Principio de la suma y la multiplicación

Notación factorial

El símbolo ***n*!**, que se lee **factorial de *n***, o simplemente ***n* factorial** representa el producto de los n números enteros consecutivos desde 1 hasta n, inclusive. Matemáticamente se expresa como:

$$n! = n(n-1)\,(n-2)\,(n-3)\ldots 1$$

Para el desarrollo de cantidades factoriales, siempre es necesario conocer las siguientes equivalencias:

$$0! = 1 \qquad 1! = 1$$

EJEMPLOS

Ejemplos

1 Desarrolla el factorial de 4.

Solución

Por el símbolo de factorial se denota que $n = 4$ y al aplicar la expresión matemática, se tiene que:

$$4! = 4\,(3)(2)(1) = 24$$

2 Desarrolla el factorial de 9.

Solución

Por el símbolo de factorial se denota que $n = 9$ y aplicando la expresión matemática, se tiene que:

$$9! = 9\,(8)(7)(6)(5)(4)(3)(2)(1) = 362\,880$$

Suma y resta de cantidades factoriales

El proceso para sumar o restar dos o más números factoriales consiste en seleccionar como factor de la operación al menor de los números factoriales dados. Después, dicho factorial se extrae como factor común, para finalmente realizar la suma de los términos no comunes.

EJEMPLOS

1 Suma $16! + 18! - 19!$.

Solución

Como 16! es el menor de los números factoriales dados, la factorización de la expresión es:

$$16! + 18! - 19! = 16! + 16!(17)(18) - 16!\,(17)(18)(19)$$

Al sacar 16! como factor común, resulta:

$$16![1 + (17)(18) - (17)(18)(19)] = 16!\,(1 + 306 - 5\,814) = -16!(5\,507)$$

El resultado de sumar $16! + 18! - 19!$ es $-16!(5\,507)$.

2 Suma $7! - 5! + 3!$.

Solución

Como 3! es el menor de los números factoriales dados, la factorización de la expresión es:

$$7! - 5! + 3! = 3!(4)(5)(6)(7) - 3!(4)(5) + 3!$$

Al sacar como factor común al 3!, resulta:

$$3![(4)(5)(6)(7) - (4)(5) + 1] = 3!(840 - 20 + 1) = 3!(821)$$

El resultado de sumar $7! - 5! + 3!$ es $3!(821)$.

Multiplicación de cantidades factoriales

La multiplicación se realiza directamente al desarrollar los factoriales.

EJEMPLOS

1 Calcula el producto de $(4!)(3!)$.

Solución

Al desarrollar los factoriales dados se tiene que su producto es:

$$(4!)(3!) = [(4)(3)(2)(1)][(3)(2)(1)] = (24)(6) = 144$$

El producto de $(4!)(3!)$ es 144.

2 Calcula el producto de (5!)(7!).

Solución

Al desarrollar los factoriales dados se tiene que su producto es:

$$(5!)(7!) = [(5)(4)(3)(2)(1)][(7)(6)(5)(4)(3)(2)(1)] = (120)(5\,040) = 604\,800$$

El producto de (5!)(7!) es 604 800.

División de factoriales

El proceso para dividir dos o más factoriales, consiste en localizar y cancelar al menor de los factoriales presente en la división. Entonces el resultado estará conformado por todos los valores no comunes en los términos presentes en la división.

EJEMPLOS

1 Determina el cociente de $\frac{19!}{17!}$.

Solución

Como el denominador 17! es el menor de los números factoriales dados, es necesario reorganizar los multiplicandos del numerador y simplificar la expresión, mediante la cancelación del factorial común, de lo que resulta:

$$\frac{19!}{17!} = \frac{(19)(18)\cancel{17!}}{\cancel{17!}} = 342 \qquad \text{El cociente de } \frac{19!}{17!} \text{ es } 342.$$

2 Determina el cociente de $\frac{3!}{7!}$.

Solución

Como el numerador 3! es el menor de los números factoriales dados, es necesario reorganizar el denominador y simplificar la expresión, de lo que resulta:

$$\frac{3!}{7!} = \frac{\cancel{3!}}{(7)(6)(5)(4)(\cancel{3!})} = \frac{1}{840} \approx 0.0012 \qquad \text{El cociente de } \frac{3!}{7!} \text{ es } \frac{1}{840} \approx 0.0012.$$

3 Determina el cociente de $\frac{9!+6!-4!}{4!+7!}$.

Solución

Como el 4! es el menor de los números factoriales dados, la factorización de la expresión es:

$$\frac{9!+6!-4!}{4!+7!} = \frac{(9)(8)(7)(6)(5)(4!)+(6)(5)(4!)-4!}{4!+(7)(6)(5)(4!)}$$

$$\frac{\cancel{4!}[(9)(8)(7)(6)(5)+(6)(5)-1]}{\cancel{4!}[1+(7)(6)(5)]} = \frac{15120+30-1}{1+210} \approx 71.796$$

El cociente de $\frac{9!+6!-4!}{4!+7!}$ es aproximadamente 71.796.

Teorema del binomio

El teorema del binomio es muy importante en el álgebra y se basa en la multiplicación de polinomios y en la ley distributiva.

Ejemplo

$(x + y)^1 = x + y$

$(x + y)^2 = x^2 + 2xy + y^2$

$(x + y)^3 = (x + y)\,(x + y)^2 = (x + y)\,(x^2 + 2xy + y^2) = x^3 + 3x^2y + 3xy^2 + y^3$

Al aplicar los mismos procedimientos se determina que:

$(x + y)^4 = x^4 + 4x^3\,y + 6x^2y^2 + 4xy^3 + y^4$

$(x + y)^5 = x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5$

$(x + y)^6 = x^6 + 6x^5y + 15x^4y^2 + 20x^3y^3 + 15x^2y^4 + 6xy^5 + y^6$

Las igualdades anteriores son casos especiales de la fórmula general mostrada a continuación:

$$(x+y)^n = x^n + nx^{n-1}y + \frac{n(n-1)}{(2)(1)}x^{n-2}y^2 + \ldots + \frac{n!}{r!(n-r)!}x^{n-1}y + \ldots + y^n$$

Si se desarrolla $(x + y)^n$, donde n es un entero positivo arbitrario, a partir de la fórmula general simplificada, se obtiene el **teorema del binomio**:

$$(x + y)^n = C(n, 0)x^n + C(n,1)x^{n-1}y + C(n,2)x^{n-2}y^2 + C(n,3)x^{n-3}y^3 + \ldots + C(n,r)\,x^{n-r}y^r + \ldots + C(n,n)y^n$$

Donde $C(n,r)$ son los coeficientes del teorema que se obtienen de la tabla siguiente; los valores de r están anotados horizontalmente y los de n verticalmente. Esta tabla se puede desarrollar hasta n y r infinito.

EJEMPLOS

Ejemplos

1 Determina el coeficiente $C(5,3)$.

Solución

Dado que $\boldsymbol{n = 5}$ y $\boldsymbol{r = 3}$, el punto de intersección de ambas líneas indica el coeficiente buscado, es decir: $\boldsymbol{C = 10}$.

A causa de este teorema, $C(n,r)$, el número de combinaciones de n objetos tomados de r, recibe el nombre de **coeficiente binomial**.

Del desarrollo de $(x + y)^n$ al numerar los términos de izquierda a derecha, pueden observarse las siguientes propiedades: (Página 296 CC)

1. El desarrollo de $(x + y)^n$ tiene $(n + 1)$ términos.
2. El primer término del desarrollo es x^n y el último es y^n.
3. Los exponentes de x disminuyen una unidad de término a término, los de y aumentan una unidad de término a término.
4. En cada término, la suma de los exponentes de x y y es igual a n.
5. Si en cualquier término se multiplica su coeficiente por el exponente de x en ese término y se divide entre el número de orden del término, el resultado es el coeficiente del siguiente término.

2 Desarrolla el binomio $(a + b)^8$.

Solución

Con el teorema del binomio y las propiedades respectivas, se tiene que: $\boldsymbol{a}$ es el primer término y b el segundo términos del binomio y dado que $n = 8$, el desarrollo de $(a + b)^8$ tiene $(n + 1)$ términos, es decir, $(8 + 1) = 9$ términos.

El primer término del desarrollo es a^n, es decir: a^8; el último término es b^n, es decir: b^8.

Los coeficientes se pueden obtener de acuerdo con la tabla dada en el renglón de $n = 8$, es decir: 1, 8, 28, 56, 70, 56, 28, 8, 1.

Los exponentes de a disminuyen una unidad de término a término, los de b aumentan una unidad de término a término, es decir:

$$a^nb^0,\ a^{n-1}b^{0+1},\ a^{n-2}b^{0+2},\ a^{n-3}b^{0+3},\ a^{n-4}b^{0+4},\ a^{n-5}b^{0+5},\ a^{n-6}b^{0+6},\ a^{n-7}b^{0+7},\ a^{n-8}b^{0+8}\ a^8,$$

De lo que resulta: $a^8, a^7b, a^6b^2, a^5b^3, a^4b^4, a^3b^5, a^2b^6, ab^7, b^8$

Al ordenar todos los pasos anteriores se obtiene:

$$(a + b)^8 = a^8 + 8a^7b + 28a^6b^2 + 56a^5b^3 + 70a^4b^4 + 56a^3b^5 + 28a^2b^6 + 8ab^7 + b^8$$

También se puede comprobar la propiedad que establece que en cada término, la suma de los exponentes de $\boldsymbol{a}$ y $\boldsymbol{b}$ es igual a $\boldsymbol{n}$.

La última propiedad se puede demostrar seleccionando el coeficiente del cuarto término (56) y multiplicándolo por el exponente de $\boldsymbol{a}$ en dicho término (5), después el producto se divide entre el número de orden del término seleccionado (4) y lo que se obtiene es el coeficiente del siguiente término (quinto), es decir:

$$\frac{(56)(5)}{4} = 70 \qquad \text{El coeficiente del quinto término es 70.}$$

3 Desarrolla el binomio $(a^2 - 3b^3)^6$.

Solución

Con la fórmula binominal y como a^2 es el primer término del binomio, $-3b^3$ el segundo término y dado que $n = 6$, el desarrollo de $(a^2 - 3b^3)^6$ es:

$$(a^2)^6 + 6(a^2)^{6-1}(-3b^3) + \frac{6(6-1)}{(2)(1)}(a^2)^{6-2}(-3b^3)^2 + \frac{6(6-1)(6-2)}{(3)(2)(1)}(a^2)^{6-3}(-3b^3)^3$$

$$+\frac{6(6-1)(6-2)(6-3)}{(4)(3)(2)(1)}(a^2)^{6-4}(-3b^3)^4 + \frac{6(6-1)(6-2)(6-3)(6-4)}{(5)(4)(3)(2)(1)}(a^2)^{6-5}(-3b^3)^5 + (-3b^3)^6$$

$$(a^2 - 3b^3)^6 = a^{12} + 6(a^2)^5(-3b^3) + \frac{6(5)}{2}(a^2)^4(9b^6) + \frac{6(5)(4)}{6}(a^2)^3(-27b^9)$$

$$+\frac{6(5)(4)(3)}{24}(a^2)^2(81b^{12}) + \frac{6(5)(4)(3)(2)}{120}(a^2)(-243b^{15}) + (729b^{18})$$

$$(a^2 - 3b^3)^6 = a^{12} - 18a^{10}b^3 + 132a^8b^6 - 540a^6b^9 + 1215a^4b^{12} - 1458a^2b^{15} + 729b^{18}$$

Teorema del binomio para exponentes fraccionarios y negativos

Si se aplica el teorema del binomio a $(1 + x)^n$, donde n es cero o un entero positivo, su desarrollo tiene $(n + 1)$ términos. Pero si n fuera un número real cualquiera diferente de un entero positivo, el desarrollo de $(1 + x)^n$ no tiene fin, es decir:

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{(2)(1)}x^2 + \frac{n(n-1)(n-2)}{(3)(2)(1)}x^3 + \frac{n(n-1)(n-2)(n-3)}{(4)(3)(2)(1)}x^4 + \ldots + \frac{n(n-1)(n-2)\ldots(n-r+1)}{r!}x^y + \ldots$$

En un desarrollo infinito se pueden escribir tantos términos como se quieran, para todo n que no sea un entero positivo si $|x| < 1$.

EJEMPLOS

1 Determina el valor aproximado de $(1.04)^{-4}$ con cuatro cifras significativas.

Solución

Como $(1.04)^{-4} = (1 + 0.04)^{-4}$, al aplicar el teorema del binomio $(1 + x)^n$ resulta:

$$1 + (-4)(0.04) + \frac{(-4)(-4-1)}{(2)(1)}(0.04)^2 + \frac{(-4)(-4-1)(-4-2)}{(3)(2)(1)}(0.04)^3 + \frac{(-4)(-4-1)(-4-2)(-4-3)}{(4)(3)(2)(1)}(0.04)^4 + \ldots$$

$$(1 + 0.04)^{-4} = 1 - 0.16 + 0.016 - 0.00128 + 0.0000896 + \ldots$$

$$(1.04)^{-4} = (1 + 0.04)^{-4} \approx 0.8548.$$

2 Determina el valor aproximado de $\sqrt{24}$ con tres cifras significativas.

Solución

Como $\sqrt{24} = (25-1)^{\frac{1}{2}} = (1-0.02)^{-\frac{1}{2}}$, al aplicar el teorema del binomio $(1 + x)^n$, resulta:

$$(25)^{\frac{1}{2}} + \frac{1}{2}(25)^{\frac{1}{2}-1}(-1) + \frac{\left(\frac{1}{2}\right)\left(\frac{1}{2}-1\right)}{(2)(1)}(25)^{\frac{1}{2}-2}(-1)^2 + \frac{\left(\frac{1}{2}\right)\left(\frac{1}{2}-1\right)\left(\frac{1}{2}-1\right)}{(3)(2)(1)}(25)^{\frac{1}{2}-3}(-1)^3 + ..$$

$$(25-1)^{\frac{1}{2}} = 5 - \frac{1}{2}(25)^{\frac{1}{2}} + \frac{1}{8}(25)^{-\frac{3}{2}} - \frac{1}{16}(25)^{-\frac{5}{2}} + \ldots$$

$$(25 - 1)^{-\frac{1}{2}} = 5 - 0.1 + 0.001 - 0.00002 + \ldots$$

$$\sqrt{24} = (25 - 1)^{-\frac{1}{2}} \approx 4.90098 \approx 4.900$$

3 Determina el valor aproximado de $(0.98)^{-\frac{1}{2}}$ con tres cifras significativas.

Solución

Como $(0.98)^{-\frac{1}{2}} = (1 - 0.02)^{-\frac{1}{2}}$, al aplicar el teorema del binomio $(1 + x)^n$, resulta:

$$1+\left(-\frac{1}{2}\right)(-0.02)+\frac{\left(\frac{-1}{2}\right)\left(\frac{-1}{2}-1\right)}{(2)(1)}(-0.02)^2+\frac{\left(\frac{-1}{2}\right)\left(\frac{-1}{2}-1\right)\left(\frac{-1}{2}-2\right)}{(2)(1)}(-0.02)^3+\ldots$$

$$(1-0.02)^{-\frac{1}{2}}=1+0.01+\frac{3}{8}(0.0004)-\frac{15}{48}(-0.000008)+\ldots$$

$$(1-0.02)^{-\frac{1}{2}}=1+0.01+0.00015+0.0000025+\ldots$$

$$(0.98)^{-\frac{1}{2}}=(1-0.02)^{-\frac{1}{2}}\approx 1.0101525\approx 1.010$$

Triángulo de Pascal

Los coeficientes $C(n,0)$, $C(n,1)$, $C(n,2)$, …, $C(n,n)$ de la fórmula del teorema del binomio, reciben el nombre de **coeficientes binomiales** para el desarrollo binomial de $(x + y)^n$. Dichos coeficientes pueden disponerse en la forma siguiente para los valores sucesivos de n:

$$\begin{array}{lc}
n=1 & C(1,0)\ C(1,1) \\
n=2 & C(2,0)\ C(2,1)\ C(2,2) \\
n=3 & C(3,0)\ C(3,1)\ C(3,2)\ C(3,3) \\
n=4 & C(4,0)\ C(4,1)\ C(4,2)\ C(4,3)\ C(4,4) \\
n=5 & C(5,0)\ C(5,1)\ C(5,2)\ C(5,3)\ C(5,4)\ C(5,5) \\
n=6 & C(6,0)\ C(6,1)\ C(6,2)\ C(6,3)\ C(6,4)\ C(6,5)\ C(6,6) \\
n=7 & C(7,0)\ C(7,1)\ C(7,2)\ C(7,3)C(7,4)\ C(7,5)C(7,6)\ C(7,7) \\
n=8 & C(8,0)\ C(8,1)\ C(8,2)\ C(8,3)\ C(8,4)\ C(8,5)\ C(8,6)\ C(8,7)\ C(8,8)
\end{array}$$

y así sucesivamente.

Si se fija el valor de cada una de las combinaciones indicadas en este arreglo se obtiene la siguiente ordenación triangular, conocida como **triángulo de Pascal**.

$$\begin{array}{lc}
n=1 & 1\quad 1 \\
n=2 & 1\quad 2\quad 1 \\
n=3 & 1\quad 3\quad 3\quad 1 \\
n=4 & 1\quad 4\quad 6\quad 4\quad 1 \\
n=5 & 1\quad 5\quad 10\quad 10\quad 5\quad 1 \\
n=6 & 1\quad 6\quad 15\quad 20\quad 15\quad 6\quad 1 \\
n=7 & 1\quad 7\quad 21\quad 35\quad 35\quad 21\quad 7\quad 1 \\
n=8 & 1\quad 8\quad 28\quad 56\quad 70\quad 56\quad 28\quad 8\quad 1
\end{array}$$

y así sucesivamente

Del triángulo de Pascal se deducen las siguientes propiedades.

1. El primero y último número de cada fila es 1.

2. Cada uno de los otros números de la ordenación se obtienen sumando los dos números que aparecen directamente encima de él, es decir:

$$\begin{array}{ll} n=5 & 1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\ n=6 & 1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1 \\ n=7 & 1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1 \end{array}$$

El 15 se obtiene al sumar $(5 + 10)$, el 21 se obtiene al sumar $(15 + 6)$, etcétera.

Cuando n es un exponente entero positivo, los coeficientes binomiales pueden extenderse indefinidamente, ya que el **triángulo de Pascal** es simétrico.

EJEMPLO

Ejemplo

Se lanzan al aire cinco monedas normales y se tiene el interés en saber cuántos soles aparecen. Puede no aparecer sol alguno, puede aparecer un sol, dos soles, tres soles, cuatro soles o cinco soles. ¿De cuántas maneras pueden ocurrir cada uno de estos resultados?

Solución

El número de maneras de obtener cero soles es $C(5,0)$, un sol es $C(5,1)$, dos soles es $C(5,2)$, tres soles es $C(5,3)$, cuatro soles es $C(5,4)$ y cinco soles $C(5,5)$, es decir, lo dan las entradas sucesivas del quinto renglón del triángulo de Pascal. Además, el número de resultados del experimento es la suma de las entradas de dicho renglón, esto es:

$$1 + 5 + 10 + 10 + 5 + 1 = 32$$

Permutación y combinación

Permutación

Una permutación es una disposición u orden donde se pueden colocar los objetos pertenecientes a un grupo determinado. Cualquier forma de ordenar una colección de objetos recibe el nombre de permutación.

Ejemplo

Si los objetos dados son tres, b, g y o, y se agrupan tomando los tres a la vez, pueden ser ordenados en las siguientes formas:

$$\left.\begin{array}{l} bgo,\ bog,\ gbo \\ gob,\ obg,\ ogb \end{array}\right\} \text{ordenados de izquierda a derecha en seis formas diferentes.}$$

Por lo tanto, una **permutación** es cada una de las posibles maneras en que pueden ser ordenados los elementos de un conjunto finito.

Permutaciones de n elementos tomados todos a la vez

Los símbolos $\boldsymbol{nPn}$, $\boldsymbol{P_n^n}$ o $\boldsymbol{P(n,n)}$ representan el número total de permutaciones posibles de n objetos distintos, tomados de $\boldsymbol{n}$ en $\boldsymbol{n}$, matemáticamente se representa como:

$$nPn = n! \quad \text{o} \quad P_n^n = n! \quad \text{o} \quad P(n, n) = n!$$

¿De cuántas formas diferentes se pueden colocar n elementos distintos de un conjunto, tomando todos a la vez y colocándolos en fila?

Para dar respuesta, se considera que se tienen n espacios vacíos y cada espacio deberá ser ocupado con objetos. El primer espacio puede estar ocupado de n maneras, el segundo de $(n-1)$ formas distintas, el tercero de $(n-2)$ maneras diferentes y así sucesivamente, hasta llegar al último espacio que será ocupado por el único objeto que se ha dejado, es decir, en una sola forma.

Por lo tanto, el número total de permutaciones posibles de los n elementos dados es:

$$nPn = n(n-1)(n-2)(n-3)\ldots 1 = n!$$

EJEMPLOS

Ejemplos

1 ¿De cuántas maneras distintas pueden ordenarse 6 personas en una fila de 6 asientos?

Solución

El primer asiento puede ser ocupado por cualquiera de las 6 personas, es decir, hay 6 formas de ocupar el primer asiento. Cuando lo anterior haya sucedido, habrá 5 formas de ocupar el segundo asiento. Después habrá 4 formas de ocupar el tercer asiento, 3 formas de ocupar el cuarto asiento, 2 maneras de ocupar el quinto asiento y solamente 1 de ocupar el último asiento.

El número de ordenaciones posibles de las 6 personas en una fila de 6 asientos es:

$$(6)(5)(4)(3)(2)(1) = 6! = 720 \text{ maneras}$$

También puede resolverse de la siguiente manera:

Datos	**Fórmula**	**Sustitución**
$n = 6$ personas	$nPn = n!$	${}_6P_6 = 6!$
$n = 6$ asientos		${}_6P_6 = 720$

Existen 720 maneras distintas de ordenar 6 personas en una fila de 6 asientos.

2 ¿De cuántas maneras pueden arreglarse 5 libros en un estante, si es posible cualquier ordenación?

Solución

Ya que es posible acomodar los 5 libros en el estante sin importar su ordenación, se tiene que:

Datos	**Fórmula**	**Sustitución**
$n = 5$ libros	$nPn = n!$	${}_5P_5 = 5!$
		${}_5P_5 = 120$

Existen 120 formas posibles de ordenar 5 libros en un estante.

Permutaciones de n diferentes elementos tomados en grupos de r a un tiempo

Los símbolos nPr, P_r^n o $P(n,r)$ representa el número total de permutaciones de n objetos distintos tomados de r en r, cuando $r \leq n$, o bien, el número total de permutaciones de n objetos colocados en r lugares, donde $r \leq n$.

Matemáticamente se representa como:

$$nPr = \frac{n!}{(n-r)!}$$

Demostración

Cuando una permutación de n objetos diferentes tomados de r en r, se dice, que es una ordenación de r objetos entre los n dados. Así, es necesario llenar r lugares con algunos de los n objetos disponibles. Para llenar el primer lugar, se tienen n posibilidades, para llenar el segundo se tienen $n-1$, el tercer lugar sólo se puede llenar con los $n-2$ objetos restantes y así sucesivamente, hasta ocupar los r lugares. Entonces, si $r < n$ se tiene:

$$nPr = n(n-1)(n-2)(n-3)\ldots(n-r+1)$$

Si la ecuación anterior se multiplica y divide entre $(n-r)!$, se obtiene:

$$nPr = \frac{n!}{(n-r)!}$$

Si $r = n$, se obtiene la ecuación que representa el número total de permutaciones posibles de n objetos, tomados de n en n, es decir:

$$nPn = n!$$

EJEMPLOS

Ejemplos

1 Determina el número de permutaciones de los enteros 1, 2, 3, 4, 5, 6 tomados de tres en tres.

Solución

Puesto que se desea determinar el número de disposiciones de 6 objetos tomados de tres en tres, es decir:

Datos

$n = 6$ enteros
$r = 3$ enteros

Fórmula

$$nPr = \frac{n!}{(n-r)!}$$

Sustitución

$${}_6P_3 = \frac{6!}{(6-3)!} = \frac{6!}{3!}$$

$${}_6P_3 = 120$$

El número de permutaciones de los 6 enteros dados, tomados de tres en tres, son 120.

2 Cinco personas suben a un autobús en el que hay doce asientos desocupados, ¿de cuántas maneras pueden sentarse?

Solución

Esto corresponde al número de disposiciones de los doce asientos tomados de cinco en cinco, es decir:

Datos

$n = 12$ asientos
$r = 5$ personas

Fórmula

$$nPr = \frac{n!}{(n-r)!}$$

Sustitución

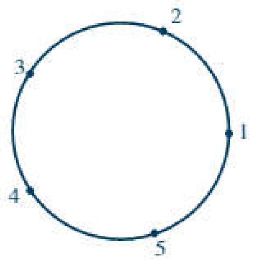
$${}_{12}P_5 = \frac{12!}{(12-5)!} = \frac{12!}{7!}$$

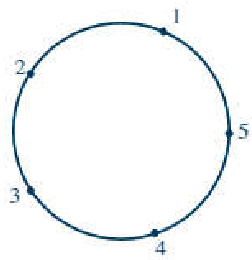
$${}_{12}P_5 = 95\,040$$

El número de permutaciones de los doce asientos tomados de cinco en cinco, son 95 040.

Permutaciones bajo condiciones diversas

Se realiza la siguiente pregunta, ¿de cuántas maneras diferentes pueden acomodarse cinco personas (1, 2, 3, 4, 5) alrededor de una mesa circular, si se consideran equivalentes los arreglos que se obtienen a partir de la rotación de la mesa? Los dos arreglos, cuya representación gráfica se muestran a continuación, se consideran iguales.





Para determinar el número de arreglos posibles bajo la condición indicada, se puede definir como fija la posición de una persona en la mesa, entonces, la silla que queda a su derecha puede ocuparse de cuatro maneras, la siguiente silla a su derecha del lugar fijo puede ocuparse de tres maneras, la tercera silla a la derecha del lugar fijo puede ocuparse de dos maneras y la última silla a la derecha del lugar fijo puede ocuparse de una sola manera. Si se aplica el principio fundamental del conteo, después de que se sienta la primera persona, las cuatro sillas restantes pueden ocuparse en: $(4)(3)(2)(1) = 4! = 24$ maneras.

El número de maneras en las que pueden acomodarse cinco personas alrededor de una mesa circular es 24.

Permutación circular

Se define como el arreglo posible de n objetos alrededor de un círculo o cualquier otra curva simple cerrada, donde uno de ellos mantiene una posición fija.

El número de permutaciones circulares P_c de n objetos se determina por la ecuación:

$$P_c = (n - 1)!$$

EJEMPLO

Ejemplo 1 ¿De cuántas maneras se puede acomodar una reunión de 9 personas alrededor de una mesa redonda?

Solución

Una persona puede sentarse en cualquier posición fija de la mesa; las otras ocho personas pueden arreglarse de $(8)(7)(6)(5)(4)(3)(2)(1) = 8!$ maneras alrededor de la misma, es decir:

Datos	**Fórmula**	**Sustitución**
$n = 9$	$P_c = (n - 1)!$	$P_c = (9 - 1)! = 8! = 40\,320$

El número de maneras en las que pueden acomodarse nueve personas alrededor de una mesa redonda es de 40 320.

2 ¿De cuántas maneras pueden sentarse siete personas alrededor de una mesa, si dos personas insisten en sentarse una al lado de la otra?

Solución

Si se consideran a las dos personas que insisten en sentarse una al lado de la otra como una sola, entonces habrá seis personas para sentarse alrededor de la mesa y lo pueden hacer de $(5)(4)(3)(2)(1) = 5!$ maneras. Pero las dos personas consideradas como una sola pueden ordenarse entre sí de $(2)(1) = 2!$ maneras. En total existen $(5!)(2!) = 240$ maneras.

El número de maneras en las que pueden sentarse siete personas alrededor de una mesa, si dos personas insisten en sentarse una al lado de la otra, es 240.

3 ¿De cuántas formas pueden sentarse 12 personas en torno a una mesa redonda si sólo se consideran sus posiciones relativas, pero de tal forma que 3 personas en particular nunca se sienten juntas?

Solución

Si las 3 personas que no deben ir juntas se consideran como 1 sola, entonces existen 11 personas se sentarán alrededor de la mesa y que lo pueden hacer de 10! formas.

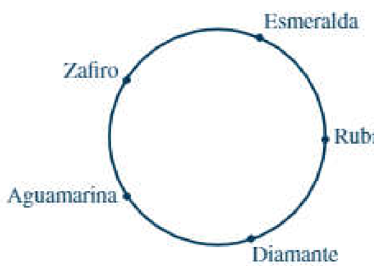
Si las 3 personas consideradas como 1 sola pueden acomodarse entre sí de 3! maneras. El número de arreglos de 11 personas sentadas alrededor de una mesa circular con 3 de ellas sentadas juntas en $(10!)(3!) = 21\,772\,800$ maneras.

Si sólo se consideran sus posiciones relativas, las 12 personas pueden sentarse al rededor de la mesa redonda de $11! = 39\,916\,800$ maneras.

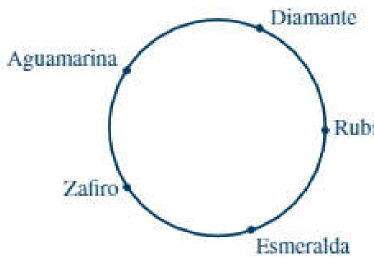
El número total de maneras en que 12 personas pueden sentarse alrededor de una mesa circular, de modo que tres de ellas no estén sentadas juntas es $39\,916\,800 - 21\,772\,800 = 18\,144\,000$.

¿De cuántas maneras pueden arreglarse cinco piedras preciosas distintas (rubí, esmeralda, zafiro, aguamarina y diamante), sobre la forma de un brazalete?

Si se supone que las piedras preciosas se disponen de la siguiente manera:



Si se levanta el brazalete y se voltea, aparecerá el arreglo de las piedras preciosas la siguiente manera:



Se hace notar que las dos representaciones gráficas son iguales y, en consecuencia, tratándose de piedras preciosas distribuidas sobre un brazalete que puede levantarse y voltearse, se concluye que existen la mitad de arreglos posibles en relación con el número de permutaciones circulares.

Si se tienen cinco piedras preciosas distribuidas sobre un brazalete, existen $\left(\frac{1}{2}\right)\left[(5-1)!\right]=\left(\frac{1}{2}\right)(4!)=12$ ordenaciones distintas.

Permutaciones sobre un anillo

Se define como el número de arreglos diferentes de n objetos distintos sobre un anillo (aro, collar, brazalete, argolla, etc.) que pueda levantarse y voltearse. Matemáticamente se expresa como:

$$P_A=\frac{1}{2}\left[(n-1)!\right]$$

EJEMPLO

Ejemplo 1 ¿En cuántas posiciones relativas pueden disponerse siete llaves sobre una argolla?

Solución

El tipo de problema presentado se refiere al número de arreglos distintos de n objetos diferentes sobre una argolla que pueda levantarse y voltearse.

Datos	Fórmula	Sustitución
$n=7$	$P_A=\frac{1}{2}\left[(n-1)!\right]$	$P_A=\frac{1}{2}[(7-1)!]=\left(\frac{1}{2}\right)(6!)$ $P_A=360$

El número de posiciones relativas en que pueden disponerse siete llaves sobre una argolla son 360.

Permutaciones de objetos que no sean todos diferentes (permutaciones con repetición)

El número de permutaciones diferentes de n objetos tomados todos a la vez; de los cuales, hay n_1 iguales entre sí, otros n_2 iguales entre sí y así sucesivamente, matemáticamente se determina como:

$$P\frac{n!}{n_1!n_2!n_3!\cdots n_r!}$$

EJEMPLOS

Ejemplos

1 ¿De cuántas maneras pueden disponerse en una fila cinco fichas rojas idénticas entre sí, seis fichas blancas también idénticas entre sí y cuatro fichas azules iguales entre sí?

Solución

Se tienen en total 15 fichas, de las cuales, 5 son rojas e idénticas entre sí, 6 son blancas e idénticas entre sí y 4 azules e idénticas entre sí. Al aplicar la fórmula para determinar las permutaciones con repetición, se tiene:

Datos

$n = 15$ fichas

$n_1 = 5$ fichas rojas

$n_2 = 6$ fichas blancas

$n_3 = 4$ fichas azules

Fórmula

$$P = \frac{n!}{n_1!n_2!n_3!\cdots n_r!}$$

Sustitución

$$P = \frac{15!}{5!6!4!}$$

$$P = 630\,630$$

Existen 630 630 maneras para disponer en una fila cinco fichas rojas idénticas entre sí, seis fichas blancas idénticas entre sí y cuatro fichas azules idénticas entre sí.

2 ¿Cuántas permutaciones pueden hacerse utilizando todas las letras de la palabra Chihuahua?

Solución

En total, la palabra Chihuahua tiene 9 letras, de las cuales, 3 son H, 2 son U, 2 son A, 1 es C y 1 es I. Al aplicar la ecuación para determinar las permutaciones con repetición, se tiene:

Datos

$n =$ 9 letras

$n_1 =$ 3 letras H

$n_2 =$ 2 letras U

$n_3 =$ 2 letras A

$n_4 =$ 1 letra C

$n_5 =$ 1 letra I

Fórmula

$$P = \frac{n!}{n_1!n_2!n_3!\cdots n_r!}$$

Sustitución

$$P = \frac{9!}{3!2!2!1!1!}$$

$$P = 15120$$

Existen 15 120 permutaciones que pueden hacerse con las letras de la palabra Chihuahua.

Pruebas ordenadas

Por lo general, los problemas de análisis combinatorio, especialmente los de probabilidad, se relacionan con la selección de un naipe (carta) de una baraja, de una persona de una población, de una ficha tomada de una urna que contiene ***n*** fichas, etcétera.

La acción de seleccionar un naipe de una baraja tras otro ***r*** veces está definida como una **prueba ordenada** (P_o) de tamaño ***r***. Las pruebas ordenadas pueden presentarse de dos maneras: con sustitución y sin sustitución.

Pruebas con sustitución. En este caso se toman r naipes de la baraja con la condición de que cada naipe seleccionado se regresa a la baraja antes de tomar el siguiente. Como en total hay n naipes diferentes, entonces hay n maneras diferentes para escoger cada naipe; de acuerdo con el principio fundamental del conteo, se tiene:

$$P_o = \overbrace{(n)(n)(n)\ldots(n)}^{r \text{ veces}} = n^r$$

Pruebas ordenadas diferentes de tamaño ***r*** con sustitución.

EJEMPLO

Si una urna contiene 7 bolas, determina el número de pruebas ordenadas de tamaño 3 con sustitución.

Solución

Como cada bola de la prueba ordenada puede seleccionarse de 7 maneras y su tamaño es 3, se tiene:

Datos	**Fórmula**	**Sustitución**
$n = 7$ bolas	$P_o = n^r$	$P_o = (7)^3 = 343$
$r =$ tamaño 3		

El número de pruebas ordenadas con sustitución es 343.

Pruebas sin sustitución. En este caso se toman r naipes, pero el naipe seleccionado no se regresa a la baraja antes de seleccionar el siguiente; entonces no hay repeticiones en la prueba ordenada. Una prueba ordenada de tamaño r sin sustitución es sencillamente una permutación r de n objetos, es decir:

$$P_n = n(n-1)(n-2)(n-3)\ldots(n-r+1) = \frac{n!}{(n-r)!} = P_r^n$$

Pruebas ordenadas diferentes de tamaño r sin sustitución tomadas de un grupo n objetos.

EJEMPLO

¿De cuántas maneras se puede seleccionar un comité de tres personas de un grupo de 35 personas sin sustitución?

Solución

Como la condición es de que no hay sustitución, entonces la primera persona del comité puede seleccionarse de 35 maneras diferentes, la segunda persona del comité puede seleccionarse de 34 formas distintas, la tercera y última persona puede seleccionarse de 33 maneras diferentes, es decir:

Datos	**Fórmula**	**Sustitución**
$n = 35$ personas $r = 3$ personas	$P_n = \dfrac{n!}{(n-r)!}$	$P = \dfrac{35!}{(35-3)!} = \dfrac{35!}{32!}$ $P_c = 39\,270$

Un comité de tres personas de un grupo de 35 personas, se puede seleccionar, de 39 270 maneras sin sustitución.

Combinaciones

Si se desea formar un comité de dos personas que tiene que seleccionarse de un grupo de tres personas, ¿cuántos comités pueden formarse?

Si $\{a, b, c\}$ representa al grupo de tres personas, entre las que se debe seleccionar un comité compuesto por dos personas únicamente; al aplicar la ecuación de permutaciones para n diferentes objetos tomados en grupos de r a un tiempo, tenemos:

$$P_2^3 = \frac{3!}{(3-2)!} = \frac{3!}{1!} = 6 \text{ comités}$$

Es decir: $\{ab, ba, ac, ca, bc, cb\}$. Por lo anterior, se puede pensar que existen seis comités de tamaño dos.

Cabe mencionar que ***ab*** y ***ba*** son permutaciones diferentes aunque contengan las mismas letras, porque **el orden es de primordial importancia para una permutación**. Sin embargo, *ab* y *ba* son indistingibles, es decir, el orden en el que se anoten los miembros del comité no presentan influencia alguna en su integración.

Por lo tanto, se concluye que puede haber solamente tres comités, cada uno de dos personas, es decir: $\{ab, ac, be\}$. Cada uno de los comités escritos se denomina **combinación** y **el orden de sus elementos carece de importancia**.

Una combinación es un conjunto no ordenado de objetos distintos.

Combinación *r* de *n* objetos

Una combinación de n objetos diferentes tomados de r es una selección de r de los n objetos sin atender a la ordenación de los mismos.

El número de **combinaciones** de n objetos tomados de r en r, donde $r < n$ se simboliza por C_r^n, nCr, $C(n,r)$ o $\left(\frac{n}{r}\right)$; matemáticamente se expresa como:

$$\left(\frac{n}{r}\right) = C_r^n = \frac{n!}{r!(n-r)!} = \frac{P_r^n}{r!}$$

EJEMPLOS

Ejemplos

1 En un grupo de 15 hombres y 10 mujeres, ¿de cuántas maneras puede formarse un comité de 3 hombres y 2 mujeres?

Solución

Los hombres pueden seleccionarse de C_3^{15} maneras y las mujeres pueden escogerse de C_2^{10} formas, entonces el comité puede seleccionarse de $(C_3^{15})\,(C_2^{10})$ maneras, es decir:

Datos

$n_1 = 15$ hombres
$r_1 = 3$ hombres
$n_2 = 10$ mujeres
$r_2 = 2$ mujeres

Fórmula

$$C_r^n = \frac{n!}{r!(n-r)!}$$

Sustitución

$$C_3^{15} = \frac{15!}{3!(15-3)!} = \frac{15!}{3!12!} = 455$$

$$C_2^{10} = \frac{10!}{2!(10-2)!} = \frac{10!}{2!8!} = 45$$

$$(C_3^{15})(C_2^{10}) = (455)(45) = 20\,475$$

El comité puede seleccionarse de 20 475 maneras.

2 ¿De cuántas maneras puede obtener una persona una mano de Bridge (13 cartas) consistente sólo de ases o naipes con figura?

Solución

Una baraja normal consta de 52 cartas y para obtener la mano de Bridge deseada se tienen 4 ases y 12 naipes con figura, que hacen un total de 16 cartas. Al aplicar la ecuación de combinaciones, se tiene:

Datos
$n = 16$ naipes
$r = 13$ naipes (Bridge)

Fórmula
$$C_r^n = \frac{n!}{r!(n-r)!}$$

Sustitución
$$C_{13}^{16} = \frac{16!}{13!(16-13)!} = \frac{16!}{13!2!} = 560$$

Una mano de Bridge que consiste sólo de ases o naipes con figuras, puede obtenerse de 560 maneras.

3 En un examen el alumno debe contestar 8 de un total de 12 preguntas y debe incluir exactamente 5 de entre las 6 primeras, ¿de cuántas maneras pueden resolver el examen?

Solución

Puesto que debe contestar 5 de las 6 primeras preguntas, lo puede hacer de C_5^6 maneras y las 3 restantes que debe contestar se deben elegir de entre las 6 últimas, esto puede hacerse de C_3^6 maneras. Entonces el examen puede resolverse de (C_5^6) (C_3^6) maneras.

Datos
$n_1 = 6$ preguntas
$r_1 = 5$ respuestas
$n_2 = 6$ preguntas
$r_2 = 3$ respuestas

Fórmula
$$C_r^n = \frac{n!}{r!(n-r)!}$$

Sustitución
$$C_5^6 = \frac{6!}{5!(6-5)!} = \frac{6!}{5!1!} = 6$$
$$C_3^6 = \frac{6!}{3!(6-3)!} = \frac{6!}{3!3!} = 20$$
$$(C_5^6)(C_3^6) = (6)(20) = 120$$

El alumno puede resolver el examen de 120 maneras.

4 ¿De cuántas maneras puede invitar a tomar té la esposa del director de una escuela a: (*a*) dos, (*b*) tres, (*c*) dos o más, esposas de ocho profesores del plantel?

Solución

a) Puede invitar a 2 de las 8 esposas de C_2^8 maneras, es decir:

Datos
$n = 8$ esposas
$r = 2$ esposas

Fórmula
$$C_r^n = \frac{n!}{r!(n-r)!}$$

Sustitución
$$C_2^8 = \frac{8!}{2!(8-2)!} = \frac{8!}{2!6!} = 28$$

Puede invitar a 2 de las 8 esposas de 28 maneras.

b) Puede invitar a 3 de las 8 esposas de C_3^8 maneras, es decir:

Datos
$n = 8$ esposas
$r = 3$ esposas

Fórmula
$$C_r^n = \frac{n!}{r!(n-r)!}$$

Sustitución
$$C_3^8 = \frac{8!}{3!(8-3)!} = \frac{8!}{3!5!} = 56$$

Puede invitar a 3 de las 8 esposas de 56 maneras.

c) Puesto que puede seleccionar dos, tres, cuatro, … u ocho esposas, entonces el número de maneras de invitarlas es $(C_2^8) + (C_3^8) + \ldots + (C_8^8)$, es decir: $28 + 56 + 70 + 56 + 28 + 8 + 1 = 247$ maneras.

Puede invitar a 2 o más de las 8 esposas de 247 maneras.

Partición

Una **partición** de un conjunto x es una subdivisión de x en subconjuntos no vacíos que son disjuntos y cuya unión es x, o sea, en la clase de subconjuntos no vacíos de x tales que cada **elemento** del conjunto x pertenece a un único subconjunto.

Los subconjuntos de una partición son denominados **células**, **bloques** o **clases de equivalencia**.

EJEMPLO

Ejemplo 1 Dado el conjunto $A = \{a, b, c, d, e, f, g, h, i\}$, determina los subconjuntos que sean una partición de A.

Solución

Los siguientes subconjuntos $[\{a, c, e\}, \{b, d, f, h\}, \{g, i\}]$ son algunas particiones del conjunto A puesto que cada elemento de A pertenece a una célula exactamente.

Nota: pueden considerarse otros subconjuntos.

Particiones ordenadas

Si una urna x contiene nueve fichas numeradas de 1 a 9, determina el número de maneras como se pueden sacar, primero dos fichas de la urna, después tres fichas y finalmente cuatro, es decir, calcula el número de **particiones ordenadas**. Del conjunto de 9 fichas, la célula $\boldsymbol{x_1}$ tiene 2 fichas, $\boldsymbol{x_2}$ tiene 3 fichas y $\boldsymbol{x_3}$, tiene 4 fichas. Estas células se denominan **particiones ordenadas**, es decir: $[\{1, 2\}, \{3, 4, 5\}, \{6, 7, 8, 9\}]$, donde cada subconjunto es una partición de $\boldsymbol{x}$.

Como al inicio hay 9 fichas en la urna, existen C_2^9 maneras de sacar las 2 primeras fichas, con el fin de obtener la primera célula x_1; en la urna quedan 7 fichas y por lo tanto, se tienen C_3^7 formas de seleccionar 3 fichas para determinar x_2; finalmente quedan 4 fichas en la urna, es decir, que hay C_4^4 maneras de obtener la última célula x_3.

Entonces existen:

$$\begin{aligned}(C_2^9)(C_3^7)(C_4^4) &= \left[\frac{9!}{2!(9-2)!}\right]\left[\frac{7!}{3!(1-3)!}\right]\left[\frac{4!}{4!(1-4)!}\right] \\ &= \left[\frac{(9)(8)(7!)}{(2!)(7!)}\right]\left[\frac{(7)(6)(5)(4!)}{(3!)(4!)}\right]\left[\frac{4!}{(4!)(0!)}\right] = (36)(35)(1) = 1260\end{aligned}$$

Se tienen 1260 particiones ordenadas diferentes del conjunto x; distribuidas en las células x_1 con dos fichas, x_2 con tres fichas y x_3 con cuatro fichas.

Dado un conjunto x formado por n elementos y sean $n_1, n_2, n_3, \ldots n_r$ enteros positivos con $n_1 + n_2 + n_3 + \ldots + n_r = n$. El número de particiones ordenadas diferentes del conjunto x, distribuidas en células con $n_1, n_2, n_3, \ldots n_r$ elementos se calcula mediante la siguiente expresión matemática

$$P_0 = \frac{n!}{n_1!n_2!n_3!\ldots n_r!}$$

La ecuación indica las particiones ordenadas (P_o) diferentes del conjunto x, de tal manera que la célula x_1 tenga n_1 elementos, x_2 esté formada por n_2 elementos, x_3 tenga n_3 elementos, …, y x_r tenga n_r elementos.

EJEMPLO

Ejemplo 1 En una clase hay 9 estudiantes, ¿de cuántas maneras los estudiantes pueden presentar 3 pruebas diferentes si a cada prueba le corresponden 3 estudiantes?

Solución

Se desea determinar el número de particiones ordenadas de los 9 estudiantes entre 3 células, que cada una consta de 3 estudiantes. Al aplicar la expresión matemática correspondiente, se tiene:

Datos	Fórmula	Sustitución
$n = 9$ estudiantes $n_1 = 3$ estudiantes $n_2 = 3$ estudiantes $n_3 = 3$ estudiantes	$P_o = \dfrac{n!}{n_1!n_2!n_3!\ldots n_r!}$	$P_o = \dfrac{9!}{(3!)(3!)(3!)} = \dfrac{362\,880}{(6)(6)(6)}$ $P_o = 1\,680$

Existen 1 680 particiones ordenadas para que los estudiantes presenten las pruebas.

EJERCICIO 19

I. Realiza lo que se indica en cada caso.

1. Menciona los principios básicos que están contenidos en las técnicas de conteo o análisis combinatorio.
2. ¿Qué es un diagrama de árbol?
3. ¿En qué consiste el principio fundamental del conteo?
4. ¿Qué significado tiene el símbolo $n!$?
5. ¿Cuál es la base del teorema del binomio?
6. Menciona las propiedades que se observan al desarrollar un binomio.
7. Explica el desarrollo del teorema del binomio para exponentes fraccionarios y negativos.
8. Explica la formación de un triángulo de Pascal.
9. Menciona las propiedades que se deducen a partir del triángulo de Pascal.
10. ¿Qué es una permutación?
11. ¿Qué significado tiene el símbolo P_n^n?
12. ¿De cuántas maneras diferentes se pueden colocar n elementos distintos de un conjunto, tomando todos a la vez y colocándolos en fila?
13. ¿Qué significado tiene el símbolo P_r^n?
14. ¿Cómo se obtiene una permutación de n objetos diferentes tomados de r en r?
15. ¿Qué es una permutación circular?
16. ¿Qué es una permutación sobre un anillo?
17. ¿En qué consiste una permutación con repetición?
18. ¿Qué es una prueba ordenada?
19. Menciona las maneras en que pueden presentarse las pruebas ordenadas.
20. ¿Qué es una prueba ordenada con sustitución? Explica.
21. ¿Qué es una prueba ordenada sin sustitución? Explica.

22. ¿Qué es una combinación?
23. ¿Cuál es la diferencia entre combinación y permutación?
24. Explica en qué consiste una combinación r de n objetos.
25. ¿Qué es una partición?

II. En plenaria, resuelvan los siguientes problemas.

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

1. Construye el diagrama de árbol para el número de permutaciones de $\{2, 4, 6, 8\}$.
2. Construye el diagrama de árbol para el procedimiento de elegir una pareja (una muchacha y un muchacho) de un grupo de dos muchachos y tres muchachas.
3. Encuentra el conjunto que resulta del producto de $\{2, 4, 6\} \times \{k, l, m\} \times \{3, 5, 7\}$, a partir de la construcción del diagrama de árbol apropiado.
4. Los equipos de béisbol Bravos de Atlanta y Expos de Montreal juegan la Serie Mundial. El primer equipo que gane cuatro juegos de un total de siete será el campeón. Representa mediante un diagrama de árbol el número de maneras como la serie se puede llevar a cabo.
5. Un caballero tiene tres pantalones, cinco camisas y cuatro corbatas. ¿De cuántas maneras puede seleccionar un pantalón, una camisa y una corbata?
6. ¿Cuántas placas distintas de circulación para automóviles pueden hacerse usando tres letras del alfabeto seguidas de cuatro dígitos?
7. Un club social tiene 40 socios, de los cuales, 24 son hombres y 16 mujeres. Se trata de elegir un presidente, un vicepresidente y un secretario. Si el presidente debe ser un hombre y el vicepresidente una mujer, ¿cuántas planillas es posible construir?
8. Si una persona puede viajar de San Luis Potosí al D. F. por autobús, tren o por avión y puede ir del D. F. a Cancún por autobús o avión, ¿de cuántas maneras diferentes puede viajar de San Luis Potosí a Cancún?
9. Se van hacer las placas de circulación para motocicletas con tres letras del alfabeto español seguidas de dos dígitos del sistema decimal. La letra intermedia debe ser una vocal y las otras dos letras no pueden ser vocales. El primero de los dos dígitos debe ser diferente de cero.
   a) ¿Cuántas de tales placas pueden hacerse si ninguna letra ni ningún dígito puede aparecer más de una vez?
   b) ¿Cuántas pueden hacerse si se permiten las repeticiones tanto de las letras como de los números?
10. Si se tiran dos dados, ¿cuál es la probabilidad de obtener:
   a) una suma de 11? *b*) una suma de 7?
11. Desarrolla el factorial de los siguientes números:
   a) 7 *b*) 11 *c*) 8 *d*) 13
12. Calcula las siguientes operaciones factoriales:
   a) $5! + 8! - 3!$
   b) $13! - 9! + 6!$
   c) $(6!)\,(9!)$
   d) $(2!)\,(3!)\,(5!)$
   e) $\dfrac{10!}{8!}$
   f) $\dfrac{7!+12!}{(2!)(3!)(4!)}$
   g) $\dfrac{9!}{3!+4!}$
   h) $\dfrac{16!}{(8!)^2}$
   i) $\dfrac{3!+7!-2!}{5!+2!}$

13. Determina el desarrollo binomial de:
    a) $(x+y)^7$
    b) $(xy-2)^4$
    c) $(5x-y^2)^5$
    d) $\left(x^2-\frac{a}{2}\right)^6$
    e) $\left(x^{-\frac{1}{3}}-y^{-\frac{1}{3}}\right)^9$
    f) $\left(a^{-\frac{2}{3}}-b^{-\frac{2}{3}}\right)^8$
14. Calcula cada una de las expresiones siguientes con cuatro cifras significativas exactas, utilizando el teorema del binomio.
    a) $(1.02)^{-2}$
    b) $(1.05)^{-5}$
    c) $(1.03)^{-4}$
    d) $\sqrt{33}$
    e) $\sqrt[4]{17}$
    f) $\sqrt[3]{120}$
15. Si el octavo renglón del triángulo de Pascal es 1, 8, 28, 56, 70, 56, 28, 8,1, determina el noveno, décimo y undécimo renglón del mismo triángulo.
16. Si se lanzan al aire cinco monedas, determina la probabilidad de que:
    a) Aparezcan exactamente dos soles.
    b) Aparezcan exactamente tres soles.
    c) No aparezca sol alguno.
    d) Aparezcan exactamente cuatro soles.
17. ¿Cuántas designaciones de tres letras para fraternidades estudiantiles pueden formarse, usando las 24 letras del alfabeto griego, si no debe repetirse ninguna letra en una sola designación?
18. ¿De cuántas maneras pueden disponerse en hilera 12 libros distintos:
    a) Cuando un libro dado debe ocupar un extremo específico?
    b) Cuando un libro dado debe ocupar uno de los extremos?
    c) Cuando un determinado par de libros deben quedar uno al lado de otro?
    d) Cuando un determinado par de libros no deben quedar juntos por ningún motivo?
19. ¿De cuántas maneras pueden acomodarse tres libros diferentes de química y cuatro libros diferentes de física en un librero, de manera que los libros de química queden siempre juntos?
20. Se quieren sentar 7 mujeres y 6 hombres en una fila de modo que los hombres ocupen los sitios pares, ¿de cuántas formas pueden sentarse?
21. De A a B hay 8 caminos y de B a C hay 6 caminos, ¿de cuántas maneras se puede ir de A a C pasando por B?
22. ¿Cuántas señales diferentes, cada una de 9 banderas colgadas en una línea vertical, pueden formarse con 4 banderas amarillas idénticas, 3 banderas rojas idénticas y 2 banderas azules idénticas?
23. ¿De cuántas maneras se puede acomodar una reunión de 13 personas alrededor de una mesa redonda?
24. Se ordenan en una fila 7 bolas verdes, 4 bolas amarillas y 5 bolas anaranjadas. Si las bolas de igual color no se distinguen entre sí, ¿de cuántas formas posibles pueden ordenarse?
25. ¿De cuántas formas pueden sentarse 9 personas alrededor de una mesa, si
    a) pueden sentarse de cualquier forma?
    b) dos personas determinadas no deben estar una al lado de la otra?
26. ¿En cuántas posiciones relativas pueden disponerse once llaves sobre un anillo?
27. ¿De cuántas maneras pueden sentarse 9 personas alrededor de una mesa redonda si sólo se consideran sus posiciones relativas, pero de tal forma que 2 personas en particular deben sentarse juntas siempre?
28. Determina cuántas permutaciones pueden hacerse utilizando todas las letras de la siguiente palabra:
    a) Estadística
    b) Matemáticas
    c) Illinois
    d) Banana
    e) Mississippi
    f) Campana
    g) Asamayama
    h) Indiana

29. Cuatro parejas van a sentarse a una mesa circular. Encuentra el número de arreglos al sentarse si cada pareja debe estar junta.
30. ¿De cuántas maneras es posible para 10 clientes comprar 5 latas de piña, 2 latas de duraznos y 2 latas de mangos si cada cliente obtiene 1 lata?
31. En una urna contiene 10 bolas, determina el número de pruebas ordenadas de tamaño 4.
    a) Con sustitución
    b) Sin sustitución
32. Un paquete contiene 15 barajas. Determina el número de pruebas ordenadas:
    a) de tamaño 8 con sustitución.
    b) de tamaño 9 con sustitución.
    c) de tamaño 9 sin sustitución.
    d) de tamaño 10 sin sustitución.
33. ¿De cuántas maneras puede seleccionarse un grupo de 35 naipes, de un conjunto de 40 naipes de baraja?
34. ¿Cuántas líneas rectas pueden trazarse pasando por ocho puntos, de los cuales, no hay cinco que estén sobre la misma recta?
35. De un conjunto de nueve hombres y uno de siete mujeres, ¿de cuántas maneras puede seleccionarse un comité formado por cinco hombres y tres mujeres.
36. ¿De cuántas maneras puede un grupo de 15 personas dividirse en:
    a) Dos grupos de nueve y seis personas?
    b) Tres grupos de siete, cinco y tres personas?
38. Encuentra el número de (*a*) combinaciones y (*b*) permutaciones de cuatro letras, que cada una pueda formarse con las letras de la palabra Tennessee.
39. Encuentra el número de elecciones que un estudiante tiene para resolver seis problemas de un grupo de once problemas.
40. Siete monedas diferentes se arrojan al aire simultáneamente. ¿De cuántas maneras pueden caer tres soles y cuatro águilas hacia arriba?
41. Un señor tiene 11 amigos de confianza.
    a) ¿De cuántas maneras puede invitar a 7 de ellos a comer?
    b) ¿De cuántas formas si 3 de ellos son casados y no asisten por separado?
    c) ¿De cuántas maneras si 4 de ellos no se llevan bien y no asisten juntos?
42. Una persona recibe una mano de póquer (cinco naipes) de una baraja de 52 cartas, ¿de cuántas maneras puede recibir:
    a) una escalera flor?
    b) un póquer?
    c) una escalera?
    d) un par de ases?
    e) un par cualquiera?
43. ¿De cuántas maneras se pueden distribuir 9 juguetes entre 4 niños, si el menor recibe 3 juguetes y cada uno de los otros niños 2 juguetes?

Verifica tus resultados en la sección de respuestas.

TEMA 8 Probabilidad para eventos

Propósito del tema

Que el estudiante:

- Aprenda y aplique la probabilidad condicional.
- Interprete y analice los resultados que puede obtener al resolver problemas de eventos independientes.
- Aplique el teorema de Bayes para calcular probabilidades en situaciones reales.

Competencias disciplinares

1. Construye e interpreta modelos matemáticos mediante la aplicación de procedimientos aritméticos, algebraicos, geométricos y variacionales, para la comprensión y análisis de situaciones reales, hipotéticas o formales.
3. Interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos o situaciones reales.
5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.
7. Elige un enfoque determinista o uno aleatorio para el estudio de un proceso o fenómeno, y argumenta su pertinencia.
8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

Contenidos que aborda el tema

Contenidos conceptuales

- Definición de probabilidad.
- Concepto de probabilidad condicional.
- Definición de eventos independientes.
- Definición del teorema de Bayes.
- Definición de selecciones al azar con o sin reemplazo.

Contenidos procedimentales

- Sintetizará evidencias obtenidas mediante la experimentación para producir conclusiones y formulará nuevas preguntas.
- Seguirá procedimientos de manera reflexiva, comprendiendo como cada uno de sus pasos contribuye al alcance de un objetivo.
- Resolverá problemas aplicando el teorema de Bayes.
- Aprenderá a identificar selecciones al azar, con o sin reemplazo.
- Usará diversas estrategias de resolución de problemas y los aplicara a situaciones reales.

Contenidos actitudinales

- Expresará sus ideas mediante el cálculo de la probabilidad para diversos eventos.
- Trabajará en equipo y respetará a sus compañeros al resolver problemas.
- Aprenderá a valorar el trabajo de sus compañeros al resolver problemas.
- Estructurará ideas y argumentos de manera clara, coherente y sintética.

Probabilidad para eventos

Probabilidad condicional

Introducción a la probabilidad

En el siglo XVII Blas Pascal y Pierre de Fermat, ambos matemáticos franceses, descubrieron la mayoría de los conceptos más importantes de la teoría de la probabilidad. Aunque comenzó con los juegos de azar, la teoría de la probabilidad ha ido ganando cada vez mayor importancia hasta que, en la actualidad, contribuye de manera esencial a facilitar los estudios de las ciencias naturales, sociales y muchos de los problemas prácticos del mundo de los negocios, la industria y el gobierno.

En los pronósticos que se hacen acerca del resultado de algo en lo que intervienen la acción y la reacción de los seres humanos, se habla en términos de lo que es más probable y no en términos de una seguridad al 100%. El dueño de una zapatería tiene que decidir cuáles son los posibles estilos de zapatos más atractivos y de moda para satisfacer a sus clientes. Cuando Fernando Valenzuela entró a lanzar un partido de la Serie Mundial de béisbol, la probabilidad de que lograra un ponche por entrada era mayor del 70%. Para decidir si una persona se lanza como candidato a la presidencia de la república (así como el partido que lo apoyará y los colaboradores de su campaña) deseará estimar la probabilidad que tiene de ganar.

La probabilidad también existe en situaciones que no pueden ser afectadas por acciones humanas. Se dice a veces, es muy probable que mañana haya un sol brillante. Existe la misma probabilidad de ganar o perder cuando se lanza una moneda al aire, es decir, el 50%.

La probabilidad está asociada con los juegos de azar y por lo general, se realizan las siguientes preguntas: ¿cuál es la probabilidad de tirar 7 u 11 en un juego de dados?, ¿cuál es la probabilidad de ganar en la ruleta con el 8 negro?, ¿cuál es la probabilidad de ganar con una mano de póquer?, entre otras.

El uso del concepto de **probabilidad** en situaciones tales como las descritas en los párrafos anteriores, está relacionado con el grado de confianza que la persona tiene de que ocurra un evento en particular. El objeto de la teoría de las probabilidades es dar un verdadero significado al concepto de la probabilidad precisada en términos aritméticos y proporcionar métodos sistemáticos para concluir en resultados numéricos. Lo anterior permite tratar con mayor efectividad las incertidumbres que intervienen en muchas decisiones que se tienen que afrontar en la vida diaria.

Definición clásica de probabilidad

Si en un suceso $\boldsymbol{E}$, existe un total de n situaciones posibles, (todas con la misma probabilidad de ocurrir) una de ellas deberá ocurrir; es decir, podrá presentarse en S de las situaciones. Entonces, la probabilidad p de que un suceso ocurra viene dada por la ecuación:

$$\text{Probabilidad del evento} = \frac{\text{Número de maneras en que puede ocurrir el evento}}{\text{Número de resultados posibles del experimento}}$$

$$p = P(E) = \frac{S}{b}$$

La probabilidad de no aparición o de no ocurrencia $\boldsymbol{q}$ del suceso, viene dada por la ecuación:

$$\begin{matrix}\text{Probabilidad del}\\ \text{no evento}\end{matrix} = \frac{\begin{pmatrix}\text{Número de resultados}\\ \text{posibles del experimento}\end{pmatrix} - \begin{pmatrix}\text{Número de maneras en que}\\ \text{puede ocurrir el evento}\end{pmatrix}}{\text{Número de resultados posibles del experimento}}$$

$$q = P(\text{no } E) = \frac{nS}{n}$$

De lo anterior, se tiene que:

$$p + q = 1 \text{ o } P(E) + P(\text{no } E) = 1$$

La probabilidad del evento p se denomina también **probabilidad de éxito** y la probabilidad del no evento q se denomina también **probabilidad de fracaso**.

EJEMPLOS

Ejemplos

1 Si se tira un dado, ¿cuál es la probabilidad de obtener un número par?

Solución

Al lanzar un dado puede obtenerse un número par de tres maneras {2, 4, 6}, las seis igualmente posibles maneras de obtenerse {1, 2, 3, 4, 5, 6}; entonces:

Datos	Fórmula	Sustitución
$S = \{2, 4, 6\}$ (tres) $n = \{1, 2, 3, 4, 5, 6\}$ (seis)	$p = \frac{S}{n}$	$p = \frac{3}{6} = \frac{1}{2}$

La probabilidad de obtener un número par es de $\frac{1}{2}$.

2 En una sola tirada de un dado, (*a*) ¿cuál es la probabilidad de obtener los números 1 o 6? y (*b*) ¿cuál es la probabilidad de no obtener los números 1 o 6?

Solución

Si el dado está bien construido, es decir, no está cargado, existen seis casos numéricos igualmente posibles que pueden presentarse {1. 2, 3, 4, 5, 6}. Puesto que sólo dos resultados son de interés (1, 6), se tiene que:

a)

Datos	Fórmula	Sustitución
$S = \{2, 4, 6\}$ (tres) $n = \{1, 2, 3, 4, 5, 6\}$ (seis)	$p = \frac{S}{n}$	$p = \frac{2}{6} = \frac{1}{3}$

La probabilidad de obtener los números 1 o 6 es de $\frac{1}{3}$.

b)

Datos	Fórmula	Sustitución
$p = \frac{1}{3}$	$q = 1 - p$	$q = 1 - \frac{1}{3} = \frac{2}{3}$

La probabilidad de no obtener los números 1 o 6 es de $\frac{2}{3}$.

Es importante mencionar que la probabilidad de un suceso es un valor numérico comprendido entre 0 y 1, o bien, porcentualmente entre 0 y 100%. Si un suceso ocurre con toda seguridad, entonces su probabilidad es 1 o 100%. Si un suceso es imposible que ocurra, entonces su probabilidad es de 0, o bien, de 0%.

Si p es la probabilidad de que un suceso ocurra y por el contrario, q es la probabilidad de que un suceso no ocurra, entonces es posible comparar ambas probabilidades evaluando la tendencia a favor o en contra de que ocurra un resultado esperado. La única diferencia que existe entre la tendencia a favor y la tendencia en contra es el orden en que se lee cada probabilidad. Así, para expresar cada tendencia se tiene lo siguiente:

a) La tendencia a favor de su aparición es $(p : q)$, que se lee $\boldsymbol{p}$ a $\boldsymbol{q}$.
b) La tendencia en contra de su aparición es $(q : p)$, que se lee $\boldsymbol{q}$ a $\boldsymbol{p}$

EJEMPLO

Ejemplo 1 ¿Cuál es la tendencia en contra de obtener los números 1 o 6 en una sola tirada de un dado bien construido?

Solución

Con base en los resultados del ejemplo anterior, se tiene que la tendencia en contra de obtener los números 1 o 6 es:

Datos	Fórmula	Sustitución
$p=\frac{1}{3}$	$q:p$	$\frac{2}{3}:\frac{1}{3}=2:1$
$q=-\frac{2}{3}$		

La tendencia en contra de obtener los números 1 o 6 es 2 a 1.

Definición de probabilidad como frecuencia relativa

La probabilidad es el estudio de experimentos **aleatorios** o libres de determinación. Si una moneda es lanzada al aire se tiene la certeza de que caerá, pero no es seguro que cuando caiga sea águila. Si el experimento de lanzar la moneda al aire n veces, sea S el número de aciertos (número de veces en que aparece águila) y sea n el número de lanzamientos hechos; entonces se puede obtener información al comparar el número total de lanzamientos contra el número de veces que aparece águila.

De acuerdo con lo anterior, la **probabilidad estimada** o **empírica** de un suceso se toma como la **frecuencia relativa** de la aparición del suceso, cuando el número de observaciones es muy grande. La probabilidad por sí misma es el **límite** de la frecuencia relativa cuando el número de observaciones crece indefinidamente. Esto da origen a los problemas de **verificación** y **confiabilidad**, temas principales de la estadística.

La ecuación que define a la probabilidad como frecuencia relativa, es:

$$p_f=\frac{S}{n}$$

EJEMPLO

Ejemplo 1 Si en 100 lanzamientos de una moneda al aire resultan 53 soles, ¿cuál es su probabilidad como frecuencia relativa?

Solución

Datos	Fórmula	Sustitución
$S=53$ soles		
$n=100$ lanzamientos	$p_f=\frac{S}{n}$	$p_f=\frac{53}{100}=0.53$

La probabilidad como frecuencia relativa de soles es de 0.53.

EJEMPLOS

Ejemplos

1 Si en otros 100 lanzamientos de la moneda al aire resultan 49 soles y si se consideran los 100 lanzamientos descritos en el ejemplo anterior, ¿cuál es la probabilidad como frecuencia relativa en el total de los 200 lanzamientos?

Solución

Datos	Fórmula	Sustitución
$S = (53 + 49)$ soles $n = 200$ lanzamientos	$p_f = \dfrac{S}{n}$	$p_f = \dfrac{53+49}{200} = 0.51$

La probabilidad como frecuencia relativa en el total de soles es de 0.51.

Si el experimento se continua y de acuerdo con la definición estadística, cada vez más y más se estaría más cerca del resultado de 0.5, que es la probabilidad determinada por la ecuación que define clásicamente a la probabilidad.

Experimento

Es todo lo que puede repetirse una y otra vez. Se define como el proceso mediante el cual se hace o mide una observación. Para determinar plenamente un experimento, debe describirse la acción que ha de ejecutarse y decir lo que se ha de observar o medir.

Los términos **intento** y **ensayo** son sinónimos de **experimento**, ya que se refieren a su ejecución. Lo que se obtiene por medio de un experimento se denomina **resultado**.

Espacio muestral

El conjunto de todos los resultados posibles de un experimento se denomina **espacio muestral** de dicho experimento. El **espacio muestral** se representa por ***S***.

Un resultado particular, es decir, un elemento del conjunto ***S***, se denomina **punto muestral** o **muestra**. El espacio muestral puede ser:

Espacio muestral finito. Es aquel que está determinado por un número específico de intentos; por ejemplo, lanzar una moneda al aire cuatro veces.

Espacio muestral infinito. Es aquel que no está determinado por un número específico de intentos, es decir, que éstos son **continuos**.

Eventos independientes

Evento

Es un conjunto de resultados, es decir, cualquier subconjunto de un espacio muestral. El **evento** se representa por ***E***. El evento que consta de una muestra simple se denomina **evento elemental**. El conjunto vacío $\varnothing$ y el espacio muestral S también son eventos; el conjunto vacío $\varnothing$ en algunas ocasiones se llama **evento imposible** y el espacio muestral S se denomina **evento cierto** o **seguro**.

EJEMPLOS

Ejemplos

1 Describe el espacio muestral, los puntos de muestreo y los eventos simples o elementales para el experimento en el que se lanzan simultáneamente dos monedas al aire, observando para cada moneda si cae águila o sol.

Solución

Para determinar todos los resultados posibles del experimento se utiliza un diagrama de árbol:

Donde A y S representan respectivamente al águila y al sol.

El espacio muestral S está constituido por: $S = \{AA, AS, SA, SS\}$.

Los puntos de muestreo son AA, AS, SA, SS.

Los eventos simples o elementales son $\{AA\}$, $\{AS\}$, $\{SA\}$, $\{SS\}$.

2 Con base en el ejemplo anterior, indica en notación de conjuntos, cada uno de los siguientes eventos:

a) Por lo menos se obtiene un sol.
b) Se obtienen dos águilas.
c) El número de soles que se obtienen es igual al número de águilas.
d) Se obtienen tres águilas.

Solución

Si el espacio muestral del experimento es $S = \{AA, AS, SA, SS\}$, se tiene:

a) El evento descrito por obtener por lo menos un sol es $E = \{AS, SA, SS\}$.
b) El evento descrito por obtener dos águilas es $E = \{AA\}$.
c) El evento descrito por obtener un número de soles igual al número de águilas es $E = \{AS, SA\}$.
d) El evento descrito por obtener tres águilas no puede ocurrir, se denomina evento imposible y se representa por el conjunto vacío $\varnothing$.

Se hace notar que en los incisos *a*, *b* y *c* el evento E es un subconjunto del espacio muestral S, por lo que se establece que dicho evento E ocurre en un intento del experimento si el resultado del intento es un elemento de E, es decir $E \subset S$.

Para el inciso *d*, se observa que el conjunto vacío $\varnothing$ es un conjunto de todo conjunto, por lo que $\varnothing \subset S$ y el $\varnothing$ es también un evento.

3 Para el experimento del ejemplo 1, determina la probabilidad de que ocurra el espacio muestral S, considerado como evento, y cada uno de los eventos simples o elementales, en un solo intento del experimento; también, la probabilidad de que ocurra cada uno de los eventos descritos por los incisos *a*, *b*, *c* y *d* del ejemplo 2.

Solución

Si el espacio muestral del experimento es $S = \{AA, AS, SA, SS\}$ y se considera como un evento, se observa que existen cuatro maneras en que el evento pueda ocurrir, por lo que su probabilidad es:

Datos	Fórmula	Sustitución
$S = \{AA, AS, SA, SS\}$ (cuatro) $n = \{AA, AS, SA, SS\}$ (cuatro)	$p = \frac{S}{n}$	$p = \frac{4}{4} = 1$

La probabilidad de que ocurra el espacio muestral, considerado como evento, es de 1.

Como existen cuatro resultados posibles del experimento y cada uno de los eventos simples o elementales $\{AA\}$, $\{AS\}$, $\{SA\}$, $\{SS\}$ puede ocurrir de una sola manera, su probabilidad es:

Datos	**Fórmula**	**Sustitución**
$S = \{AA\}, \{AS\}, \{SA\}, \{SS\}$ (uno) $n = \{AA, AS, SA, SS\}$ (cuatro)	$p = \dfrac{S}{n}$	$p = \dfrac{1}{4} = 0.25$

La probabilidad de que ocurra cada uno de los eventos simples o elementales es de 0.25.

Como existen cuatro resultados posibles del experimento y el evento descrito por obtener por lo menos un sol $E = \{AS, SA, SS\}$ puede ocurrir de tres maneras, su probabilidad es:

Datos	**Fórmula**	**Sustitución**
$S = \{AS, SA, SS\}$ (tres) $n = \{AA, AS, SA, SS\}$ (cuatro)	$p = \dfrac{S}{n}$	$p = \dfrac{3}{4} = 0.75$

La probabilidad de que ocurra el evento descrito por obtener por lo menos un sol es de 0.75.

Dado que existen cuatro resultados posibles del experimento y que el evento descrito por obtener dos águilas $E = \{A, A\}$ es también uno de los eventos simples o elementales que puede ocurrir de una sola manera, su probabilidad es:

Datos	**Fórmula**	**Sustitución**
$S = \{AA\}$ (uno) $n = \{AA, AS, SA, SS\}$ (cuatro)	$p = \dfrac{S}{n}$	$p = \dfrac{1}{4} = 0.25$

La probabilidad de que ocurra el evento descrito por obtener dos águilas es de 0.25.

Como hay cuatro resultados posibles del experimento y el evento descrito por el número de soles obtenidos es igual al número de águilas $E = \{AS, SA\}$ puede ocurrir de dos maneras, su probabilidad es:

Datos	**Fórmula**	**Sustitución**
$S = \{AS, SA\}$ (dos) $n = \{AA, AS, SA, SS\}$(cuatro)	$p = \dfrac{S}{n}$	$p = \dfrac{2}{4} = 0.5$

La probabilidad de que ocurra el evento descrito por el número de soles obtenidos es igual al número de águilas es de 0.5.

Como hay cuatro resultados posibles del experimento y el evento descrito por obtener tres águilas $E = \varnothing$, no puede ocurrir, su probabilidad es:

Datos	**Fórmula**	**Sustitución**
$S = \varnothing$ (cero) $n = \{AA, AS, SA, SS\}$ {cuatro}	$p = \dfrac{S}{n}$	$p = \dfrac{0}{4} = 0$

La probabilidad de que ocurra el evento descrito para obtener tres águilas, es de 0.

Con base en este último ejemplo, se pueden formalizar los siguientes enunciados básicos de la probabilidad:

- La probabilidad de cada evento simple o elemental es un número mayor o igual a cero.
- La suma de las probabilidades de los eventos simples o elementales de un espacio muestral es igual a la unidad.
- Si un experimento tiene n resultados igualmente probables, la probabilidad de cada uno de los n eventos simples o elementales igualmente probables es de $\frac{1}{n}$.
- La probabilidad del conjunto vacío ($\varnothing$) es igual a cero.

PROBLEMAS DE APLICACIÓN

Aplicación de las técnicas de conteo en probabilidad

1 En un cajón hay 8 diferentes resistencias que se requieren utilizar en orden para completar un circuito electrónico. ¿Cuál es la probabilidad de que en extracciones aleatorias sin sustitución, las resistencias se obtengan en el orden que se requieren?

Solución

Primero se determina el número total de posibles resultados:

Datos	Fórmula	Sustitución
$n = 8$	$P_n^n = n!$	$P_8^8 = 8! = 40\ 320$

La probabilidad deseada es:

Datos	Fórmula	Sustitución
$S = 1$ resistencia $n = 40\ 320$ maneras	$p = \frac{S}{n}$	$p = \frac{1}{40\ 320} \approx 0.0000248$

La probabilidad de que en extracciones aleatorias sin sustitución se obtenga una resistencia en el orden que se requieren es de ≈ 0.0000248.

2 Una urna contiene 6 bolas blancas idénticas entre sí y 4 bolas negras idénticas entre sí. Se necesitan primero las 6 bolas blancas y luego las 4 bolas negras, ¿cuál es la probabilidad de extraer aleatoriamente las bolas en el orden requerido?

Solución

Primero se determina el número total de posibles resultados:

Datos	Fórmula	Sustitución
$n = 6 + 4 = 10$ bolas $n_1 = 6$ bolas blancas $n_2 = 4$ bolas negras	$P = \frac{n!}{n_1!n_2!n_3!\ldots n_1!}$	$P = \frac{10!}{6!4!} = 210$

La probabilidad deseada es:

Datos	Fórmula	Sustitución
$S = 1$ bola $n = 210$ maneras	$p = \frac{S}{n}$	$p = \frac{1}{210} \approx 0.00476$

La probabilidad de extraer aleatoriamente las bolas en el orden requerido es de ≈ 0.00476.

3 ••• En un centro comercial de computadoras se recibieron 14 equipos nuevos, 9 del modelo 486 y 5 del modelo 386. Si se venden cinco equipos:

a) ¿Cuál es la probabilidad de que las cinco computadoras vendidas sean 3 del modelo 486 y 2 del modelo 386?

b) ¿Cuál es la probabilidad de que las cinco computadoras vendidas sean del mismo modelo?

Solución

a) El número total de maneras de elegir 5 computadoras entre las 14 es:

Datos

$n = 14$ computadoras

$r = 5$ computadoras

Fórmula

$$P_r^n = \frac{n!}{r!(n-r)!}$$

Sustitución

$$P_5^{14} = \frac{14!}{5!(14-5)!} = 2002$$

El número de maneras de seleccionar 3 computadoras de entre 9 es C_3^9 y el número de maneras de elegir 2 computadoras de entre 5 es C_2^5. Entonces la probabilidad deseada es:

Datos

$s = (C_3^9)\,(C_2^5) = (84)(10) = 840$

$n = C_5^{14} = 2\,002$

Fórmula

$$p = \frac{S}{n}$$

Sustitución

$$p = \frac{(C_3^9)(C_2^5)}{C_5^{14}} = \frac{(84)(10)}{2002} = \frac{840}{2002}$$

$$p \approx 0.4195$$

La probabilidad de que las cinco computadoras vendidas sean 3 del modelo 486 y 2 del modelo 386 es de ≈ 0.4195.

b) Dado que el número de maneras de elegir 5 computadoras de entre las 14 es 2002, entonces el número de formas de elegir 5 computadoras entre 9 es C_5^9 y el número de formas de seleccionar 5 computadoras entre 5 es (C_5^5). La probabilidad deseada es:

Datos

$S = C_5^9 + C_5^5 = 126 + 1 = 127$

$n = 2\,002$

Fórmula

$$p = \frac{S}{n}$$

Sustitución

$$p = \frac{C_5^9 + C_5^5}{2\,002} = \frac{126+1}{2\,002} = \frac{127}{2\,002}$$

$$p \approx 0.0634$$

La probabilidad de que las cinco computadoras vendidas sean del mismo modelo es de ≈ 0.0634.

EJERCICIO 20

I. Realiza lo que se indica en cada caso.

Escribe los números correspondientes	
Competencias genéricas	
Competencias disciplinares	

1. Describe con tus propias palabras lo que entiendes por probabilidad.
2. ¿Cuáles son las aplicaciones actuales de la probabilidad?
3. Cita la definición clásica de probabilidad.
4. Escribe la ecuación de la probabilidad de un evento.
5. Escribe la ecuación de la probabilidad de fracaso.
6. ¿Cuál es el valor numérico de la probabilidad de un suceso?
7. ¿Cuál es la probabilidad de un suceso posible?
8. ¿Cuál es la probabilidad de un suceso imposible?
9. Explica cada una de las tendencias que presenta la probabilidad para que un suceso ocurra.
10. Define la probabilidad como frecuencia relativa.

11. Escribe la ecuación que define a la probabilidad como frecuencia relativa.
12. ¿Qué es un experimento?
13. ¿Qué es un espacio muestral?
14. Explica cada una de las formas en las que un espacio muestral puede presentarse.
15. Define evento.
16. ¿Qué es un evento simple o elemental?
17. Cita los principales enunciados básicos de la probabilidad.

II. Resuelve en pareja, los siguientes problemas y discute los resultados con el resto del grupo.

1. Se lanzan tres monedas al aire simultáneamente y para cada moneda se registra si aparece águila o sol. ¿Cuál es la probabilidad de que ocurra cada uno de los siguientes eventos?
   a) Que se obtengan dos águilas.
   b) Que se obtenga al menos un águila.
   c) Que no se obtenga ninguna águila.
2. Un dado ordinario se tira sobre una mesa una sola vez y se observa el número que aparece en su cara superior. ¿Cuál es la probabilidad de que el número de la cara superior sea:
   a) un 5 o un 6?
   b) un número par menor que 5?
   c) diferente de 2?
   d) un número impar?
   e) un número primo?
3. Se ponen en una urna tres papeletas, una de ellas lleva el número 1, la segunda el número 2 y la tercera el número 3. Si se extraen de la urna dos papeletas al azar, una después de la otra y sin regresarlas, haz una lista de los resultados posibles del experimento. Determina la probabilidad de que la suma de los dos números escritos en las papeletas sea 5. Determina la probabilidad de que ambos números sean menores de 3.
4. ¿Cuál es la probabilidad de seleccionar un as de una baraja de 52 cartas?
5. ¿Cuál es la probabilidad de obtener por lo menos un sol al lanzar al aire cinco monedas?
6. ¿Cuál es la probabilidad de seleccionar un rey, el siete de diamantes o el cuatro de corazones en una sola extracción de una baraja de 52 cartas?
7. De una caja que contiene 8 canicas verdes, 5 amarillas y 3 blancas se extrae una al azar. Determina la probabilidad de que sea:
   a) Verde *b*) Amarilla *c*) Blanca *d*) No amarilla *e*) Verde o blanca
8. Determina la probabilidad de la aparición de águila en el siguiente lanzamiento de una moneda, si en 1000 lanzamientos previos aparecieron 574 soles.
9. Determina la probabilidad de la aparición de un número par en la siguiente tirada de un dado, si en 100 tiradas sucesivas aparecieron 44 veces un número impar.
10. Una tabla de mortalidad de 1990 en la República Mexicana indica que de 10 978 586 personas vivas a la edad de 15 años, solamente 8 402 319 llegan a los 65 años de edad. Determina la probabilidad de que una persona de 15 años alcance la edad de 65 años.
11. Si una persona tiene un billete de cincuenta pesos, otro de cien y otro de doscientos en su cartera, describe el espacio muestral de un experimento en el que se sacan dos billetes de la cartera, uno tras otro. Determina el evento simple o elemental en el que los billetes sacados sean el de cincuenta y el de doscientos.

12. De una bolsa que contiene 9 dados de color rojo y 6 de color negro se saca un dado al azar.
    a) ¿Cuál es el espacio muestral del experimento?
    b) ¿Cuál es el evento simple o elemental en el que el dado sacado es de color negro?
    c) ¿Cuál es la probabilidad de que el dado sacado sea de color rojo?
13. Se hace un juego de etiquetas metálicas numeradas en el que no hay 2 etiquetas con el mismo número. El número marcado en cada etiqueta es un número de dos dígitos y el dígito de las decenas es diferente de cero. Determina el espacio muestral del experimento y la probabilidad de que una etiqueta seleccionada al azar contenga un número divisible entre 5.
14. Se sacan dos naipes al azar de una baraja de 52 naipes. Encuentra la probabilidad de que:
    a) Los dos sean del palo de diamantes.
    b) Uno sea del palo de corazones y el otro del palo de tréboles.
    c) Uno sea un as y el otro un rey.
    d) Los dos sean figuras (rey, reina y caballero).
    e) Los dos sean números.
15. ¿Cuál es la probabilidad de tirar un total de 7 puntos con un par de dados bien equilibrados?
16. Un vendedor de colchones tiene 15 colchones nuevos, seis del estilo *A*, cuatro del estilo *B* y cinco del estilo *C*. ¿Cuál es la probabilidad de vender tres colchones del mismo estilo si las ventas de estilos diferentes se hacen al azar?
17. De 12 concursantes en un evento femenino cuatro participantes tienen ojos verdes. Si se eligen tres finalistas al azar, ¿cuál es la probabilidad de que:
    a) las tres tengan ojos verdes?
    b) ninguna tenga ojos verdes?
    c) por lo menos una tenga ojos verdes?
18. Una clase extraordinaria consta de 8 alumnos y de 14 alumnas. Si se elige al azar un grupo de cuatro, determina la probabilidad de:
    a) seleccionar cuatro alumnos.
    b) seleccionar dos alumnos y dos alumnas.
    c) Seleccionar cuatro alumnas.

Verifica tus resultados en la sección de respuestas.

Tipos de eventos y cálculo de probabilidades

Si para un experimento dado un evento E es un subconjunto del espacio muestral S, de tal manara que E sea la unión de un número de eventos simples o elementales; también la probabilidad p del evento E será la suma de las probabilidades de los eventos simples o elementales y es igual a la unidad. Es decir,

$$0 \leq p(E) \leq 1$$

Si E_1 y E_2 son eventos y $E_1 \subset E_2$, entonces todo punto de muestreo que aparezca en E_1 aparecerá también en E_2 y posiblemente en algunos otros puntos. En consecuencia, la probabilidad del evento E_2 es por lo menos tan grande como la probabilidad del evento E_1.

$$p(E_1) \leq p(E_2)$$

EJEMPLO

Ejemplo 1 Sea el experimento de lanzar dos monedas al aire, para el cual, el espacio muestra es $S = \{AA, AS, SA, SS\}$; sean también los eventos $E_1 = \{AA\}$ (ambas monedas caigan águila) y $E_2 = \{AA, AS, SA\}$ (obtener por lo menos una vez águila), verifica que:

a) $p(E_1) \leq p(E_2)$.

b) $0 \leq p(E) \leq 1$.

Solución

a) De la información se denota que $E_1 \subset E_2$, por lo que las probabilidades respectivas son:

Datos	**Fórmula**	**Sustitución**
$S_1 = E_1 = \{AA\}$ (uno)	$p(E_1) = \dfrac{S_1}{n}$	$p(E_1) = \dfrac{1}{4} = 0.25$
$S_2 = E_2\ \{AA, AS, SA\}$ (tres)	$p(E_2) = \dfrac{S_2}{n}$	$p(E_2) = \dfrac{S_2}{n}$
$n = (AA, AS, SA, SS)$ (cuatro)		

$$p(E_1) \leq p(E_2)$$

b) Por definición axiomática, se tiene que $p(E) = p(E_1) + p(E_2)$, es decir:

$$p(E) = \frac{1}{4} + \frac{3}{4} = \frac{4}{4} = 1$$

$$0 \leq p(E) \leq 1$$

Eventos mutuamente excluyentes

Dos eventos E_1 y E_2 son **mutuamente excluyentes** si no pueden ocurrir simultáneamente, es decir, E_1 y E_2 son mútuamente excluyentes sí y sólo sí sus respectivos conjuntos son disjuntos, es decir, sí y sólo sí $E_1 \cap E_2 = \varnothing$.

Ejemplo

Si se tira un dado ordinario, su espacio muestral es $S = \{1, 2, 3, 4, 5, 6\}$. Sean los eventos $E_1 = \{1, 3, 5\}$ (obtener un número impar) y $E_2 = (2, 4, 6)$ (obtener un número par), se observa que son eventos mutuamente excluyentes ya que satisfacen la condición $E_1 \cap E_2 = \varnothing$.

Si para el mismo experimento se tienen los eventos $E_3 = \{1, 3, 5\}$ (obtener un número impar) y $E_4 = \{2, 3, 5\}$ (obtener un número primo), se observa que no son mutuamente excluyentes ya que $E_3 \cap E_4 = \{3, 5\}$ no es un conjunto vacío, es decir, ambos eventos podrían ocurrir en un solo intento del experimento.

Si se sabe que un evento E es la unión de dos eventos E_1 y E_2, para determinar la probabilidad E a partir de las probabilidades de E_1 y E_2, se establece que:

a) En el caso de que E_1 y E_2 sean mutuamente excluyentes ($E_1 \cap E_2 = \varnothing$) se tiene la ecuación:

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) = p(E)$$

b) En el caso de que E_1 y E_2 no sean mutuamente excluyentes ($E_1 \cap E_2 \neq \varnothing$) se tiene la ecuación:

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2) = p(E)$$

EJEMPLOS

Ejemplos

1 Se saca una sola canica de una bolsa que contiene 8 canicas verdes, 9 canicas amarillas y 11 canicas anaranjadas, ¿cuál es la probabilidad del evento E de que la canica sea de color verde o amarilla?

Solución

Si se considera que E es la unión de los eventos E_1, la canica es verde, y E_2, la canica es amarilla; como E_1 y E_2 son eventos mutuamente excluyentes se tiene que las probabilidades de dichos eventos son:

Datos	Fórmula	Sustitución
$S_1 = E_1 = 8$ canicas verdes $S_2 = E_2 = 9$ canicas amarillas	$p(E_1) = \frac{S_1}{n}$	$p(E_1) = \frac{8}{28}$
$n = 8 + 9 + 11 = 28$ canicas en total.	$p(E_2) = \frac{S_2}{n}$	$p(E_2) = \frac{9}{28}$

Si se aplica la ecuación para cuando E_1 y E_2 son eventos mutuamente excluyentes resulta:

$$p(E) = p(E_1 \cap E_2) = p(E_1) + p(E_2) = \frac{8}{28} + \frac{9}{28} = \frac{17}{28}$$

La probabilidad del evento E de que la canica sea verde o amarilla es de $\frac{17}{28}$.

2 En el experimento de lanzar simultáneamente un dado y una moneda una sola vez, resulta que el espacio muestral es $S = \{1A, 2A, 3A, 4A, 5A, 6A, 1S, 2S, 3S, 4S, 5S, 6S\}$. Sean $E_1 = \{1A, 2A, 3A, 4A, 5A, 6A\}$ el evento en que la moneda cae con águila hacia arriba y $E_2 = \{2A, 2S, 4A, 4S, 6A, 6S\}$ el evento en que el dado exhibe un número par hacia arriba. Determina la probabilidad del evento E, que es la unión de los eventos E_1 y E_2.

Solución

Como los eventos E_1 y E_2 no son mutuamente excluyentes, es decir, $E_1 \cap E_2 = \{2A, 4A, 6A\}$, entonces las probabilidades de dichos eventos son:

Datos	Fórmula	Sustitución
$S_1 = E_1 = \{1A, 2A, 3A, 4A, 5A, 6A\}$ (seis)	$p(E_1) = \frac{S_1}{n}$	$p(E_1) = \frac{6}{12}$
$S_2 = E_2 = \{2A, 2S, 4A, 4S, 6A, 6S\}$ (seis)	$p(E_2) = \frac{S_2}{n}$	$p(E_2) = \frac{6}{12}$
$S_3 = E_1 \cap E_2 = \{2A, 4A, 6A\}$ (tres)	$p(E_1 \cap E_2) = \frac{S_3}{n}$	$p(E_1 \cap E_2) = \frac{3}{12}$
$n = \{1A, 2A, 3A, 4A, 5A, 6A, 1S, 2S, 3S, 4S, 5S, 6S\}$ (doce)		

Al aplicar la ecuación para cuando E_1 y E_2 no son mutuamente excluyentes resulta:

$$p(E) = p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2) = \frac{6}{12} + \frac{6}{12} - \frac{3}{12} = \frac{9}{12} = \frac{3}{4} = 0.75$$

La probabilidad del evento E es de 0.75.

También se puede determinar directamente puesto que $E_1 \cup E_2 = \{1A, 2A, 3A, 4A, 5A, 6A, 2S, 4S, 6S\}$, donde:

Datos	Fórmula	Sustitución
$S = E_1 \cup E_2$ (nueve elementos). $n =$ doce elementos.	$p(E_1 \cup E_2) = \frac{S}{n}$	$p(E_1 \cup E_2) = \frac{9}{12} = 0.75$

Eventos complementarios

El complemento de un conjunto A con relación a un universo S es el conjunto de todos los elementos de S que no son elementos de A. El complemento de A se representa como A'.

Ejemplo

Si $S = \{2, 4, 6, 8, 10, 12\}$ y $A = \{2, 4, 6\}$, entonces el complemento A' es $\{8, 10, 12\}$.

Sea S el espacio muestral de un experimento. Si E' es el complemento de un evento E con relación al espacio muestral S, entonces E y E' se denominan **eventos complementarios**. Un evento E no ocurre en un intento del experimento si el resultado del intento es un elemento de E'. Debe observarse que si E y E' son eventos complementarios en un espacio muestral S, de la definición del complemento de un conjunto se deduce que $E \cup E' = S$ y $E \cap E' = \varnothing$, por consiguiente:

$$p(E \cup E') = p(E) + p(E') \text{ y como } p(E \cup E') = p(S) = 1 \text{ se tiene que:}$$
$$p(E) + p(E') = 1.$$

EJEMPLOS

Ejemplos

1 Sea el experimento de lanzar un dado ordinario, cuyo espacio muestral es $S = \{1, 2, 3, 4, 5, 6\}$. Sean los eventos $E_1 = \{1, 3, 5\}$ (obtener un número impar) y $E_2 = \{2, 4, 6\}$ (obtener un número par), determinar la probabilidad de E_1'.

Solución

De la información del problema se deduce que $E_2 = E_1'$, por lo que evidentemente se tiene:

Datos	**Fórmula**	**Sustitución**
$S_1 = E_1 = \{1, 3, 5\}$ (tres) $S_1' = E_1' = \{2, 4, 6\}$ (tres) $n = \{1, 2, 3, 4, 5, 6\}$ (seis)	$p(E_1) = \frac{S_1}{n}$ $p(E_1') = \frac{S_1'}{n}$	$p(E_1) = \frac{3}{6} = 0.5$ $p(E_1') = \frac{3}{6} = 0.5$

$\therefore$ La probabilidad del complemento E_1' es de 0.5.

En este caso se comprueba que $P(E_1) + P(E_1') = 1$.

2 Sea el experimento de lanzar tres monedas una sola vez, cuyo espacio muestral es $S = \{AAA, AAS, ASA, ASS, SAA, SAS, SSA, SSS\}$. ¿Cuál es la probabilidad de que aparezca por lo menos una vez águila?

Solución

Sea E el evento de que aparezca por lo menos un águila. Se observa que el evento complementario de E es E', en el que no aparecen águilas. Por consiguiente $E' = \{SSS\}$ y su probabilidad es:

Datos	**Fórmula**	**Sustitución**
$S = E' \{SSS\}$ (uno) $n =$ ocho elementos	$p(E') = \frac{S}{n}$	$p(E') = \frac{7}{8}$

Al aplicar la ecuación $p(E) + p(E') = 1$ se obtiene:

$$p(E) = 1 - p(E') = 1 - \frac{1}{8} = \frac{7}{8}$$

La probabilidad de que aparezca por lo menos una vez águila es de $\frac{7}{8}$.

Teoremas y axiomas de probabilidad

Sea S un espacio muestral de un experimento, sea E la clase de eventos y sea P una función de valores reales definida en E.

Por consiguiente p se denomina **función de probabilidad** y $p(E)$ es la **probabilidad del evento** E sí y sólo sí se satisfacen los siguientes axiomas:

1. Si E es un evento de S, entonces la probabilidad del evento E es: $0 \leq p(E) \leq 1$.
2. Si el espacio muestral S se obtuvo de un experimento aleatorio, entonces: $p(S) = 1$.
3. Si E_1 y E_2 son eventos mutuamente excluyentes, entonces: $p(E_1 \cup E_2) = p(E_1) + p(E_2)$.
4. Si $E_1, E_2, E_3, \ldots$ son eventos mutuamente excluyentes e infinitos, entonces:
$$p(E_1 \cup E_2 \cup E_3 \cup \ldots \cup E_i) = p(E_1) + p(E_2) + p(E_3) + \ldots$$
5. Si $E_1, E_2, E_3, \ldots, E_n$ son eventos mutuamente excluyentes y finitos, entonces:
$$p(E_1 \cup E_2 \cup E_3 \ldots \cup E_n) = p(E_1) + p(E_2) + p(E_3) + \ldots + p(E_n)$$

Ahora se considera una serie de teoremas que se deducen directamente de los axiomas anteriores.

1. La probabilidad de un evento imposible $\varnothing$ es igual a cero: $p(\varnothing) = 0$.
2. Si E' es el complemento de un evento E, entonces: $p(E') = 1 - p(E)$.
3. Si $E_1 \subset E_2$, entonces: $p(E_1) \leq p(E_2)$.
4. Si E_1 y E_2 son dos eventos, entonces la probabilidad de la diferencia de E_1 relativa a E_2 es:
$$p(E_1 \cap E_2') = p(E_1) - p(E_1 \cap E_2)$$
5. Si E_1 y E_2 son dos eventos, entonces la probabilidad de la unión de dos eventos es:
$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

Corolario 1. Para los eventos E_1, E_2 y E_3, la probabilidad de su unión es:

$$p(E_1 \cup E_2 \cup E_3) = p(E_1) + p(E_2) + p(E_3) - p(E_1 \cap E_2) - p(E_1 \cap E_3) - p(E_2 \cap E_3) + p(E_1 \cap E_2 \cap E_3)$$

EJEMPLOS

Ejemplos

1 De una urna que contiene 7 boletos rojos, 5 boletos azules y 6 boletos verdes se extrae uno al azar. Determina la probabilidad de que sea:

a) Rojo *b*) Azul *c*) Verde *d*) No rojo *e*) Rojo o azul

Solución

Sea E_1 el evento de extracción de un boleto rojo, sea E_2 el evento de extracción de un boleto azul y E_3 el evento de extracción de un boleto verde.

a) **Datos**	**Fórmula**	**Sustitución**
$S = 7$ boletos rojos $n = 7 + 5 + 6 = 18$ boletos en total	$p(E_1) = \dfrac{S}{n}$	$p(E_1) = \dfrac{7}{18}$

La probabilidad de extraer un boleto de color rojo es de $\dfrac{7}{18}$.

b) $S = 5$ boletos azules $n = 18$ boletos en total	$p(E_2) = \dfrac{S}{n}$	$p(E_2) = \dfrac{5}{18}$

La probabilidad de extraer un boleto de color azul es de $\dfrac{5}{18}$.

c) $S = 6$ boletos verdes
$n = 18$ boletos en total

$$p(E_3) = \frac{S}{n} \qquad p(E_3) = \frac{1}{3}$$

La probabilidad de extraer un boleto de color verde es de $\frac{1}{3}$.

d) Sea E_1' el evento complementario de E_1, que se refiere a la no extracción de un boleto rojo, entonces:

$$p(E_1') = 1 - p(E_1) = 1 - \frac{7}{18} = \frac{11}{18}$$

La probabilidad de no extraer un boleto de color rojo es de $\frac{11}{18}$.

e) Sean E_1 la extracción de un boleto rojo y E_2 la extracción de un boleto azul, eventos mutuamente excluyentes, entonces:

Datos	**Fórmula**	**Sustitución**
$p(E_1) = \frac{7}{18}$	$p(E_1 \cup E_2) = p(E_1) + p(E_2)$	$p(E_1 \cup E_2) = \frac{7}{18} + \frac{5}{18} = \frac{2}{3}$
$p(E_2) = \frac{5}{18}$		

La probabilidad de no extraer un boleto de color rojo o azul es de $\frac{2}{3}$.

2 De 150 alumnos 90 estudian matemáticas, 80 estudian física y 50 estudian matemáticas y física. Si se selecciona un estudiante al azar determina la probabilidad de que sea:

a) Un estudiante de matemáticas o física. *b*) Que no estudie matemáticas ni física.

Solución

Sean los eventos E_1 estudiantes de matemáticas, E_2 estudiantes de física y $(E_1 \cap E_2)$ estudiantes de matemáticas y física, sus probabilidades son:

a)

Datos	**Fórmula**	**Sustitución**
$S_1 = E_1 = 90$	$p(E_1) = \frac{S_1}{n}$	$p(E_1) = \frac{90}{150} = \frac{3}{5}$
$S_2 = E_2 = 80$	$p(E_1) = \frac{S_2}{n}$	$p(E_2) = \frac{80}{150} = \frac{8}{15}$
$S_3 = E_1 \cap E_2 = 50$	$p(E_1 \cap E_2) = \frac{S_3}{n}$	$p(E_1 \cap E_2) = \frac{50}{150} = \frac{1}{3}$
$n = 150$		

Al aplicar la ecuación para cuando E_1 y E_2 no son mutuamente excluyentes resulta:

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2) = \frac{3}{5} + \frac{8}{15} - \frac{1}{3} = \frac{4}{5}$$

La probabilidad de que sea un estudiante de matemáticas o física es de $\frac{4}{5}$.

b) Sea $(E_1' \cap E_2') = (E_1 \cup E_2)'$ el evento complementario de $(E_1 \cap E_2)$, que se refiere a que el alumno no estudie matemáticas ni física, su probabilidad es:

$$p(E_1' \cap E_2') = p(E_1 \cup E_2)' = 1 - p(E_1 \cup E_2) = 1 - \frac{4}{5} = \frac{1}{5}$$

La probabilidad de que el alumno no estudie matemáticas ni física es de $\frac{1}{5}$.

EJERCICIO 21

I. Realiza lo que se indica en cada caso.

1. ¿Cuándo se tienen dos eventos mutuamente excluyentes?
2. ¿Cuál es la condición para establecer que dos eventos son mutuamente excluyentes?
3. Escribe las ecuaciones de los diferentes casos para determinar la probabilidad E, a partir de las probabilidades de los eventos E_1 y E_2.
4. ¿A qué se le denomina eventos complementarios?
5. Escribe los principales axiomas de la probabilidad.
6. Con base en los axiomas de la probabilidad, demuestra cada uno de los respectivos teoremas.

II. Resuelve los siguientes problemas y discute tus resultados con el resto del grupo.

Escribe los números correspondientes

Competencias genéricas

Competencias disciplinares

1. Sea el experimento de lanzar dos monedas y un dado.
   a) Describe el espacio muestral del experimento.
   b) Expresa los siguientes eventos: E_1 = obtener dos soles y un número primo, E_2 = que se obtenga un 6 y E_3 = que se obtenga exactamente un sol y un número primo.
   c) Describe el evento en que E_1 y E_2 suceden, solamente sucede E_2 y que ocurra E_2 o E_3.
   d) Determina las probabilidades de los respectivos eventos.
2. Tres autos X, Y y Z intervienen en una carrera; X tiene doble posibilidad de ganar que Y y Y tiene doble posibilidad de ganar que Z. ¿Cuáles son las respectivas probabilidades de ganar?, ¿cuál es la probabilidad de que Y o Z ganen?
3. Sea E_1 el suceso de extraer un as de una baraja de 52 cartas y E_2 el suceso de extraer una reina de una baraja de 52 cartas. ¿Cuál es la probabilidad de extraer un as o una reina en un solo intento?
4. Sea E_1 el evento obtener un número impar en una sola tirada de un dado y E_2 el evento obtener un número primo. ¿Cuál es la probabilidad de obtener un número impar o primo o ambos números?
5. Una bolsa contiene 6 botones blancos y 4 botones negros; otra bolsa contiene 5 botones blancos y 7 botones negros. Si se extrae un botón de cada bolsa, determina la probabilidad de que:
   a) Ambos sean blancos.
   b) Ambos sean negros.
   c) Uno sea blanco y el otro negro.
6. Si se lanza un dado en una sola ocasión, determina la probabilidad de que:
   a) Se obtenga un 2 o 6.
   b) Se obtenga cualquier número impar.
7. Entre los 300 profesores del departamento de ciencias exactas hay 250 titulados, 100 del total se dedican a las clases de matemáticas y 80 de los 250 se dedican a las clases de matemáticas. Si se selecciona al azar un profesor, ¿cuál es la probabilidad de que no sea titulado y no se dedique a las clases de matemáticas?
8. Manuel recibe un premio de $1 000 000 y piensa distribuirlo en educación, ropa y viajes. Las probabilidades de las tres opciones son 40%, 25% y 15%, respectivamente.
   a) ¿Cuál es la probabilidad de que logre hacer una de estas opciones?
   b) ¿Cuál es la probabilidad de que olvide estas tres opciones y gaste el dinero en otra cosa?
9. Si se tira un dado siete veces, determina la probabilidad de que:
   a) Se obtenga un 5 solamente en los dos primeros tiros.
   b) Se obtenga un 5 exactamente en dos de las siete tiradas.
10. Si se forman 12 personas en fila, ¿cuál es la probabilidad de que dos de ellas, seleccionadas con anterioridad, queden juntas?

11. Una encuesta sobre 500 estudiantes de una o más asignaturas de geometría analítica, física y química durante un semestre reveló los siguientes números de estudiantes en las asignaturas indicadas.

Geometría analítica	329	Geometría analítica y física	83
Física	186	Geometría analítica y química	217
Química	295	Física y química	63

¿Cuántos alumnos estaban estudiando

a) geometría analítica pero no química?

b) física pero no geometría analítica?

c) geometría analítica o química pero no física?

d) las tres asignaturas?

e) geometría analítica pero no física o química?

f) química pero no física?

Escribe los números correspondientes

Competencias genéricas

Competencias disciplinares

12. Dado un experimento en el que $p(E_1 \cap E_2) = \frac{1}{4}$, $p(E_1 \cup E_2) = \frac{7}{8}$ y $p(E_1') = \frac{5}{8}$ determina:

a) $p(E_1)$ *b*) $p(E_2)$ *c*) $p(E_1 \cap E'_2)$

13. Sean los eventos $p(E_1) = \frac{1}{2}$, $p(E_2) = \frac{1}{2}$ y $p(E_1 \cup E_2) = \frac{2}{3}$ determina:

a) $p(E_1')$ *d*) $p(E_1' \cap E_2')$ *f*) $p(E_1 - E_2) = p(E_1 \cap E_2')$

b) $p(E_2')$ *e*) $p(E_1' \cup E_2')$ *g*) $p(E_2 - E_1) = p(E_1' \cap E_2)$

c) $p(E_1 \cap E_2)$

14. Sean los eventos E_1 y E_2 con $p(E_1) = \frac{1}{2}$, $p(E_2') = \frac{5}{8}$ y $p(E_1 \cup E_2) = \frac{3}{4}$ determina:

a) $p(E_1 \cap E_2)$ *b*) $p(E_1' \cap E_2')$ *c*) $p(E_1' \cup E_2')$

15. La probabilidad de que A resuelva un problema es $\frac{1}{3}$ y la de que B resuelva el mismo problema es $\frac{2}{3}$.
Si tanto A como B intentan resolver el problema, ¿cuál es la probabilidad de que dicho problema se resuelva?

Verifica tus resultados en la sección de respuestas.

Probabilidad condicional

Espacio de muestreo reducido

En el experimento de lanzar simultáneamente un dado y una moneda una sola vez, resulta el espacio muestral $S = \{1A, 2A, 3A, 4A, 5A, 6A, 1S, 2S, 3S, 4S, 5S, 6S\}$ y los eventos $E_1 = \{1A, 2A, 3A, 4A, 5A, 6A\}$ (que la moneda caiga águila hacia arriba) y E_2 $\{2A, 2S, 4A, 4S, 6A, 6S\}$ (que el dado exhiba un número par hacia arriba). Si se realiza la siguiente pregunta: ¿cuál es la probabilidad del evento E_2, dado que el evento E_1 ya ocurrió?

Para dar respuesta, dado que el evento E_1 ya ha ocurrido, los únicos resultados posibles son los seis elementos $\{1A, 2A, 3A, 4A, 5A, 6A\}$ de dicho evento. De los seis resultados, en tres elementos si se sabe puede aparecer el dado con un número par en su cara superior, por lo que la probabilidad del evento E_2 es:

Datos	Fórmula	Sustitución
$S = \{2A, 4A, 6A\}$ (tres) $n = E_1 = \{1A, 2A, 3A, 4A, 5A, 6A\}$ (seis)	$p\left(\frac{E_2}{E_1}\right) = \frac{S}{n}$	$p\left(\frac{E_2}{E_1}\right) = \frac{3}{6} = 0.5$

El símbolo $p\left(\frac{E_2}{E_1}\right)$ indica la probabilidad de que ocurra el evento E_2 cuando ya ha ocurrido en evento E_1

La probabilidad de que ocurra el evento E_2 cuando ya ha ocurrido el evento E_1 es de 0.5.

Como conclusión, para determinar la probabilidad de un evento E_2 sabiendo que ya ha ocurrido un evento E_1 no se considera como espacio muestral a S, sino que se considera al evento E_1 (subconjunto de S) como el **espacio muestral reducido**.

EJEMPLO

Ejemplo 1 En el experimento en el que se tiran dos dados, si la suma de los números que aparecen es mayor que 6, ¿cuál es la probabilidad de que uno y sólo uno de los dados caiga con el 4 hacia arriba?

Solución

Por el principio fundamental del conteo se tienen (6) (6) = 36 resultados posibles del experimento y en consecuencia, un espacio muestral de 36 elementos. Sea el evento E_1 la suma de los números del tiro que sea mayor que 6 y sea el evento E_2 que uno y sólo uno de los dados caiga con el 4 hacia arriba, es decir:

$E_1 = \{(1,6), (2,5), (2,6), (3,4), (3,5), (3,6), (4,3), (4,4), (4,5), (4,6), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}$

$E_2 = \{(1,4), (2,4), (3,4), (4,1), (4,2). (4,3), (4,4), (4,5), (4,6), (5,4), (6,4)\}$

Si se considera que el evento E_1 ya ha ocurrido, entonces el **espacio muestral reducido** es E_1 y consta de 21 elementos (eventos simples o elementales), de los cuales, 6 tienen uno y sólo uno de los dados con el 4 hacia arriba. La probabilidad para el evento $\left(\frac{E_2}{E_1}\right)$ es:

Datos	Fórmula	Sustitución
$S = \{(3,4), (4,3), (4,5), (4,6), (5,4), (6,4)\}$ (seis) $n = E_1 = 21$ eventos simples o elementales.	$p\left(\frac{E_2}{E_1}\right) = \frac{S}{n}$	$p\left(\frac{E_2}{E_1}\right) = \frac{6}{21} = \frac{2}{7}$

La probabilidad del evento E_2, dado que ya ocurrió el evento E_1, es $\frac{2}{7}$.

Probabilidad de E_2 dado E_1

Con base en el experimento y los eventos del ejemplo anterior, determina la probabilidad del evento $E = E_1 \cap E_2$.

El evento $E_1 \cap E_2 = \{(3,4), (4,3), (4,5), (4,6), (5,4), (6,4)\}$ contiene seis eventos simples o elementales y por consiguiente su probabilidad es:

Datos	Fórmula	Sustitución
$S = E_1 \cap E_2 =$ seis elementos $n = 36$ elementos en total	$p(E_1 \cap E_2) = \frac{S}{n}$	$p(E_1 \cap E_2) = \frac{6}{36} = \frac{1}{6}$

También la probabilidad del evento E_1 es:

Datos	Fórmula	Sustitución
$S = E_1 = 21$ elementos. $n = 36$ elementos	$p(E_1) = \frac{S}{n}$	$p(E_1) = \frac{21}{36} = \frac{7}{12}$

Del ejemplo anterior se sabe que $p\left(\frac{E_2}{E_1}\right) = \frac{2}{7}$ y al relacionar estos tres resultados, se tiene que:

$$p\left(\frac{E_2}{E_1}\right) \times p(E_1) = p(E_1 \cap E_2)$$

Lo anterior conduce a la siguiente definición: Sean los eventos E_1 y E_2 de un espacio muestral S que tiene la propiedad de que la probabilidad $p(E_1) \neq 0$ y sea $p\left(\frac{E_2}{E_1}\right)$ la probabilidad del evento E_2, dado que ya ha ocurrido el evento E_1.

La probabilidad de $p\left(\frac{E_2}{E_1}\right)$ se obtiene por medio de:

$$p\left(\frac{E_2}{E_1}\right) = \frac{p(E_1 \cap E_2)}{p(E_1)} \qquad \text{si } p(E_1) \neq 0$$

Eventos independientes

Se establece que dos eventos son **independientes** si la ocurrencia de uno de ellos no afecta de ninguna manera la ocurrencia o no ocurrencia del otro.

E_1 y E_2 son eventos independientes si y sólo si:

$$P\left(\frac{E_2}{E_1}\right) = p(E_1) \qquad \text{y} \qquad p\left(\frac{E_2}{E_1}\right) = p(E_2)$$

Además, si E_1 y E_2 son eventos independientes, entonces:

$$P(E_1 \cap E_2) = P(E_1) \times P(E_2)$$

Esta ecuación se denomina también **teorema de la multiplicación para probabilidad condicional.**

EJEMPLOS

Ejemplos

1 Sea el experimento de lanzar simultáneamente un dado y una moneda una sola vez, del que resulta el espacio muestral $S = \{1A, 2A, 3A, 4A, 5A, 6A, 1S, 2S, 3S, 4S, 5S, 6S\}$. Sean $E_1 = \{1A, 2A, 3A, 4A, 5A, 6A\}$ el evento en que la moneda cae con águila hacia arriba y $E_2 = \{2A, 2S\}$ el evento en que el dado exhibe un 2 hacia arriba y suponiendo que son eventos independientes, ¿cuál es la probabilidad de $p(E_1 \cap E_2)$?

Solución

Primero se determina la probabilidad de los eventos E_1 y E_2:

Datos	Fórmula	Sustitución
$S_1 = E_1 = 6$ elementos $S_2 = E_2 = 2$ elementos $n = 12$ elementos	$p(E_1) = \frac{S_1}{n}$	$p(E_1) = \frac{6}{12} = 0.5$
	$p(E_2) = \frac{S_2}{n}$	$p(E_2) = \frac{2}{12} = \frac{1}{6}$

Al aplicar la ecuación para cuando E_1 y E_2 son eventos independientes resulta:

$$p(E_1 \cap E_2) = p(E_1) \cdot p(E_2) = \left(\frac{1}{2}\right)\left(\frac{1}{6}\right) = \frac{1}{12}$$

La probabilidad de $E_1 \cap E_2$ es de $\frac{1}{12}$.

También se puede determinar directamente puesto que $E_1 \cap E_2 = \{2A\}$, donde:

$S = E_1 \cap E_2 = 1$ elemento $n = 12$ elementos	$p(E_1 \cap E_2) = \frac{S}{n}$	$p(E_1 \cap E_2) = \frac{1}{12}$

2 Si se lanzan seis monedas simultáneamente ¿cuál es la probabilidad de obtener tres águilas y tres soles?

Solución

Seis monedas pueden lanzarse de $(2)(2)(2)(2)(2)(2) = (2)^2 = 64$ maneras, por lo que el espacio muestral del experimento consta de 64 elementos o eventos simples.

El número de maneras para obtener tres águilas y tres soles cuando se lanzan seis monedas una vez es igual al número de permutaciones de seis objetos, tres de las cuales son semejantes de una manera (águilas) y las otras tres de las cuales son semejantes de una segunda manera (son soles), es decir:

Datos	**Fórmula**	**Sustitución**
$n = 6$ monedas $n_1 = 3$ águilas $n_2 = 3$ soles	$p = \dfrac{n!}{n_1!n_2!n_3!\ldots n_1!}$	$p = \dfrac{6!}{3!3!} = 20$

El evento E de lanzar las seis monedas y obtener tres águilas y tres soles es la unión de 20 eventos simples o elementos (mutuamente excluyentes). Dado que el espacio muestral consta de 64 elementos, la probabilidad del evento E es:

Datos	**Fórmula**	**Sustitución**
$S = 20$ elementos $n = 64$ elementos en total	$p(E) = \dfrac{S}{n}$	$p(E) = \dfrac{20}{64} = \dfrac{5}{16}$

La probabilidad de obtener tres águilas y tres soles al lanzar seis monedas de una sola vez es de $\frac{5}{16}$.

La probabilidad de ocurrencia de E_1, E_2 y E_3 es igual a la probabilidad de E_1 por la probabilidad de que ocurra E_2, dado que ha ocurrido E_1, por la probabilidad de que ocurra E_3, dado que ha ocurrido E_1 y E_2. Para sucesos independientes, se tiene la ecuación:

$$p(E_1 \cap E_2 \cap E_3) = p(E_1)\, p(E_2)\, p(E_3)$$

Eventos dependientes

Cuando la ocurrencia de un evento E_1 afecta la probabilidad de ocurrencia de un evento E_2, entonces se dice que E_1 y E_2 son eventos dependientes.

La probabilidad de dos eventos dependientes es igual a la probabilidad de un evento multiplicado por la probabilidad condicional del otro, es decir:

$$P(E_1 \cap E_2) = P(E_1) \times P\left(\frac{E_2}{E_1}\right) \quad \text{y} \quad P(E_1 \cap E_2) = P(E_2) \times P\left(\frac{E_1}{E_2}\right)$$

EJEMPLO

Ejemplo 1

Una caja contiene 5 fichas rojas y 4 fichas azules. Sea el evento E_1 la primera ficha extraída es azul y el evento E_2 la segunda ficha extraída es azul, en extracciones sin reemplazamiento, en este caso los eventos E_1 y E_2 son dependientes. Determina la probabilidad de que la primera ficha extraída sea azul, de que la segunda ficha extraída sea azul y de que ambas fichas sean azules.

Solución

Dado que los eventos E_1 y E_2 son dependientes, sus probabilidades son:

Datos	**Fórmula**	**Sustitución**
$S = 4$ fichas azules $n = 5 + 4 = 9$ fichas en total	$p(E_1) = \dfrac{S}{n}$	$p(E_1) = \dfrac{4}{9}$

La probabilidad de que la primera ficha extraída sea azul es de $\frac{4}{9}$.

La probabilidad de que la segunda ficha extraída sea azul, dado que la primera extracción fue de ficha azul, es:

Datos

$S = 3$ fichas azules

$n = 5 + 3 = 8$ fichas en total

Fórmula

$$p\left(\frac{E_2}{E_1}\right) = \frac{S}{n}$$

Sustitución

$$p\left(\frac{E_2}{E_1}\right) = \frac{3}{8}$$

La probabilidad de que la segunda ficha extraída sea azul es de $\frac{3}{8}$.

La probabilidad de que ambas fichas sean azules, es:

Datos

$p(E_2) = \frac{4}{9}$

$p(E_2, E_1) = \frac{3}{8}$

Fórmula

$$p(E_1 \cap E_2) = p(E_1) \times p(E_2 \cap E_1)$$

Sustitución

$$p(E_1 \cap E_2) = \left(\frac{4}{9}\right)\left(\frac{3}{8}\right) = \frac{1}{6}$$

La probabilidad de que ambas fichas sean azules es de $\frac{1}{6}$.

Teorema de Bayes

Es una ecuación que permite determinar probabilidades condicionales. Se aplica a eventos diferentes donde se sabe que al menos uno de ellos ha ocurrido.

Sean E_1, E_2 y E_3 tres eventos diferentes de los que al menos uno de ellos se sabe que ha ocurrido. Si se supone que con cualquiera de éstos puede presentarse otro evento F que ha ocurrido también y si todas las probabilidades $p(E_1)$, $p(E_2)$, $p(E_3)$ y $p\left(\frac{F}{E_1}\right)$, $p\left(\frac{F}{E_2}\right)$, $p\left(\frac{F}{E_3}\right)$ son conocidas, se tiene la fórmula:

$$p\left(\frac{E_1}{F}\right) = \frac{p(E_1) \times p\left(\frac{F}{E_1}\right)}{p(E_1) \times p\left(\frac{F}{E_1}\right) + p(E_2) \times p\left(\frac{F}{E_2}\right) + p(E_3) \times p\left(\frac{F}{E_3}\right)}$$

Se pueden obtener ecuaciones similares para $p\left(\frac{E_2}{F}\right)$ y $p\left(\frac{E_3}{F}\right)$.

Si un evento puede ocurrir en más de una forma, entonces la probabilidad de que ocurra en una forma particular será igual a la razón de la probabilidad de que se presente la forma respecto a la probabilidad de que ocurra.

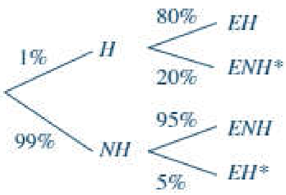
EJEMPLOS

Ejemplos

1 Para determinar si una persona tiene hepatitis se le realiza un examen de sangre de cierto tipo. La aceptación de este procedimiento se basa en lo siguiente: entre personas con hepatitis, el 80% de los exámenes de sangre descubren la enfermedad, pero el 20% fallan al hacerlo. Entre personas sin hepatitis, el 5% de los diagnósticos indican error como casos de hepatitis y el 95% de los exámenes indican el diagnóstico correcto. Si se toma una persona cualquiera de un grupo, de los cuales, un 1% tiene hepatitis y que en un examen de sangre muestra que esa persona tiene hepatitis, ¿cuál es la probabilidad de que realmente tenga la enfermedad?

Solución

Por medio de un diagrama de árbol se tiene que:



H – persona con hepatitis
NH – persona que no tiene hepatitis
EH – ENH examen de hepatitis correcto
EH^* - ENH^* – examen de hepatitis con error

Datos

$p\,(H) = 0.01$

$p\left(\frac{EH}{H}\right) = 0.8$

$p\,(NH) = 0.99$

$p\left(\frac{EH^*}{NH}\right) = 0.05$

Fórmula

$$p\left(\frac{H}{EH}\right) = \frac{p(H)\,p\left(\frac{EH}{H}\right)}{p(H)\,p\left(\frac{EH}{H}\right) + p(NH)\,p\left(\frac{EH^*}{NH}\right)}$$

Sustitución

$$p\left(\frac{H}{EH}\right) = \frac{(0.01)(0.8)}{(0.01)(0.8) + (0.99)(0.05)}$$

$$p\left(\frac{H}{EH}\right) = \frac{0.008}{0.0575} \approx 0.139$$

La probabilidad de que realmente tenga hepatitis es aproximadamente de 13.9%.

2 Tres máquinas X, Y y Z producen, respectivamente, 45%, 35% y 20% del número total de artículos de una fábrica de juguetes. Los porcentajes de defectos de producción de dichas máquinas son 2%, 4% y 6%, respectivamente; si se selecciona al azar un artículo, determina la probabilidad de que el artículo sea defectuoso.

Solución

Sea A el evento de que un artículo sea defectuoso, al aplicar el teorema de la multiplicación para la probabilidad condicional resulta:

Datos

$p\,(X) = 0.45$

$p\left(\frac{A}{X}\right) = 0.02$

$p\,(Y) = 0.35$

$p\left(\frac{A}{Y}\right) = 0.04$

$p\,(Z) = 0.2$

$p\left(\frac{A}{Z}\right) = 0.06$

Fórmula

$$p(A) = p(X)\,p\left(\frac{A}{X}\right) + p\,(Y)\,p\left(\frac{A}{Y}\right) + p\,(Z)\,p\left(\frac{A}{Z}\right)$$

Sustitución

$p\,(A) = (0.45)\,(0.02) + (0.35)\,(0.04) + (0.2)\,(0.06)$
$p\,(A) = 0.009 + 0.014 + 0.012 = 0.035$

La probabilidad de que el artículo sea defectuoso es de 3.5%.

3 Con base en la información del ejemplo anterior, si se selecciona un artículo al azar y resulta ser defectuoso; determina la probabilidad de que el artículo fue producido por la máquina Y.

Solución

Se trata de determinar $p\left(\frac{Y}{A}\right)$ por el teorema de Bayes, tenemos:

Datos

$p\,(X) = 0.45$

$p\left(\frac{A}{X}\right) = 0.02$

$p\,(Y) = 0.35$

$p\left(\frac{A}{Y}\right) = 0.04$

$p\,(Z) = 0.2$

$p\left(\frac{A}{Z}\right) = 0.06$

Fórmula

$$p\left(\frac{Y}{A}\right) = \frac{p(Y)p\left(\frac{A}{Y}\right)}{p(X)p\left(\frac{A}{X}\right) + p(Y)p\left(\frac{A}{Y}\right) + p(Z)p\left(\frac{A}{Z}\right)}$$

Sustitución

$$p\left(\frac{Y}{A}\right) = \frac{(0.35)(0.04)}{(0.45)(0.02) + (0.35)(0.04) + (0.2)(0.06)} = 0.4$$

La probabilidad de que el artículo fue producido por la máquina Y es del 40%.

EJERCICIO 22

Escribe los números correspondientes

Competencias genéricas

Competencias disciplinares

I. Realiza lo que se indica en cada caso.

1. ¿Qué es un espacio muestral reducido?
2. ¿Cuál es el significado del símbolo $p\left(\frac{E_2}{E_1}\right)$?
3. Explica qué son los eventos independientes.
4. Escribe la ecuación del teorema de la multiplicación para probabilidad condicional.
5. ¿Qué es un evento dependiente?
6. Escribe las ecuaciones de probabilidad para dos eventos dependientes.
7. ¿En qué consiste el teorema de Bayes?
8. Escribe las ecuaciones de la probabilidad de tres eventos diferentes por medio del teorema de Bayes.

II. Con ayuda de tu profesor resuelve los siguientes problemas.

1. Sea el experimento en el que se tiran dos dados, si la suma de los números que aparecen es menor que 8, ¿cuál es la probabilidad de que uno y sólo uno de los dados caiga con el 3 hacia arriba?
2. Determina $E = E_1 \cap E_2$ para los datos del problema anterior.
3. Sea el experimento de lanzar un par de dados ordinarios, si la suma de los números que cayeron hacia arriba es 6, determina la probabilidad de que uno de los dados haya caído en 5.
4. Se seleccionan al azar dígitos desde 1 hasta 9, si la suma es par, determina la probabilidad de que ambos números sean pares.
5. A una persona se le reparten 6 naipes del palo de corazones de una baraja de 52 cartas. Si se le dan tres cartas más, determina la probabilidad de que por lo menos una de las cartas adicionales sea también del palo de corazones.

6. Se reparten 13 naipes de una baraja de 52 cartas a cuatro personas (A, B, C y D).
   a) Si B no tiene ases, determina la probabilidad de que su compañero A tenga exactamente dos ases.
   b) Si A y B juntos tienen nueve naipes del palo de tréboles, determina la probabilidad de que C y D tengan cada uno 2 naipes del palo de tréboles.
7. Un grupo consta de 15 niños y 8 niñas. Si se seleccionan cuatro integrantes del grupo al azar, ¿cuál es la probabilidad de que todos sean niños?
8. Una caja contiene 11 lápices amarillos y 6 lápices azules. Se extraen 5 lápices de la caja uno tras otro. Determina la probabilidad de que los 3 primeros lápices sean amarillos y los 2 siguientes sean azules.
9. En una escuela, el 40% de los estudiantes reprobó matemáticas, 30% reprobó física y 20% reprobó las dos asignaturas. Se elige un estudiante al azar.
   a) Si reprobó física, ¿cuál es la probabilidad de que también haya reprobado matemáticas?
   b) Si reprobó matemáticas, ¿cuál es la probabilidad de que también haya reprobado física?
   c) ¿Cuál es la probabilidad de que haya reprobado matemáticas o física?
10. La urna X contiene nueve papeletas numeradas de 1 a 9 y la urna Y contiene siete papeletas numeradas de 1 a 7. Se selecciona una urna al azar y se extrae una papeleta S, el número es impar. Determina la probabilidad de que la papeleta proceda de la urna X.
11. Una caja contiene 5 canicas verdes y 9 amarillas. Se extrae una canica de la caja y se reemplaza por una del otro color. Se extrae de la caja una segunda canica.
    a) Determina la probabilidad de que la segunda canica sea verde.
    b) Si ambas canicas son del mismo color, ¿cuál es la probabilidad de que las dos sean amarillas?
12. Sea E_1 el evento de que una familia tenga niños de ambos sexos y sea E_2 el evento de que la familia tenga a lo sumo una niña.
    a) Demuestra que E_1 y E_2 son eventos independientes si una familia tiene tres hijos.
    b) Demuestra que E_1 y E_2 son eventos dependientes si una familia tiene dos hijos.
13. La probabilidad de que un hombre viva 15 años más es $\frac{1}{4}$ y la probabilidad de que su esposa viva 15 años más es $\frac{1}{3}$. Determina la probabilidad de que:
    a) Ambos estén vivos dentro de 15 años.
    b) Al menos uno esté vivo en 15 años.
    c) Ninguno esté vivo en 15 años.
    d) Solamente la esposa esté viva en 15 años.
14. La caja A contiene 12 artículos de los cuales, 4 son defectuosos y la caja B contiene 9 artículos de los cuales, 3 son defectuosos. Se extrae al azar un artículo de cada caja.
    a) ¿Cuál es la probabilidad de que ambos artículos sean defectuosos?
    b) ¿Cuál es la probabilidad de que sólo uno de ellos sea defectuoso?
    c) Si se satisface el inciso, ¿cuál es la probabilidad de que el artículo defectuoso sea de la caja A?
15. Las probabilidades de que tres caballeros ganen una partida de carambola son: $\frac{2}{7}$, $\frac{3}{5}$ y $\frac{1}{3}$, respectivamente cada uno tiene su turno una sola vez.
    a) Determina la probabilidad de que exactamente uno de ellos gane la partida.
    b) Si solamente uno gana la partida, ¿cuál es la probabilidad de que sea el segundo caballero?

Verifica tus resultados en la sección de respuestas

Autoevaluación

1. En la sala de pediatría de un hospital, el 60% de los pacientes son niñas. De los niños el 35% son menores de 24 meses. El 20% de las niñas tienen menos de 24 meses. Un pediatra que ingresa a la sala selecciona un infante al azar.

   a) Determina el valor de la probabilidad de que sea menor de 24 meses.

   b) Si el infante resulta ser menor de 24 meses. Determina la probabilidad que sea una niña.

2. Un médico cirujano se especializa en cirugías estéticas. Entre sus pacientes, el 20% se realizan correcciones faciales, un 35% implantes mamarios y el restante en otras cirugías correctivas. Se sabe además, que son de género masculino el 25% de los que se realizan correcciones faciales, 15% implantes mamarios y 40% otras cirugías correctivas. Si se selecciona un paciente al azar:

   a) Determina la probabilidad de que sea de género masculino

   b) Si resulta que es de género masculino, determina la probabilidad que se haya realizado una cirugía de implantes mamarios.

3. Un médico dispone de tres equipos electrónicos para realizar ecosonogramas. El uso que le da a cada equipo es de 25% al primero, 35% al segundo y 40% al tercero. Se sabe que los aparatos tienen probabilidades de error de 1%, 2% y 3% respectivamente. Un paciente busca el resultado de una ecografía y observa que tiene un error. Determina la probabilidad de que se ha usado el primer aparato.

4. Una empresa que fabrica camisetas posee tres máquinas, A, B y C, producen el 45%, 30% y 25%, respectivamente, del total de las piezas producidas en la fábrica. Los porcentajes de producción defectuosa de estas máquinas son del 3%, 4% y 5% respectivamente.

   a) Seleccionamos una camiseta al azar; calcula la probabilidad de que salga defectuosa.

   b) Tomamos, al azar, una camiseta y resulta ser defectuosa; calcula la probabilidad de haber sido producida por la máquina B.

   c) ¿Qué máquina tiene la mayor probabilidad de haber producido una camiseta defectuosa?

5. Tenemos tres urnas: A con 3 bolas rojas y 5 negras, B con 2 bolas rojas y 1 negra y C con 2 bolas rojas y 3 negras. Escogemos una urna al azar y extraemos una bola. Si la bola ha sido roja, ¿cuál es la probabilidad de haber sido extraída de la urna A?

Respuestas de algunos reactivos de los distintos ejercicios propuestos

Ejercicio 1

I. 1. La estadística surge en la antigüedad, cuando Moisés contó el número de soldados israelitas menores de veinte años; tiempo después los egipcios la emplean para calcular los impuestos recaudables de las cosechas.

3. Porque fue el primero en aplicar los métodos estadísticos en la investigación de problemas educativos y sociales; contribuyó en la elaboración de los primeros censos oficiales europeos, así como al desarrollo de la igualdad y similitud de datos estadísticos entre naciones.

5. *a*) John Graunt, contribuyó en estudios para calcular la tasa de mortalidad producida por la peste en Londres y España, durante el siglo XVI.

b) De Moivre, durante el siglo XVII realizó estudios de probabilidad, desarrollando en 1730 la curva de distribución normal.

c) Gottfried Achenwall, en 1748 llevó a cabo estudios de población complementándolos posteriormente con teoría de probabilidades aplicándolo a aspectos demográficos (natalidad, mortalidad, criminalidad, educación y salud).

d) Florence Nightingale, reconoce la importancia de la estadística en los negocios.

e) Laplace y Gauss, a principios del siglo XIX ambos realizaron estudios sobre el cálculo de probabilidades en astronomía, observando una curva de distribución denominada campana de distribución de Gauss para errores, estudio aplicable también a productos.

f) Ronald A. Fisher, identificó la distribución *F*, gracias a sus estudios realizados en productos agrícolas.

II. 1. Según el diccionario de la lengua española, el término estadista corresponde a un técnico o especialista en estadística.

3. En el momento en que todo movimiento o fenómeno social intenta predecirse o explicarse a partir de números puede caer en errores y encasillamientos que limiten la visión del científico, puede restar importancia a un problema o lo puede magnificar sin contemplar otros elementos cualitativos, o simplemente puede dejar fuera contextos fundamentales.

Ejercicio 2

I. 1. Es la ciencia o arte que estudia y analiza los datos que son susceptibles de expresión numérica para llegar a conclusiones que permitan tomar decisiones y pronosticar las consecuencias de las mismas.

3. *a*) Dirección de personal, cuando en una empresa se aplican test de habilidades o destrezas a sus trabajadores.

b) Departamento de producción, al realizar un estudio donde se muestre los logros en materia de producción de la empresa, para mejorar la maximización de tiempo o la calidad del producto.

c) Departamento de ventas, la estadística se emplea para conocer el resultado de ganancias, pérdidas, proyectos a largo o corto plazo, estudios de mercado, la demanda de sus productos.

d) Consejo administrativo, al hablar de reducir costos, inversión, disminuir pérdidas.

e) Control de calidad, para maximizar la calidad de un producto, aplicación efectiva de los insumos.

f) Departamento de diseño, para la interpretación de datos en los estudios de mercado, el impacto y gustos de los productos ente los consumidores.

Ejercicio 3

I. 1. La estadística se divide para su estudio en dos ramas, que son: **la estadística descriptiva y la estadística inferencial** siendo ambas funciones del análisis estadístico.

3. Se puede censar tu salón de clases sobre:

a) Color preferencial de pantalón.

b) Horario preferencial de la asignatura de matemáticas.

c) Característica de los métodos y técnicas sobre la enseñanza de las matemáticas utilizadas por los docentes.

d) Las estaturas y pesos de tus compañeros.

e) Gusto por los tipos de música… etcétera.

5. Se puede inducir sobre:

a) Juegos de azar.

b) Pronosticar ciclones o huracanes.

c) Investigar a los grupos por egresar de tu plantel si van a continuar estudiando o si van a trabajar, el tipo de carrera a seguir, la ciudad donde van a continuar sus estudios, tipo de trabajo, industria, etcétera.

II. 1. La importancia radica en que refleja el impacto de los productos o proyectos tendrá en la población receptora, el estudio de mercadeo refleja qué tan viable será lanzar un nuevo producto en venta, mientras que en la industria médica por ejemplo tomando un grupo de muestra que empleará el producto indicará sus efectos secundarios o si el precio es accesible, la presentación o si es efectivo realmente.

Ejercicio 4

I. 1. Entre los elementos básicos de la estadística, encontramos fundamentalmente el de población; muestra variable (continua y discreta) y los tipos de datos; todos ellos con el único fin de fortalecer la comprensión de los procedimientos estadísticos.

3. Se define como el conjunto de observaciones que representan la totalidad de las características a examinar de una población.
5. Existen dos clases de población las cuales se identifican por finita e infinita.
7. Las variables son cantidades a las que se les asignan un número ilimitado de valores.
9. Los datos continuos son aquellos que provienen de una variable continua, por ejemplo, las medidas dan lugar a dichos datos.
11. Los datos originales son aquellos que son recopilados por nosotros mismos; es decir, que son comprobables en forma rigurosa. Los datos indirectos son aquellos que son recopilados de enciclopedias, libros de registro, sucesos grabados en audio y video.
13. El redondeo de datos es una técnica útil, ya que tiene la finalidad de reducir al mínimo los errores acumulados por dicha práctica, sobre todo cuando la cantidad de datos es muy grande.
15. La letra griega mayúscula Σ que se llama **sigma**, matemáticamente simboliza a la sumatoria.

II. 1. Discreta 7. Continua 13. Discreta
3. Discreta 9. Discreta 15. Continua
5. Discreta 11. Discreta

III. 1. 2.06 9. 278.6 17. 0.740
3. 14.0336 11. 0.1 19. 13.000
5. 67.760 13. 5.78
7. 125.458 15 103.0

IV. 1. *a*)453.495 *b*) 453.014 *c*) 454.014

3. *a*) $\sum_{p=1}^{11} Z_p = Z_1+Z_2+Z_3+Z_4+Z_5+Z_6+Z_7+Z_8 +Z_9+Z_{10}+Z_{11}$

c) $\sum_{k=1}^{11}(3-y_k)^2=(3-y_1)^2+(3-y_2)^2+(3-y_3)^2$

e) $\sum_{k=1}^{3} F_1X_1 = F_1X_1+F_2X_2+F_3X_3+F_4X_4+F_5X_5+F_6X_6+F_7X_7 +F_8X_8+F_9X_9+F_{10}X_{10}$

g) $\sum_{k=1}^{4}(x_k-y_k)^2=(x_1-y_1)^2+(x_2-y_2)^2+(x_3-y_3)^2+(x_4-y_4)^2$

5. *a*) $\sum_{k=1}^{6} S_k = -2$

c) $\sum_{k=1}^{6} S_kT_k = 14$

e) $\sum_{k=1}^{6} T_k^3 = 0$

g) $\sum_{k=1}^{6}(S_k + T_k^3)^2 = 342$

i) $\sum_{k=1}^{6}(S_k + T_k)^3 = -1274$

Ejercicio 5

I. 1. Consiste en recolectar una serie de datos que deben ser organizados, con base en un orden numérico y en subgrupos, de acuerdo con las características comunes que presentan.
3. Creciente o ascendente.
5. Se define como la diferencia que existe entre el mayor y el menor de los datos.

II. 1. *a*) 60, 62, 63, 68, 72, 74, 75, 77, 78, 78, 85, 88, 90, 93, 94, 95
Rango = 35
b) 10, 16, 19, 24, 29, 36, 39, 48, 52, 54
Rango = 44
c) 1, 2, 5, 7, 9, 11, 13, 16, 18, 20 Rango = 19

Ejercicio 6

I. Desarrolla los siguientes ejercicios

1. Frecuencia de clase precisa el número de datos pertenecientes a una clase o categoría.
3. La ordenación tubular de los datos en clase o categorías, conjuntando las clases con sus respectivas frecuencias, se denomina **distribución de frecuencias** o **tabla de frecuencias**.
5. Se define como el intervalo de clase que, al menos teóricamente, no tiene límite inferior ni límite superior.
7. Son los límites verdaderos de clase y constan de un límite real inferior y de un límite real superior, donde se ubican los datos que dan lugar a los intervalos de clase o categorías.
9. Se define como la diferencia entre los límites reales de clase que forman el intervalo de clase o categoría.
11. Se define como el punto medio de un intervalo de clase y se representa por (X).
13. • Determinar el recorrido o rango entre los datos registrados.
 • Se divide el rango en un número razonable de intervalos de clase que tengan el mismo tamaño o anchura. También se recomienda utilizar intervalos de clase abiertos. Por lo general, el número de intervalos de clase se selecciona entre 5 y 20, dependiendo de la cantidad de datos registrados, los intervalos de clase se eligen también de manera que las marcas de clase o puntos medios sean coincidentes con los datos observados realmente, lo cual permite disminuir el **error de agrupamiento**.
 • Determinar las frecuencias de clase; es decir, contar el número de observaciones que caen dentro de cada intervalo de clase.
15. Es ordenar y resumir los datos en una tabla que indique su clase y frecuencia.

II. Desarrolla los siguientes ejercicios

1. *a*) 8 clases o categorías
b) 6.30 – 6.70
c) 7.30
d) 6.20
e) 8.25 – 8.75
f) 0.4
g) (x) = 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0
h) 5
i) 7.30 – 7.70

III. 1.

Intervalo de clase	Marca de clase	Frecuencia de clase
33-40	36.5	3
41-48	44.5	3
49-56	52.5	4
57-64	60.5	6
65-72	68.5	10
73-80	76.5	12
81-88	84.5	5
89-96	92.5	5
		48

3.

Intervalo de clase	Marca de clase x	Frecuencia
302 - 314	308	3
315 - 327	321	10
328 - 340	334	5
341 - 353	347	6
354 - 366	360	10
367 - 379	373	6
380 - 392	386	7
393 - 405	399	3
		50

5.

Intervalo de clase	Marca de clase x	Frecuencia
0.542 - 0.547	0.544	3
0.548 - 0.553	0.550	7
0.554 - 0.559	0.556	6
0.560 - 0.565	0.562	8
0.566 - 0.571	0.568	6
0.572 - 0.577	0.574	10
0.578 - 0.583	0.580	8
0.584 - 0.589	0.586	7
0.590 - 0.595	0.592	8
		63

Ejercicio 7

I. 1. La representación de información por medio de gráficos es realmente una exposición artística que no sólo tiene la finalidad de presentar datos sino que también el de enfocar las ideas y propósitos deseados.

3. Las gráficas de líneas, las gráficas de barras, los pictogramas, los gráficos circulares, histogramas, polígonos de frecuencia, etcétera.

5. Los datos acumulativos y datos instantáneos.

7. En las gráficas de sucesiones cronológicas, la variable que representa el tiempo se ubica sobre el eje horizontal, la otra variable se coloca en el eje vertical; la relación de ambas variables se hace cuando se unen por líneas rectas para dar lugar a la curva descriptiva.

   Se advierte que el **cero** se ubica en el eje vertical y siempre debe representarse si es necesario interrumpir el eje vertical, esto se hace notar por una línea en zigzag.

   - Las unidades de las variables deben sobresalir claramente y la curva se traza más gruesa que los ejes, para que resalte.
   - La longitud de los ejes se seleccionan de manera que la gráfica quede equilibrada a lo largo y ancho.
   - Los títulos se escriben en la parte superior del gráfico, los letreros y notas se escriben por abajo del eje horizontal; si hay que destacar puntos específicos de la curva, éstos deben indicarse con notas al pie del eje horizontal.
   - Siempre debe citarse la fuente informativa.

9. Datos ordenados o sucesivos y datos que expresan cantidades de conteo.

11. Se define como la representación de datos por medio de símbolos, donde su forma insinúa la especie del dato estadístico.

13. Se define como la representación de datos distribuidos en forma porcentual; es decir, el círculo se divide en sectores (rebanadas de pastel) que son equivalentes en tamaño a las frecuencias porcentuales correspondientes.

15. La construcción de una distribución de frecuencias tiene como objetivo principal condensar grandes conjuntos de datos y hacer notar sus propiedades en forma gráfica.

17. Se define como la forma de representar gráficamente una distribución de frecuencias y que básicamente consta de una sucesión de rectángulos que tienen sus bases sobre el eje horizontal y longitud igual a la anchura de los intervalos de clase, sus alturas son proporcionales a las frecuencias de clase que se ubican sobre el eje vertical.

19. *a*) En un gráfico de barras las alturas de las mismas van en relación a la variable ubicada sobre el eje vertical, mientras que en los histogramas las superficies de los rectángulos (barras) son proporcionales a las frecuencias de clase.

    b) En un diagrama de barras, éstas se grafican separadas, es decir, dejando espacios entre cada una de ellas, mientras que en los histogramas, los rectángulos (barras) se representan en forma consecutiva.

21. Para construir un polígono de frecuencias, se trazan sucesivamente sobre el eje horizontal las marcas de clase y las frecuencias correspondientes sobre el eje vertical, luego los puntos que resultan se unen por segmentos de recta. Se hace notar que se deben agregar clases con frecuencia cero en ambos extremos de la distribución con el fin de enlazar el diagrama al eje horizontal.

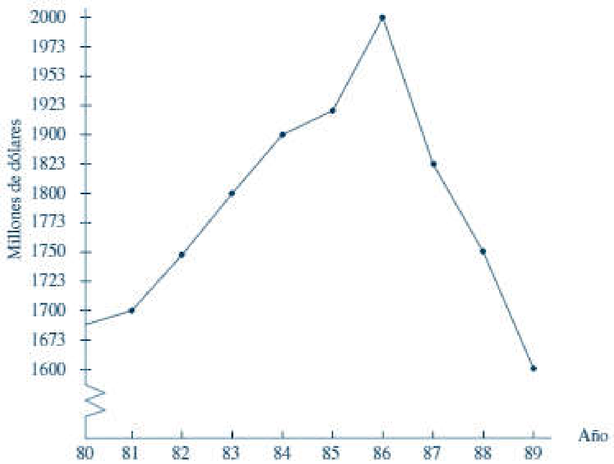
23. Una tabla de distribución de frecuencia relativa o porcentual resulta al sustituir las frecuencias de las clases por las frecuencias relativas.

25. Las gráficas de distribuciones de frecuencia relativa se pueden derivar del histograma y del polígono de frecuencias, lo anterior se obtiene al cambiar en el eje vertical la frecuencia de clase por frecuencia relativa, conservándose exactamente el mismo gráfico. Los gráficos resultantes se denominan **histogramas porcentuales o de frecuencias relativas** y **polígonos porcentuales o de frecuencias relativas**, respectivamente.

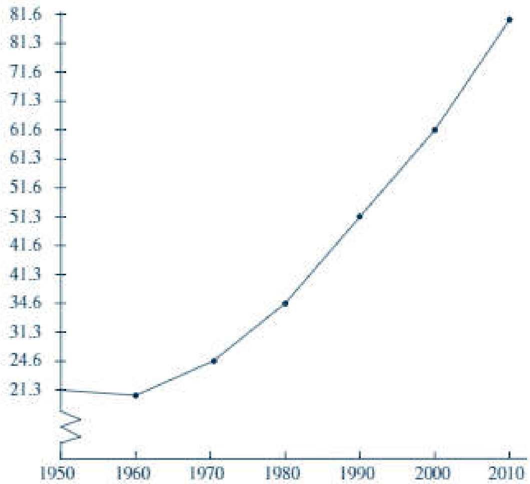
27. La frecuencia acumulada hasta un determinado intervalo de clase, se define como la frecuencia total de todos los valores menores que el límite real superior de clase del intervalo de clase considerado.

29. Es el grupo de pequeños segmentos rectos que se obtienen de un grupo de población muy grande, la curva que se forma con los datos se denomina curvas de frecuencia o curva de frecuencias relativas.

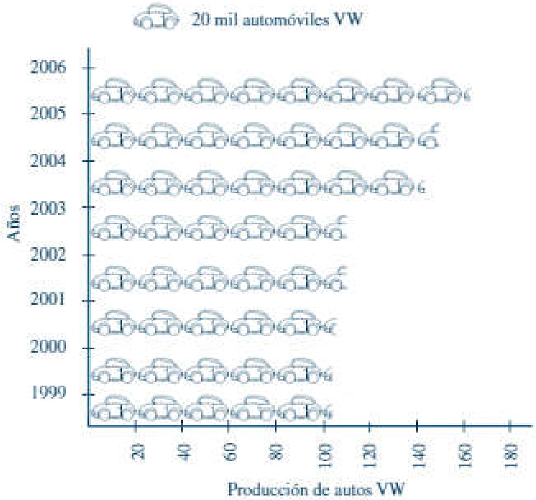
II. 1. Exportación de café mexicano (1980—1989)



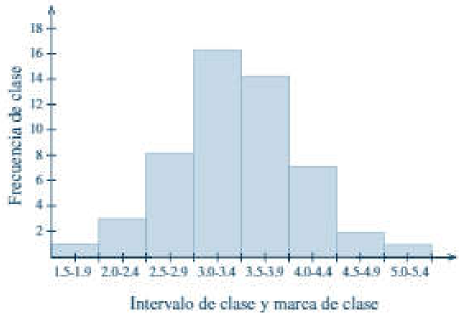
3.



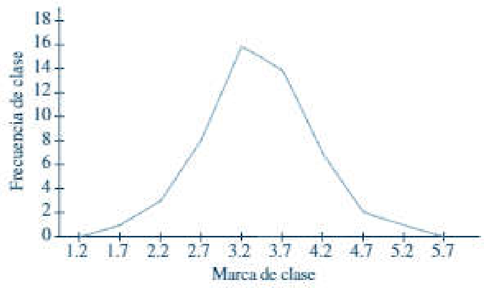
5.



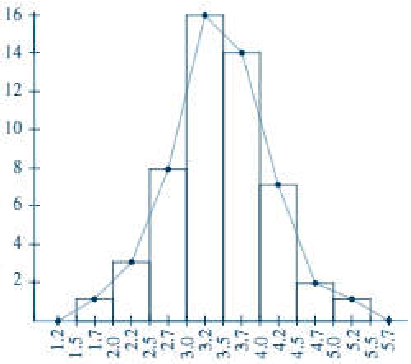
7. *a*)



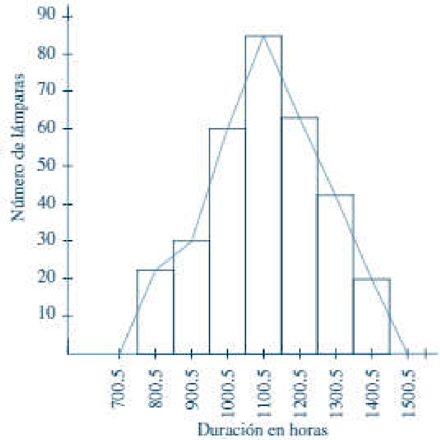
b)



c)



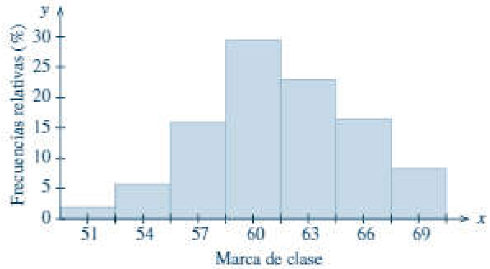
9.



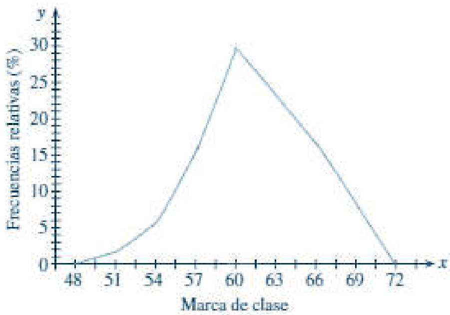
11. *a*)

Intervalos de clase	Marca de clase (x)	Frecuencias de clase	Frecuencia relativa (%)
50 - 52	51	2	1.8
53 - 55	54	6	5.5
56 - 58	57	17	15.6
59 - 61	60	32	29.4
62 - 64	63	25	22.9
65 - 67	66	18	16.5
68 - 70	69	9	8.3
		109	100

b)



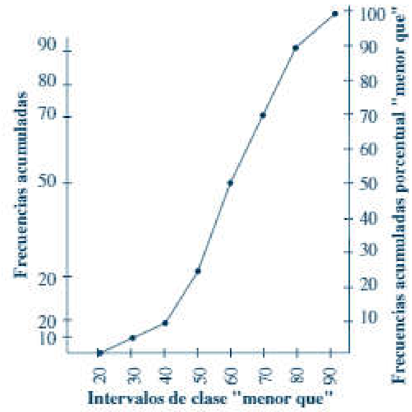
c)



d)

Intervalos de clase	Frecuencia acumulada	Frecuencias acumulada porcentual
Menor que 50	0	0.0
Menor que 53	2	1.8
Menor que 56	8	7.3
Menor que 59	25	22.9
Menor que 62	57	52.2
Menor que 65	82	75.2
Menor que 68	100	91.7
Menor que 71	109	100.0

e)



15. *a*) El tamaño o anchura del segundo intervalo de clase es de 200; el del quinto intervalo de clase es de 3000.

b) Cinco.

c) Dos.

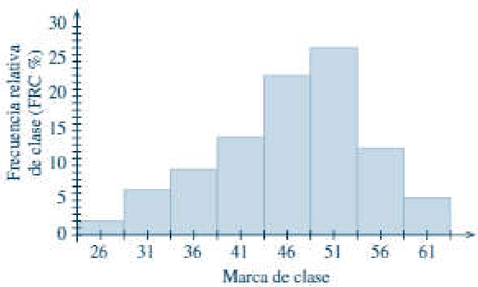
d) Aproximadamente el 4% ganó más de 8,000 y el 84% ganó menos de 3,000.

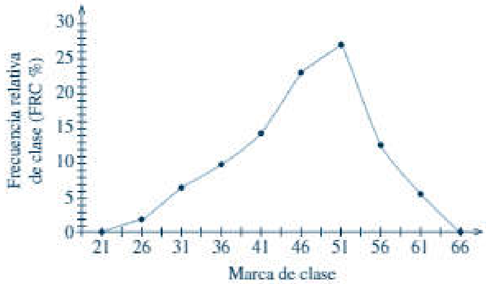
e) 19%.

17. *a*)

Longitudes	Marca de clase (x)	Número de plantas	Frecuencia relativa de clase (FRC %)
24 - 28	26	4	2.0
29 - 33	31	13	6.5
34 - 38	36	19	9.5
39 - 43	41	28	14.0
44 - 48	46	46	23.0
49 - 53	51	54	27.0
54 - 58	56	25	12.5
59 - 63	61	11	5.5
		200	

b)





c)

Longitudes	Frecuencias acumuladas o más	Frecuencias acumuladas porcentual o más
Más que 24	200	100.0
Más que 29	196	98.0
Más que 34	183	91.5
Más que 39	164	82.0
Más que 44	136	68.0
Más que 49	90	45.0
Más que 54	36	18.0
Más que 59	11	5.5
Más que 64	0	0.0

d)

EJERCICIO 8

I. 1. Se define como la raíz cuadrada del cuadrado de un conjunto de elementos numéricos $(x_1, x_2, x_3, \ldots, x_N)$ dado.

3. La media cuadrática se simboliza por las siglas R.M.S. o por la expresión $\sqrt{x^{-2}}$

5. Se aplica en procesos físicos, por ejemplo en el cálculo de la velocidad cuadrática media de las moléculas de un gas.

II. 1. *a*) $34.79 > 34.66 > 34.55 > 34.43$

b) $34.7850 > 34.6666 > 34.5465 > 34.4252$

c) $337.1 > 337 > 336.95 > 336.9$

d) $12.9384 > 12.6428 > 12.3264 > 11.9979$

e) $18 = 18 = 18 = 18$

f) $\pi = \pi = \pi = \pi$

g) $5.68 > 5.1 > 4.42 > 3.75$

h) $a = a = a = a = a$

3. $57.72 > 57.582 > 57.443 > 57.303$

5. *a*) $38.92 > 37.725 > 36.385 > 34.906$

b) $7.2941 > 7.2471 > 7.1996 > 7.1519$

c) $12.12 > 12.10 > 12.08 > 12.06$

d) $263.8878 > 252.415 > 232.12 > 224.382$

e) $61.68 > 60.33 > 58.9 > 57.26$

f) $19.6395 > 18.85 > 17.9344 > 16.878$

g) $0.6071 > 0.6070 > 0.6069 > 0.6068$

h) $39.6776 > 37.9375 > 35.9309 > 33.61$

EJERCICIO 9

I. 1. Se define como el **valor medio** que divide a un conjunto de datos que se ordenan de acuerdo con su magnitud en dos partes iguales, es decir, es aquel valor central que deja por debajo igual número de elementos que por arriba de él.

3. La mediana de un conjunto de elementos **par** se determina por la media aritmética de los dos elementos medios.

5. $$\text{Mediana} = L_1 + \left[\frac{\frac{N}{2} - (\Sigma f)}{f \text{ mediana}}\right] C$$

7. La mediana de una distribución de frecuencias relativas acumuladas, se determina por medio de una ojiva porcentual resultante de la distribución dada, donde la **mediana** es la abscisa (x) correspondiente a un punto sobre la ojiva, cuya ordenada es el 50%.

II. 1. *a*) $x \approx 7.83$; Mediana $= 8$; Moda $= 8$; $G \approx 7.41$; $H \approx 6.64$

c) $x \approx 10.76$; Mediana $= 10$; Moda $= 10$; $G \approx 10.54$; $H \approx 10.32$

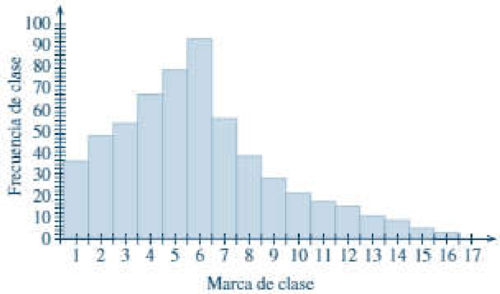
3. *a*) $x = 68$ *c*) Moda $= 64$ *e*) $H = 63.65$

5. *a*) 15.85

b) 15.85

c) 14.75

d)



e) 15.36

f) 12.102

g) 7.04

7. *a*) 7%

b) 25.61%.

Ejercicio 10

I. 1. Es el valor que se presenta con mayor frecuencia, es decir, es el valor más común de un conjunto de elementos numéricos dado.

3. Se fundamenta en encontrar primeramente la clase modal (es la clase en la cual se localiza el valor de la moda) que se caracteriza por tener la máxima frecuencia de la distribución. El siguiente paso es determinar la diferencia absoluta entre la frecuencia de la clase modal y la frecuencia de la clase premodal o anterior; así mismo, la diferencia absoluta entre la frecuencia de la clase modal y la frecuencia de la clase postmodal o siguiente. Lo anterior es con el fin de observar hacia dónde se concentra el valor modal.

II. 1. *a*) 200.70
b) 201
c) 202

3. *a*) 80.75
b) 82
c) 82

5. 4 años

III. (8), (5), (11), (1), (7), (3), (10), (6), (12), (9)

Ejercicio 11

I. 1. Son medidas que se fundamentan en las divisiones proporcionales que pueden hacerse en datos agrupados o sin agrupar.

3. El primer cuartil (Q_1) es el valor que indica en el cual o por debajo del cual queda el 25% o un cuarto de todos los datos. El segundo cuartil (Q_2) es la medida igual a la mediana, es decir, es el valor que indica en el cual o por debajo del cual queda el 50% o la mitad de todos los datos. El tercer cuartil (Q_3), es el valor que indica en el cual o por debajo del cual queda el 75% o las tres cuartas partes de todos los datos.

5. Se define como los intervalos dentro de los cuales quedan proporcionalmente repartidos los datos sin agrupar o agrupados de una distribución en diez partes iguales.

7. $D_k = L_1 + \left[\dfrac{\dfrac{KN}{10} - (\Sigma f)_1}{fK}\right] C$

9. El primer percentil (P_1), es el valor que indica en el cual o por debajo del cual queda el 1% o un centésimo de todos los datos. El segundo percentil (P_2), es el valor que indica en el cual o por debajo del cual queda el 2% o dos centésimos de todos los datos. El tercer percentil (P_3), es el valor que indica en el cual o por debajo del cual queda el 3% o tres centésimos de todos los datos. Análogamente y en forma sucesiva tenemos que los percentiles $P_4, P_5, \ldots, P_{99}$ son los valores que respectivamente indican en el cual o por debajo del cual quedan los porcentajes o centésimas partes de todos los datos.
El vigésimo quinto percentil (P_{25}) es la medida igual al primer cuartil (Q_1), es decir, es el valor que indica en el cual o por debajo del cual queda el 25% o un cuarto de todos los datos.
El quincuagésimo percentil (P_{50}) es la medida igual a la mediana, es decir, es el valor que indica en el cual o por debajo del cual queda el 50% o la mitad de todos los datos.
El septuagésimo quinto percentil (P_{75}) es la media igual al tercer cuartil (Q_3), es decir, es el valor que indica en el cual o por debajo del cual queda el 75% o tres cuartas partes de todos los datos.
Los percentiles $P_{10}, P_{20}, P_{30}, P_{40}, P_{50}, P_{60}, P_{70}, P_{80}$ y P_{90}, son las medidas iguales a los deciles $D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8$ y D_9, respectivamente.

11. Los cuantiles (cuantiles, deciles y percentiles) se determinan geométricamente en gráficos de histogramas de frecuencia porcentual, en polígonos de frecuencia porcentual y en ojivas porcentuales.

II. 1. $Q_1 = 38$, $Q_2 = 64.5$, $Q_3 = 85$; $D_1 = 24.5$, $D_3 = 40.5$, $D_5 = 64.5$, $D_7 = 85$, $D_9 = 111.5$; $P_5 = 19$, $P_{15} = 26$, $P_{45} = 54$, $P_{65} = 81$, $P_{85} = 97$ y $P_{95} = 123$

3. *a*) $Q_2 = 30.5$, $Q_3 = 33$; $D_1 = 18$, $D_3 = 27$, $D_6 = 31$, $D_9 = 39$; $P_{20} = 23$, $P_{40} = 29$, $P_{70} = 33$, $P_{80} = 34$, $P_{95} = 39$
b) $Q_2 = 240$, $Q_3 = 255$; $D_1 = 218$, $D_3 = 230$, $D_6 = 242.5$, $D_9 = 257$; $P_{20} = 227$, $P_{40} = 232.5$, $P_{80} = 256.5$, $P_{95} = 259$
c) $Q_2 = 62$, $Q_3 = 78$; $D_1 = 42$, $D_3 = 57$, $D_6 = 67$, $D_9 = 89$; $P_{20} = 52$, $P_{40} = 59$, $P_{70} = 76$, $P_{80} = 82$, $P_{95} = 93$

5. $Q_1 \approx 9.82$, $Q_2 \approx 10.2$; $D_3 \approx 9.9$, $D_6 \approx 10.31$, $D_9 \approx 10.96$; $P_{35} \approx 10$, $P_{50} \approx 10.2$, $P_{65} \approx 10.4$, $P_{80} \approx 10.7$

7. $Q_1 \approx 0.258$, $Q_2 \approx 0.261$, $Q_3 \approx 0.264$; $D_1 \approx 0.2547$, $D_3 \approx 0.259$, $D_5 \approx 0.261$, $D_7 \approx 0.264$
$D_9 \approx 0.269$; $P_5 \approx 0.252$, $P_{15} \approx 0.256$, $P_{20} \approx 0.257$, $P_{35} \approx 0.2592$, $P_{40} \approx 0.260$, $P_{45} \approx 0.2603$, $P_{60} \approx 0.2622$, $P_{65} = 0.263$, $P_{85} \approx 0.267$, $P_{75} \approx 0.2714$, $P_{99} = 0.275$

Ejercicio 12

I. 1. Para explicar el grado de dispersión o variación de los datos que se expanden alrededor de una medida central.

3. Es la medida de dispersión más simple que se define como la diferencia que existe entre el mayor y el menor de los datos de un conjunto.

5. La diferencia entre los valores encontrados se debe a que los límites de clase se extienden siempre más que las marcas de clase, que tienden a eliminar los casos más extremos. También, por lo general, se piensa que cuanto mayor es el valor del rango, mayor es la dispersión o variación de los datos.

7. El rango o amplitud también presenta desventajas, ya que depende solamente de los valores mayor y menor, por lo que es notoria su delicadeza a los valores desproporcionados que pueda haber en un conjunto.

II. 1. *a*) Rango = 0.28 *c*) Rango = 15
b) Rango = 27 *d*) Rango = 14

Ejercicio 13

I. 1. Se define como la media aritmética de los valores absolutos de las desviaciones de las variables respecto de la media aritmética; matemáticamente se expresa por la ecuación:

$$\text{Desviación media (M.D.)} = \frac{\sum |X - \bar{X}|}{N}$$

$\bar{X}$ = Media aritmética
N = Número total de elementos del conjunto

3. Si la distribución de frecuencias presenta forma simétrica y gráficamente tiene forma de campana o normal, donde se hace notar que el 57.5% de las observaciones caen en el intervalo $\bar{X} \pm M.D.$, es decir, que más de la mitad de las observaciones se concentran dentro de un intervalo de una unidad de la desviación media o promedio

de desviación a uno y otro lado de la media aritmética; incluso si es moderadamente asimétrica

5. Se define como la diferencia entre el tercer cuartil (Q_3) y el primer cuartil (Q_1); matemáticamente se expresa por la ecuación:

$$\text{Rango intercuartílico} = Q_3 - Q_1$$

7. *a)* No toma en cuenta a todos los valores de una distribución de frecuencias y puede ser que los valores por abajo de (Q_1) o por arriba de (Q_3) estén demasiado compactos o demasiado dispersos, por lo que el valor del rango intercuartílico es el mismo.
   b) Cuando sólo se conoce el valor del rango intercuartílico, no es posible precisar la ubicación de una cualquiera de las observaciones que están dentro de la distribución.
   c) Ambas medidas, no tienen propiedades que les permitan participar en las relaciones matemáticas que se emplean en la estadística.
9. Se define como la mitad del rango entre percentiles 10-90

$$\text{Rango semipercentílico} = \frac{P_{90} - P_{10}}{2}$$

II. 1. *a)* 3.95, 4.09, 3.91, 3.98, 4.03, 3.89, 4.05, 3.90, 4.13, 4.01, 3.85, 3.97

$$\bar{X} = 3.9799999 \approx 3.98$$

$$M.D. = 0.068 \approx 0.07$$

3.

a) $M.D. = \frac{900}{72} \approx 12.5$

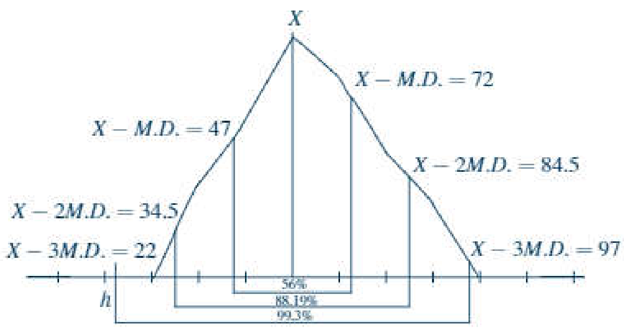
b) $M.D. = \frac{895.76}{72} \approx 12.44$

c) $\bar{X} \pm M.D. = 56.25\%$

$\bar{X} \pm 2M.D. = 88.19\%$

$\bar{X} \pm 3M.D. = 99.3\%$

d)



5. *a)* $Q = 61$

$Q_{SI} = 30.5$

b) $Q = 10.5$

$Q_{SI} = 5.25$

EJERCICIO 14

I. 1. La varianza de un conjunto de elementos se define como la media aritmética de los cuadrados de las desviaciones respecto a la media aritmética.

3. Para un conjunto de elementos numéricos, la desviación típica se define como la raíz cuadrada del cuadrado medio de las desviaciones a la media, es decir, es la raíz cuadrada de la **varianza**.
5. • Si ***a*** es cualquier promedio diferente de la media aritmética, la desviación típica o estándar se puede definir por la expresión matemática $S = \sqrt{\frac{\sum (X-a)^2}{N}}$. De todas las desviaciones típicas resultantes, siempre será mínima para cuando $\boldsymbol{a} = \bar{X}$.
   • Si se tienen distribuciones normales, se hace notar que:
   a) El 62.27% de las observaciones quedan comprendidas entre $\bar{X} \pm S$; es decir, es el valor de la desviación típica o estándar a ambos lados de la media aritmética.
   b) El 95.45% de las observaciones quedan comprendidas entre $\bar{X} \pm 2S$; es decir, es el doble de valor de la desviación típica o estándar a ambos lados de la media aritmética.
   c) El 99.73% de las observaciones quedan comprendidas entre $\bar{X} \pm 3S$; es decir, es el triple de valor de la desviación típica o estándar a ambos lados de la media aritmética
   Si las distribuciones moderadamente asimétricas los porcentajes mencionados suelen mantenerse aproximados.
   • Si suponemos dos conjuntos de elementos numéricos de N_1 y N_2 datos o dos distribuciones de frecuencias totales, con variaciones respectivas de S_1^2 y S_2^2 y que además presenta la misma media aritmética $(\bar{X})$; en tal caso la varianza combinada para ambos conjuntos o ambas distribuciones, se expresa por la siguiente ecuación matemática:

$$S^2 = \frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2}$$

La expresión anterior es la media aritmética ponderada de las varianzas de dichos conjuntos de datos o distribuciones de frecuencias, la misma puede ser para más de 3 conjuntos de datos o distribuciones de frecuencias, siempre y cuando tengan la misma media aritmética.

7. Siempre que determinamos la desviación típica o estándar en datos agrupados, se obtiene algo de error que generalmente se debe al error de agrupar los datos en clases. Para concordar con la realidad de la información y eliminar el **error**, se emplea la **varianza corregida**.
9. La corrección Sheppard para la varianza se aplica en distribuciones continuas donde los **sesgos o colas** tienden en forma gradual a cero.

II. 1. *a)* $S^2 \approx 0.006683$, $S = 0.0818$ *b)* $S^2 = 94.667$, $S = 9.7296$
c) $S^2 = 23.75$, $S = 4.8734$ *d)* $S^2 = 13.16$, $S = 3.6276$

3. *a)* $S^2 = 0.0066833$, $S = 0.08175$ *b)* $S^2 = 94.67$, $S = 9.73$
c) $S^2 = 23.75$, $S = 4.8734$ *d)* $S^2 = 13.16$, $S = 3.63$

5. *a)* Si $a = \bar{x}$, $S = 4.383$; Si $a < \bar{x}$, $S = 4.57$
c) Si $a = \bar{x}$, $S = 3.99$; Si $a > \bar{x}$, $S = 4.19$; Si $a > x$, $S \approx 6.6225$

7. 2 *a*) $\bar{x} \pm S = \approx 67.49\%$; $\bar{x} \pm 2S \approx 90.55\%$; $\bar{x} \pm 3S = 100\%$
   4 *a*) *a*) $\bar{x} \pm S = 59.8\%$; $\bar{x} \pm 2S \approx 95.83\%$; $\bar{x} \pm 3S = 100\%$
   a) $\bar{x} \pm S = 64.16\%$; $\bar{x} \pm 2S \approx 97.27\%$; $\bar{x} \pm 3S = 100\%$

9. 2 *a*) Para x, tenemos que $74 = 74$
   Para S, tenemos que $966 = 966$
   4 *a*) Para $\bar{x}$, tenemos que $-3 = -3$
   Para S, tenemos que $63 = 63$
   9) Para $\bar{x}$, tenemos que $-108 = -108$
   Para S, tenemos que $326 = 326$

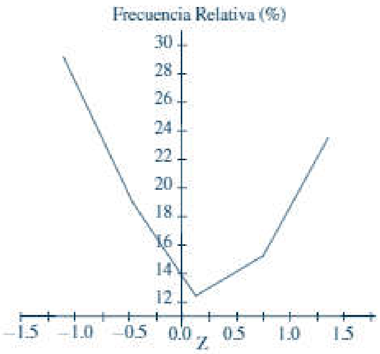
11. *a*) $S^2 = 406673.213$ *b*) Válida
    $S = 637.709$
    c) $S_c^{\,2} = 406670.21$
    $S_c = 637.706$

Los resultados pueden variar, según se haya seleccionado el valor de A.

Ejercicio 15

I. 1. Respectivamente son: $0.7979 = \dfrac{4}{5}$ y $0.6745 = \dfrac{2}{-3}$
   3. Se expresan en las mismas unidades con las que se miden sus elementos variables.
   5. Coeficiente de variación $(V) = \dfrac{S}{\bar{x}}$
   7. Una desventaja del coeficiente de variación se presenta cuando el valor de la media aritmética ($\bar{x}$) está próximo a cero.

II. 1. *a*) No son válidas.
   b) El coeficiente de variación no corregido de la distribución es 70%. El coeficiente de variación no corregido de la distribución es 69.43%.
   c)



3. *a*) Al compararse, se observa que los de cálculo integral presentan la mayor dispersión absoluta.
   b) Al compararse los coeficientes de variación se observa que los de cálculo integral presentan la mayor dispersión relativa con 12.93%.

5.
$$V_Q = \frac{\dfrac{Q_3 - Q_1}{2}}{\dfrac{Q_3 + Q_1}{2}} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

a) $V_Q = 15.60\%$ *b*) $V_Q = 4.02\%$

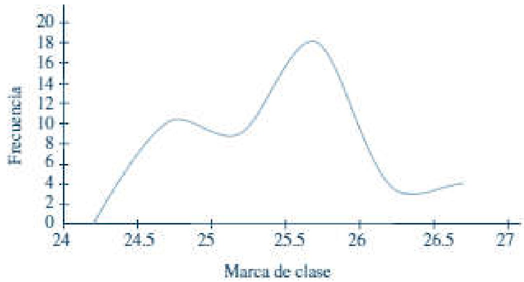
7. *a*) Z de las estaturas es aproximadamente de 0.49.
   Z de los pesos es aproximadamente de 0.51.
   b) La medida de mayor desviación es la de los pesos.

9. 5*a*) No son válidas
   5*b*) No son válidas

Ejercicio 16

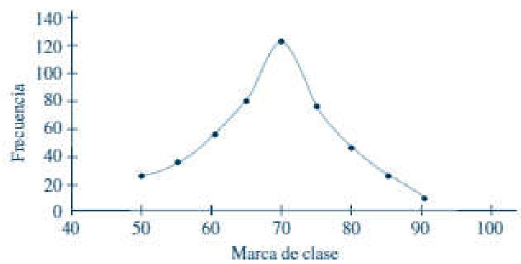
I. 1. Son indicadores que dan información acerca de la manera en que los datos de encuentran contenidos dentro de una distribución.
   3. La asimetría positiva es aquella que gráficamente se encuentra a la derecha y la asimetría negativa es aquella que gráficamente se encuentra a la izquierda.
   5. Si $f_s = 0$ significa que la distribución es simétrica.
   Si $f_s > 0$ significa que la distribución tiene sesgo positivo.
   Si $f_s < 0$ significa que la distribución tiene sesgo negativo.
   7. Es una distribución que contiene una gran concentración de datos en la zona central.
   9. Es una distribución que contiene una baja conglomeración de datos en su región central.

II. 1. *a*)



   b) 25.51
   c) 25.7
   d) 25.7
   e) La distribución es asimétrica

3. *a*)



   b) 68.89
   c) 70
   d) 70
   f) La distribución es simétrica

Ejercicio 17

I. 1. Gráfica que contiene una dispersion de puntos.

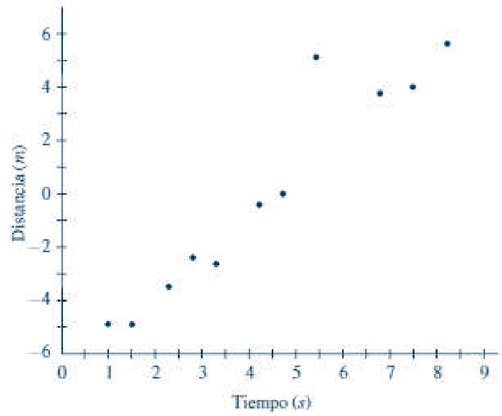
3. Es una cantidad que permite conocer qué tan bien se ajusta un conjunto de pares ordenados a un comportamiento determinado.

5. Los valores que puede tomar son:
Cuando $r = 0$, no existe correlación lineal alguna entre las variables que se estudian; en cambio, si $r = 1$, o bien, $r = -1$, entonces se dice que existe una correlación lineal perfecta entre las variables. Los valores ubicados entre 1 y -1 indican una correlación que no es exacta. Por lo tanto, mientras el valor de r esté más cerca de 1 y -1, la correlación entre las variables se semejará más al comportamiento de una línea recta.

7. Es la recta que mejor se ajusta a la dispersión de puntos en un plano cartesiano.

9. Indica qué tan confiable es la recta de regresión elaborada para un conjunto de datos en una dispersión.

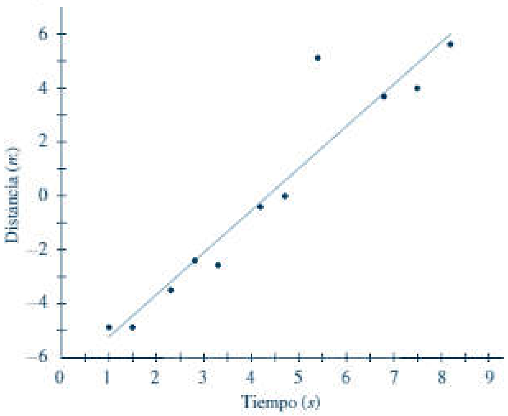
II. 1. *a*)



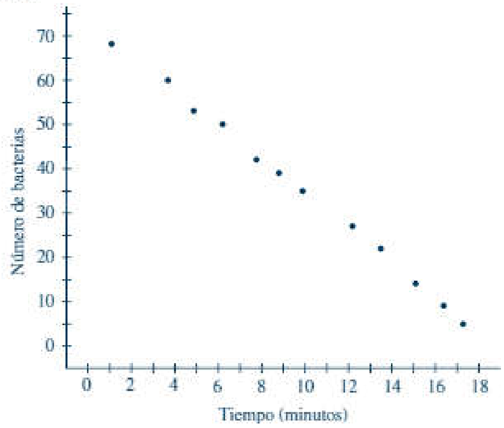
b) El coeficiente de correlación es 0.95311.

c) La ecuación de la recta de regresión es $y = 1.57x - 6.83$

d)



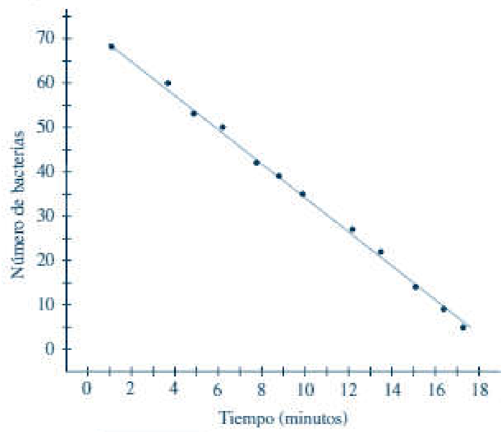
3. *a*)



b) El coeficiente de correlación es -0.9989.

c) La ecuación de la recta de regresión es $y = -3.89x + 73.2$.

d)



e) $S = \sqrt{\dfrac{\sum (Y - Y')}{n-2}} = 0.184$

f) El intervalo de confianza está definido mediante:

$$y = -3.89x + 73.2 \pm 0.184$$

Así, el intervalo de confianza queda definido por las rectas:
$y = -3.89x + 73.384$
$y = -3.89x + 73.016$

Ejercicio 18

I. 1. Un conjunto es el agrupamiento de elementos que no necesariamente implica la existencia de una característica común y evidente o algún tipo de colectividad natural.

3.

a) $C =$ {Guatemala, Belice, Panamá, Nicaragua, Costa Rica, El Salvador, Honduras, Panamá, Cuba, Santo Domingo, Puerto Rico, Haití}

b) $S =$ {Filosofía, Física Moderna, Análisis Matemático, Historia}

c) $A =$ {enero, febrero, marzo, abril, mayo. junio, julio, agosto, septiembre, octubre, noviembre, diciembre}

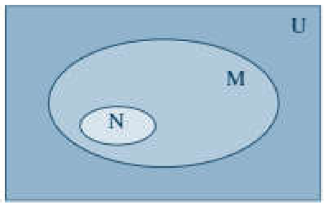
d) $S =$ {martes, miércoles}

e) $P =$ {24, 22, 20, 18, 16, 14, 12, 10, 8, 6, 4, 2}

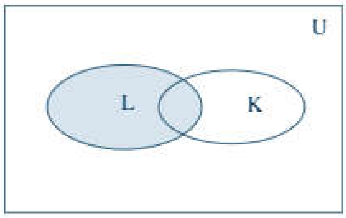
5. *a*) $\{x|x$ son los tres últimos meses del año$\}$.
   b) $P = \{x|x$ es las 5 primeras letras del alfabeto$\}$
   c) $\{x|x$ son los colores del arco iris$\}$
   d) $Q = \{x|x$ son las ramas de las matemáticas$\}$
   e) $\{x|x$ son los integrantes del conjunto musical **The Beatles**$\}$.
7. *a*) Definido.
   b) Se describe o enumerado.
   c) Se describe o enumerado.
   d) definido
   e) Se describe o enumerado.
9. Para indicar **pertenencia** se usa el símbolo $\in$.
11. El conjunto que contiene a los elementos de los conjuntos que se estén considerando en un análisis cualquiera, se denomina **conjunto universal** y se representa por la letra mayúscula **U**.
13. *a*) Sí son iguales.
   b) No son iguales
   c) Sí son iguales.
   d) No son iguales
   e) Sí son iguales.
15. *a*) Sí
   b) No
   c) Sí
   d) No
   e) Sí
17. Cuando un conjunto cualquiera ***A*** en el que todos sus elementos son también miembros de otro conjunto ***B***, se dice que el conjunto ***A*** es subconjunto del conjunto ***B***.
19. *a*) $B = \{1, 3, 5, 7, 9\}$ es decir $B \subset A$
   c) $D = \{2, 3, 5, 7\}$ es decir $D \subset A$
21. Un conjunto potencial es el número de subconjuntos resultantes de cualquier conjunto dado, se obtienen por la expresión $T = 2^n$.
23. *a*) finito *d*) infinito
   b) finito *e*) infinito
   c) infinito *f*) finito

II. 1. *a*) $A \cup B = \{1, 2, 3, 4, 5, 6, 7, 9\}$
   $A \cap B = \{1, 3, 5, 7\}$
   $A - B = \{2, 4, 6\}$
   b) $A \cup B =$ {Brasil, México, Colombia, Venezuela, Cánada, Perú, Argentina, Ecuador}
   $A \cap B =$ {México, Venezuela}
   $A - B =$ {Brasil, Colombia, Cánada}
   c) $A \cup B =$ {a, b, c, d, e, f, 1, 3, 5, 7, 9}
   $A \cap B = \{\ \}$
   $A - B =$ {a, b, c, d, e, f}
   d) $A \cup B =$ {rojo, verde, azul, café, amarillo, gris}
   $A \cap B =$ {rojo, café}
   $A - B =$ {verde, azul}
   e) $A \cup B =$ {Pedro, Héctor, David, Álvaro, Manuel, Alberto, Samuel, Rubén}
   $A \cap B =$ {Héctor, Álvaro}
   $A - B =$ {Pedro, Manuel, David}
3. *a*) $A \times B = \{(1,1), (1,3), (1,5), (1,7), (1,9), (2,1), (2,3), (2,5), (2,7), (2,9), (3,1), (3,3), (3,5), (3,7), (3,9), (4,1), (4,3), (4,5), (4,7), (4,9), (5,1), (5,3), (5,5), (5,7), (5,9), (6,1), (6,3), (6,5), (6,7), (6,9), (7,1), (7,3), (7,5), (7,7), (7,9)\}$
   b) $A \times B =$ {(Brasil, Perú), (Brasil, Argentina), (Brasil, México), (Brasil, Ecuador), (Brasil, Venezuela), (México, Perú) , (México, Argentina), (México, México), (México, Ecuador), (México, Venezuela), (Colombia, Perú), (Colombia, Argentina), (Colombia, México), (Colombia, Ecuador), (Colombia, Venezuela), (Venezuela, Perú) , (Venezuela, Argentina), (Venezuela, México), (Venezuela, Ecuador), (Venezuela, Venezuela), (Cánada, Perú), (Cánada, Argentina), (Cánada, México), (Cánada, Ecuador), (Cánada, Venezuela)}
   c) $A \times B =$ {(a,1), (a,3), (a,5), (a,7), (a,9), (b,1), (b,1), (b,3), (b,5), (b,7), (b,9), (c,1), (c,3), (c,5), (c,7), (c,9), (d,1), (d,3), (d,5), (d,7), (d,9), (e,1), (e,3), (e,5), (e,7), (e,9), (f,1), (f,3), (f,5), (f,7), (f,9)}
   d) $A \times B =$ {(rojo, amarillo), (rojo, rojo), (rojo, café), (rojo, gris), (verde, amarillo), (verde, verde), (verde, café), (verde, gris), (azul, amarillo), (azul, azul), (azul, café), (azul, gris), (café, amarillo), (café, café), (café, café), (café, gris)}
   e) $A \times B =$ {(Pedro, Héctor), (Pedro, Alberto), (Pedro, Samuel), (Pedro, Álvaro), (Pedro, Rubén), (Héctor, Héctor), (Héctor, Alberto), (Héctor, Samuel), (Héctor, Álvaro), (Héctor, Rubén), (David, Héctor), (David, Alberto), (David, Samuel), (David, Álvaro), (David, Rubén), (Álvaro, Héctor), (Álvaro, Alberto), (Álvaro, Samuel), (Álvaro, Álvaro), (Álvaro, Rubén), (Manuel, Héctor), (Manuel, Alberto), (Manuel, Samuel), (Manuel, Álvaro), (Manuel, Rubén)}
5. $\{(a, a), (a, s), (s, a), (s, s)\}$
7. $\{x|x$ sea todos los elementos del conjunto A y $B\}$
9. $\{x|x$ sea todos los elementos del conjunto A no contenidos en el conjunto $B\}$.

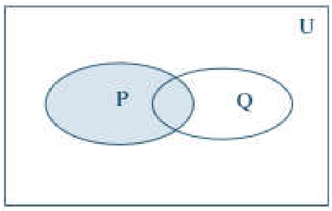
III. 1.
   a) $M \cup N'$



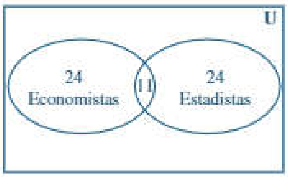
   b) $K' \cap L$



   c) $(P' \cup Q)'$



3. *a*) $(A \cup B) \cap C' = \{2, 4, 6, 10, 13\}$
   b) $A' \cap (B' \cap C') = \{7, 8, 12\}$
   c) $(A \cup B)' \cap C' = \{7, 8, 12\}$
   d) $A \cup (B \cap C') = \{3,4,5,6,10,11,13\}$
   e) $(A' \cap B') \cap C' = \{7, 8, 12\}$
   f) $[(A \cup B) \cup C]' = \{7,8,12\}$
5.



Ejercicio 19

I. 1. El diagrama de árbol, el principio fundamental del conteo, la notación factorial, teoremas del binomio, triángulo de Pascal, las permutaciones, las combinaciones y las particiones.

3. El principio fundamental del conteo establece que una operación puede efectuarse independientemente de n, maneras diferentes y si continuando el procedimiento una segunda operación puede efectuarse independientemente de n_2 maneras diferentes y si, después de realizadas, una tercera operación puede efectuarse independientemente de n_3 maneras diferentes y así sucesivamente (para un número finito arbitrario de operaciones), entonces el número total de maneras de las cuales pueden efectuarse todas las operaciones en el orden indicado es el producto $n_1\, n_2 n_3 \ldots n_r$.

5. La base del teorema del binomio es la multiplicación de polinomios y la ley distributiva.

7. Cuando n es un número real cualquiera diferente de un entero positivo, el desarrollo de $(1+x)^n$ no tiene fin, es decir:

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{(2)(1)}x^1 + \frac{n(n-1)(n-2)}{(3)(2)(1)}x^3 +$$

$$\frac{n(n-1)(n-2)(n-3)}{(4)(3)(2)(1)}x^4 + \ldots + \frac{n(n-1)\ldots(n-r+2)}{(r-1)!}x^{r-1} + \ldots$$

Se puede escribir tantos términos como queramos de un desarrollo infinito, para todo n que no sea un entero positivo si $|x| < 1$.

9. 1. El primero y el último número de cada fila es 1.
   2. Cada uno de los otros números de la ordenación se obtienen sumando los dos números que aparecen directamente encima de él.

11. Representa el número total de permutaciones posibles de n objetos distintos, tomados de n en n.

13. Representa el número total de permutaciones de n objetos distintos, tomados de r en r, siendo $r \leq n$.

15. Se define como el arreglo posible de n objetos alrededor de un círculo (o cualquier otra curva simple cerrada), donde uno de ellos mantiene una posición fija.

17. Consiste en un número de permutaciones diferentes de n objetos, tomados todos a la vez, de los cuales hay n_1 igual entre sí, otros n_2 iguales entre sí, etcétera.

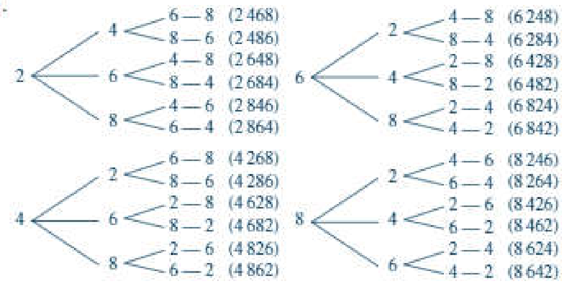
19. Pruebas con sustitución y pruebas sin sustitución.

21. Una prueba ordenada de tamaño r sin sustitución es sencillamente una permutación r de n objetos.

23. En una combinación carece de importancia el orden de sus elementos, mientras que en una permutación el orden de sus elementos es de primordial importancia.

25. Una partición de un conjunto x es una subdivisión de x entre subconjuntos no vacíos que son disjuntos y cuya reunión es x.

II. 1.



Total: 24

3. {(2, *k*, 3), (2, *k*, 5), (2, *k*, 7), (2, 1, 3), (2, 1, 5), (2, 1, 7), (2, *m*, 3), (2, *m*, 5), (2, *m*, 7), (4, *k*, 3), (4, *k*, 5), (4, *k*, 7), (4, 1, 3), (4, 1, 5), (4, 1, 7), (4, *m*, 3), (4, *m*, 5), (4, *m*, 7), (6, *k*, 3), (6, *k*, 5), (6, k, 7), (6, 1, 3), (6, 1,5), (6,1, 7), (6, *m*, 3), (6, m, 5), (6, *m*, 7)}.

5. Puede escoger un pantalón, una camisa y luego una corbata de 60 maneras.

7. Se pueden construir 132 480 planillas diferentes.

9. *a*) Se pueden construir 207 900 placas diferentes.
   b) Se pueden construir 217 800 placas diferentes.

11. *a*) 5 040 *c*) 40 320
    b) 39 916 800 *d*) 6 227 020 800

13. *a*) $x^7 + 7x^6y + 21x^5y^2 + 35\,x^4y^3 + 35x^33y^4 + 21x^2y^5 + 7xy^6 + y^7$.
    c) $3125x^5 - 3125x^4y^2 + 1250x^3y^4 - 250x^2y^6 + 25xy^8 - y^{10}$.
    e) $x^3 - 9x^{\frac{8}{3}}\, y^{\frac{1}{3}} + 36x^{\frac{7}{3}}y^{\frac{2}{3}} - 84x^2y^1 + 126x^{\frac{5}{3}}y^{\frac{1}{3}} - 126x^{\frac{4}{3}}y^{\frac{5}{3}}$
    $+84x^1y^2 - 36x^{\frac{2}{3}}\, y^{\frac{7}{3}} + 9x^{\frac{1}{3}}\, y^{\frac{8}{3}} - y^3$.

15. 1, 9, 36, 84, 126, 126, 84, 36, 9, 1.
    1, 10, 45, 120, 210, 252, 210, 120, 45, 10, 1.
    1, 11, 55, 165, 330, 462, 462, 330, 165, 55, 11, 1.

17. 12 144 designaciones.

19. Se puede acomodar de 144 maneras.

21. Hay 48 maneras.

23. El número de maneras en las que pueden acomodarse 13 personas alrededor de una mesa redonda es de 4 790 001 600.

25. *a*) pueden sentarse de 40 320 maneras.
    b) 30 240 maneras

27. El número de maneras en las que pueden sentarse 9 personas alrededor de una mesa, si dos personas en particular deben sentarse siempre juntos es 10 080.

29. 144 maneras.

31. *a*) 10 000 *b*) 5 040

33. 658 008 maneras.

35. El comité puede relacionarse de 4410 maneras.

37. *a*) 5 005 *b*)360 360

39. 462 elecciones.

41. *a*) 330 maneras
    b) 165 formas
    c) 330 maneras

43. 7 560 maneras.

Ejercicio 20

I. 1. La probabilidad es un hecho que puede suceder o no suceder y que está condicionada a situaciones del azar y de la naturaleza.

3. Un suceso E, donde hay un total de n situaciones posibles, todas igualmente probables, de las cuales una deberá ocurrir, es decir, puede presentarse en s de las situaciones.

5.
$$\text{Probabilidad del no evento} = \frac{(\text{Número de resultados posibles del experimento}) - (\text{Número de maneras en que puede ocurrir el evento})}{\text{Número de resultados posibles del experimento}}$$

7. La probabilidad de un suceso posible es 1 o 100%.

9. *a*) En favor de la aparición ($p : q$), que se lee p a q.
   b) En contra de su aparición ($q : p$), que se lee q a p.

11. $P_f = \dfrac{S}{n}$

13. El conjunto de todos los resultados posibles de un experimento, se denomina **espacio muestral** de dicho experimento.

15. Es un conjunto de resultados, es decir, es cualquier subconjunto de un espacio muestral

17. La probabilidad de cada evento simple o elemental es un número mayor o igual a cero, La suma de las probabilidades de los eventos simples o elementales de un espacio muestral es igual a la unidad. Si un experimento tiene n resultados igualmente probables, la probabilidad de cada uno de los n eventos simples o elementales igualmente probables es $1/n$.
La probabilidad de un evento E es igual a la suma de las probabilidades de los eventos simples o elementales de los que E es la unión.
La probabilidad del conjunto vacío ($\varnothing$) es igual a cero.

II. 1. *a*) $\dfrac{3}{8}$ *b*) $\dfrac{7}{8}$ *a*) $\dfrac{1}{8}$

3. Los resultados posibles del experimento son: {(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)}.
La probabilidad de que la suma de los dos números escritos en las papeletas sea 5, es: $\dfrac{1}{3}$.
La probabilidad de que ambos números sean menores de 3, es: $\dfrac{1}{3}$.

5. $\dfrac{31}{32}$

7. *a*) $\dfrac{1}{2}$ *d*) $\dfrac{11}{16}$
   b) $\dfrac{5}{16}$ *e*) $\dfrac{11}{16}$
   c) $\dfrac{3}{16}$

9. 0.56

11. El espacio muestral es: {(5, 10), (5, 20), (10, 5), (10, 20), (20, 5), (20, 10)}. El evento simple en que los billetes sacados sean el de cinco y el de veinte, es: {5, 20} y {20, 5}.

13. $S = \{10, 11, 12,..., 19, 20, 21,..., 29, 30, 31,..., 39, 40, 41,..., 49, 50, 51,..., 59, 60, 61,..., 69, 70, 71,..., 79, 80, 81,..., 89, 90, 91,..., 98, 99\}$

$$P = \frac{1}{5} = 0.2$$

15. $p = \dfrac{1}{6} = 0.2$

17. *a*) $p = 0.018$ *b*) $p = 0.254$ *c*) $p = 0.018$

Ejercicio 21

I. 1. Dos eventos E_1 y E_2 son **mutuamente exclusivos** si no pueden ocurrir simultáneamente

3. *a*) En caso de que E_1 y E_2 son mutuamente exclusivos ($E_1 \cap E_2 = \varnothing$) se tiene la ecuación:
$$p(E_1 \cup E_2) = p(E_1) + p(E_2) = p(E)$$
   b) En caso de que E_1 y E_2 no son mutuamente exclusivos ($E_1 \cap E_2 \neq \varnothing$)se tiene la ecuación:
$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2) = p(E)$$

5. • Si E es un evento de S entonces la probabilidad del evento E es: $0 \leq p(E) \leq 1$
• Si el espacio muestral S se obtiene de un experimento aleatorio entonces: $p(S) = 1$.
• Si E_1 y E_2 son eventos mutuamente exclusivos entonces: $p(E_1 \cup E_2) = p(E_1) + p(E_2)$.
• Si $E_1, E_2, E_3..., E_i$) son eventos mutuamente exclusivos e infinitos entonces: $p(E_1 \cup E_2 \cup E_3... \cup E_i) = p(E_1) + p(E_2) + p(E_3) + ... + p(E_i)$.
• Si $E_1, E_2, E_3..., E_n$ son eventos mutuamente exclusivos y finitos entonces: $p(E_1 \cup E_2 \cup E_3... \cup E_n) = p(E_1) + p(E_2) + p(E_3) + ... + p(E_n)$.

II. 1. *a*) $S = (AA1, AA2, AA3, AA4, AA5, AA6, AS1, AS2, AS3, AS4, AS5, AS6, SA1, SA2, SA3, SA4, SA5, SA6, SS1, SS2, SS3, SS4, SS5, SS6)$.
   b) $E_1 = \{SS2, SS3, SS5\}$; $E_2 = \{AA6, AS6, SA, SS6\}$.
   $E_3 = \{AS2, AS3, AS5, SA2, SA3, SA5\}$.
   c) $E = \{SS2, SS3, SS5, AA6, AS6, SA6, SS6\}$.
   $E = \{AA6, AS6, SA6, SS6\}$.
   $E = \{AA6, AS6, SA6, SS6, AS2, AS3, AS5, SA2, SA3, SA5\}$.
   d) $p(E) = \dfrac{7}{24}$; $p(E) = \dfrac{1}{6}$; $p(E) = \dfrac{5}{12}$

3. $p = \dfrac{2}{3}$

5. $p = \dfrac{1}{4}$; $p = \dfrac{7}{30}$; $p = \dfrac{31}{60}$

7. La probabilidad de que no sea titulado y no se dedique a las clases de Matemáticas es de 0.1.

9. *a*) $p = \left(\dfrac{1}{6}\right)^2\left(\dfrac{5}{6}\right)^5 = \dfrac{3125}{279936}$
   b) $p = 21\left(\dfrac{1}{6}\right)^2\left(\dfrac{5}{6}\right)^5$

11. *a*) 298 *d*) 447
    b) 181 *e*) 29
    c) 298 *f*) 478

13. *a*) $p(E_{1'}) = 0.5$ o 50%
    b) $p(E_{2'}) = 0.5$ o 50%
    c) $p(E_1 \cap E_2) = 0.33$ o 33%
    d) $p(E_{1'} \cap E_{2'}) = 0.33$ o 33%

e) $p(E_1 \cap E_2) = 0.66$ o 66%

f) $p(E_1 - E_2) = \frac{1}{6}$

g) $p(E_2 - E_1) = \frac{1}{6}$

15. $p = \frac{7}{9}$

Ejercicio 22

I. 1. Para determinar la probabilidad de un evento E_2 sabiendo que ya ha ocurrido un evento E_1 no se considera como espacio muestral S, sino que se considera al evento E_1 (Subconjunto de S) como el **espacio muestral reducido**.

3. Se establece que dos eventos son **independientes** si la ocurrencia de uno de ellos no afecta de ninguna manera la ocurrencia o no ocurrencia del otro.

5. Cuando la ocurrencia de un evento E_1 afecta a la probabilidad de ocurrencia de un evento E_2 entonces se dice que E_1 y E_2 son eventos dependientes.

7. Consiste en una ecuación que permite determinar probabilidades condicionales. Se aplica a eventos diferentes donde al menos uno de ellos se sabe que ha ocurrido.

"Si un evento puede ocurrir en más de una forma, entonces la probabilidad de que ocurra en una forma particular será igual a la razón de la probabilidad de que se presente la forma respecto a la probabilidad de que ocurra".

$$p(E_1 / F) = \frac{p(E_1)\,p(F / E_1)}{p(E_1)\,p(F / E_1) + p(E_2)\,p(F / E_2) + p(E_3)\,p(F / E_3)}$$

II. 1. $p(E_2 / E_1) = \frac{2}{7}$

3. $p(E_2 / E_1) = \frac{2}{15}$

5. $p = \frac{6041}{15\,180}$

7. $p = \frac{39}{253}$

9. a) $p = \frac{2}{3}$; b) $p = \frac{1}{2}$; c) $p = \frac{1}{2}$

11. a) $p = \frac{33}{98}$; b) $p = \frac{3}{5}$

13. a) $p = \frac{1}{12}$; b) $p = \frac{1}{2}$; c) $p = \frac{1}{2}$; d) $p = \frac{1}{4}$;

15. $p = \frac{16}{35}$

Visítenos en:
www.pearsonenespañol.com

ISBN 978-607-32-2783-4