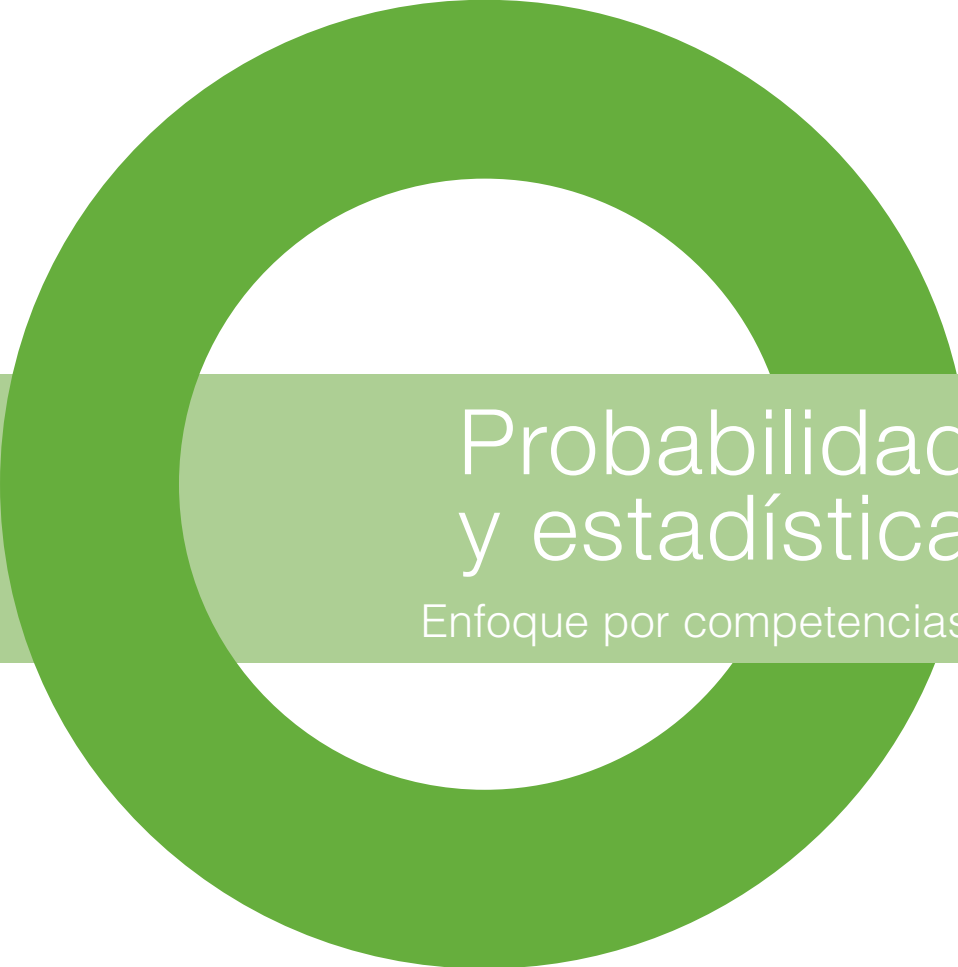




Probabilidad y estadística

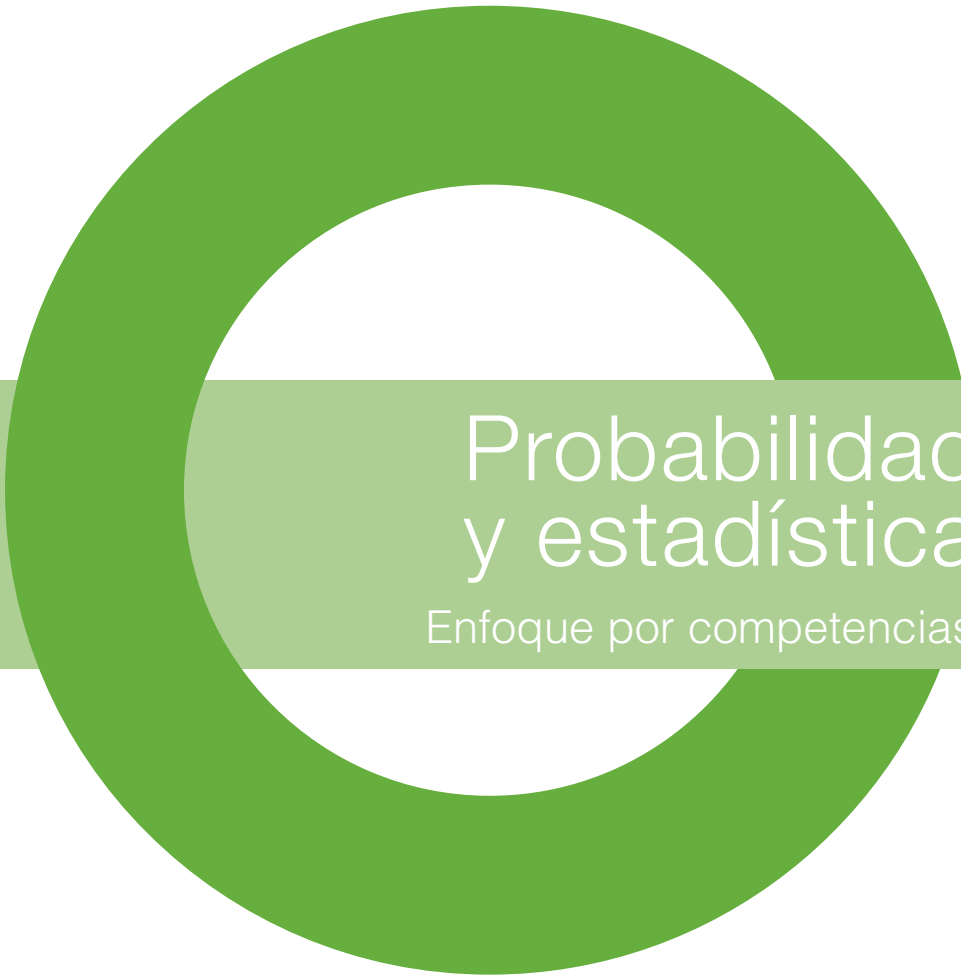
Enfoque por competencias

Ana Laura Gutiérrez Banegas



Probabilidad y estadística

Enfoque por competencias



Probabilidad y estadística

Enfoque por competencias

Ana Gutiérrez Banegas

Instituto Tecnológico Autónomo de México (ITAM)

Revisión técnica

Javier Sabaria León

Instituto Politécnico Nacional

Eduardo Martínez Márquez

Dirección General de Bachillerato



MÉXICO • BOGOTÁ • BUENOS AIRES • CARACAS • GUATEMALA • MADRID • NUEVA YORK
SAN JUAN • SANTIAGO • SAO PAULO • AUCKLAND • LONDRES • MILÁN • MONTREAL
NUEVA DELHI • SAN FRANCISCO • SINGAPUR • ST. LOUIS • SIDNEY • TORONTO

Coordinadora editorial: Alejandra Martínez Ávila
Editor sponsor: Sergio G. López Hernández
Supervisora de producción: Marxa de la Rosa Pliego
Diseño de portada: José Palacios Hernández

Probabilidad y estadística

Enfoque por competencias

Prohibida la reproducción total o parcial de esta obra,
por cualquier medio, sin la autorización escrita del editor.



Educación

DERECHOS RESERVADOS © 2012, respecto a la primera edición por:
McGRAW-HILL / INTERAMERICANA EDITORES S.A. DE C.V.
A Subsidiary of The McGraw-Hill Companies, Inc.

Punta Santa Fe,
Prolongación Paseo de la Reforma 1015,
Torre A, piso 17,
Colonia Desarrollo Santa Fe,
Delegación Álvaro Obregón,
C.P. 01376, México D.F.

Miembro de la Cámara Nacional de la Industria Editorial Mexicana, Reg. Núm. 736

ISBN: 978-607-15-0712-9

1234567890
Impreso en México

1098765432
Printed in Mexico

Dedicatoria



A mis alumnos del Tec de Monterrey.
A mis profesores del ITAM.
A mi familia.
A Piotr Wisniewski.
A mis amigos de la “hora de la crepa”
A Gustavo Fondevila
A mis jefas Tere Coteria, Sandra Viscarra y Linda Medina
¡Gracias por sus consejos y todo su apoyo!

Contenido

Prólogo	X
Rúbrica de evaluación	XII

Unidad 1	Recolección de datos	2
	<i>Ubícate</i>	3
	Fundamentos de la estadística	4
	La estadística y su utilidad	5
	Clasificación de la estadística	6
	Términos básicos	7
	Fuentes	9
	Tipos de variables	9
	Censo y muestreo	10
	Métodos de muestreo	10
	Muestreo aleatorio simple	11
	Muestreo aleatorio sistemático	11
	Muestreo aleatorio estratificado	12
	Muestreo por conglomerados	12
	<i>Ejercicios</i>	13
	EVALUACIÓN SUMATIVA. UN ESCALÓN HACIA ARRIBA	16
	PROYECTO 1. TRANSPORTE PÚBLICO	17

Unidad 2	Presentación de datos	18
	<i>Ubícate</i>	19
	Distribuciones de frecuencias y gráficas	20
	Distribuciones de frecuencias	21
	Histograma	25
	Polígono de frecuencias	25
	Ojiva	25
	Hoja de cálculo para graficar	27
	Diagrama de barras	29
	Diagrama circular o de pastel	30
	<i>Ejercicios</i>	34
	EVALUACIÓN SUMATIVA. UN ESCALÓN HACIA ARRIBA	41
	PROYECTO 2. REMESAS INTERNACIONALES RECIBIDAS EN MÉXICO	43

Unidad 3	Medidas de tendencia central y de dispersión	44
	<i>Ubícate</i>	45
	Medidas de tendencia central y de dispersión	46
	Medidas de tendencia central para datos no agrupados	47
	Media	47
	Mediana	48
	Moda	49
	Sesgo	50
	Medidas de dispersión para datos no agrupados	51
	Rango	51

Varianza	52
Desviación estándar	53
Medidas de tendencia central para datos agrupados	57
Media, mediana y moda aproximadas	57
Medidas de dispersión para datos agrupados	57
Varianza aproximada	57
Desviación estándar	58
Ejercicios	60
EVALUACIÓN SUMATIVA. UN ESCALÓN HACIA ARRIBA	68
PROYECTO 3. CALIDAD DEL AIRE Y CAMBIO CLIMÁTICO	70

Unidad 4 Probabilidad y teoría de conjuntos 72

Ubícate	73
Probabilidad	74
Teoría de conjuntos	75
Descripción de conjuntos	75
Conjunto universal	76
Subconjunto	77
Conjunto vacío	78
Cardinalidad de un conjunto	78
Operaciones con conjuntos	79
Probabilidad	88
Conceptos básicos	88
Enfoques de la probabilidad	88
Ejercicios	94
EVALUACIÓN SUMATIVA. UN ESCALÓN HACIA ARRIBA	100
PROYECTO 4. VIOLENCIA EN EL NOVIAZGO	102

Unidad 5 Técnicas de conteo 104

Ubícate	105
Introducción	106
Diagrama de árbol	107
Técnicas de conteo	109
Notación factorial	109
Principio multiplicativo	109
Principio aditivo	111
Permutaciones	111
Combinaciones	112
Ejercicios	115
EVALUACIÓN SUMATIVA. UN ESCALÓN HACIA ARRIBA	120
PROYECTO 5. RUTA ÓPTIMA	121

Unidad 6 Probabilidad conjunta 122

Ubícate	123
Probabilidad conjunta	124

Composición de eventos	125
Definición y representación de eventos	125
Probabilidad de la unión	125
Probabilidad de la intersección	125
Probabilidad del complemento	126
Probabilidad de eventos disjuntos o mutuamente excluyentes	126
Probabilidad de la diferencia	127
Regla de la adición para eventos no excluyentes	128
Regla de la adición para eventos disjuntos o mutuamente excluyentes	129
Independencia de eventos	130
Visualización de eventos: tablas de contingencia y árboles de probabilidad	130
Tabla de contingencia	130
Árbol de probabilidad	131
Probabilidad condicional	132
Teorema de Bayes	135
Ejercicios	138
EVALUACIÓN SUMATIVA. UN ESCALÓN HACIA ARRIBA	144
PROYECTO 6. Sobrepeso y obesidad	145

Unidad 7

Distribución de probabilidad de variables aleatorias discretas 146

<i>Ubícate</i>	147
Distribución de probabilidades de variables aleatorias discretas	148
Variable aleatoria discreta	149
Distribución de probabilidad	149
Media y desviación estándar	152
Media o valor esperado	152
Varianza y desviación estándar	152
Distribución binomial	153
Función de probabilidad	154
Media y desviación estándar	157
Distribución de Poisson	157
Función de probabilidad	158
Media y desviación estándar	161
Aproximación de la distribución binomial a la distribución de Poisson	161
Ejercicios	163
EVALUACIÓN SUMATIVA. UN ESCALÓN HACIA ARRIBA	168
PROYECTO 7. HÁBITOS CULTURALES	169

Unidad 8

Distribución de probabilidad de variables aleatorias continuas 170

<i>Ubícate</i>	171
Función de densidad	172
Variable aleatoria continua	173
Función de densidad	173
Media y desviación estándar de una variable continua	173

Distribución normal	176
Función de densidad	176
Distribución normal estándar	177
Función de densidad	178
Área bajo la curva	178
Aproximación de la normal a la binomial	182
Ejercicios	184
EVALUACIÓN SUMATIVA. UN ESCALÓN HACIA ARRIBA	190
PROYECTO 8. RESERVAS DE PETRÓLEO CRUDO	191

Unidad 9 Análisis de datos de dos variables 192

Ubícate	193
Análisis de datos de dos variables	194
Representación	195
Tabla de contingencia	195
Diagrama de dispersión	195
Análisis de correlación	196
Covarianza	196
Coeficiente de correlación	199
Coeficiente de determinación	202
Regresión lineal simple	202
Método de mínimos cuadrados	204
Ejercicios	210
EVALUACIÓN SUMATIVA. UN ESCALÓN HACIA ARRIBA	220
PROYECTO 9. ACCIDENTES DE TRÁNSITO	221

Unidad 10 La sumatoria 222

Ubícate	223
Propiedades de la sumatoria	224
Notación	225
Propiedades de la sumatoria	225
Sumatorias especiales	226
Ejercicios	229
EVALUACIÓN. UN ESCALÓN HACIA ARRIBA	231

Apéndices 232

A. Breve historia de la estadística	232
B. Tabla de distribución acumulada normal $N(0, 1)$	236
C. Instalación del complemento Análisis de datos de Excel	238
Bibliografía y sitios de interés en internet	241

Prólogo

Este libro que tienes en las manos surge fundamentalmente de una inquietud personal por escribir un texto que presentara ejemplos, ejercicios y casos sobre situaciones reales, más apegadas a lo que sucede tanto en México como en Latinoamérica para mostrar a las personas que la probabilidad y la estadística no son materias aburridas ni temibles.

El objetivo de la obra, sustentada en el enfoque por competencias, es ofrecerte una visión práctica de la probabilidad y la estadística, aunque sin dejar de lado la teoría. Por ello, se presentan sus aplicaciones en distintos ámbitos, como deportes, economía, demografía y mercadotecnia, entre otros. Al respecto, cabe destacar que en la mayoría de los ejercicios que se presentan se utilizan datos reales.

Es importante señalar que el libro abarca los temas correspondientes a dos semestres de bachillerato (Probabilidad y estadística I y Probabilidad y estadística II), pero resultará útil también a quienes cursan estudios “más allá” del bachillerato que tengan curiosidad por aprender acerca de estas dos áreas de la matemática.

La estructura de cada bloque permite:

- Revisar, por medio de exámenes de diagnóstico denominados “Ubícate”, cuánto sabes acerca de los conceptos necesarios para aprender los temas expuestos en cada bloque.
- Conocer, por medio de anécdotas o pasajes curiosos, a las personas que fueron desarrollando la probabilidad y la estadística.
- Comprender los conceptos mediante la teoría y los ejemplos, los cuales se resuelven paso a paso.
- Aprender el uso de tecnologías, fundamentalmente la hoja de cálculo (Excel de Microsoft®).
- Reforzar los conocimientos adquiridos a lo largo de los bloques por medio de la sección de ejercicios correspondiente.
- Probar tus conocimientos, mediante las evaluaciones al final de cada bloque.
- Reflexionar sobre algún dilema ético, presentado al final de cada bloque con un Proyecto real, al tiempo que resuelves problemas relativos a la estadística o a la probabilidad.

El libro se forma de los bloques siguientes:

- **1. Recolección de datos.** Se trata de un bloque introductorio, en el que se exponen los conceptos básicos empleados a lo largo del texto y se ofrece una explicación breve de algunos tipos de muestreo usados para obtener datos.
- **2. Presentación de datos.** En este bloque se explica cómo se pueden presentar, mediante gráficas y tablas, los datos recolectados.
- **3. Medidas de tendencia central y de dispersión.** En este bloque se presentan las medidas más importantes para caracterizar un conjunto de datos.
- **4, 5 y 6. Probabilidad, técnicas de conteo y probabilidad conjunta, respectivamente.** A lo largo de estos tres bloques se explican los conceptos que sustentan la teoría de probabilidad. Se hace, asimismo, una revisión de la teoría de conjuntos.
- **7 y 8. Distribuciones de probabilidad de variables aleatorias discretas y continuas, respectivamente.** Debido a su importancia y para revisarlas con mayor detalle, se dedica un bloque exclusivo a las distribuciones de probabilidad de variables aleatorias discretas y otro a las variables aleatorias continuas. En estos dos bloques se sigue revisando el cálculo de probabilidades para situaciones específicas.

- ▶ **9. Análisis de datos de dos variables.** En este último bloque se presenta una introducción al método de regresión lineal simple.
- ▶ **10. Sumatoria.** Si bien este no es un tema común en los planes de estudio, en muchas fórmulas de la estadística y de la probabilidad se aplica esta operación, de modo que para comprenderla cabalmente se decidió incluir este bloque dedicado a la notación sigma. Debido a su importancia se propone como bloque opcional al programa de estudio.

Hacia el final del libro se incluyen tres apéndices. En el *A* se relata una breve historia de la estadística; en el *B* se incluye la tabla de distribución acumulada normal estándar, y en el *C* se explica cómo instalar complementos de Excel necesarios para realizar algunas operaciones estadísticas. Cabe señalar que en la OLC se presentan los resultados de los ejercicios impares. Al final hallarás bibliografía y ligas de internet que pueden resultar interesantes.

Cabe resaltar que este libro no sólo se centra en el estudio de la probabilidad y la estadística y en la resolución de problemas, sino que intenta hacer reflexionar a las personas sobre las situaciones que se plantean en los ejercicios, permitiendo el trabajo en equipo, el intercambio de opiniones, la tolerancia y el respeto entre los participantes, elementos importantes para el proceso de aprendizaje. Espero lograrlo.

Ana Gutiérrez Banegas

Rúbrica de evaluación

La rúbrica siguiente permite llevar a cabo una evaluación multidimensional (no centrada en los conocimientos adquiridos) que se sugiere realizar a medida que se avance en el estudio de cada bloque.

Aspecto por evaluar	4 (100%)	3 (80%)	2 (50%)	1 (20%)
Conocimientos (20 %)	Conocimientos muy bien aprendidos y relacionados con los del bloque. Considera y discute planteamientos alternativos relacionados con diversos aspectos del tema.	Establece correctamente los conceptos centrales y su relación con los conocimientos académicos del bloque.	Establece los conceptos centrales.	Identifica sólo algunos conceptos centrales.
Investigación (20%)	Analiza y sintetiza información de múltiples fuentes proporcionando referencias. Valora la veracidad de las distintas fuentes. Analiza la relación de la información con el contexto global y específico del problema por indagar.	Analiza datos de múltiples fuentes, incluidas referencias relacionadas directamente con el contexto del problema.	Analiza datos de varias fuentes y proporciona las referencias correspondientes.	Considera información extraída de una sola fuente sin indicar las referencias.
Conceptos (20%)	Todos los conceptos estadísticos son correctos y están bien fundamentados.	Todos los conceptos estadísticos son correctos.	Algunos conceptos estadísticos son confusos o incorrectos.	Conceptos estadísticos confusos o incorrectos.
Profundidad (20%)	Plantea correctamente el problema y justifica el proceso, apoyándose en los conceptos estadísticos correspondientes de manera correcta.	Plantea la mayor parte del problema, y justifica el proceso, apoyándose en los conceptos estadísticos correspondientes de manera casi correcta.	Plantea parte del problema y justifica una parte del proceso, pero en general los conceptos estadísticos se tratan de manera poco correcta.	Superficial. No planteamiento ni se justifica el proceso. Se trata el problema de manera incorrecta.
Solución (20%)	Analiza múltiples opciones apoyadas en la investigación. Identifica claramente las suposiciones involucradas en el análisis y dentro del contexto del problema.	Presenta diversas opciones y basado en la investigación discute los motivos para escoger una solución.	Presenta una solución fundamentada en la investigación y considera el grado de adecuación de la solución al contexto del escenario.	Presenta una solución apoyada de alguna manera en la investigación realizada.

Rúbrica de evaluación para los proyectos

Esta actividad se evaluará, mediante el reporte final elaborado por equipo de acuerdo con la rúbrica de evaluación que a continuación se presenta y tomando en cuenta el trabajo desarrollado en el proceso.

Rúbrica de evaluación para la discusión

Criterio / Nivel de desempeño	Siempre	Casi siempre	A veces	Nunca
Refleja el esfuerzo por entender el problema.				
Refleja el esfuerzo por integrar la información.				
Presenta argumentos sólidos en sus respuestas.				
Demuestra comprensión acerca del papel de la ética.				
Aporta conclusiones al debate y al caso.				



Competencias

- Estructura ideas y argumentos de manera clara, coherente, sintética, relacionadas con la estadística descriptiva e inferencial
- Argumenta el uso de la estadística descriptiva e inferencial en la solución de un problema
- Elige una técnica de recolección de datos para el estudio de una población y argumenta su pertinencia
- Expresa ideas y conceptos sobre las relaciones entre los datos recolectados de una población para determinar o estimar su comportamiento
- Analiza críticamente los factores que influyen en su toma de decisiones en la organización de datos
- Ordena los datos de una población de acuerdo con categorías, jerarquías y relaciones
- Analiza las relaciones entre las variables y los datos en un proceso social o natural para determinar o estimar el comportamiento de la población de estudio

1. Recolección de datos

Contenido

- La estadística descriptiva y la inferencial y sus aplicaciones en diversos contextos
- Las técnicas de recolección de datos como herramientas en el análisis de una población
- La noción de variabilidad en el comportamiento de un conjunto de datos

Por cierto...

De acuerdo con la historia sagrada, José y María acudieron a Belén porque allí se realizaba un censo de población, de manera que podemos afirmar, sin temor a equivocarnos, que desde el principio de nuestra era la estadística estaba ya en pleno auge. De hecho, aun antes los egipcios hacían gran acopio de datos.

Todo surgió no como un fervor numérico extraordinario, sino como un medio para controlar impuestos... y poder cobrarlos. Vaya, lo de siempre.¹

Ubícate

En equipo, realiza las actividades siguientes:

1. Según la *Encuesta Nacional de la Juventud 2005*,² en promedio, los jóvenes mexicanos entre los 15 y los 19 años:
 - Tienen su primer trabajo a los 16.4 años
 - Salen de casa de sus padres por primera vez a los 18.7 años
 - Tienen su primera relación sexual a los 17.5 años
 - Usaron métodos anticonceptivos a los 19

Discute con tus compañeros cuál es la importancia que tienen esos cuatro aspectos en la vida de cada uno de ustedes e indiquen si es conveniente contar con ese tipo de datos y por qué.

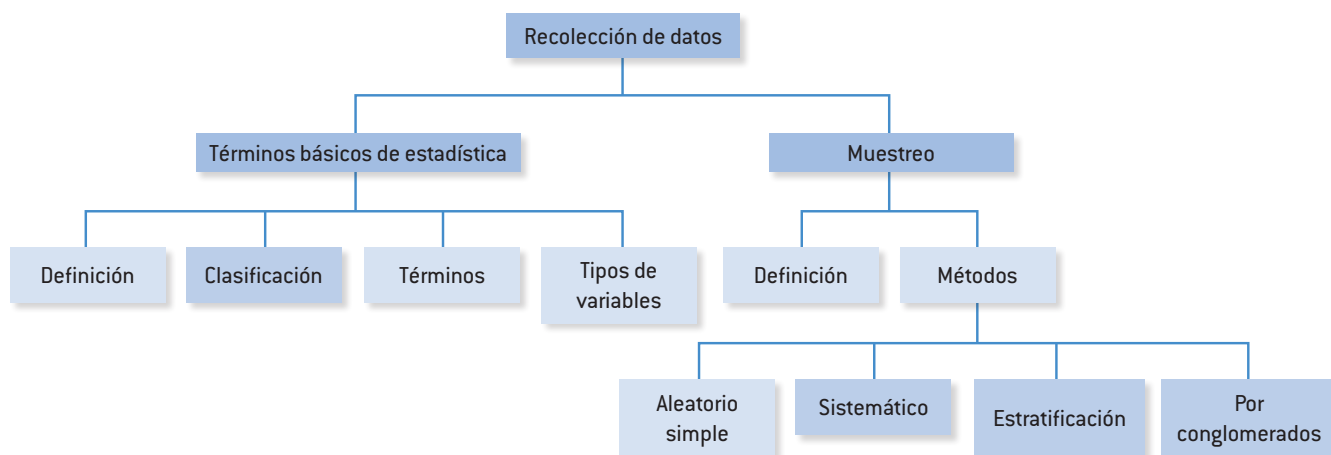
2. Planteen una hipótesis acerca de cómo se obtuvieron esos resultados en la *Encuesta Nacional de la Juventud*. Indiquen qué tipo de preguntas consideran que se plantearon, en qué instrumento se plasmaron (entrevista o cuestionario), así como a quiénes y a cuántos jóvenes se les hicieron.
3. A partir de la hipótesis y la conclusión obtenidas en las dos actividades anteriores, escriban una definición de encuesta, citen alguna de sus características e indiquen cuál puede ser su utilidad. Léanla en voz alta frente a sus compañeros de clase.
4. Investiga qué es un cuestionario y una encuesta, e indica sus diferencias, así como las ventajas y desventajas de cada uno de estos instrumentos.

¹ Alsina, Claudi, *El club de la hipotenusa*, México, 2008, Ariel.

² *Encuesta Nacional de la Juventud 2005*, México, Instituto Nacional de la Juventud, 2006.

Fundamentos de la estadística

Este bloque es de introducción a la estadística, y en él se presenta una visión general de los instrumentos que ofrece esta ciencia para la recolección de datos. Es muy importante señalar que *si no hay datos, sencillamente no es posible trabajar con las herramientas que brinda la estadística*. En el esquema siguiente se muestran los temas que se abordarán en el bloque.



Desempeños

Al terminar de estudiar este bloque y realizar todas las actividades propuestas habrás aprendido a:

- Valorar la estadística como una herramienta matemática que te permite organizar, resumir datos y transmitir resultados de forma comprensible.
- Distinguir las ramas de la estadística para identificar su aplicación en diferentes situaciones.
- Reconocer las características de una población y las técnicas de recolección de datos para aplicarlas en situaciones hipotéticas.
- Valorar las ventajas que tiene el empleo de las diversas técnicas de muestreo para el análisis de los datos de una población.
- Comprender, identificar y describir las variables como atributos de interés de los datos provenientes de una población o de una muestra para reconocer su comportamiento y sus diferencias.

La estadística y su utilidad

Desde tiempos lejanos el ser humano ha usado números, ya sea para conocer su entorno o para describir su región. Con la aparición y desarrollo de instrumentos como calculadoras y computadoras, el uso de números se ha vuelto algo cotidiano. Es así que actualmente vivimos rodeados de cifras y la información que recibimos va acompañada de ellas. Si leemos el periódico podemos encontrar múltiples ejemplos: cuando revisamos la sección sobre la Ciudad de México es común encontrar notas como “las estadísticas de la PGJ indican que durante el año pasado se presentaron 15 mil 625 denuncias por asalto, lo que representó un promedio de 42.8 delitos cometidos diariamente”. O bien, hemos escuchado el problema de obesidad en niños, acerca del que se nos dice: “en promedio la ingesta calórica diaria recomendada para un niño en edad escolar es de mil 800 calorías, sin embargo, son entre 2 mil y 2 mil 500 las que consume cada día”. Si vemos los resultados de encuestas es común leer o escuchar frases como “8 de cada 10 mascotas consumen alimentos X”, o “3 de cada 4 personas usan cierta marca de champú”.

Otros ejemplos de la vida diaria relacionados con la estadística se refieren a la cantidad de habitantes que hay en un país; las preferencias electorales de los candidatos a puestos de elección popular; las estadísticas a la mitad de un partido de fútbol (o de cualquier otro deporte), o el número de accidentes automovilísticos. La realidad es así: estamos rodeados de números y la estadística es la ciencia que se “encarga” de ellos. De allí la importancia de entender y aplicar la estadística.

Pero, ¿qué es la estadística?

Para comprender qué es y para qué sirve la estadística explicaremos primero sus orígenes.

La palabra *estadística* proviene del vocablo latino *status* (“estado, condición”), utilizado en la Edad Media para denotar un Estado político, es decir, para describir verbalmente las características principales de un Estado; cuántas personas había en una ciudad feudal, o cuántas cabezas de ganado, entre otras cosas, con el objetivo de cobrar impuestos. Posteriormente, a fin de tener descripciones más concisas fue necesario expresarlas de forma numérica.

Con el tiempo, la estadística empezó a usarse para describir y analizar datos no necesariamente relacionados con las características del Estado: por ejemplo, para conocer el número de nacimientos y de muertes. En el siglo XVIII comenzó a desarrollarse la estadística teórica con el propósito de establecer orden en el uso de los métodos existentes para recopilar información y para desarrollar nuevos métodos conforme fueran requiriéndose, es decir, se formaliza la estadística.

ESTADÍSTICA

Actualmente, la estadística se define como la ciencia que brinda los instrumentos (métodos y técnicas) para recopilar, organizar, presentar, analizar e interpretar información que apoye los procesos de toma de decisiones en cualquier ámbito.

A partir de la definición de estadística podemos deducir algunas de las razones de su utilidad:

- Describir cómo es una población, ya sea mediante el análisis de toda ella o de sólo una parte.
- Presentar y describir la información de manera numérica o gráfica, de modo que sea comprensible.
- Obtener pronósticos confiables que nos permitan tomar las mejores decisiones.



En educación, gracias a la estadística es posible saber, por ejemplo, cuántos jóvenes se hallan inscritos en cada nivel escolar y cuántos no, y este conocimiento puede emplearse para elaborar políticas educativas.

Como se advierte, la estadística resulta muy útil, y gracias a ello se ocupa en diversas áreas, entre otras:

- En deportes para medir el desempeño de los equipos, o bien, de los jugadores, así como el número de aficionados que se tendrían en una campaña.
- En finanzas, por ejemplo, se utiliza para identificar tendencias (comportamientos) en la tasa de interés a lo largo del tiempo, lo que permite tomar la decisión de invertir o no.
- En los procesos de manufactura se emplean métodos estadísticos para mejorar la calidad de los productos, así como para determinar los factores que afectan el proceso de producción.
- En mercadotecnia, la estadística es útil para estimar, por ejemplo, el porcentaje de personas que prefieren cierto producto, lo que a menudo desemboca en la elección de una estrategia publicitaria específica para promoverlo.
- En psicología, se pueden hacer estudios para determinar el coeficiente intelectual o el nivel de ansiedad en una población a partir de un grupo de estudio.
- En educación, permite determinar, entre otros, el grado de conocimientos y habilidades que tienen en ciencias los alumnos de cierto grado educativo, y de esta forma emprender acciones para mejorarlo o mantenerlo.
- En biología, se pueden levantar censos para establecer la cantidad de individuos de cierta especie que viven en un ecosistema.

Clasificación de la estadística

En general, la estadística se clasifica en dos grandes ramas: descriptiva e inferencial.

La *estadística descriptiva* se basa en:

- La recolección de datos por medio de encuestas o de otras técnicas.
- La caracterización de datos por medio del cálculo de medidas estadísticas como la media (promedio).
- La presentación de datos mediante tablas y gráficas.

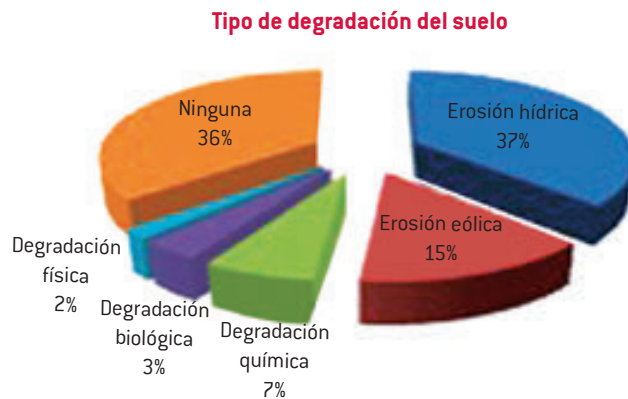
En capítulos posteriores explicaremos con más detalle cada uno de estos aspectos.

Ejemplo 1

Uno de los graves problemas actuales es la degradación de la superficie terrestre. A continuación se presenta, una tabla y en una gráfica de pastel la información referente a México.

Tipo de degradación	Porcentaje respecto a la superficie total
Erosión hídrica	37
Erosión eólica	14.9
Degradación química	6.8
Degradación biológica	3.6
Degradación física	1.8
Ninguna	36

Fuente: SEMARNAT. CONAGUA. Estadísticas del Agua en México, 2006.



¿Qué observaciones harías al ver la información de esta manera? Ninguna
 ¿Qué información te llama la atención? ¿Por qué

Por su parte, la *estadística inferencial* comprende los métodos y las técnicas utilizadas para que a partir del estudio de una parte de la población (muestra) se hagan generalizaciones, predicciones o estimaciones con el fin de tomar las decisiones correctas.

El director de una universidad en la que están inscritos 2400 estudiantes desea conocer cuántos alumnos estarían dispuestos a tomar cursos de verano a fin de planear cuántos se impartirán en la institución. ¿Cómo puede obtener esta información?

Ejemplo 2

Solución

El director tiene tres opciones:

- Consultar la programación del verano anterior y replicarla.
- Preguntar a cada uno de los 2400 alumnos sobre sus intenciones de estudiar en el verano.
- Tomar un subconjunto de todos los alumnos, o sea, una muestra de la población, e indagar si ellos se inscribirían a un curso de verano. Sobre los resultados obtenidos con la muestra haría una generalización.

En este ejemplo, lo mismo que en muchos casos reales, la mejor opción es trabajar únicamente con una muestra, pues se ahorra dinero y tiempo. También con ayuda de la estadística se determinan las mejores muestras.

Términos básicos

Toda disciplina, científica o humanística, tiene un conjunto de términos que la distinguen y cuyo conocimiento resulta fundamental para comprenderla. La estadística no es la excepción, de modo que a continuación revisaremos sus términos básicos, lo que nos permitirá establecer un marco de referencia para comprender los temas siguientes.

Datos son valores o características que pueden ser observados, contados o medidos. Algunos ejemplos de datos son las calificaciones de un grupo de alumnos, las respuestas de alguna pregunta hecha en Facebook, el número de boletos vendidos para un concierto, el número de artículos producidos en una fábrica y la cantidad de gasolina que se consume diariamente, entre otros.

Variable es una característica o atributo que puede adquirir distintos valores.

Población También llamada *universo*, es el conjunto finito o infinito de personas u objetos que presentan características comunes (figura 1).

Población = todos los elementos



Muestra = elementos seleccionados de la población

Figura 1 Población vs muestra.

Ejemplo 3

Si se desea conocer el porcentaje de personas que tienen credencial para votar, ¿cuál es la población?

Solución

La población es el conjunto de personas mayores de 18 años que tienen credencial de elector vigente.

La *población finita* es la que posee o incluye un número limitado de observaciones o datos, por ejemplo, el número de personas que están en un estadio de fútbol. Por su parte, la *población infinita* es la que, hipotéticamente, no tiene límite en cuanto al número de observaciones que pueden realizarse, es decir, el número de elementos se considera muy grande, por ejemplo, el conjunto de los números naturales.

Muestra Es un subconjunto o una parte representativa de una población (figura 1).

Muestreo Procedimiento mediante el cual se obtiene una o varias muestras representativas de la población; en otras palabras, el muestreo es el conjunto de técnicas que se tienen para elegir los elementos que compondrán una muestra. Más adelante se revisarán brevemente los métodos de muestreo que más se utilizan.

Experimento Es la operación que consiste en observar los resultados en ciertas condiciones. Lanzar una moneda al aire, establecer el número de alumnos con promedio mayor o igual a 9.5, realizar una encuesta sobre ingresos y gastos son, entre otros, ejemplos de un experimento estadístico.

Parámetro Es una medida que se calcula para describir una característica de la población.

Estadístico Es una medida que se calcula para describir una característica de la muestra. Dicho con otras palabras, es la estimación de un parámetro.

Ejemplo 4

En un estudio se quiere conocer el ingreso promedio en México. ¿Cuál sería la población? ¿Cuál sería el parámetro y cuál el estadístico?

Solución

La población serían todas las personas que viven en México y que perciben un salario o un sueldo, es decir, lo que se conoce como *población económicamente activa* (PEA). En este caso,

el parámetro es el ingreso promedio de toda la PEA. Como ya se explicó, es costoso preguntar a toda la población, así que se obtiene una muestra: algunas personas del centro del país, algunas del sur y otras del norte, por ejemplo, y de allí se calcula el ingreso promedio, que es el estadístico.

Fuentes

Para poder llevar a cabo un estudio estadístico de una población, primero se debe tener muy claro qué se quiere analizar para recolectar los datos adecuados. Ahora, la recolección se puede hacer recurriendo a diversas fuentes. En términos generales, las fuentes de donde se obtienen los datos pueden clasificarse en *primarias* o *secundarias*:

- **Primarias**, mediante la observación o realización de experimentos, encuestas o cuestionarios.
- **Secundarias**, como las bases de datos ya existentes, por ejemplo, la del INEGI o BANXICO.

Tipos de variables

Hemos visto que una variable es una característica o atributo que puede adquirir distintos valores. De acuerdo con el valor que adoptan las variables se clasifican en cualitativas y cuantitativas (figura 2).

Las variables *cualitativas*, también conocidas como *variables categóricas* o *de atributo*, son las características no numéricas del fenómeno que se estudia.

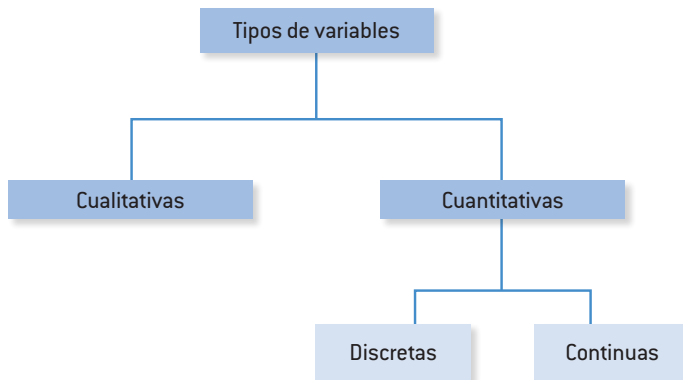


Figura 2 Tipos de variables.

En un cuestionario se hacen preguntas sobre género, estado civil, medio de transporte y religión. Estos son ejemplos de variables categóricas porque:

- a. Género. Las respuestas posibles son femenino o masculino.
- b. Estado civil. Se indica soltero, casado, divorciado o viudo.
- c. Medio de transporte. Las opciones son automóvil propio, taxi, metrobús, colectivo o metro, entre otros.
- d. Religión. Se responde católico, protestante, musulmán u otra.

Ejemplo 5

Por otro lado, como su nombre lo indica, las *variables cuantitativas* describen de forma numérica la característica que se estudia, por lo cual se conocen también como *variables numéricas*. Como se muestra en la figura 2, las variables cuantitativas se clasifican en discretas y continuas.

Las *variables discretas* son las que asumen ciertos valores después de haber realizado un conteo; por ejemplo, el número de habitaciones en una casa, el número de goles anotados en un partido o el tabulador de salarios según el nivel de estudios.

A su vez, las *variables continuas* surgen de un proceso de medición y pueden asumir cualquier valor en un rango específico. Entre los ejemplos de variables continuas se cuentan la cantidad diaria de kilómetros recorridos en auto, que puede encontrarse en el intervalo $[0, \infty]$, la cantidad de agua que consume una persona en una hora, que podría estar en el intervalo $[0, 10]$ litros, o bien, la estatura de una persona adulta, que puede hallarse en el intervalo $(0.75, 2.30)$ metros.

Censo y muestreo

El término *censo* se refiere a la obtención de datos de cada uno de los miembros que forman una población. A medida que una población crece, se requiere más dinero y tiempo para realizar ese estudio. Por ejemplo, en México el censo poblacional se realiza cada diez años, ya que implica el desplazamiento y levantamiento de cuestionarios en cada rincón del país.

En todo estudio, lo ideal sería llevar a cabo un censo, pues así se conocería muy bien a la población, pero cuando no se tienen los recursos necesarios se recurre al muestreo. Como recordarás, una muestra es un subconjunto extraído de una población que se selecciona para su estudio. El *muestreo* es, por consiguiente, un conjunto de técnicas empleadas para seleccionar la “mejor” muestra posible, es decir, la que se considera que representa “mejor” a la población.

El muestreo se centra en obtener, por medio de métodos diversos, una muestra de toda la población para luego, una vez que se cuenta con los resultados acerca de la o las características que se desean estudiar, hacer una proyección respecto de toda la población.

Métodos de muestreo

El proceso de muestreo comienza con la identificación de fuentes adecuadas de datos, como las listas de población, directorios y mapas, entre otras, que son llamadas *marcos*. De éstos se extrae la muestra. Cabe señalar que si el marco es inadecuado, entonces las muestras serán también inadecuadas y nuestras estimaciones serán “malas”.

Hay dos tipos básicos de muestras:

- *Muestra no probabilística*: los elementos se eligen sin tener en cuenta su probabilidad de ocurrencia, es decir, sin tener en cuenta que suceda cierto resultado. Por ejemplo, una muestra obtenida mediante una red social, en la cual sólo intervienen las personas que usan ese medio. Este tipo de muestras son convenientes porque resultan rápidas y de bajo costo, pero existe una falta de precisión en las estimaciones.
- *Muestra probabilística*: los elementos que la componen se eligen de acuerdo con las probabilidades de ocurrencia, esto es, existe un trabajo estadístico previo a la selección de la muestra.

El proceso de muestreo puede realizarse por etapas y en cada una de ellas es posible aplicar un método distinto. Todo depende de la situación en estudio, del presupuesto con que se cuente y de la variabilidad que se quiera aceptar en los resultados.

A continuación revisaremos de forma general los métodos de muestreo más utilizados para obtener muestras probabilísticas.

Muestreo aleatorio simple

El muestreo aleatorio simple es la técnica de muestreo más elemental, al punto que constituye la base de otras técnicas. La muestra obtenida con este método resulta de una selección hecha de manera tal que cada elemento de la población tiene la misma oportunidad (probabilidad) de resultar seleccionado, como se advierte en el ejemplo.

Algo común es recurrir a una urna llena de “papelitos” que representan cada elemento de una población para elegir de ahí una muestra. Es necesario señalar que ello no es muy conveniente, pues si no se revuelven perfectamente bien los “papelitos”, entonces habrá sesgos. Lo más apropiado es emplear una tabla de números aleatorios.

En una empresa se realizará una encuesta sobre el clima organizacional para conocer si los 2500 empleados que la forman tienen un ambiente agradable de trabajo. Por cuestiones de tiempo y costos no es posible censar a los 2500 trabajadores. ¿Cómo se usaría la técnica del muestreo aleatorio simple para obtener una muestra?

Ejemplo 6

Solución

Se decide tomar una muestra de 125 personas y hacer una encuesta con ellas. Para asegurar que todas las personas tengan la misma probabilidad de ser elegidas sin importar el puesto que tengan en la empresa, se pide que cada una escriba su nombre en una pequeña hoja de papel y la deposite en una urna. Se revuelven y se extraen estos “papelitos” hasta completar el tamaño de la muestra de 125. A ellos se aplica el cuestionario sobre clima organizacional y el resultado se extiende a toda la población.

Muestreo aleatorio sistemático

De acuerdo con este método, para obtener la muestra requerida primero se divide el tamaño de la población (N) entre el tamaño de muestra (n) deseado y el resultado obtenido (k) se redondea al entero más cercano. Luego, para seleccionar la muestra se elige al azar el primer elemento y los subsiguientes se escogen cada k elementos.

Cabe precisar que el muestreo sistemático tiene un pequeño inconveniente: si existe algún patrón (comportamiento u ordenamiento bajo algún criterio, como orden alfabético o por fecha de nacimiento) en la lista de la población, existirán errores de selección que afectarán los resultados.

El gerente de producción de una fábrica desea saber si sus productos cumplen los estándares de calidad que se tienen establecidos. Como resulta muy costoso revisar artículo por artículo, decide tomar una muestra de 100 productos. ¿Cómo puede obtenerla? Considera que la producción diaria es de 2000 artículos y se opta por emplear el muestreo sistemático.

Ejemplo 7

Solución

En este caso:

$$\frac{N}{n} = k \quad \begin{array}{l} \text{Se divide la población (N)} \\ \text{entre la muestra requerida} \end{array}$$

$$\frac{2000}{100} = 20 \quad \text{Se sustituyen valores}$$

$$k = 20 \quad \text{Cada 20 elementos se elegirá uno para la muestra}$$

Ahora, en la selección del primer artículo se utiliza el muestreo aleatorio simple: se elige un artículo de los primeros 20, por ejemplo, el 16. Entonces, se comienza por el artículo 16, y cada 20 artículos después se seleccionará uno para la muestra { 16, 36, 56, 76, ... }.

Muestreo aleatorio estratificado

Esta técnica se utiliza cuando una población está dividida en grupos, llamados *estratos*, formados con base en cierta característica, pues así se garantiza que cada miembro de la población esté en un y solamente un estrato.

Después se toma una muestra de cada estrato y se hacen comparaciones entre ellas.

Merece la pena indicar que para que este tipo de muestreo ofrezca una buena precisión es necesaria la homogeneidad de los elementos en cada estrato, así como la heterogeneidad entre los estratos.

Ejemplo 8

Retomemos el ejemplo del muestreo aleatorio simple (ejemplo 6). Los directivos desean conocer con mayor detalle el clima organizacional que prevalece en la empresa, por tanto, deciden tomar muestras por departamento, es decir, considerar cada departamento como un estrato. De esta manera, las opiniones dentro de cada estrato serán homogéneas (los empleados de cada departamento se enfrentan a los mismos problemas), y serán heterogéneas para cada estrato, (los empleados pueden tener distinta percepción del clima organizacional, según si están en ventas, mercadotecnia, etcétera).

Muestreo por conglomerados

El muestreo por conglomerados consiste en dividir una población en grupos o conglomerados usando cierto tipo de límite, por ejemplo, geográfico. Posteriormente, se seleccionan conglomerados al azar y se recolecta una muestra eligiendo en forma aleatoria elementos de cada uno de ellos.

El muestreo por conglomerados resulta de gran utilidad para reducir el costo del muestreo cuando la población está dispersa en una zona geográfica extensa.

Ejemplo 9

Supón que se desea estimar el índice de criminalidad en México. ¿Cómo podrías hacerlo?

Solución

Lo ideal sería realizar un censo, pero como esto resultaría muy costoso e implicaría mucho tiempo, se puede utilizar el muestreo por conglomerados utilizando los municipios (o las delegaciones, en el caso del Distrito Federal) como conglomerados y, enseguida, elegir los municipios según su número de habitantes, es decir, seleccionarlos proporcionalmente a su tamaño.

Ejercicios

I. TEORÍA Y ALGO MÁS

1. *Completa los espacios en blanco:*

La estadística se define como la ciencia para **a.** _____, **b.** _____, **c.** _____, **d.** _____, **e.** _____ información para apoyar la toma de decisiones. La estadística se clasifica en dos grandes ramas: **f.** _____ **g.** _____.

El objetivo de la estadística inferencial es _____ sobre poblaciones a partir de una muestra.

Los tipos de variables son: **a.** _____, **b.** _____. Las variables cuantitativas se clasifican en **c.** _____, **d.** _____.

El muestreo es un procedimiento para obtener una **a.** _____. Los métodos de muestreo principales son: **c.** _____, **d.** _____, **e.** _____, **f.** _____.

II. PIENSO, ... ¿Y LUEGO?

2. *Ecología.* (Alba, “Numeralia”, en *Petróleo*, 2010). En la revista *Equilibrio* de enero 2011 se informa que:

700 millones de litros fueron derramados en el Golfo de México en el accidente del pozo Ixtoc-1, en 1979

45% de la electricidad se genera con petróleo y 95% del transporte depende de él

8031 toneladas de hidrocarburos pierde Pemex al año entre derrames y fugas

De acuerdo con la información estadística que se presenta, elabora un par de conjeturas con tus compañeros acerca de lo que implican los “números” presentados respecto de:

a. Los daños ecológicos.

b. Los efectos económicos.

Comparte con tus compañeros, oralmente o por escrito tus conjeturas. Obtén una conclusión acerca de cómo interpretar la información estadística.

3. *Educación*. De acuerdo con un buen número de analistas de la política y la economía, uno de los grandes retos para que el país se supere es elevar el nivel educativo. La Organización para la Cooperación y el Desarrollo Económico (OCDE) informa que México es la nación que menos recursos asigna a infraestructura educativa (Alba: 2011): “Cerca de 95% [del presupuesto asignado a la educación] se destina a servicios personales (salarios, honorarios y prestaciones), mientras que arriba de 4% a gastos de operación, por lo que queda únicamente 0.03% para invertirlos en infraestructura”. El gasto en infraestructura en Alemania es de 6.7%, el de Japón de 9.9% y el de Estados Unidos de 11.2%.

¿Qué opinas de estas cifras? ¿Qué aspectos considerarías indispensables para mejorar la educación en el país? Investiga en qué rubros se necesita una mayor inversión, así como en qué áreas de la educación en México se requiere poner mayor énfasis. Organiza una discusión con tus compañeros de grupo y redacten un ensayo sobre la situación educativa en México y lo que necesitamos para tener un mayor nivel académico.

4. *Seguridad* (Mitofsky: 2010). De acuerdo con un estudio acerca de la percepción de la inseguridad realizado en marzo de 2010 por la agrupación México Unido contra la Delincuencia AC y la empresa Consulta Mitofsky, prácticamente cuatro de cada cinco ciudadanos consideran que las condiciones en las que vivimos hoy son menos seguras que las que teníamos hace un año en México.

El 18% que califica como mejor la seguridad es el peor porcentaje notificado en la serie de 17 meses desde noviembre de 2008.

Conforme mayor es la edad del ciudadano es más negativa su percepción de las condiciones de inseguridad y por regiones el norte del país y la región centro es donde peores opiniones se vierten.

¿Qué te dicen estas cifras? ¿A qué conclusiones puedes llegar? ¿Qué aspectos consideras importantes modificar para cambiar esta percepción?

III. PURO DEPORTE

5. *Publicidad*. Forma un equipo y junto con tus compañeros busca en internet anuncios o videos donde se aplique la estadística. Elaboren un breve informe donde indiquen cuáles son los datos que se presentan y cómo consideran que se obtuvieron. Expongan frente al grupo sus descubrimientos.
6. *Medios de comunicación*. Una empresa de telefonía quiere realizar un estudio para conocer las necesidades de sus usuarios. Al revisar su base de datos encuentra las variables siguientes:

- a. Número de teléfonos fijos en el hogar. ()
- b. Número de llamadas nacionales al mes. ()
- c. Duración promedio (en minutos) de las llamadas nacionales al mes. ()
- d. Monto de la factura mensual por concepto de llamadas nacionales. ()
- e. Número de llamadas de larga distancia al mes. ()
- f. Duración promedio (en minutos) de las llamadas de larga distancia al mes. ()
- g. Monto de la factura mensual por concepto de llamadas de larga distancia. ()
- h. Número de llamadas a celular al mes. ()
- i. Duración promedio (en minutos) de las llamadas a celular al mes. ()
- j. Monto de la factura mensual por concepto de llamadas a celular. ()
- k. Si cuenta con servicio de internet. ()
- l. Si cuenta con servicio de televisión por cable. ()

Escribe en el paréntesis, en cada caso, si las variables son cualitativas (c) o cuantitativas (q). Si son cuantitativas, determina si son discretas (qd) o continuas (qc). ¿Qué otras variables consideras que sería importante que tuviera en cuenta esta compañía para satisfacer las necesidades de sus usuarios?

7. *Demografía.* Organizaciones como el Instituto Nacional de Geografía e Informática (INEGI) realizan encuestas para conocer las características principales de la población. Uno de los aspectos que ha intentado establecer, por ejemplo, es cómo son los hogares en México. Para ello, ha buscado determinar:

a. Tipo de vivienda (particular o colectiva).	b. Número de cuartos.
c. Número de dormitorios.	d. Si cuenta con cocina.
e. Número de ocupantes.	f. Tipo de material en las paredes y el techo.
g. Tipo de material en los pisos.	h. Si cuenta con drenaje.
i. Si cuenta con agua entubada.	j. Si cuenta con energía eléctrica.
k. Si cuenta con lavadora.	l. Si cuenta con computadora.
m. Si la vivienda no es particular, monto de la renta.	

Indica si las variables son cualitativas (c) o cuantitativas (q). Si son cuantitativas, determina si son discretas (qd) o continuas (qc). ¿Crees que faltaría alguna otra variable por considerar para formar un modelo de hogar en nuestro país? ¿Cuál y de qué tipo sería?

8. *Mercadotecnia.* El gerente de ventas de una tienda departamental desea realizar una encuesta en el área metropolitana para identificar el tipo de mujeres que compran en su establecimiento. Por equipos, (a) describan la población (y la muestra) de interés, (b) enumeren las variables que consideren relevantes para el estudio, y (c) indiquen de qué tipo de variables se trata.

Mediante una lluvia de ideas con los otros equipos, determinen las variables “óptimas”. Posteriormente, y de nuevo con sus equipos originales, elaboren el borrador de un cuestionario que conste de seis preguntas. Presenten todo su trabajo a sus compañeros de clase.

9. *Educación.* El director de una escuela está preocupado por elevar el nivel académico y mejorar la infraestructura del plantel. Para ello, decide realizar una encuesta. A la escuela asisten 500 alumnos que cursan el primer año de preparatoria, 600 el segundo y 450 el tercero.

- a. ¿Crees que sería útil una muestra aleatoria simple? Justifica oralmente tu respuesta.
b. ¿Crees que sería mejor una muestra estratificada? Justifica oralmente tu respuesta e indica qué estratos tomarías.

10. *Impuestos.* Uno de los problemas que se presentan en el país es la evasión de impuestos. Para revisar si los contribuyentes están al corriente en el pago de sus contribuciones fiscales se decide tomar una muestra nacional de personas físicas. En la base de datos se tiene el Registro Federal de Contribuyentes (RFC) respectivo.

Escribe un texto breve donde expliques qué consideras mejor: tomar una muestra aleatoria simple o una muestra sistemática. Léelo frente a tus compañeros de clase y, con la guía del profesor, obtengan conclusiones.

11. *Comercio.* Una empresa estudia la posibilidad de abrir nuevos supermercados en el área metropolitana de Guadalajara. Dentro del proceso de investigación de mercados se debe realizar una encuesta a los posibles clientes del área de estudio. ¿Qué tipo de muestreo propondrías? Explícalo por escrito.

Evaluación sumativa

Un escalón hacia arriba

Contesta los reactivos siguientes, que te permitirán saber si has aprendido lo que se esperaba.

1. En una encuesta realizada a los empleados de una empresa se les pidió que proporcionaran los datos siguientes:
 - Nombre.
 - Teléfono particular.
 - Sexo.
 - Edad (años cumplidos).
 - Tiempo (minutos) que tardan en trasladarse de su casa al trabajo.
 - Medio de transporte que utilizan para llegar al trabajo.
 - Gasto semanal en transporte.

Identifica las posibles variables de interés para analizarse e indica si son cualitativas (c) o cuantitativas (q). Si son cuantitativas, indica si son discretas (qd) o continuas (qc).

2. Supón que en un hospital se desea realizar una encuesta para conocer la opinión de las personas sobre el servicio que se les brinda. Como no se cuenta con tiempo ni presupuesto suficiente para realizar un censo se decide tomar una muestra. En esta situación, ¿se trata de una población finita o infinita? ¿Qué tipo de muestreo sugerirías utilizar? Explica oralmente por qué.
3. En una línea de producción donde diariamente se generan 1000 artículos se desea revisar el control de calidad. Para saber si se cumple con los estándares establecidos se tomará una muestra un día cualquiera: se elige el artículo y se revisa.
Para esta situación, ¿se trata de una población finita o infinita? ¿Qué tipo de muestreo sugerirías emplear? Explica por escrito tus respuestas.
4. En una preparatoria se desea conocer si los hombres o las mujeres consumen más alimentos de los llamados *chatarra*. ¿Qué tipo de muestreo sugieres utilizar? Justifica por escrito tu respuesta.
5. En una de las delegaciones políticas del Distrito Federal se han recibido quejas vecinales respecto al servicio de limpieza en varias colonias. Para formarse una idea más clara de la situación, se ha decidido levantar una encuesta entre los habitantes. ¿Qué tipo de muestreo propondrías emplear? Justifica oralmente tu respuesta.

Proyecto 1

TRANSPORTE PÚBLICO

Uno de los graves problemas que enfrentan las ciudades actualmente es el crecimiento desmedido del parque de vehículos automotores, situación que ocasiona congestionamientos viales, mayor contaminación y, en cierta medida, enfermedades nerviosas.

Para las elecciones federales y a fin de aumentar su popularidad, el candidato del partido xyz al gobierno de la ciudad quiere conocer las necesidades de las personas en materia de transporte público y decide realizar una encuesta sobre el tema.

Los resultados le ayudarán a su equipo a esbozar las propuestas de campaña, ya sea para mejorar y ampliar la red de transporte público, o bien, realizar nuevas obras de infraestructura, como son segundos pisos, túneles, la construcción de un nuevo periférico, etcétera.

Supón que has sido contratado por el candidato del partido xyz para realizar la encuesta, con los resultados que se obtengan debes escribir un informe donde abordes estos aspectos:

- Marco muestral (población objetivo).
- Técnica de muestreo (justifica tu elección).
- Cuestionario donde incluyas apartados sobre calidad del servicio, tiempo de recorridos, costos y seguridad, entre otros, indicando el tipo de variables que se utilizarán en cada pregunta, así como las poblaciones correspondientes.

Para conocer más de cerca el problema, tú y tu equipo de trabajo utilizan los diferentes transportes públicos de la ciudad durante una semana.

Al cabo de este tiempo te das cuenta que tanto el servicio como la seguridad en las unidades del transporte público son muy deficientes. Por consiguiente, en la encuesta deseas hacer énfasis en estos dos rubros, pero si entregas el cuestionario para mostrar realmente al candidato la situación de la red de transporte público, implicaría el mejoramiento de ésta a costa de un aumento en las tarifas del transporte público, o bien, la creación de un nuevo impuesto, lo que le restaría votos.

Si sesgas la encuesta, por ejemplo, realizando preguntas “capciosas” para que las personas no demuestren su descontento, el candidato podría hacer propuestas para crear nuevas vialidades, con lo que dejaría a un lado el mejoramiento del transporte público, y ello a su vez causaría accidentes en la red de transporte público, además de un aumento del parque vehicular.

Preguntas para dirigir la discusión

- ¿Qué harías, modificarías las preguntas de manera que las personas se confundan y contesten lo que genere más simpatía o votos?
- ¿Qué beneficios/pérdidas económicos y sociales se obtendrían en cada situación?
- ¿Qué principios éticos estarían en juego en cada caso?
- ¿Cuál es la responsabilidad como profesional?
- ¿Cuál consideras que es la responsabilidad social que se tiene como ciudadano? ¿Y como servidor público?



Competencias

- Formula y resuelve problemas aplicando las reglas que le permitan agrupar datos en una distribución de frecuencias
- Organiza los datos en una distribución de frecuencias a partir de la variabilidad estadística observada y argumenta su pertinencia
- Expresa ideas y conceptos mediante representaciones lingüísticas, matemáticas o gráficas relacionadas con la distribución de frecuencias de datos procedentes de una muestra
- Argumenta la solución obtenida de un problema relacionado con la distribución de frecuencias de datos procedentes de una muestra
- Construye e interpreta diferentes representaciones gráficas para la comprensión y el análisis de las situaciones reales, hipotéticas y formales
- Maneja las tecnologías de la información y la comunicación para obtener tablas de distribución de frecuencias a partir de los datos de una población y expresar las conclusiones de dicho proceso

2. Presentación de datos

Contenido

- Reglas para determinar el número de clases y la amplitud de intervalo en una serie de datos provenientes de una población o muestra
- Representación tabular de los datos en categorías mutuamente excluyentes procedentes de una población o muestra
- Representación gráfica y análisis de los datos mediante histogramas, polígonos de frecuencias y polígonos de frecuencias acumuladas
- Otras representaciones tabulares y gráficas de un conjunto de datos como herramienta en el análisis de una población

Por cierto...

Florence Nightingale (Florencia, 1820-Londres, 1910) se preocupó por la gente y su bienestar, y a su entender, el gobierno sólo podría favorecer ese bienestar si basaba sus políticas en la información estadística. El primer informe estadístico de Nightingale fueron sus “Notas sobre cuestiones que afectan la salud, la eficiencia y la gestión hospitalaria en el ejército británico, basadas sobre todo en la experiencia durante la última guerra (Crimea)”.

Uno de los aspectos novedosos que incluyó fue el uso de diagramas estadísticos, pues su intención era que los lectores comprendieran su mensaje, es decir, “alegró su informe con diagramas”.¹

Ubícate

Antes de proseguir con el estudio del bloque, resuelve los ejercicios siguientes. En caso de tener algún problema, te recomendamos que repases tus notas y libros de matemáticas para que te sea más fácil el estudio de este bloque.

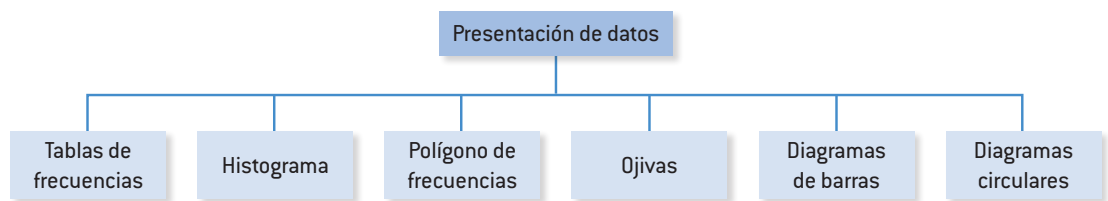
1. ¿Qué es el plano cartesiano y cuáles son sus componentes?
2. Traza un plano cartesiano y localiza en él los puntos siguientes:
 - $P(3, 1)$
 - $R(-4, 3)$
 - $Q(-2, -2)$
 - $S(1, -3)$
3. Expresa en decimales y en porcentaje las fracciones siguientes:
 - a. $\frac{1}{4}$
 - c. $\frac{32}{40}$
 - b. $\frac{2}{3}$
 - d. $\frac{26}{50}$
4. En una urna hay 11 pelotas; tres son negras, cinco rojas y el resto, amarillas.
 - a. ¿Qué proporción de las pelotas son negras?
 - b. ¿Cuánto representan las rojas del total?
5. Si un círculo tiene 360°
 - a. ¿A cuántos grados equivale el 33% del círculo?
 - b. ¿Qué porcentaje representan 72° ?

¹ I. B. Cohen, *El triunfo de los números*, Alianza Editorial, Madrid, 2005.

Distribuciones de frecuencias y gráficas

En el bloque anterior expusimos conceptos básicos de la teoría de probabilidades y de estadística, así como una breve explicación de los métodos de muestreo, que sirven para la recolección de datos.

Una vez que se obtienen los datos, el siguiente paso es elegir cómo se van organizar, ya sea para su análisis o presentación. Ahora veremos las herramientas estadísticas más comunes para presentar datos y hacer más comprensible la información. Esto se logra mediante la elaboración de tablas o gráficas. En el esquema siguiente se muestran los temas que abordaremos.



Cabe señalar que los histogramas, los polígonos de frecuencias y las ojivas se utilizan cuando se tienen datos numéricos. Por otra parte, las gráficas de barras y las circulares suelen emplearse cuando se cuenta con datos categóricos. Finalmente, las distribuciones de frecuencias se utilizan para ambos casos.

Hay que resaltar, también, que las gráficas utilizadas no sólo sirven para presentar los datos, sino para expresar la idea que se desea destacar. Por eso hay que ser muy cuidadosos para elegir la gráfica adecuada.

Desempeños

Al terminar de estudiar este bloque y realizar todas las actividades propuestas habrás aprendido a:

- Organizar y presentar los datos obtenidos en una distribución de frecuencias.
- Presentar una distribución de frecuencias en un histograma, un polígono de frecuencias y un polígono de frecuencias acumuladas.
- Construir representaciones tabulares y gráficas después de reconocer el tipo de agrupación de datos al que pertenecen, para obtener una mejor comprensión del comportamiento de la población del objeto de estudio.

Distribuciones de frecuencias

Una vez que se han recolectado los datos (observaciones), es importante caracterizar (describir) cómo se comportan, lo que se logra determinando la frecuencia con que ocurren esos valores mediante distribuciones de frecuencias conocidas también como *tablas de frecuencias* y de gráficas.

Las distribuciones de frecuencias permiten observar cómo se distribuye un conjunto de datos. Por tanto, es posible observar dónde se encuentra la mayor concentración si éstos están muy dispersos o no, por ejemplo. Dicho con otras palabras, las distribuciones de frecuencias son tablas en las que se muestran todos los posibles valores de una variable junto con el número de veces que cada uno de ellos es observado.

En el ejemplo siguiente construiremos una tabla de frecuencias.

Se entrevistó a 20 jóvenes para conocer el número de refrescos de cola que beben al día. Las respuestas obtenidas son las siguientes:

5	2	2	4	5
1	1	4	2	1
3	5	1	0	5
0	3	2	3	4

Con esta información construye una tabla de frecuencias.

Ejemplo 1

El número de refrescos que se toman en un día es un ejemplo de *dato discreto*, ya que los valores se obtuvieron mediante un proceso de conteo.

Solución

El primer paso para elaborar la tabla de frecuencias consiste en ordenar los datos de menor a mayor, de esta forma:

0 0 1 1 1 1 2 2 2 2 3 3 3 4 4 4 5 5 5 5

La variable de estudio es el número de refrescos de cola, y los valores son 0, 1, 2, 3, 4 y 5. Posteriormente, se cuenta cuántas personas toman 0 refrescos, cuántas 1, y así sucesivamente. A esto se le conoce como *frecuencia absoluta* (f_a) o *frecuencia de clase*.

Número de refrescos	f_a
0	2
1	4
2	4
3	3
4	3
5	4
Total	20

La interpretación verbal de esta tabla de frecuencias es la siguiente: de las 20 personas entrevistadas (columna de la derecha), dos no beben refresco de cola, cuatro toman uno al día, tres beben tres al día, también tres beben cuatro al día y, finalmente, cuatro toman cinco refrescos de cola al día.

A partir de una distribución de frecuencias como la del ejemplo anterior también pueden calcularse las *frecuencias relativas* (f_r). Para ello sólo hay que dividir cada una de las frecuencias absolutas entre el total de datos y multiplicarlo por cien para presentarlo como porcentaje. De esta manera, la tabla de frecuencias quedaría como sigue:

Número de refrescos	f_a	f_r
0	2	10%
1	4	20%
2	4	20%
3	3	15%
4	3	15%
5	4	20%
Total	20	100%

$\frac{4}{20} \cdot 100 = 20\%$

Ejemplo 2

El sabor del refresco es una variable o *dato categórico*, es decir, no es un valor numérico, sino una característica de lo que se quiere analizar.

Se entrevistó a 30 personas para saber qué sabor de refresco prefieren. Las respuestas fueron las que se muestran en la tabla siguiente.

A partir de esta información, elabora la distribución de frecuencias correspondiente, calculando tanto las frecuencias absolutas como las relativas.

Toronja	Cola	Manzana	Toronja	Cola
Toronja	Cola <i>light</i>	Cola	Limón	Manzana
Naranja	Naranja	Cola <i>light</i>	Limón	Cola
Toronja	Cola	Limón	Toronja	Cola <i>light</i>
Cola <i>light</i>	Cola	Limón	Cola <i>light</i>	Cola
Manzana	Cola	Naranja	Cola	Cola <i>light</i>

Solución

En este ejemplo nos interesa conocer el sabor de refresco preferido de 30 personas, por tanto, primero identificamos las categorías de interés:

Sabor
Cola
Cola <i>light</i>
Limón
Manzana
Naranja
Toronja

Después contamos el número de respuestas que obtuvo cada uno de estos sabores, es decir, calculamos la frecuencia absoluta $\{f_a\}$:

Sabor	f_a
Cola	9
Cola <i>light</i>	6
Limón	4
Manzana	3
Naranja	3
Toronja	5
Total	30

Por último, determinamos la frecuencia relativa $\{f_r\}$:

Sabor	f_a	f_r
Cola	9	30%
Cola <i>light</i>	6	20%
Limón	4	13.33%
Manzana	3	10%
Naranja	3	10%
Toronja	5	16.67%
Total	30	100%

Interpretemos las dos primeras filas de la tabla: de 30 personas entrevistadas acerca del sabor de refresco que más les gusta, nueve respondieron que el de Cola, es decir, 30% de los encuestados prefieren ese sabor. Por otra parte, 20% de las personas prefieren cola *light*. Esto corresponde a seis de las 30 personas entrevistadas. Ahora interpreta por escrito o de manera oral el resto de las filas.

Cuando los datos son continuos, es decir, provenientes de un proceso de medición es necesario modificar la distribución de frecuencias, puesto que los datos deben organizarse en clases.

El proceso para construir una distribución de frecuencias para datos continuos es el siguiente:

1. Se ordenan los datos de menor a mayor.
2. Se determinan el valor máximo y mínimo del conjunto de datos y se calcula la amplitud o rango con esta fórmula:

$$\text{Amplitud o rango} = \text{Valor máximo} - \text{Valor mínimo}$$

3. Se establece el número de clases (intervalos) de acuerdo con la fórmula:

$$\text{Núm. de clases} = 1 + 3.3 \log(n)$$

donde $\log(n)$ es el logaritmo del número de datos.

Es aconsejable que una distribución de frecuencias tenga cinco clases o más ya que así puede determinarse mejor el comportamiento de los datos.

4. Se calcula el ancho que debe tener cada clase:

$$\text{Ancho de clase} = \frac{\text{Rango}}{\text{Núm. de clases}} = \frac{1.65}{5} = 0.33$$

El ancho de clase se refiere a la longitud que existe entre el límite inferior y el límite superior de una clase.

5. Se ubican los límites de cada clase.
6. Se determina la *marca de clase*, que es el punto medio entre el límite inferior y el superior.
7. Por último, se cuenta el número de observaciones que corresponden a cada clase, es decir, la frecuencia absoluta. También se puede calcular la frecuencia relativa (f_r).

Se entrevistó a 20 jóvenes para conocer qué cantidad [en litros] de refresco de cola beben al día. Los resultados se presentan a continuación. Con ellos construye una distribución de frecuencias.

0.33	1.65	0.99	1.32	1.32
0	1.65	0.33	1.65	0.66
0.99	0.66	1.65	0.33	0.33
1.32	0.99	0.66	0.66	0

Recuerda que el logaritmo log de un número es el exponente al que debe elevarse la base para obtener dicho número, en otras palabras, la operación contraria de la exponenciación.

Ejemplo 3

La cantidad [en litros] de refresco es un *dato continuo*, ya que los resultados provienen de un proceso de medición.

Solución

Primero ordenamos los datos de menor a mayor:

0 0 0.33 0.33 0.33 0.33 0.66 0.66 0.66 0.66 0.99 0.99 0.99 1.32 1.32 1.32 1.65 1.65 1.65 1.65

En este caso, como la variable de estudio es la cantidad de refresco que beben al día 20 jóvenes, determinamos intervalos para establecer la frecuencia absoluta $\{f_a\}$. El procedimiento es el siguiente:

1. Identifica el valor máximo y mínimo del conjunto de datos y calcula la amplitud o rango:

$$\text{Amplitud o rango} = \text{Valor máximo} - \text{Valor mínimo}$$

$$\text{Amplitud o rango} = 1.65 - 0 = 1.65$$

2. Determina el número de clases (intervalos), según esta fórmula:

$$\text{Núm. de clases} = 1 + 3.3 \log(n) = 1 + 3.3 \log(20) \approx 5.29$$

(donde la n representa el número de datos). Por tanto, debemos crear cinco intervalos de clase.

3. Calculamos ahora el ancho que debe tener cada clase:

$$\text{Ancho de clase} = \frac{\text{Rango}}{\text{Núm. de clases}} = \frac{1.65}{5} = 0.33$$

4. Hallamos los límites de cada clase:

Observa que la diferencia entre el límite inferior y el superior equivale al ancho de clase. Además, el límite inferior corresponde al límite superior de la clase anterior.

Clase	Límites inferiores	Límites superiores
1	0	0.33
2	0.33	0.66
3	0.66	0.99
4	0.99	1.32
5	1.32	1.65

5. Determinamos la marca de clase:

Clase	Límites inferiores	Límites superiores	Marca de clase
1	0	0.33	0.165
2	0.33	0.66	0.495
3	0.66	0.99	0.825
4	0.99	1.32	1.155
5	1.32	1.65	1.485

6. Ahora debemos contar el número de observaciones que corresponden a cada clase, o sea, la frecuencia absoluta $\{f_a\}$. Aquí también podemos calcular la frecuencia relativa $\{f_r\}$.

Clase	Límites inferiores	Límites superiores	Marca de clase	f_a	f_r
1	0	0.33	0.165	6	30%
2	0.33	0.66	0.495	4	20%
3	0.66	0.99	0.825	3	15%
4	0.99	1.32	1.155	3	15%
5	1.32	1.65	1.485	4	20%
Total				20	100%

Histograma

Un *histograma* es una gráfica de columnas que representa la distribución de frecuencias de datos continuos. Se utiliza para ver cómo se distribuyen estos últimos, así como qué clases tienen mayor concentración de datos.

TRAZO DE UN HISTOGRAMA

Para trazar un histograma se siguen estos pasos:

1. Se dibujan los ejes coordenados.
2. En el eje x se localizan las marcas de cada clase. Como cada columna ejemplifica un intervalo de clase, la base debe coincidir con el ancho de clase.
3. En el eje y se refiere la frecuencia (absoluta o relativa), así que la altura de cada columna corresponde a ésta.

Polígono de frecuencias

Un *polígono de frecuencias* es una gráfica de línea que sirve para representar la distribución de frecuencias de datos continuos. Igual que el histograma, se utiliza para ver la forma de la distribución de los datos, así como la ubicación de la mayor concentración de éstos.

Como ya se explicó, un histograma y un polígono de frecuencias sirven para ver la forma de la distribución de los datos. Cada persona decide cómo presentar la gráfica, si con columnas o con línea.

TRAZO DE UN POLÍGONO DE FRECUENCIAS

Para graficar un polígono de frecuencias se siguen estos pasos:

1. Se trazan los ejes coordenados.
2. Se dibuja un punto por cada coordenada. Las coordenadas son pares ordenados donde la abscisa es la marca de clase (el eje x) y la ordenada es la frecuencia absoluta o relativa (el eje y).
3. Se unen los puntos.
4. Se cierran los extremos con el eje horizontal.

Ojiva

Una *ojiva* es una gráfica de línea donde se representa la frecuencia relativa acumulada.

TRAZO DE UNA OJIVA

Para graficar una ojiva se siguen estos pasos:

1. Se calcula la frecuencia relativa acumulada.
2. Se dibujan los ejes coordenados.
3. Se dibuja un punto por cada coordenada. Las coordenadas son marca de clase en el eje x , y la frecuencia relativa acumulada en el eje y .
4. Se unen los puntos.

Es importante señalar que la ojiva siempre es una curva ascendente. No toma valores negativos ni valores superiores a 100.

Retomemos el ejemplo de la entrevista a 20 jóvenes para conocer la cantidad de litros de refresco de cola que beben al día. Los resultados se presentan en una tabla de frecuencias. Con esa

Ejemplo 4

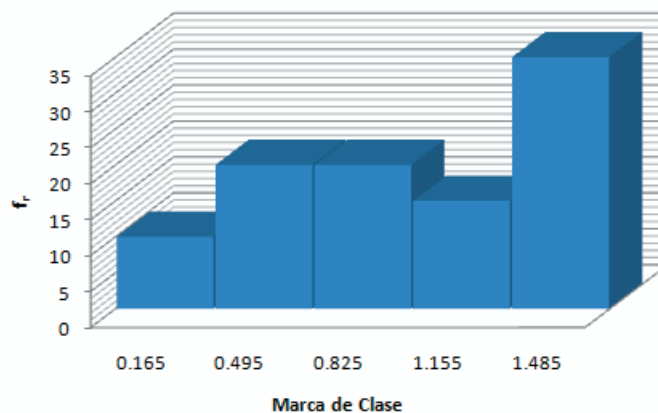
información, elabora el histograma de frecuencias relativas, el polígono de frecuencias relativas y la ojiva correspondientes. Asimismo, escribe tus conclusiones.

Clase	Límites inferiores	Límites superiores	Marca de clase	f_a	f_r
1	0	0.33	0.165	2	10%
2	0.33	0.66	0.495	4	20%
3	0.66	0.99	0.825	4	20%
4	0.99	1.32	1.155	3	15%
5	1.32	1.65	1.485	7	35%
Total				20	100%

Solución

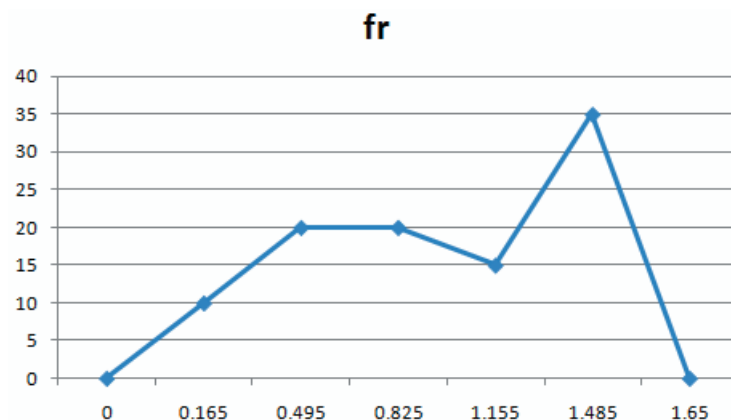
Recuerda, la marca de clase es el punto medio entre el límite inferior y el superior de cada intervalo.

Tracemos primero el histograma. Para ello, tomamos los datos de las columnas Marca de clase y f_i :



Como se puede observar en la gráfica, la clase con mayor concentración es la de 1.485 l (la columna más alta) y la clase con menor concentración es la de 0.165 l (la columna más baja). Es decir, 35% de los entrevistados toman en promedio 1.485 l de refresco, mientras que 10% bebe en promedio 0.165 l.

Para el polígono de frecuencias, tomamos los datos de las columnas Marca de clase y f_r , pero se realiza una gráfica de línea.



Al observar el polígono de frecuencias, llegamos a las mismas conclusiones que en el histograma: 35% de los entrevistados toman en promedio 1.485 l de refresco y 10% bebe en promedio 0.165 l.

Antes de trazar la ojiva debemos calcular la frecuencia relativa acumulada ($f_{r\text{ acum}}$):

Marca de clase	f_a	f_r	$f_{r\text{ acumulada}}$
0		0	0
0.165	2	10%	10%
0.495	4	20%	30%
0.825	4	20%	50%
1.155	3	15%	65%
1.485	7	35%	100%
Total	20	100%	

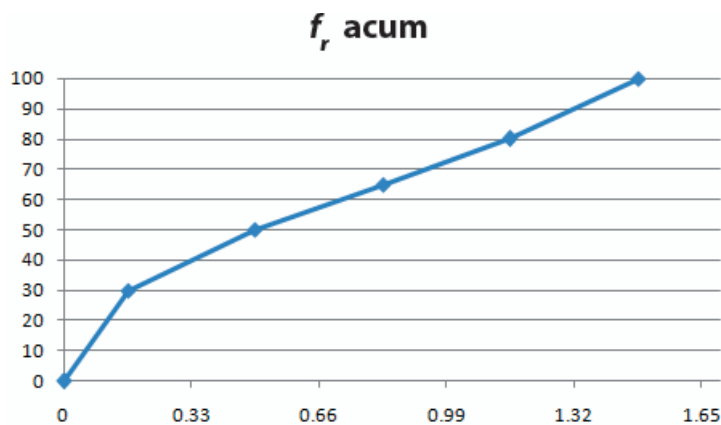
En el primer renglón, la $f_{r\text{ acum}}$ es igual a la f_r . A partir de la clase 2, la $f_{r\text{ acum}}$ es la $f_{r\text{ acum}}$ anterior más la f_r de la clase en que sitúa. Así, en la clase 2,

$$f_{r\text{ acum}} = 10 + 20 = 30$$

en la clase 3,

$$f_{r\text{ acum}} = 30 + 20 = 50$$

Ahora, para elaborar la ojiva, tomamos los datos de las columnas Marca de clase y $f_{r\text{ acum}}$ y trazamos una gráfica de línea:



De la ojiva podemos concluir que 30% de los entrevistados consumieron menos de 0.4125 l de refresco. La mitad, es decir, 50% consumieron menos de 0.825 l; en tanto, 65% consumieron menos de 1.2375 l.

Hoja de cálculo para graficar

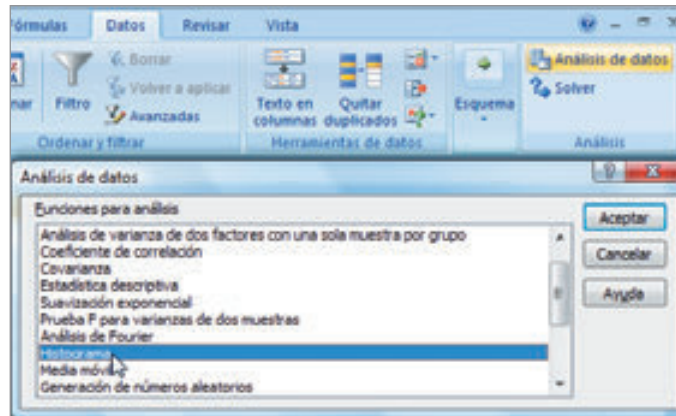
Para elaborar un histograma y una ojiva en Excel sigue estos pasos:

1. En la Barra de herramientas de Excel haz clic en la ficha Datos:



2. Enseguida haz clic en Análisis de datos² para desplegar la ventana Análisis de datos, donde debes hacer clic en Histograma:

² Si en tu versión de Excel no aparece la opción Análisis de datos, revisa el anexo C, donde se detallan los pasos que debes seguir para instalarlo.



3. Después, en el cuadro de diálogo que aparece debes indicar los rangos de datos de la manera siguiente:

Indica los datos de la muestra

Indica los límites inferiores de cada clase, excepto el último

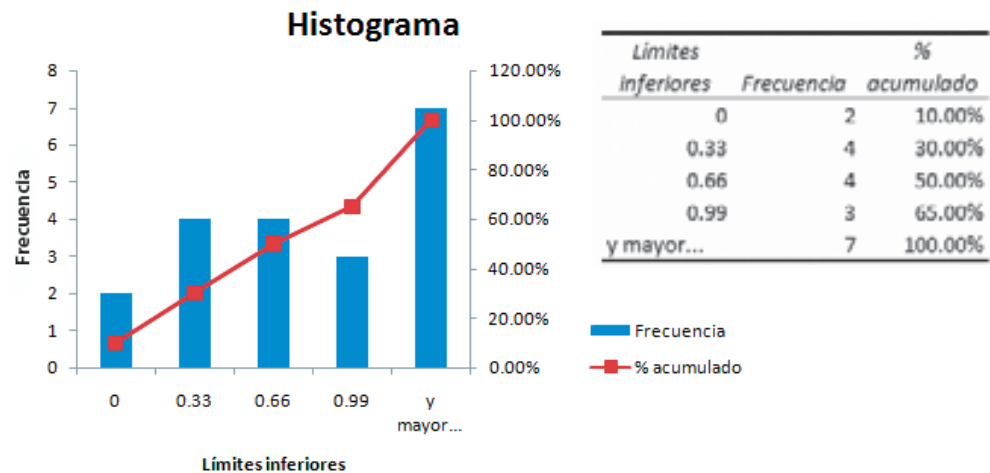
No olvides marcar Porcentaje acumulado y crear gráfico

Señala una celda de la hoja de cálculo (debes tener espacio libre a la derecha y abajo de la celda).

Clase	Límites inferiores	Límites superiores
1	0	0.33
2	0.33	0.66
3	0.66	0.99
4	0.99	1.32
5	1.32	1.65

Si quieres, puedes dejar vacío el cuadro de selección Rango de clases y Excel determina las clases automáticamente. No necesariamente coinciden con las que calculaste en forma manual.

4. Haz clic en Aceptar y listo, aparece el histograma:



Excel genera el histograma (correspondiente al eje izquierdo) y la ojiva (respecto del eje derecho). Además, proporciona una tabla de frecuencias con la frecuencia absoluta y la relativa acumulada.

Diagrama de barras

Un *diagrama de barras* se utiliza para representar una distribución de frecuencias de datos discretos o categóricos. En el eje x se indican los valores o las categorías y la altura de las barras corresponden a la frecuencia absoluta.

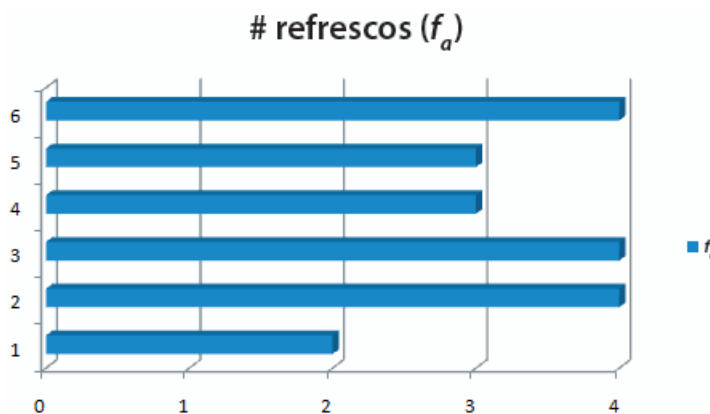
Se entrevistó a 20 jóvenes para conocer cuántos refrescos de cola se beben en un día. La distribución de frecuencias se presenta en la tabla siguiente. Traza un diagrama de barras a partir de la información que se muestra.

Ejemplo 5

# refrescos de cola	f_a	f_r
0	2	10
1	4	20
2	4	20
3	3	15
4	3	15
5	4	20
Total	20	100

Solución

Tomamos la frecuencia absoluta para graficar. El diagrama de barras queda así:

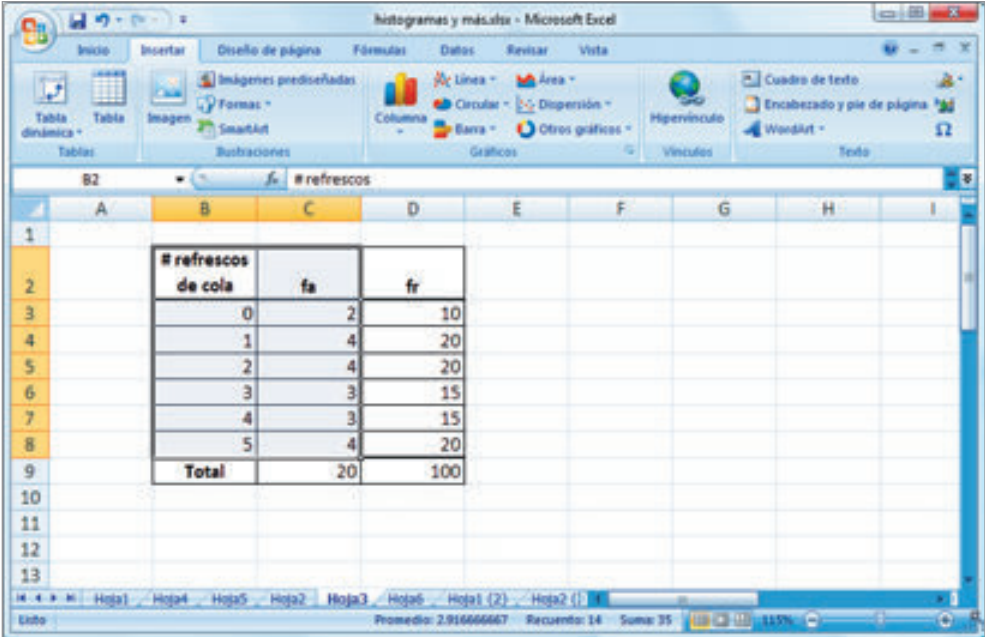


Para diferenciar el diagrama de barras de un histograma es común que en el primero se deje una separación entre cada barra, o bien, realizar el diagrama con barras horizontales, como en este ejemplo.

Interpretación: de acuerdo con el diagrama de barras podemos afirmar que cuatro personas consumen un refresco al día y que también cuatro beben dos o cinco refrescos de cola al día. Además, podemos decir que dos personas no consumen refrescos de cola.

Para realizar en Excel una gráfica de barras similar a la anterior sigue estos pasos:

1. Copia la distribución de frecuencias en una hoja de cálculo.
2. Selecciona las columnas “# refrescos de cola” y “ f_a ”.
3. Haz clic en la ficha Insertar:



histogramas y más.xlsx - Microsoft Excel

	A	B	C	D	E	F	G	H	I
1									
2		# refrescos de cola	fa	fr					
3		0	2	10					
4		1	4	20					
5		2	4	20					
6		3	3	15					
7		4	3	15					
8		5	4	20					
9		Total	20	100					
10									
11									
12									
13									

Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista

Tabla dinámica Tablas Imagen Imágenes prediseñadas Formas SmartArt Ilustraciones

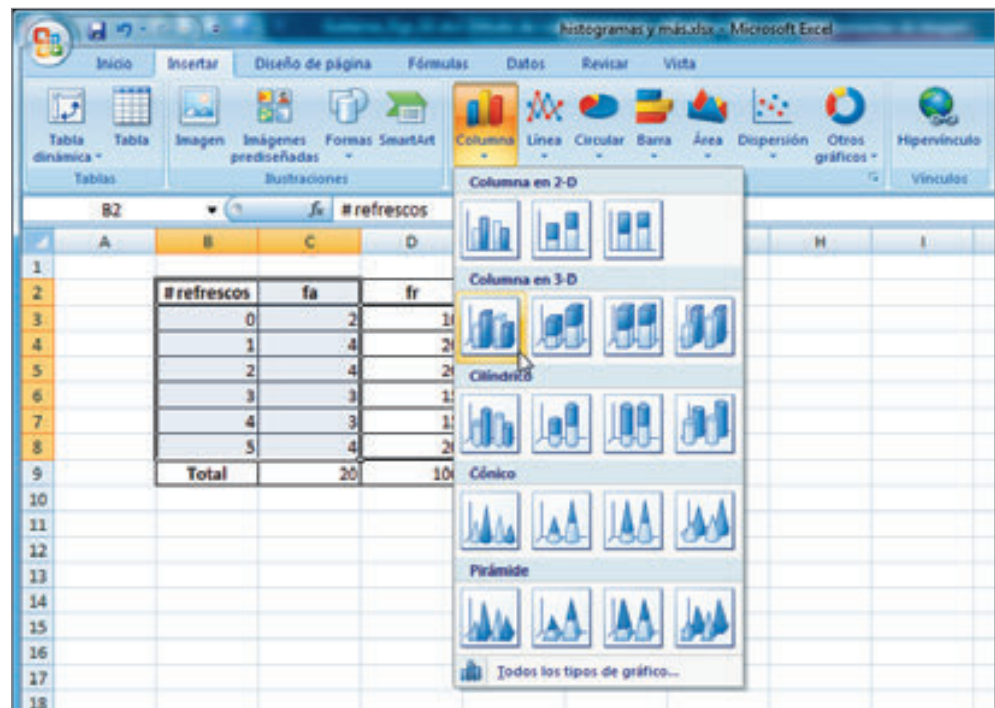
Columna Línea Área Gráficos Circular Barra Dispersion Otros gráficos

Hipervínculo Vinculos Cuadro de texto Encabezado y pie de página WordArt Texto

Hoja1 Hoja4 Hoja5 Hoja2 Hoja3 Hoja6 Hoja1 (2) Hoja1 (1)

Ente Promedio: 2.916666667 Recuento: 14 Suma: 35 155%

4. En la barra de herramientas se presentan las opciones de gráficos que ofrece el programa, haz clic en el botón Barra, después en una de las opciones que se despliegan y listo, aparece la gráfica de barras.



Se aconseja elegir gráficas sencillas, es decir, de rectángulos en dos o tres dimensiones (2-D o 3-D). Esto hace que su análisis sea más simple.

Diagrama circular o de pastel

Un *diagrama circular o de pastel*, igual que el diagrama de barras, se usa para representar una distribución de frecuencias de datos discretos o categóricos y, como su nombre lo indica, hay que trazar un círculo, en el que luego hay que dibujar divisiones (rebanadas) que representan la frecuencia relativa.

A partir de los datos del ejemplo anterior [5], representa con un diagrama circular el porcentaje de los 20 jóvenes que beben cierta cantidad de refrescos de cola en un día.

Ejemplo 6

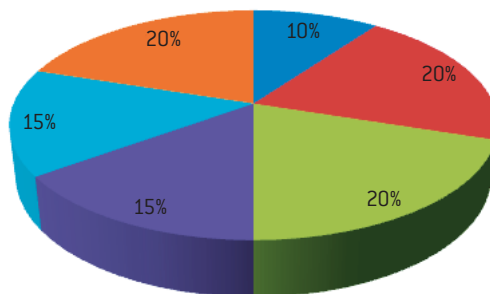
Solución

A partir de la frecuencia relativa (f_r) se realiza un diagrama circular:

% de personas que beben refrescos

f_r

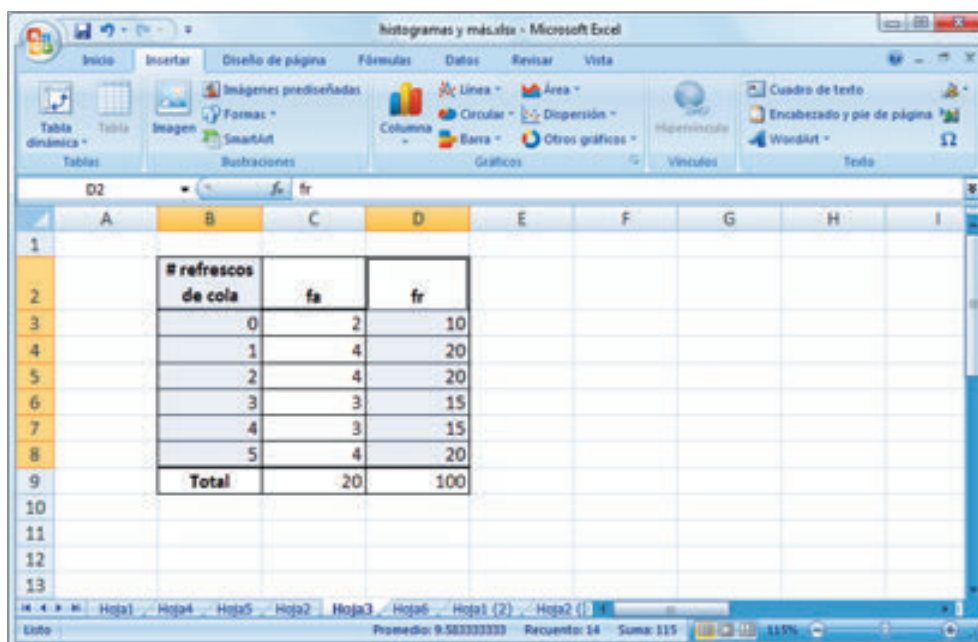
0 1 2 3 4 5



Interpretación: de acuerdo con la gráfica, se observa que 20% de las personas consumen dos refrescos al día, 20% beben cinco refrescos al día y 10% de las personas no consumen refrescos.

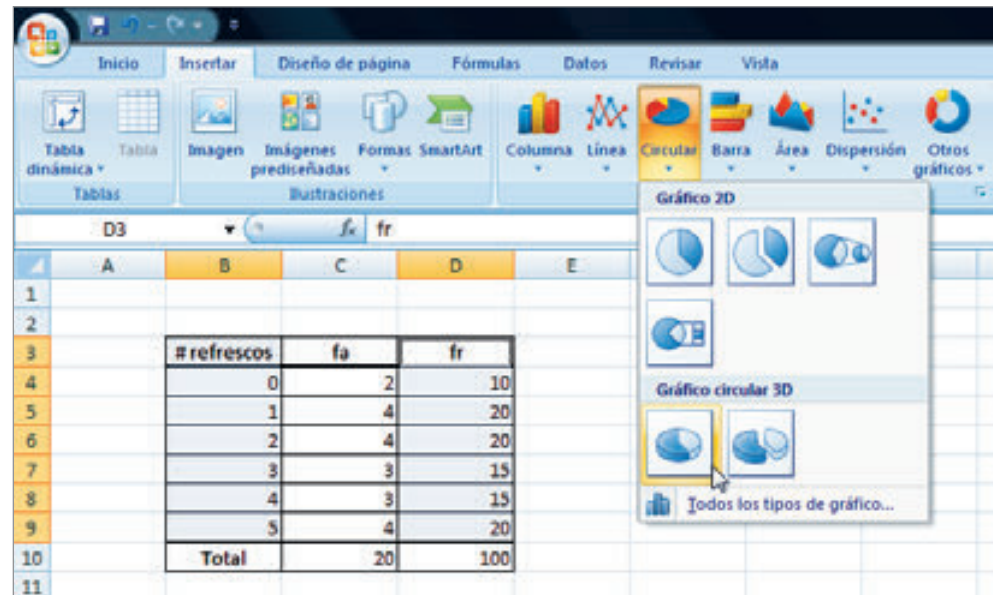
Para realizar con Excel la gráfica como diagrama circular o de pastel sigue estos pasos:

1. Copia la distribución de frecuencias en una hoja de cálculo.
2. Selecciona las columnas “# refrescos de cola” y “ f_r ”.
3. Haz clic en la ficha Insertar y después en el botón Circular:



Elige una de las opciones de gráficas que ofrece el programa y listo.

Se aconseja elegir gráficas sencillas, es decir, círculos en dos o en tres dimensiones (2-D o 3-D), ya que así es más simple analizar la gráfica.

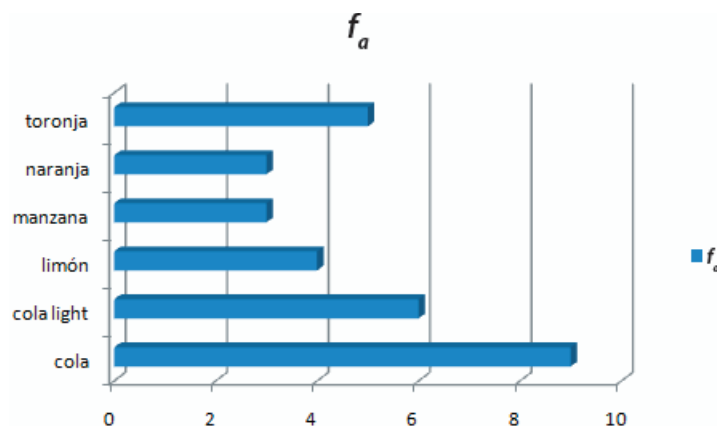


Ejemplo 7

Se entrevistó a 30 personas para conocer su sabor de refresco favorito. En la tabla de frecuencias siguiente se registran los resultados. Con base en ella, traza un diagrama de barras y uno circular. Interpreta.

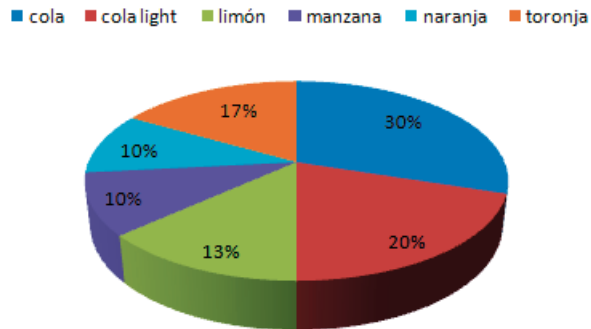
Sabor	f_a	f_r
Cola	9	30.00
Cola <i>light</i>	6	20.00
Limón	4	13.33
Manzana	3	10.00
Naranja	3	10.00
Toronja	5	16.67
Total	30	100

Solución



Interpretación: de acuerdo con la muestra, el refresco que tiene una mayor preferencia entre los consumidores es el de cola, después el de cola *light* y luego el de toronja. Los refrescos de manzana y naranja son los que menos se consumen. Enseguida aparece el diagrama circular.

% de preferencias en refrescos



Interpretación: de acuerdo con la muestra, 30% de las personas prefiere refrescos de cola, 20% consume refrescos de cola light, 17% toma refrescos de toronja. En contraparte, 10% de las personas consume refrescos tanto de naranja como de manzana.

Ejercicios

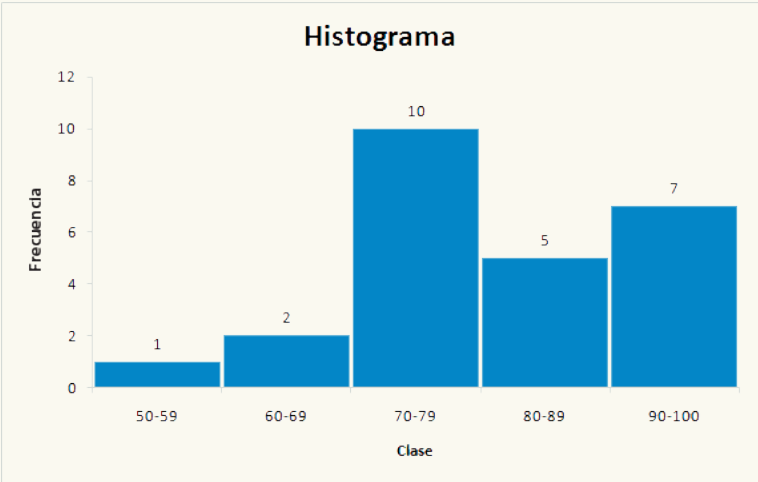
I. TEORÍA Y ALGO MÁS

1. Relaciona las columnas:

a. Histograma	()	Es un gráfico de línea donde se representa generalmente la frecuencia relativa acumulada
b. Distribución de frecuencias	()	Son los valores obtenidos mediante un proceso de conteo
c. Ojiva	()	Se utiliza para representar una distribución de frecuencias (absolutas) de datos discretos o categóricos
d. Polígono de frecuencias	()	Representa el número de veces que se observó cierto valor o cierta categoría
e. Diagrama de barras	()	Son los valores obtenidos mediante un proceso de medición
f. Diagrama circular o de pastel	()	Representa el porcentaje de cierto valor o categoría
g. Frecuencia absoluta	()	Se utiliza para representar una distribución de frecuencias (relativas) de datos discretos o categóricos
h. Frecuencia relativa	()	Tablas que muestran todos los posibles valores de una variable, junto con el número de veces que cada uno de ellos es observado
i. Datos discretos	()	Es una gráfica de columnas que representa la distribución de frecuencias de datos continuos
j. Datos continuos	()	Es una gráfica de línea que se utiliza para ver la forma de la distribución, así como dónde se localiza la mayor concentración de datos

II. PIENSO, ... ¿Y LUEGO?

2. En el histograma siguiente se muestran las calificaciones del primer examen de estadística:

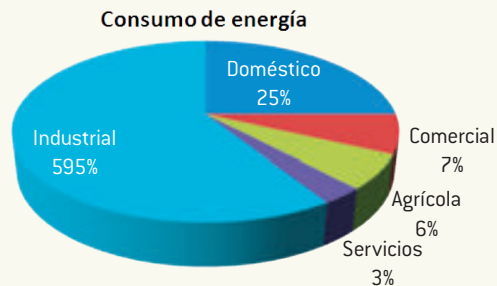


a. ¿Cuántos estudiantes presentaron el examen?

b. ¿Cómo se distribuyen las calificaciones?

c. ¿Cuántos estudiantes obtuvieron una calificación de menos de 70? ¿Cuántos más de 90?

3. En el diagrama siguiente se muestra el consumo de energía eléctrica en México por sector:

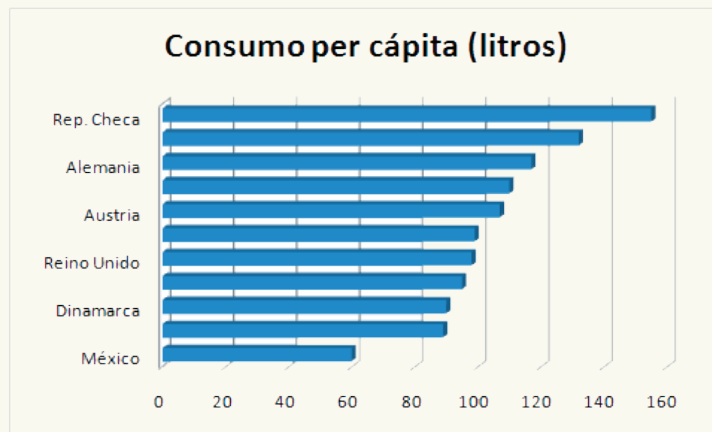


a. ¿Cuál es el sector que consume más energía?

b. ¿Cuál es el sector que consume menos energía?

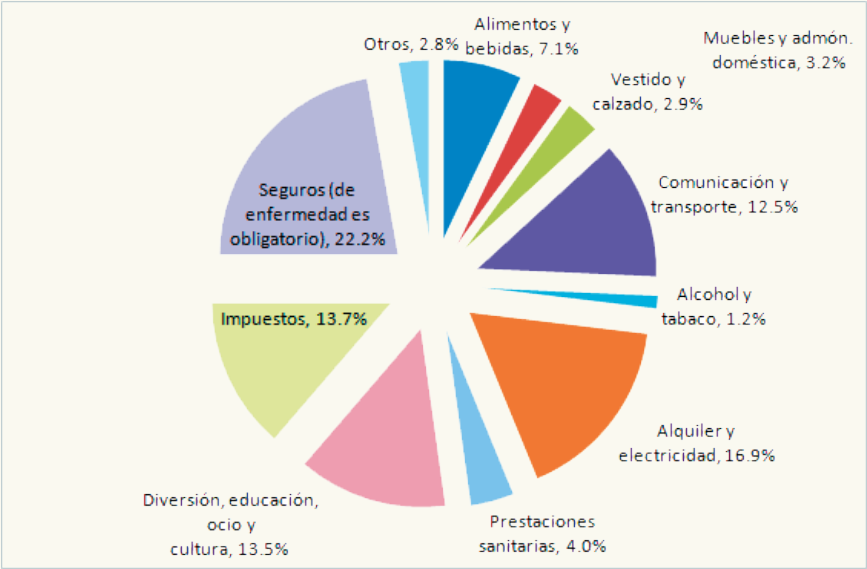
c. ¿Qué consecuencias consideras que tiene el consumo excesivo de energía eléctrica?

4. *Mitos y realidades.* En México se tiene la idea de que los alemanes son los que consumen más cerveza en el mundo. Según las empresas cerveceras, México es el país más rentable. Pero la realidad es otra. ¿A qué conclusiones llegas con base en el diagrama siguiente?



Fuente: Euromonitor internacional

5. *Economía.* Suiza tiene una de las economías más prósperas y modernas del mundo. Está basada en una mano de obra sumamente calificada y con un PIB per cápita de los más altos del orbe. En el 2005, los suizos gastaban así su sueldo (Tapia, Oliver y Cortés, Juan José: 2010):



Discute con tus compañeros si crees que los mexicanos gastamos en la misma proporción que los suizos.

III. PURO DEPORTE

6. *Economía.* De acuerdo con la Encuesta Nacional de Ingresos y Gastos de los Hogares (INEGI, 2008), los rubros en que gastan los mexicanos se presentan en la tabla siguiente. Con esta información traza un diagrama de barras y uno circular. Compara esta distribución de gastos con la presentada en el ejercicio anterior.
Discute con tus compañeros qué diferencias encuentras entre México y Suiza.

Ingreso y gasto	
Gasto corriente total promedio trimestral por hogar, por grandes rubros de gasto	
(Pesos a precios constantes de 2008)	
Grandes rubros de gasto	2008
Alimentos, bebidas y tabaco	7,389.00 ^a
Vestido y calzado	1,155.00 ^a
Vivienda y combustibles	2,207.00 ^a
Artículos y servicios para la casa	1,316.00 ^a
Cuidados de la salud	685
Transporte y comunicaciones	4,052.00 ^a
Educación y esparcimiento	2,974.00 ^a
Cuidados personales	1,530.00 ^a
Transferencia de gasto	676
Autoconsumo	305
Remuneraciones en especie	545
Transferencias en especie	2,352.00 ^a
Estimación del alquiler de la vivienda	4,091.00 ^a
NOTA: Para calcular el gasto promedio se toma en cuenta el total de hogares a nivel nacional.	
FUENTE: INEGI. Encuesta Nacional de Ingresos y Gastos de los Hogares.	

7. *Deporte*. Se entrevistó a 30 personas para que indicaran cuál es su equipo de fútbol preferido. Los resultados fueron los siguientes:

Pumas	América	Chivas	Pumas	Chivas
Chivas	América	Toluca	Toluca	Pumas
Santos	Necaxa	América	América	América
Pumas	Santos	Pumas	Pumas	Pumas
América	Chivas	América	Pumas	Necaxa
Toluca	Chivas	América	Toluca	Santos

- Construye la distribución de frecuencias correspondiente.
 - Construye un diagrama de barras y uno circular.
 - Discute con tus compañeros la información que se desprende de las gráficas anteriores.
8. *Tecnología*. Se preguntó a 30 jóvenes cuántas horas dedicaban cada día a navegar en internet (el tiempo que dedican a sesiones de *chat* y a las redes sociales quedan incluidos). Los resultados son los siguientes:

2.5	4.5	5.8	2.2	5.7
5.3	4.0	0.9	2.5	2.3
3.2	4.9	2.7	3.7	3.8
4.9	2.8	3.6	4.2	3.1
7.2	1.9	5.9	3.3	2.9
1.1	1.7	2.6	3.1	4.4

- ¿Qué tipo de datos se maneja en este problema?
- Construye la distribución de frecuencias correspondiente.
- Traza un histograma y una ojiva.
- ¿A qué conclusiones llegaron tú y tus compañeros? ¿Crees que es recomendable pasar tanto tiempo frente a una computadora?

9. *Seguridad.* De acuerdo con la Asociación Mexicana de Instituciones de Seguros (AMIS), en los primeros siete meses del año 2010, 55% de los autos se robaron cuando estaban estacionados, en tanto que 45% restante se hurtaron con violencia. Además, a continuación se muestran los vehículos más robados:

Vehículos con mayor número de robos en enero-julio 2010 en el país		
Marca	Tipo	# de unidades
Nissan	Tsuru	7,912
Nissan	Pick up	2,264
Nissan	Sentra	1,370
Volkswagen	Jetta 4a. Generación	1,091
Volkswagen	Bora	1,016
General Motors	Pick up	873
Ford	Pick up	599
Nissan	Estaquitas 35 Ton	445
Volkswagen	Pointer	411
Volkswagen	Sedán	405
Fuente: AMIS		

- ¿Qué tipo de gráficas recomiendas para mostrar esta información? ¿Por qué?
 - Construye una tabla de frecuencias por tipo de vehículo.
 - Realiza un diagrama de barras.
 - Construye una tabla de frecuencias por marca de automóvil y, con base en ella, traza un diagrama de barras.
 - Realiza un diagrama circular o de pastel.
 - Con tus compañeros, propongan medidas para disminuir el robo de autos.
10. *Demografía.* De los 7.7 millones de habitantes que tiene Suiza, 21% es de origen extranjero. Como ejemplo de su diversidad, en Suiza se hablan cuatro lenguas oficiales. (Tapia y Cortés: 2010).

Idioma	# Parlantes
Alemán	4 920 952
Francés	1 575 941
Italiano	494 413
Otros	695 268
Romance	38 626

- Traza un diagrama de barras y uno circular para representar la información anterior. Escribe un texto breve en el que interpretes los diagramas.
- Realiza los diagramas correspondientes para responder: ¿qué otras lenguas se hablan en México? ¿Cuáles son las que más se hablan? Consulta www.inali.gob.mx, o bien, www.inegi.org.mx.

11. *Ocio y diversión.* La información siguiente se refiere al número de películas exhibidas en México durante 2007, según el país de procedencia de la película:

Películas exhibidas	Distribución por país de origen			
	Estados Unidos de América	México	Otro país	Coproducción
98,911	74,085	7,616	9,792	7,418

- Realiza un diagrama circular a partir de la información anterior. ¿Cuáles son tus conclusiones?
 - Con tus compañeros indiquen cuántas películas ven al mes y de qué países proceden. Realicen los diagramas correspondientes. ¿Te has puesto a pensar qué sucedería si no se realizaran películas mexicanas?
12. *Nutrición.* El número de calorías correspondiente a los alimentos tipo sándwich que se sirven en un negocio de comida rápida se presenta a continuación:

250	790	470	510	500	330
300	750	580	740	420	260
440	770	400	540	530	330
390	380	340	460	530	260
410	360	270	510	630	330

- Realiza un histograma usando seis clases. Describe la forma de la distribución de los datos.
 - Traza una ojiva y expón tus conclusiones.
 - Forma un equipo de trabajo e investiga cuál es la ingesta calórica diaria de un mexicano. Compara ese resultado con los incisos anteriores. ¿Has pensado qué sucede si consumes alimentos con muchas calorías?
13. *Meteorología.* Los datos siguientes representan las temperaturas promedio en la República Mexicana durante octubre de 2010.

Temperatura media (°C) Octubre 2010							
27.4	29.5	23.5	32.3	29.6	24.8	34.7	23.0
28.5	28.1	27.0	24.8	34.3	26.2	32.3	28.1
32.1	30.5	21.6	28.7	29.9	32.5	30.7	31.4
31.8	33.3	26.9	26.9	29.9	29.9	31.1	26.1
Fuente: http://smn.cna.gob.mx							

- Construye una distribución de frecuencias donde se presenten las frecuencias absolutas y las relativas acumuladas.
- Traza un histograma y una ojiva. Analízalas.

14. *Demografía.* A continuación se presenta la esperanza de vida (promedio en años) para cada uno de los estados de la República Mexicana, tanto para hombres como para mujeres:

Entidad federativa	2010	
	Hombres	Mujeres
Aguascalientes	73.5	77.8
Baja California	73.8	78.6
Baja California Su	73.8	78.6
Campeche	73.1	77.6
Coahuila de Zarag	72.8	77.4
Colima	73.3	78.3
Chiapas	72.1	76.7
Chihuahua	73.8	78.2
Distrito Federal	73.8	78.8
Durango	72.5	77
Guanajuato	73.5	77.9
Guerrero	71.6	76.1
Hidalgo	72.8	77.5
Jalisco	73.4	77.9
México	73.6	78.5
Michoacán de Ocampo	72.9	77.2

FUENTE: CONAPO (www.conapo.gob.mx)

Entidad federativa	2010	
	Hombres	Mujeres
Morelos	73.6	78.7
Nayarit	73.1	77.5
Nuevo León	73.2	78.1
Oaxaca	72.3	77
Puebla	73	78
Querétaro	73.3	78.2
Quintana Roo	74.2	78.8
San Luis Potosí	73.1	77.8
Sinaloa	72.7	77
Sonora	73	78
Tabasco	72.8	77.2
Tamaulipas	73	77.7
Tlaxcala	74	78.5
Veracruz de Ignacio de la Llave	71.9	76.6
Yucatán	73.1	77.7
Zacatecas	73.7	78.2

- a. Elabora una distribución de frecuencias para hombres y otra para mujeres.
 - b. Traza un histograma y una ojiva para cada género. Compáralas. ¿A qué conclusiones puedes llegar?
 - c. ¿Qué influye para que aumente o disminuya la esperanza de vida?
15. *Sociedad.* En un estudio se desea conocer cuáles son los apellidos más comunes en México. Para ello se les preguntó a 100 personas su apellido y los resultados son los siguientes:

Ramírez	García	González	Rodríguez	Pérez	Martínez	García	López	Gutiérrez	López
Rodríguez	Martínez	Rodríguez	Hernández	García	García	González	Hernández	García	Gutiérrez
Hernández	López	Pérez	González	Martínez	Ramírez	García	Rodríguez	Hernández	Sánchez
Hernández	González	Rodríguez	Sánchez	Hernández	González	Gutiérrez	Martínez	González	Rodríguez
Rodríguez	Martínez	Ramírez	Pérez	Sánchez	Hernández	Hernández	González	García	Hernández
Rodríguez	López	Sánchez	Rodríguez	Ramírez	López	Martínez	García	García	Hernández
González	López	Martínez	López	García	Sánchez	García	Hernández	Flores	Pérez
González	González	López	Martínez	González	Martínez	Hernández	Ramírez	López	Sánchez
Ramírez	Gutiérrez	Sánchez	Flores	García	Martínez	García	Flores	Pérez	López
Hernández	Martínez	García	Hernández	Flores	Pérez	González	Hernández	Sánchez	Martínez

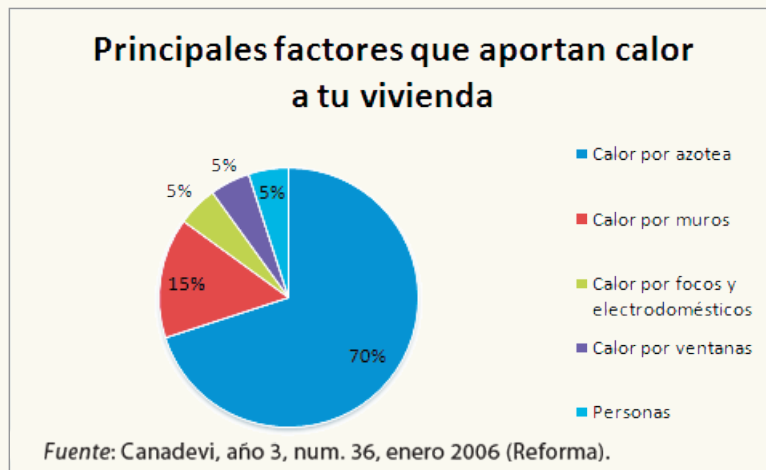
- a) Construye una tabla de frecuencias.
- b) ¿Qué tipo de gráficas utilizarías para representar esta información? ¿Por qué?
- c) Realiza las gráficas que propones en el inciso b y expón de manera oral las conclusiones correspondientes.

Evaluación sumativa

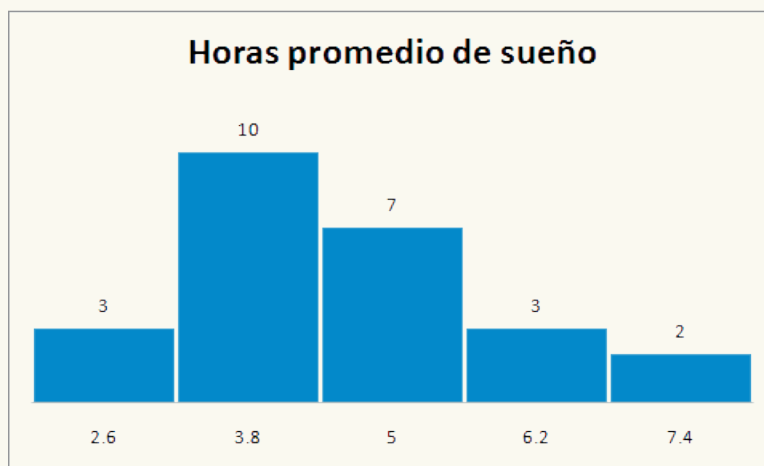
Un escalón hacia arriba

Contesta los siguientes reactivos que te permitirán saber si has comprendido lo expuesto en este bloque. ¡Adelante!

1. La Asociación de Empresas para el Ahorro de la Energía en la Edificación (AEAE) busca fomentar la cultura del ahorro de energía y desarrollar la comodidad en las edificaciones para reducir el consumo de energía eléctrica, gas y otros combustibles. Para ello, mediante un estudio se determinaron los factores siguientes que aportan calor a la vivienda. ¿A qué conclusiones llegas?



2. Durante la semana de exámenes, los alumnos suelen desvelarse. Se les preguntó cuántas horas dormían y éstos son los resultados:



- a) ¿Cuántos alumnos contestaron?
- b) ¿Cuál es la clase más alta y cuál es la más baja?
- c) ¿Hacia dónde está la concentración de la distribución? ¿Qué significa esto?

3. Se les preguntó a 15 personas cuál es su presentación favorita de una bebida. Las respuestas fueron:

Tetra pak	Botella retornable	Tetra pak
Lata	Botella de plástico	Lata
Lata	Lata	Tetra pak
Botella de plástico	Lata	Botella retornable
Botella retornable	Tetra pak	Botella retornable

- a) Construye una distribución de frecuencias.
 b) Traza un diagrama de barras y uno circular. Escribe tus conclusiones.
4. Se realizó una encuesta a 35 jóvenes entre 19 y 21 años, para que respondieran cuántos “amigos” tienen en su red social (*Facebook*). Las respuestas son:

327	429	470	550
340	250	650	326
530	250	302	280
206	645	350	956
360	1067	635	553
730	630	250	1000
240	1000	420	150
160	200	150	550
400	330	115	

- a) Construye una distribución de frecuencias.
 b) Grafica un polígono de frecuencias y una ojiva. ¿A qué conclusiones llegas? Preséntalas por escrito.

Proyecto 2

REMESAS INTERNACIONALES RECIBIDAS EN MÉXICO

Las remesas son cantidades de dinero enviadas por emigrantes a sus países de origen. En algunas naciones, las remesas se consideran la principal fuente de ingresos de la economía local.

Los migrantes han demostrado que se adaptan a condiciones severas: cambian de empleo, trabajan largas horas, reducen sus gastos, se mudan y hasta usan sus ahorros para seguir mandando dinero a sus familias. Volver a su país de origen es, generalmente, la última opción.

El envío de remesas llega al punto que los bancos nacionales establecen agencias en diversas ciudades de Estados Unidos para captar esta forma de negocio abierto por los inmigrantes. Incluso los gobiernos han comenzado a movilizarse para aprovechar el nacionalismo todavía vigente en las personas originarias, ofreciendo una especie de facilidades de coinversión en obras de tipo social y trasladando las soluciones pendientes a una parte de contribución de los ahora emigrantes.

Supón que el gobierno federal de México desea enviar al gobierno de Estados Unidos un estudio sobre la situación de los migrantes y te encargan realizar el análisis descriptivo de las remesas que éstos envían a México. Para ello debes presentar los resultados mediante gráficas y tablas. Consulta el archivo [remesas.xls](#), donde se presenta la información según el medio de transferencia. ¿A qué conclusiones llegas?

Al hacer tu estudio, te percatas de que las remesas por transferencia electrónica son superiores a las realizadas por *Money Order*, efectivo o especie, además de que se han mantenido constantes los envíos en los últimos años. Esto representa que, a pesar de la vigilancia en la frontera, los mexicanos siguen cruzándola y permanecen en Estados Unidos para buscar “mejores” ingresos. Estás en la disyuntiva de presentar la información tal cual la obtuviste, con el problema de que el gobierno estadounidense puede rastrear las transferencias electrónicas, localizar y deportar a los migrantes. Las consecuencias serían que los migrantes regresen a México y el gobierno mexicano no tiene la capacidad de emplearlos. Si modificas los datos, tal vez Estados Unidos pueda promover nuevas leyes a favor de los migrantes, pero el gobierno mexicano resentirá una disminución de sus ingresos.

Preguntas para dirigir la discusión

- ¿Qué harías? ¿Modificar las cifras? ¿Exponer las cifras como las presenta BANXICO?
- ¿Qué beneficios económicos y sociales se obtendrían en cada situación?
- ¿Qué pérdidas económicas y sociales habría en cada escenario?
- ¿Qué principios éticos estarían en juego en cada caso?
- ¿Cuál es la responsabilidad como profesionista?
- ¿Cuál crees que es la responsabilidad social que tienes como ciudadano?
- ¿Qué recomendaciones le harías al gobierno federal en cada escenario?
- ¿Cómo te sentirías si perdieras tu modo de vida y no tuvieras otra opción más que migrar?



Competencias

- Explica e interpreta el valor de las distintas medidas de una población, para la comprensión y el análisis del comportamiento de la misma
- Analiza las relaciones entre dos o más medidas de una población para determinar su comportamiento
- Formula y resuelve problemas matemáticos aplicando diferentes enfoques sobre las medidas de centralización o variabilidad para determinar algunas características de la población de estudio
- Estructura argumentos acerca del comportamiento de una población considerando las medidas provenientes de la misma
- Maneja tecnologías de la información para obtener y expresar medidas de tendencia central o variabilidad en diversas situaciones
- Interpreta el comportamiento de una población a partir de los resultados obtenidos utilizando tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos

3. Medidas de tendencia central y de dispersión

Contenido

- Las medidas de centralización y variabilidad para datos agrupados y sin agrupar, así como las relaciones entre ellas
- El comportamiento de una población a partir de las medidas estadísticas

Por cierto...

Adolphe Quetelet (Bélgica, 1796-1874) negaba el papel del azar y creía que la estadística se podía aplicar a las ciencias sociales, lo que llamó *physique sociale*. Un concepto básico de la física social de Quetelet es la del “hombre medio”, una persona con características físicas e intelectuales que servía de promedio. Le llamaba la atención que, al considerar algún atributo medible, como la altura o el peso, la mayoría de los miembros de una población determinada arrojaba datos cercanos a la *media*. Asimismo, introdujo el concepto de *homme type*, es decir, el *hombre típico* que caracteriza a un pueblo o una nación. Y llegó a creer que las *desviaciones* con respecto a este promedio eran accidentes (Cohen: 2005).

Ubícate

Antes de abordar los temas de medidas de tendencia central y de dispersión, contesta los ejercicios siguientes, que te permitirán evaluar si tienes los conocimientos que te facilitarán el estudio de este bloque.

1. Abel, un estudiante muy aplicado, sufre actualmente “mal de amores” y no le fue bien en los exámenes del periodo. Sus calificaciones fueron:

- Literatura: 8 ➤ Física: 6 ➤ Matemáticas: 6 ➤ Historia: 7
- Biología: 7 ➤ Inglés: 9

¿Cuál fue su calificación promedio?

2. A partir de los valores siguientes, realiza una tabla de frecuencias que muestre los límites de cinco clases, marcas de clase y la frecuencia absoluta (f_a):

8 4 5 10 8 3 1 9 4 2 10 5

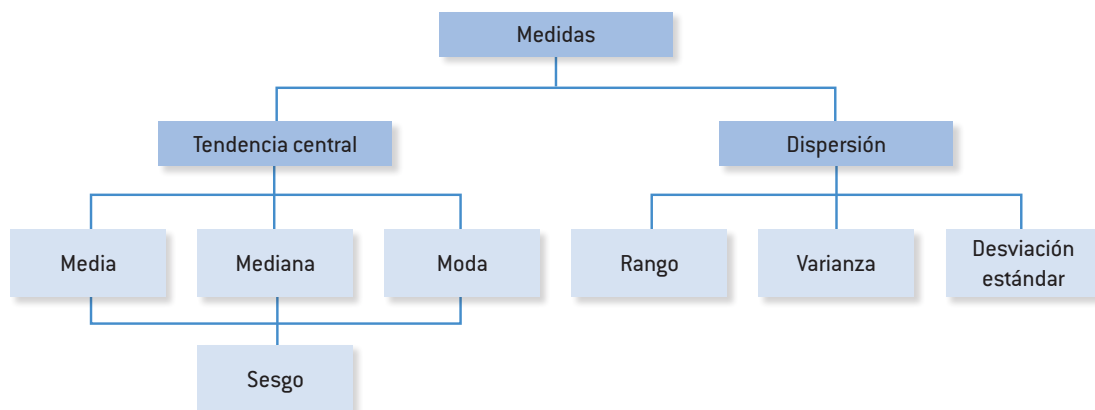
3. Calcula el promedio de: 1, 2, 3, 4, 5
4. Aplica las propiedades de la sumatoria y las sumas especiales para determinar el valor de V :

$$V = \frac{\sum_{k=1}^5 (k-3)^2}{4}$$

Si no tuviste problemas o dudas con los ejercicios anteriores, ahora sí comienza con este bloque.

Medidas de tendencia central y de dispersión

En el bloque 2 se estudiaron los histogramas, que nos permiten de forma visual conocer la forma de la distribución de los datos. En este bloque se determinará cómo es una distribución de forma numérica, a través de las medidas de tendencia central y de dispersión. En el esquema siguiente se muestran los temas que abordaremos.



Desempeños

Al terminar de estudiar este bloque y realizar todas las actividades propuestas habrás aprendido lo siguiente:

- Calcular las medidas de centralización en diversas situaciones a partir del conocimiento de los diferentes tipos de agrupación de datos para interpretarlos y analizarlos a través de las mismas.
- Calcular las medidas de variabilidad en diversas situaciones a partir del conocimiento de los diferentes tipos de agrupación de datos para interpretarlos y analizarlos a través de las mismas.
- Interpretar el comportamiento de una población a partir de las medidas de centralización y variabilidad de una muestra.

Medidas de tendencia central para datos no agrupados

Recordarás que en el bloque 2 revisamos las distribuciones de frecuencias. Si las examinamos, notaremos que en la mayoría de los conjuntos de datos se presenta cierta tendencia de agrupamiento alrededor de un valor central.

Las medidas de tendencia central permiten obtener valores que representen el punto central de los datos, es decir, determinar el valor más representativo de la variable que estamos analizando.

Las medidas de tendencia central más utilizadas son la media, la mediana y la moda.

Media

Una medida de tendencia central que comúnmente se usa es la *media*, también conocida como *valor promedio* o simplemente *promedio*. Se calcula de esta manera:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Aunque es un valor fácil de calcular, debemos tener en cuenta que si hay valores extremos o aberrantes (*outliers*), la media no será una medida representativa del centro de la distribución de datos.

Hay que tener cuidado con la notación que se utiliza. En el caso de la media, si se trata del valor poblacional se denota con la letra griega μ . Si es una muestra, entonces se utiliza $\bar{x}$, como en esta fórmula.

Calcula las medias de las muestras siguientes:

Muestra 1	3	4	6	3	5	5
Muestra 2	3	6	4	5	7	56

Ejemplo 1

Solución

Para la muestra 1:

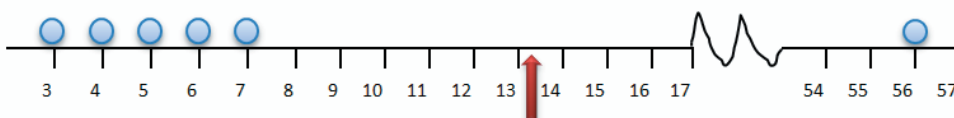
$$\bar{x} = \frac{3+4+6+3+5+5}{6} = \frac{26}{6} = 4.3333$$

Podemos realizar un diagrama de puntos para cada muestra. Un diagrama de puntos permite ver la distribución de los datos, donde cada punto en la recta numérica representa cada uno de los valores de la muestra.



Para la muestra 2:

$$\bar{x} = \frac{3+6+4+5+7+56}{6} = \frac{81}{6} = 13.5$$



Con este ejemplo podemos notar que las dos muestras son del mismo tamaño (seis datos) y la mayoría de los datos son cercanos. Pero en la muestra 2 hay un dato "alejado" de los demás. Esto causa que la media cambie, es decir, la media es sensible a valores extremos.

Para utilizar Excel, ingresa los datos (uno por celda), ya sea en una fila o en una columna. En otra celda, inserta la función =PROMEDIO() y selecciona los datos:

VA									
	A	B	C	D	E	F	G	H	I
1									
2	Muestra 1	3	4	6	3	5	5	=PROMEDIO(B2:G2)	
3	Muestra 2	3	6	4	5	7	56	PROMEDIO(número1, [número2], ...)	

Oprime la tecla Intro (Enter) para obtener los mismos resultados:

							Media
Muestra 1	3	4	6	3	5	5	4.3333
Muestra 2	3	6	4	5	7	56	13.5

Mediana

La mediana es el valor que se encuentra exactamente a la mitad de un arreglo ordenado, es decir, el conjunto de datos ordenados de menor a mayor. De lo anterior se deduce que la mitad de las observaciones son menores o iguales a la mediana.

Para determinar el valor de la mediana primero se obtiene el arreglo ordenado de los datos con los que estemos trabajando y luego se identifica la *mediana*. Para calcular la mediana debemos considerar dos casos:

- Si el número de datos (n) es impar, la mediana es el valor de en medio.
- Si el número de datos (n) es par, la mediana es el promedio de las dos observaciones de en medio.

Al contrario de la media, a la mediana no la afectan valores extremos o aberrantes, ya que sólo considera la posición del valor central.

Ejemplo 2

Calcula las medianas de las muestras siguientes:

Muestra 1	3	6	4	5	7	56	
Muestra 2	29	13	33	25	17	9	6

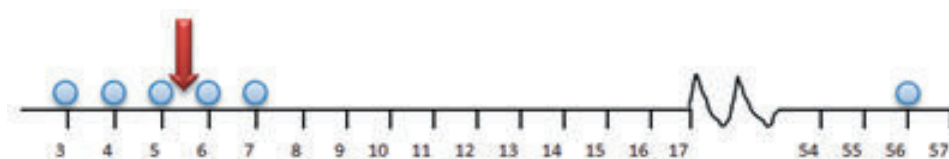
Solución

Primero ordenamos en forma ascendente los datos; para la muestra 1:

Muestra 1	3	4	5	6	7	56
-----------	---	---	---	---	---	----

Como n (el número de datos de la muestra) es par, los valores centrales son 5 y 6. Entonces,

$$\text{Mediana} = \frac{5+6}{2} = 5.5$$

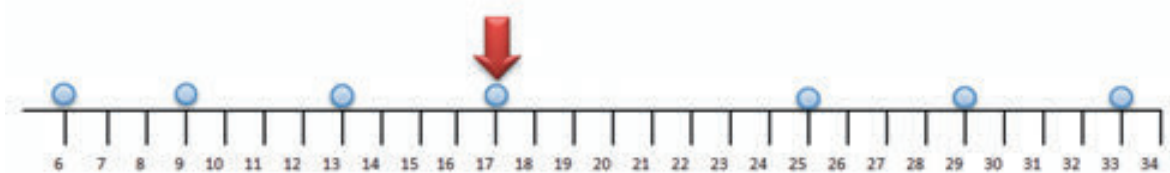


Para la muestra 2:

Muestra 2	6	9	13	17	25	29	33
-----------	---	---	----	----	----	----	----

Como n es impar, entonces el valor central se localiza en la posición $\frac{n+1}{2} = \frac{7+1}{2} = 4$; por tanto,

Mediana = 17



Para calcular la mediana con Excel, ingresa los datos (uno por celda), ya sea en una fila o en una columna. En otra celda, inserta la función `=MEDIANA()` y luego selecciona los datos (en este caso no es necesario ordenarlos):

DISTR.BINOM									
X ✓ f. =MEDIANA(C2:I2)									
	A	B	C	D	E	F	G	H	I
1									
2		Muestra 1	3	6	4	5	7	56	
3		Muestra 2	29	13	33	25	17	9	6

Oprime la tecla Intro para obtener los mismos resultados:

									Mediana
	Muestra 1	3	6	4	5	7	56		5.5
	Muestra 2	29	13	33	25	17	9	6	17

Moda

La *moda* es el valor que más se repite en una lista de datos, es decir, el que se presenta con mayor frecuencia. Puede darse el caso que haya dos modas en una lista de datos, lo que recibe el nombre de *bimodal*. Si hay más de dos modas, se denomina *multimodal*.

Si hay más de dos modas en un conjunto de datos, no se tendrá una buena descripción de la distribución de éstas.

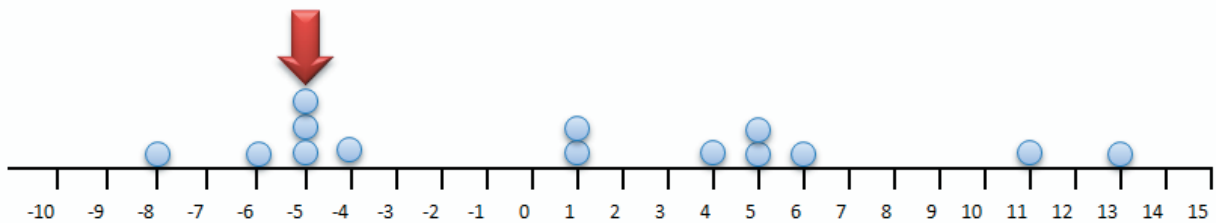
Calcula las modas de las muestras siguientes:

Muestra 1	1	-5	13	4	-8	-4	5	6	1	11	5	-5	-6	-5
Muestra 2	12	-3	0	13	15	8	13	-10	-9	6	8	-8	-8	-7

Ejemplo 3

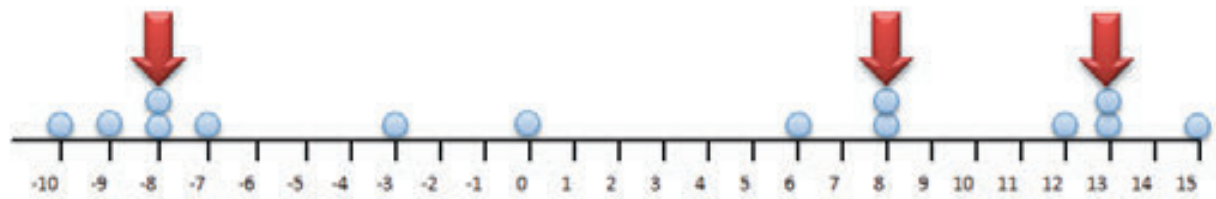
Solución

Para la muestra 1:



Al observar el diagrama, notamos que el valor que tiene mayor frecuencia es -5 : tiene tres repeticiones. Por tanto, es una distribución *unimodal* y,
 $\text{moda} = -5$

Para la muestra 2:



Al observar el diagrama, advertimos que es una distribución multimodal, pues son tres valores $\{-8, 8 \text{ y } 13\}$ los que tienen un mayor número de repeticiones $\{2\}$. Por consiguiente,
 $\text{moda} = -8, 8, 13$

Para utilizar Excel, ingresa los datos (uno por celda), ya sea en una fila o en una columna. En otra celda, inserta la función `=MODA()` y luego selecciona los datos:

VA		=MODA(B2:O2)															
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1																	Moda
2	Muestra 1	1	-5	13	4	-8	-4	5	6	1	11	5	-5	-6	-5	=MODA(B2:O2)	
3	Muestra 2	12	-3	0	13	15	8	13	-10	-9	6	8	-8	-8	-7	13	

Después oprime la tecla Intro y verás los resultados siguientes:

Hay que tener cuidado al usar la función `=MODA()` en Excel, pues sólo devuelve la primera moda que encuentra.

																	Moda
Muestra 1	1	-5	13	4	-8	-4	5	6	1	11	5	-5	-6	-5			-5
Muestra 2	12	-3	0	13	15	8	13	-10	-9	6	8	-8	-8	-7			13

Sesgo

El sesgo describe cómo es la distribución de los datos, ya que indica hacia dónde tienden a concentrarse éstos. Una distribución puede ser:

- *Simétrica*, si la mayor concentración de datos se localiza en el centro de la distribución.
- *Sesgada a la derecha*, si la mayor concentración de datos está a la izquierda de la distribución.
- *Sesgada a la izquierda*, si la mayoría de los datos están concentrados a la derecha.

Esto se puede determinar gráficamente, o bien, comparando la media, la moda y la mediana, de allí la importancia de estas tres medidas de tendencia central:

Sesgado a la izquierda

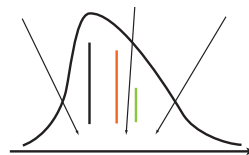
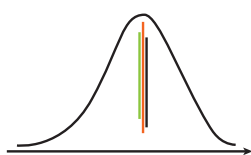
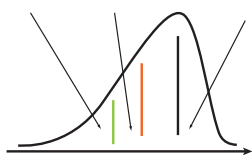
Simétrico

Sesgado a la derecha

Media < Mediana < Moda

Media = Mediana = Moda

Media < Mediana < Moda



Medidas de dispersión para datos no agrupados

Es importante también analizar cuán cercanos o lejanos están los datos respecto, por ejemplo, al valor medio. Para determinar esto se recurre a las llamadas *medidas de dispersión* o *de variabilidad*; de ellas, las medidas más importantes son el rango, la varianza y la desviación estándar.

De esta manera, si dos grupos de datos tienen el mismo centro, este centro es más descriptivo para el grupo que presente menor variabilidad.

Rango

El rango, a veces también denominado *recorrido*, es la medida de dispersión más fácil de determinar, ya que sólo depende de dos valores. Se calcula de esta manera:

$$R = X_{\text{máx}} - X_{\text{mín}}$$

donde R es el rango o recorrido.

$X_{\text{máx}}$ es el valor máximo del arreglo ordenado.

$X_{\text{mín}}$ es el valor mínimo del arreglo ordenado.

Algunas desventajas que presenta el rango son:

- Ignora la distribución de los datos, es decir, no considera si es unimodal o multimodal, o cuál es el sesgo.
- Los valores aberrantes influyen en el valor del rango.

Calcula los rangos de las muestras siguientes:

Muestra 1	2	2	-1	4	6	-2	3	2	4
Muestra 2	5	6	3	0	-2	2	1	-1	3

Ejemplo 4

Solución

Para la muestra 1:

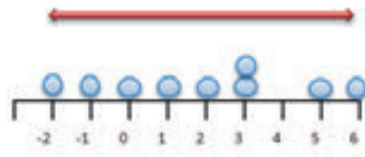
$$R = X_{\text{máx}} - X_{\text{mín}} = 6 - [-2] = 8$$



Para la muestra 2:

$$R = X_{\text{máx}} - X_{\text{mín}} = 6 - [-2] = 8$$

Con este ejemplo podemos notar que las dos muestras tienen el mismo rango o recorrido, así como los mismos valores máximos y mínimos, pero las distribuciones son distintas. La moda de la muestra 1 es 2, en tanto que la moda de la muestra 2 es 3.



Para utilizar Excel, ingresa los datos (uno por celda), ya sea en una fila o en una columna. En otra celda, inserta la función =MIN(), para obtener el valor mínimo, y luego selecciona los datos:

										Mín.	Máx.	Rango
Muestra 1	2	2	-1	4	6	-2	3	2	4	=MIN(U3:AC3)		
Muestra 2	5	6	3	0	-2	2	1	-1	3	MIN(número1, [número2], ...)		

En otra celda inserta la función =MAX(), que entrega el valor máximo de los datos:

										Mín.	Máx.	Rango
Muestra 1	2	2	-1	4	6	-2	3	2	4	-2	=MAX(U3:AC3)	
Muestra 2	5	6	3	0	-2	2	1	-1	3	-2	MAX(número1, [número2], ...)	

Por último, se realiza la resta correspondiente:

										Mín.	Máx.	Rango
Muestra 1	2	2	-1	4	6	-2	3	2	4	-2	6	=AE3-AD3
Muestra 2	5	6	3	0	-2	2	1	-1	3	-2	6	

Oprime la tecla Intro para obtener los resultados siguientes:

										Mín.	Máx.	Rango
Muestra 1	2	2	-1	4	6	-2	3	2	4	-2	6	8
Muestra 2	5	6	3	0	-2	2	1	-1	3	-2	6	8

Varianza

La varianza es la medida de dispersión más importante, pues tiene como base el promedio aritmético de las desviaciones (distancia de un valor con respecto a la media) elevado al cuadrado. En otras palabras, muestra cuán alejados o cuán cercanos están los datos respecto a la media.

Para calcular la varianza de la población o poblacional se utiliza esta fórmula:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{N}$$

- donde σ^2 es la varianza poblacional.
- x_i es el valor i del conjunto de datos.
- μ es la media poblacional.
- N es el tamaño de la población (número de observaciones).

Una de las grandes ventajas de la varianza es que si un dato está muy cerca de la media, al elevar esa distancia al cuadrado se hará más pequeña. Si el dato está muy lejos de la media, la distancia al cuadrado entre ese dato y la media se hará mayor. De esta manera hace más evidente si una distribución tiene una dispersión alta o baja.

Si se desea calcular la varianza de la muestra o muestral, entonces se usa:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

donde s^2 es la varianza muestral.

x_i es el valor i del conjunto de datos.

$\bar{x}$ es la media muestral.

n es el tamaño de la muestra (número de observaciones).

Desviación estándar

La desviación estándar es la medida de dispersión más utilizada, ya que el resultado se expresa en las mismas unidades que los datos originales. La desviación estándar es la raíz cuadrada de la varianza.

Para calcular la desviación estándar de la población se utiliza:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

donde σ es la desviación estándar poblacional.

x_i es el valor i del conjunto de datos.

μ es la media poblacional.

N es el tamaño de la población (número de observaciones).

Si se desea calcular la desviación estándar de la muestra o muestral, entonces se usa:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

donde s es la desviación estándar muestral.

x_i es el valor i del conjunto de datos.

$\bar{x}$ es la media muestral.

n es el tamaño de la muestra (número de observaciones).

En el cálculo de la desviación estándar muestral se divide entre $n - 1$, y no entre N , como en el caso del cálculo de la desviación estándar poblacional, porque la media que se utiliza es la de la muestra; por tanto, hay que realizar un pequeño ajuste.

Calcula la varianza y la desviación estándar de la muestra siguiente:

Muestra	5	6	3	0	-2	-1	3
---------	---	---	---	---	----	----	---

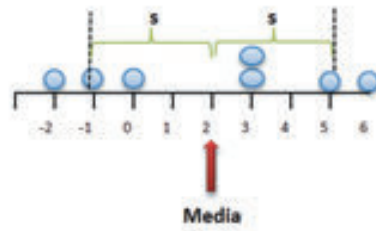
Ejemplo 5

Solución

Primero, calculamos la media:

$$\bar{x} = \frac{5+6+3+0-2-1+3}{7} = 2$$

$$\bar{x} = 2$$



Después, calculamos la varianza de la muestra:

$$s^2 = \frac{\sum_{i=1}^7 (x_i - 2)^2}{7 - 1} = \frac{(5-2)^2 + (6-2)^2 + (3-2)^2 + (0-2)^2 + (-2-2)^2 + (-1-2)^2 + (3-2)^2}{6} = 9.3333$$

Esto es,

$$s^2 = 9.3333$$

Finalmente, obtenemos la raíz cuadrada de la varianza para conocer la desviación estándar:

$$s = \sqrt{9.3333} = \pm 3.0551$$

Si queremos usar Excel, podemos emplear las funciones estadísticas =VAR() y =DESVEST() como sigue:

INTERSECCION.EJE										
A	B	C	D	E	F	G	H	I	J	K
									Varianza	Desv. Estándar
Muestra	5	6	3	0	-2	-1	3	=VAR(C4:I4)		
									VAR(número1, [número2], ...)	

INTERSECCION.EJE										
A	B	C	D	E	F	G	H	I	J	K
									Varianza	Desv. Estándar
Muestra	5	6	3	0	-2	-1	3	9.3333	=DESVEST(C4:I4)	
									DESVEST(número1, [número2], ...)	

Para llegar a los mismos resultados:

Importante: procura utilizar el mayor número posible de decimales para que tus resultados sean lo más exacto que se pueda. Si se te dificulta, entonces lo aconsejable es que uses por lo menos cuatro decimales.

									Varianza	Desv. Estándar
Muestra	5	6	3	0	-2	-1	3		9.3333	3.0551

Si quieres calcular la varianza y la desviación estándar de la población, las funciones son =VARP() y =DESVESTP().

Se entrevistó a 20 jóvenes para conocer qué cantidad (en litros) de refresco de cola beben al día. Los resultados son éstos:

Ejemplo 6

0.33	1.65	0.99	1.32	1.32
0	1.65	0.33	1.65	0.66
0.99	0.66	1.65	0.33	0.33
1.32	0.99	0.66	0.66	0

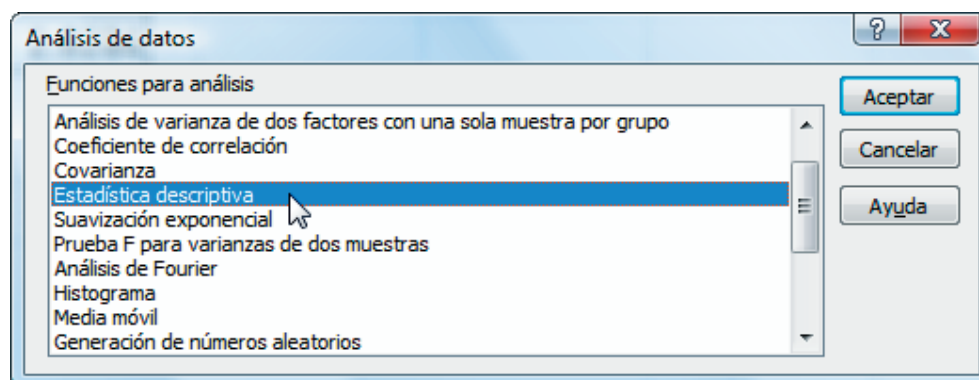
Describe la distribución de los datos, analizando tanto las medidas de tendencia central como las de dispersión.

Solución

Podemos calcular cada una de las medidas, ya sea de forma manual o con las funciones de Excel, que proporciona la herramienta Estadística descriptiva, que aparece al hacer clic en el botón Análisis de datos de la ficha Datos. Si este complemento no aparece instalado en tu versión, revisa el apéndice correspondiente al final del libro para saber cómo instalarlo.



Para usar esa herramienta primero debes acomodar los datos en una sola fila o columna. Después, en la ventana Análisis de datos busca la opción Estadística descriptiva, selecciónala y haz clic en Aceptar.



Aparece la ventana siguiente:

Indica los datos. Recuerda ponerlos en una sola fila o columna, lo que debes indicar en el botón de abajo

Se puso una etiqueta para saber a qué se refieren los datos. Esto se indica marcando la casilla Rótulos en la primera fila

Marca siempre esta casilla

Indica una celda para que se muestre ahí el resultado

Haz clic en Aceptar y entonces Excel devuelve:

<i>Cantidad consumida (litros)</i>	
Media	0.8745
Error típico	0.1250
Mediana	0.8250
Moda	0.3300
Desviación estándar	0.5592
Varianza de la muestra	0.3127
Curtosis	-1.2519
Coefficiente de asimetría	0.0377
Rango	1.6500
Mínimo	0.0000
Máximo	1.6500
Suma	17.4900
Cuenta	20

La varianza no se interpreta en el contexto del problema, ya que las unidades serían litros al cuadrado, algo que no tiene sentido.

De acuerdo con la muestra:

Muchas veces es difícil establecer la relación de la media, la mediana y la moda. Excel da el Coeficiente de asimetría, que indica cómo es el sesgo de la distribución: si el valor es positivo, el sesgo es hacia la derecha; si es negativo, hacia la izquierda, y si es cercano a cero, la distribución es casi simétrica.

- La cantidad media de refresco de cola que bebe una persona es 0.8745 litros.
- La mitad de las personas (50%) beben menos (o más) de 0.8250 litros.
- La cantidad típica (la que más se repite) de refresco consumida es 0.33 litros.
- La dispersión del consumo de refresco con respecto a la cantidad media es de ± 0.5592 litros.
- La diferencia entre la persona que más consume refresco y la que menos es de 1.65 litros.
- Como la moda < mediana < media, la distribución está sesgada a la derecha, es decir, la mayor concentración de datos es a la izquierda. Esto indica que la mayoría de las personas consumen poco refresco.
- La muestra es de 20 personas.

Medidas de tendencia central para datos agrupados

El término *datos agrupados* se refiere a cuando tenemos los datos divididos por clases y contamos únicamente con la frecuencia de cada una de ellas, es decir, cuando tenemos una tabla de frecuencias. En esta sección se explican tres medidas de tendencia central para datos agrupados: la media, la mediana y la moda aproximadas.

Media, mediana y moda aproximadas

La *media aproximada* se obtiene al sumar todos los productos de la frecuencia por la marca de clase y dividir entre el total de datos, es decir,

$$\bar{x}_a = \sum_{i=1}^n \frac{f_i \cdot x_i}{n}$$

donde $\bar{x}_a$ es la media aproximada.

f_i es la frecuencia absoluta de la clase i .

x_i es la marca de clase i .

n es el número de observaciones.

Por su parte, la *mediana aproximada* es:

- La marca de clase que esté en la posición $\frac{n+1}{2}$, si el número de datos (n) es impar.
- O bien, el promedio de las marcas de clase ubicadas en los lugares $\frac{n}{2}$ y $\frac{n}{2} + 1$, si el número de datos (n) es par.

Finalmente, la *moda aproximada* es la marca de clase con la frecuencia más alta.

En un conjunto de datos se puede presentar el caso que haya dos modas, lo que se llama *distribución bimodal*. Si tiene más de dos modas, recibe el nombre de *multimodal*.

Medidas de dispersión para datos agrupados

También es importante determinar la dispersión en una distribución de frecuencias. Para ello se revisarán la varianza y la desviación estándar aproximadas.

Varianza aproximada

Si se desea calcular la varianza de la muestra, entonces se usa:

$$s_a^2 = \frac{\sum_{i=1}^n (f_i \cdot x_i^2) - n\bar{x}_a^2}{n - 1}$$

donde s_a^2 es la varianza aproximada.

f_i es la frecuencia absoluta de la clase i .

x_i es la marca de clase i .

$\bar{x}_a$ es la media aproximada.

n es el número de observaciones.

Desviación estándar

Si se desea calcular la desviación estándar de la muestra, entonces se usa:

$$s_a^2 = \sqrt{\frac{\sum_{i=1}^n (f_i \cdot x_i^2) - n\bar{x}_a^2}{n-1}}$$

donde s_a^2 es la varianza aproximada.

f_i es la frecuencia absoluta de la clase i .

x_i es la marca de clase i .

$\bar{x}_a$ es la media aproximada.

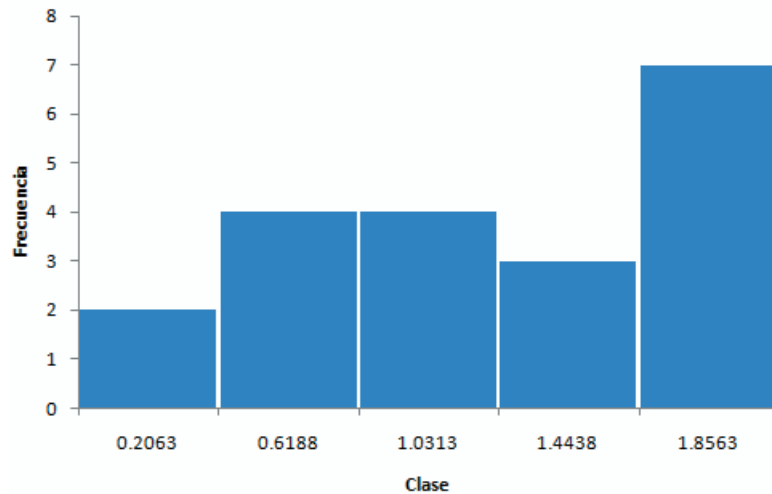
n es el número de observaciones.

Ejemplo 7

De acuerdo con el ejemplo 6, a continuación se presenta la tabla de frecuencias y el histograma correspondiente a las cantidades consumidas de refresco de cola:

	Clase	Marca de clase	f_a	$f_a \text{ acum}$
0	0.4125	0.2063	2	2
	0.4125	0.825	4	6
	0.825	1.2375	4	10
	1.2375	1.65	3	13
	1.65	2.0625	7	20

Histograma



Con esta información, calcula las medidas de tendencia central y de dispersión aproximadas. ¿A qué conclusiones llegas?

Solución

Para la media aproximada:

$$\begin{aligned}\bar{x}_a &= \frac{(0.2063 \cdot 2) + (0.6188 \cdot 4) + (1.0313 \cdot 4) + (1.4438 \cdot 3) + (1.8563 \cdot 7)}{20} \\ &= \frac{24.3375}{20} = 1.2169\end{aligned}$$

Lo que indica que una persona bebe, en promedio, 1.2169 litros.

Para la mediana aproximada, como el tamaño de la muestra es un número par ($n = 20$), entonces la mediana se encuentra entre $\frac{20}{2}$ y $\frac{20}{2} + 1$, es decir, en la posición 10.5. Nos fijamos en la columna de la frecuencia acumulada, donde se encuentra el 10.5:

Clase		Marca de clase	f_a	$f_{a \text{ acum}}$
0	0.4125	0.2063	2	2
0.4125	0.825	1.6188	4	6
0.825	1.2375	1.0313	4	10
1.2375	1.65	1.4438	3	13
1.65	2.0625	1.8563	7	20

Entonces:

$$x_{med} = \frac{1.0313 + 1.4438}{2} = 1.2375$$

lo que indica que 50% de las personas beben menos (o más) de 1.2375 litros.

Para la moda, como es la marca de clase con mayor frecuencia absoluta, entonces la cantidad típica consumida de refresco es de 1.8563 litros.

Para la varianza aproximada, calculamos las columnas siguientes,

Marca de clase (X)	f_a	$f_a \cdot X$	$f_a \cdot X^2$
0.2063	2	0.4125	0.0851
0.6188	4	2.475	1.5314
1.0313	4	4.125	4.2539
1.4438	3	4.33125	6.2532
1.8563	7	12.99375	24.1196
Suma =		24.3375	36.2433

Sustituimos en la fórmula,

$$s_a^2 = \frac{\sum_{i=1}^n (f_i \cdot x_i^2) - n \bar{x}_a^2}{n - 1} = \frac{(36.2433) - 20(1.2169)^2}{20 - 1} = \frac{36.2433 - 29.6157}{19} = 0.3488$$

Para la desviación estándar aproximada sólo calculamos la raíz cuadrada de la varianza:

$$s = \sqrt{0.3488} = 0.5906$$

Lo que significa que la dispersión respecto al consumo promedio de refrescos de cola es de ± 0.5906 litros. En otras palabras, una persona toma en promedio 1.2169 litros de refresco, pero podría tomar entre 0.6263 y 1.8075 litros.

Ejercicios

I. TEORÍA Y ALGO MÁS

1. Completa las definiciones que se presentan a continuación y busca en la Sopa de letras las respuestas correctas.

Las medidas de tendencia _____ y de _____ permiten describir una distribución de datos.

La _____ mide la dispersión de los datos respecto al valor medio de la distribución.

La _____ es el promedio de las observaciones.

La _____ es el valor central de una distribución.

La _____ es el valor con mayor frecuencia en una distribución.

El _____ mide la variabilidad de una distribución, utilizando el valor máximo y el mínimo.

El _____ indica cómo es la distribución, mostrando hacia dónde se localiza la mayor concentración de los datos.

La _____ mide la dispersión de los datos respecto al valor medio de la distribución. Su desventaja es que el resultado está expresado en unidades al cuadrado.

A	B	H	G	N	M	E	D	I	A	A	X	U	Q
M	R	G	K	T	Z	C	E	J	Z	Ñ	A	O	M
U	A	L	L	N	V	C	L	O	N	N	Q	W	M
R	N	J	R	A	H	A	Z	N	A	L	G	B	O
V	G	K	A	M	R	O	J	I	I	D	R	T	D
X	O	T	D	T	S	X	D	J	R	L	P	W	A
P	V	W	N	C	M	E	T	G	A	O	F	H	T
Ñ	C	E	A	V	M	U	S	X	V	S	T	I	R
W	C	N	T	G	L	K	Q	G	T	R	G	U	L
Y	D	E	S	V	I	A	C	I	O	N	B	Y	K
P	D	B	E	Z	E	Y	J	U	I	P	V	H	E
R	Y	B	M	F	S	A	M	K	L	Ñ	F	N	B
C	I	C	D	I	S	P	E	R	S	I	O	N	A
Z	O	X	V	S	T	F	Q	Y	U	O	E	T	R

II. PIENSO, ... ¿Y LUEGO?

2. El 16 de diciembre de 2010 apareció una nota en el diario *Reforma* de la Ciudad de México con este titular: “Atrapa alcoholímetro 160 ebrios por noche”, firmada por el reportero Ricardo Rivera; enseguida se presenta un resumen:

Por manejar en estado de ebriedad, un *promedio* de 160 conductores fueron remitidos a un juez cívico y 16 mil 595 automóviles fueron arrastrados a un corralón entre el 1o. de enero y el 11 de diciembre de este año, señalan cifras de la Secretaría de Seguridad Pública del DF (SSP-DF).

Los infractores, quienes rebasaron el límite de 0.40 de alcohol expirado en la prueba de alcoholemia, sumaron 23 mil 954 en las 150 jornadas del dispositivo realizado en la ciudad.

Las estadísticas de la SSP local señalan que durante este año se han realizado 478 mil 85 entrevistas y 86 mil 624 pruebas de alcoholemia...

Cifras de la dependencia señalan que desde septiembre de 2003 a la fecha el Programa Conduce sin Alcohol ha funcionado durante mil 181 jornadas, en las que han sido sancionadas 76 mil 704 personas.

De acuerdo con las estadísticas que se presentan en la nota, discute con tus compañeros de clase si ha funcionado este programa porque ha reducido el consumo de alcohol. En caso de que en tu ciudad no se aplique este programa, comenta con tus compañeros si funcionaría en tu entidad. También determina con ellos qué otras acciones crees que podría tomar el gobierno para evitar el exceso en el consumo de alcohol, así como para prevenir los accidentes automovilísticos.

3. *Delincuencia.* Uno de los graves problemas que enfrenta México en la actualidad es la violencia. El número de asesinatos en el país se ha incrementado en más de 40% en 2010 respecto a 2009. Las estadísticas muestran lo siguiente:

Asesinatos	
Media	347.91
Mediana	78.50
Moda	50.00
Desviación estándar	630.22
Varianza de la muestra	397 182.80
Coeficiente de asimetría	3.23
Rango	3 050.00
Mínimo (Yucatán)	
Máximo (Chihuahua)	3 050.00
Suma total	11 133.00
Estados	32

Fuente: Anuario 2010, diario *Reforma*

En clase, analicen los números anteriores. ¿Qué conclusiones se pueden obtener de esto?, ¿qué medidas propondrías al gobierno para evitar que siga creciendo el número de asesinatos?, ¿crees que sólo con educación se puede salir de esta situación?

III. PURO DEPORTE

4. *Sólo para practicar.* En la muestra siguiente obtén las medidas de tendencia central y de dispersión y di si presenta sesgo:

10	4	-3	-1	9	5	-5	0	-5	7	1	-9	-1	0	1
----	---	----	----	---	---	----	---	----	---	---	----	----	---	---

5. *Sólo para practicar.* De acuerdo con la distribución de frecuencias que sigue, calcula las medidas de tendencia central y de dispersión y di si la muestra presenta sesgo:

Clase		Marca de clase	f_a
2	13.5	7.8	3
13.5	25.0	19.3	6
25.0	36.5	30.8	7
36.5	48.1	42.3	5
48.1	59.6	53.8	2

6. *Transporte*. Los habitantes de la Ciudad de México destinan mucho tiempo a desplazarse de su casa a su lugar de trabajo. Una persona registró durante 13 días el tiempo (en minutos) que hacía en auto de la zona de Cuernavaca a Santa Fe, si salía a las 6:15 AM. Los resultados son los siguientes:

33	140	83	55	70
45	43	90	65	72
50	75	105		

- Determina las medidas de tendencia central y de dispersión, considerando que es una muestra. Interpretalas.
 - ¿Cómo es la forma de la distribución de los datos?
 - Junto con tus compañeros, expón cuáles son los principales problemas de tránsito en las ciudades, si crees que el transporte público es eficiente y seguro, y qué soluciones propones frente a este problema.
7. *Deportes*. A la Fórmula 1 (F1) se le conoce como la *reina del deporte motor*. En la competencia no sólo interviene la preparación física y mental del piloto, sino que es también un “juego” de estrategia. Las escuderías invierten millones de dólares en el desarrollo de los monoplazas (autos). En la temporada 2010 de F1, se determinó que en cada carrera se otorgara puntos a los diez primeros pilotos, de esta manera:

Posición	1	2	3	4	5	6	7	8	9	10
Puntos	25	18	15	12	10	8	6	4	2	1

A continuación se presentan las posiciones obtenidas por dos de los mejores pilotos: Fernando Alonso (campeón de 2005 y 2006) y Sebastian Vettel (campeón de 2010 y 2011):

Piloto	Escudería	Gran premio																	
		Baréin	Australia	Malasia	China	España	Mónaco	Turquía	Canadá	Europa	Gran Bretaña	Alemania	Hungría	Bélgica	Italia	Singapur	Japón	Corea	Brasil
Fernando Alonso	Ferrari	1	4	13	4	2	6	8	3	8	14	1	2	Ret*	1	1	3	1	3
Sebastián Vettel	Red Bull	4	Ret	1	6	3	2	Ret*	4	1	7	3	3	Ret*	4	2	1	Ret	1

*Ret: retiro, no suma puntos

Fuente: www.formula1.com

- Determina la media, la mediana y la moda del puntaje obtenido por cada piloto. Interpreta los resultados.
- Calcula la desviación estándar y la forma de la distribución de los puntos durante la temporada de 2010. Interpreta este resultado.

8. **Economía.** El sector turístico en México es una de las principales fuentes de ingresos en el país, de ahí la importancia de la inversión en infraestructura. En la tabla siguiente se presenta la cantidad de dinero que se ha invertido en este sector:

Inversión canalizada por FONATUR al desarrollo de infraestructura turística (en millones de pesos)

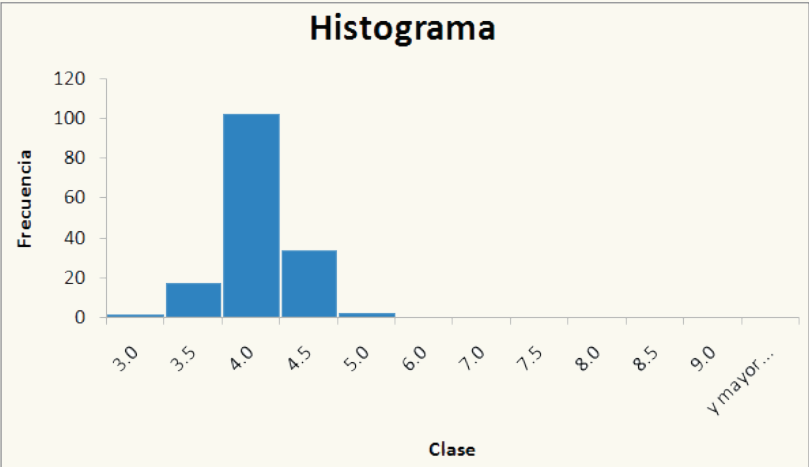
Año	Cancún	Ixtapa	Los Cabos	Loreto	Huatulco	Palenque	Nayarit	Marina
2009*	168.4	293.0	70.3	148.5	372.4	2.1	30.6	32.0

* Datos al mes de agosto

Fuente: Secretaría de Turismo

- Escribe un pequeño informe sobre la inversión canalizada por FONATUR en los principales destinos turísticos.
 - Según la página de SECTUR (www.turismo.gob.mx), “en el año 2030, México será un país líder en la actividad turística. Para lograrlo, la estrategia del nuevo gobierno se propone: reconocer al turismo como pieza clave del desarrollo económico de México, diversificar los productos turísticos y desarrollar nuevos mercados, impulsar a las empresas turísticas para ser competitivas en el nivel nacional e internacional, desarrollar el turismo respetando los entornos naturales, culturales y sociales”. ¿Cómo creen tú y tus compañeros que se podría lograr esta meta? ¿Consideran que sólo consiste en invertir en infraestructura? ¿Qué opinan acerca de que se abrieran hoteles nuevos para obtener dinero a costa, por ejemplo, de afectar los ecosistemas?
9. **Geofísica.** Un sismo o temblor es un movimiento vibratorio que se origina en el interior de la Tierra y se propaga en todas direcciones en forma de ondas. México es un país donde frecuentemente ocurren sismos y, como ejemplo, en la distribución de frecuencias siguiente se muestran los movimientos telúricos ocurridos del 1o. al 21 de diciembre de 2010 en el país:

Clase		Marca de clase	f_a
3.0	3.5	3.3	2
3.5	4.0	3.8	18
4.0	4.5	4.3	103
4.5	5.0	4.8	34
5.0	5.5	5.3	3
6.0	6.5	6.3	1
7.0	7.5	7.3	0
7.5	8.0	7.8	0
8.0	8.5	8.3	0
8.5	9.0	8.8	0
9.0	9.5	9.3	
		Total	161



- a) A partir de esta información, calcula las medidas de tendencia central aproximadas e interprétalas.
 - b) Determina y analiza las medidas de dispersión aproximadas.
 - c) En equipo, expón los resultados y comenta con tus compañeros de grupo qué se debe hacer antes, durante y después de un sismo y qué consecuencias trae el no tomar conciencia de que México es un país sísmico.
10. *Deportes.* Cada cuatro años se lleva a cabo uno de los acontecimientos deportivos más esperados por los aficionados: el campeonato mundial de futbol. En 2010 se realizó por primera vez en Sudáfrica. En la tabla siguiente se presenta la información sobre los goles anotados y recibidos por selección:

Equipo	Partidos jugados	Goles a favor	Goles en contra
Alemania	7	16	5
Holanda	7	12	6
Uruguay	7	11	8
Argentina	5	10	6
Brasil	5	9	4
España	7	8	2
Portugal	4	7	1
República de Corea	4	6	8
Ghana	5	5	4
Estados Unidos	4	5	5
Eslovaquia	4	5	7
Japón	4	4	2
Costa de Marfil	3	4	3
Italia	3	4	5
México	4	4	5
Paraguay	5	3	2
Eslovenia	3	3	3
Sudáfrica	3	3	5

(continúa)

(continuación)

Nigeria	3	3	5
Chile	4	3	5
Inglaterra	4	3	5
Dinamarca	3	3	6
Australia	3	3	6
Nueva Zelanda	3	2	2
Serbia	3	2	3
Grecia	3	2	5
Camerún	3	2	5
Suiza	3	1	1
Francia	3	1	4
RDP de Corea	3	1	12
Argelia	3	0	2
Honduras	3	0	3

Fuente: <http://es.fifa.com>

A partir de lo que aprendiste en este bloque, escribe un informe sobre el número de goles anotados y recibidos durante todo el Mundial. No olvides indicar cuál selección fue la que más/menos goles anotó y cuál fue la que menos/más goles recibió. Además, comenta con tus compañeros qué beneficios trae la realización de un acto de esta magnitud en un país como Sudáfrica, y qué necesitaría México para ganar la copa del mundo.

11. **Demografía.** La esperanza de vida al nacer es una estimación de los años que una persona podría llegar a vivir en cierto país. Además, este dato es uno de los factores que determina el índice de desarrollo humano (la calidad de vida en un país). Se obtuvo una muestra de los países que tienen un muy alto índice de desarrollo humano y otra de los que presentan un bajo índice. Esta información se resume en la tabla siguiente:

Esperanza de vida al nacer (2007)

Países con un muy alto índice de desarrollo humano	Esperanza de vida (años)	Países con un bajo índice de desarrollo humano	Esperanza de vida (años)
Alemania	79.8	Afganistán	43.6
Australia	81.4	Chad	48.6
Canadá	80.6	Costa de Marfil	56.8
Emiratos Árabes Unidos	77.3	Etiopía	54.7
España	80.7	Gambia	55.7
Estados Unidos	79.1	Mali	48.1
Francia	81.0	Mozambique	47.8
Grecia	79.1	República Centroafricana	46.7
Italia	81.1	República Democrática del Congo	47.6
Japón	82.7	Ruanda	49.7
Países Bajos	79.8	Senegal	55.4
Reino Unido	79.3	Sierra Leona	47.3
República Checa	76.4	Zambia	44.5
República de Corea	79.2		
Singapur	80.2		
Suecia	80.8		
Suiza	81.7		

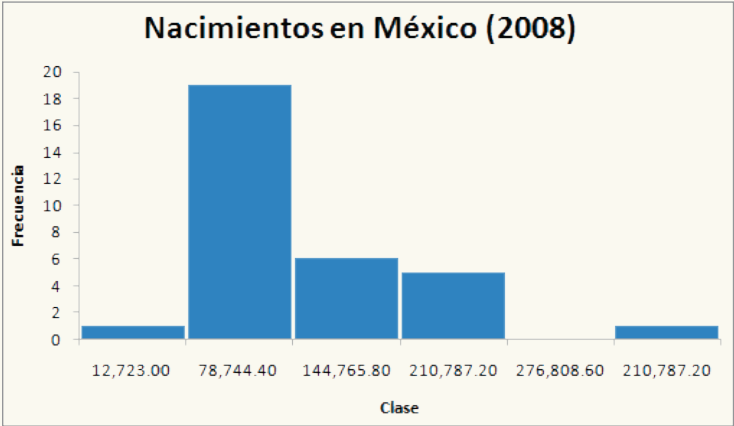
Fuente: <http://hdr.undp.org/es/desarrollohumano/>

En equipo, calcula las medidas de tendencia central y de dispersión. ¿A qué conclusiones llegas? Además, con tus compañeros de clase, determinen qué factores son importantes para alcanzar una mayor esperanza de vida, qué características tienen los países con un muy alto/bajo índice de desarrollo humano y para qué sirve calcular la esperanza de vida de un país.

12. *Demografía.* A continuación se presenta la distribución de frecuencias del número de nacimientos en México por entidad federativa durante 2008:

Clase		Marca de clase	f_a
12 723.00	78 744.40	45 733.70	1
78 744.40	144 765.80	111 755.10	19
144 765.80	210 787.20	177 776.50	6
210 787.20	276 808.60	243 797.90	5
276 808.60	210 787.20	243 797.90	0
210 787.20	276 808.60	243 797.90	1

Fuente: www.inegi.org.mx



- a) Determina los principales estadísticos que describen la distribución de nacimientos y luego interprétalos.
- b) ¿Crees que las campañas de planificación familiar han dado buenos resultados en el control de la natalidad? Explica oralmente por qué.
13. *Música.* Durante 2010, los diez artistas o grupos que más dinero obtuvieron por concepto de conciertos fueron los siguientes:

Artista	Ingreso (millones de dólares)
AC/DC	177
Bon Jovi	201.1
Lady Gaga	133.6
Metallica	110.1
Michael Bublé	104.2
Paul McCartney	93
Roger Waters	89.5

(continúa)

(continuación)

The Eagles	92.3
U2	160.9
Walking with dinosaurs	104.1

Fuente: *Reforma*, 30 de diciembre de 2010.

- a) ¿La información se considera una muestra o una población? ¿Por qué?
 - b) Determina los principales estadísticos que describen la distribución de los ingresos y luego interprétalos.
- 14. Ecología.** Una de las situaciones preocupantes que estamos viviendo es la falta de agua. Todo el proceso de extraer y distribuir el agua a las ciudades es muy costoso. Para remediar un poco el problema se han hecho campañas para crear conciencia en las personas de que no desperdicien el agua. Con tus compañeros de equipo, pregunten a 20 personas cuántos minutos tardan en bañarse. Con la información que reúnan, calculen las medidas de tendencia central y de dispersión, y luego tracen un histograma e interpreten los resultados. Tengan en cuenta que al dejar la llave abierta se gastan, en promedio, 25 litros de agua por minuto. ¿Qué otras acciones harían para disminuir el gasto de agua?

Evaluación sumativa
Un escalón hacia arriba

Contesta los reactivos siguientes que te permitirán establecer cuánto has aprendido de lo expuesto en este bloque.

1. A partir de los datos siguientes, calcula la media, la mediana, la moda, el rango, la varianza y la desviación estándar:

4	5	8	3	1
---	---	---	---	---

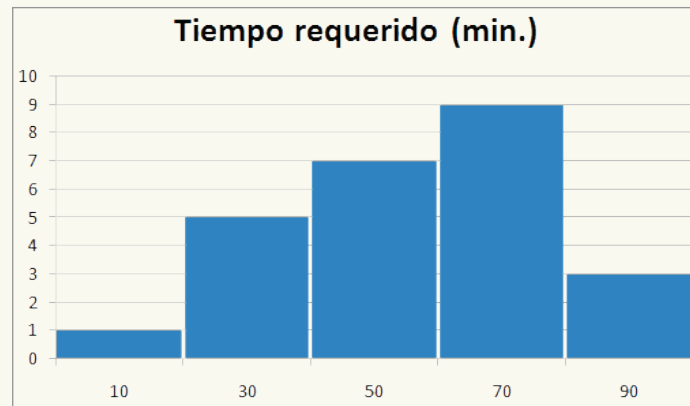
2. Debido a que se desea dar mayor apoyo al teatro en México, se realizó un estudio en todo el país para conocer el número de asistentes a espectáculos teatrales. Los resultados son los siguientes:

Asistentes a obras teatrales (2007)	
Media	153 902
Mediana	53 922
Moda	0
Desviación estándar	441 306
Varianza poblacional	194 750 968 004
Coficiente de asimetría	5
Rango	2 573 874
Mínimo (Morelos, Nayarit)	0
Máximo (Distrito Federal)	2 573 874
Suma	4 924 873
Cuenta	32

Fuente: estadísticas de cultura, INEGI

- a) Interpreta las medidas de tendencia central, así como las medidas de dispersión.
b) ¿Cómo es la forma de la distribución? Justifica tu respuesta.
3. Se preguntó a 25 jóvenes cuánto tiempo (en minutos) tardaban en comprar sus boletos por internet durante el primer día que éstos salían a la venta. Los resultados son:

Clase		Marca de clase	f_a	$f_a \text{ acum}$
0	20	10	1	1
20	40	30	5	6
40	60	50	7	13
60	80	70	9	22
80	100	90	3	25
		Total	25	



A partir de esta información determina media, mediana, moda, varianza y desviación estándar aproximadas.

Proyecto 3

CALIDAD DEL AIRE Y CAMBIO CLIMÁTICO

Del 29 de noviembre al 10 de diciembre de 2010 se llevó a cabo en Cancún la COP16, es decir, la 16ª edición de la Conferencia de las Partes de la Convención Marco de Naciones Unidas sobre el Cambio Climático. En esa reunión se expuso la importancia de tomar medidas y acciones para combatir el cambio climático, así como la contribución que el sector del transporte tiene en las emisiones contaminantes de manera local y global.

El sector transporte es el “más grande y de mayor crecimiento en el mundo, en términos de consumo de energía y emisiones de gases de efecto invernadero (GEI). México produce aproximadamente 18% del total de GEI, y el autotransporte representa cerca de 90% del consumo de energía de estos gases del sector” (Centro de Transporte Sustentable, 2010). Para contrarrestar los efectos de los GEI, se debe exigir que el transporte que entre en circulación sea de alta eficiencia para disminuir las emisiones.

La Ciudad de México es una de las más grandes y de las más contaminadas del mundo. Su nivel de polución se mide con el Índice Metropolitano de la Calidad del Aire (IMECA). Este índice sirve para informar cada hora a los habitantes sobre la calidad del aire en la Ciudad de México y su Zona Metropolitana, es decir, si es buena o representa un riesgo para la salud.

En el IMECA se cuentan contaminantes como ozono (O₃), partículas menores a diez micrómetros (PM10), dióxido de azufre (SO₂), dióxido de nitrógeno (NO₂) y monóxido de carbono (CO).

A continuación se presentan los niveles de la calidad del aire, las recomendaciones y los riesgos que se pueden presentar respecto a la calidad respectiva:

<i>Si la calidad del aire es...</i>	<i>y el IMECA se encuentra en el intervalo:</i>	<i>Recomendaciones y riesgos posibles:</i>
Buena	0 a 50	<i>Recomendaciones:</i> situación adecuada para llevar a cabo actividades al aire libre.
Regular	51 a 100	<i>Recomendaciones:</i> se pueden realizar actividades al aire libre. <i>Riesgos:</i> posibles molestias en niños, adultos mayores y personas con enfermedades respiratorias o cardiovasculares
Mala	101 a 150	<i>Recomendaciones:</i> evite las actividades al aire libre, esté atento a la información sobre la calidad del aire. Acuda al médico si presenta problemas respiratorios o cardíacos. <i>Riesgos:</i> posibles efectos adversos a la salud en particular en niños, adultos mayores y personas con enfermedades respiratorias o cardiovasculares.
Muy mala	151 a 200	<i>Recomendaciones:</i> evite salir de casa y mantenga las ventanas cerradas, no realice actividades al aire libre, esté atento a la información de la calidad de aire. Acuda al médico si presenta síntomas respiratorios o cardíacos. <i>Riesgos:</i> efectos adversos a la salud de la población en general. Se agravan los síntomas en niños, adultos mayores y personas con enfermedades cardiovasculares o respiratorias.

Extremadamente mala	>200	<i>Recomendaciones:</i> proteja su salud, no salga de casa o permanezca en lugares cerrados, esté atento a la información de la calidad de aire. Siga las recomendaciones de Protección Civil y las autoridades del sector salud. Acuda al médico si presenta síntomas respiratorios o cardíacos. <i>Riesgos:</i> efectos graves a la salud de la población en general. Se pueden presentar complicaciones en niños, adultos mayores y personas con enfermedades cardiovasculares o respiratorias.
---------------------	------	--

Supón que has sido contratado para elaborar el informe estadístico de la calidad del aire para cada zona de la Ciudad de México y tomas una muestra del IMECA del año 2010 (*archivo:imecas2010.xls*). Escribe un breve informe de las conclusiones a las que llegas.

Al terminar tu informe te percatas de que la calidad del aire no es buena y con base en ello debes llevar a cabo una presentación ante las autoridades correspondientes con las conclusiones y las recomendaciones para mejorar la calidad del aire en la zona metropolitana de la Ciudad de México.

Si te das cuenta de que la calidad del aire es mala, podrías recomendar el mejoramiento del programa “Hoy no circula”, es decir, que los automóviles, independientemente de si son nuevos o no, no circulen un día a la semana, lo cual causaría descontento entre la clase media y alta, y ocasionaría un aumento en las ventas de autos. La otra opción sería crear un impuesto variable según la cantidad de emisiones de GEI por parte de la industria, que originaría un conflicto con las industrias establecidas en el Valle de México.

Pero si decides tomar una muestra que sea en las horas con menos imeca, las autoridades correspondientes estarían contentas porque, al parecer, los programas ambientales habrían funcionado y la zona metropolitana sería una zona “verde”.

Preguntas para dirigir la discusión:

- ¿Qué harías: utilizar la muestra del archivo proporcionado o tomar otra muestra “representativa”?
- ¿Qué beneficios/pérdidas económicos, sociales y de salud se obtendrían en cada situación?
- ¿Qué principios éticos estarían en juego en cada caso?
- ¿Cuál crees que es la responsabilidad social que se tiene como ciudadano?
- ¿Qué otras recomendaciones harías a las autoridades en cada escenario para que la ciudadanía, el sector empresarial y la clase política estuvieran de acuerdo?

Reflexiona sobre el párrafo siguiente:

La tecnología no va a contrarrestar los efectos nocivos de un consumo irrestricto e irresponsable, podrá ayudar, pero si las consideraciones del cambio de cultura en cuanto al uso indiscriminado no se modifican, no será posible lograr transformaciones importantes. La sociedad en su conjunto, valiéndose de armas como la ciencia, logrará los avances y mejoras que se requieren para combatir la amenaza climática que afecta al mundo entero (Centro de Transporte Sustentable, 2010).



Competencias

- Expresa ideas y conceptos mediante representaciones lingüísticas y matemáticas relacionadas con las operaciones básicas de la teoría de conjuntos y la probabilidad y sus aplicaciones
- Maneja las tecnologías de la información para el análisis de resultados obtenidos en las operaciones de conjuntos y en las de probabilidad
- Sigue instrucciones y procedimientos de manera reflexiva y ordena información de acuerdo con categorías, jerarquías y relaciones con base en los conceptos básicos de la teoría de conjuntos y la probabilidad
- Sintetiza evidencias obtenidas mediante la experimentación para producir conclusiones y formular nuevas preguntas relacionadas con la teoría de conjuntos y la probabilidad
- Estructura ideas y argumentos de manera clara, coherente y sintetiza los resultados de la teoría de conjuntos y la probabilidad
- Define metas y da seguimiento a sus procesos de construcción de conocimiento sobre la teoría de conjuntos y la probabilidad

4. Probabilidad y teoría de conjuntos

Contenido

- Teoría de conjunto como base de la probabilidad
- La probabilidad y su aplicación

Nadie puede predecir con certeza qué cara de una moneda no manipulada quedará hacia arriba cuando caiga al suelo. Aunque la moneda acabe de caer diez veces seguidas en cara, eso no mejora ni un ápice nuestra capacidad de predecir con certeza la siguiente tirada. Sin embargo, sí podemos predecir con certeza que, si se tira la moneda diez millones de veces, prácticamente la mitad de las tiradas serán caras y la otra mitad serán cruces.

De hecho, a finales del siglo XIX, el estadístico Karl Pearson (Londres 1857-1936) tuvo la paciencia de tirar una moneda 24 000 veces. Obtuvo cara en 12 012 de las tiradas. En cierto modo, esto es, en esencia, la teoría de la probabilidad. Esta disciplina proporciona información precisa acerca de los resultados recogidos en un gran número de experimentos; no es capaz de predecir el resultado de un experimento específico. (Livio, 2009)

Ubícate

Antes de comenzar con el estudio de la probabilidad, contesta este examen para determinar si tienes los conocimientos necesarios para comprender cabalmente este bloque.

1. Especifica los miembros de cada uno de los siguientes conjuntos:
 - a) Los estados de la República que tienen litoral con el Golfo de México.
 - b) Los números enteros entre -5 y 1 , ambos inclusive.
2. Selecciona la descripción más adecuada para cada uno de los conjuntos:

a) $\{w, x, y, z\}$	b) $\{1, 2, 3, 4, 5, 6\}$
i) Cuatro consonantes del alfabeto.	i) Seis números.
ii) Cuatro letras del alfabeto.	ii) Los números naturales entre 1 y 6 , inclusive.
iii) Las cuatro últimas letras del alfabeto.	iii) Seis números enteros positivos.
3. Convierte los números decimales siguientes en fracciones:
 - a) 0.25
 - b) 0.125
 - c) 0.3333
4. Convierte las fracciones siguientes en números decimales. Exprésalos con cuatro decimales:

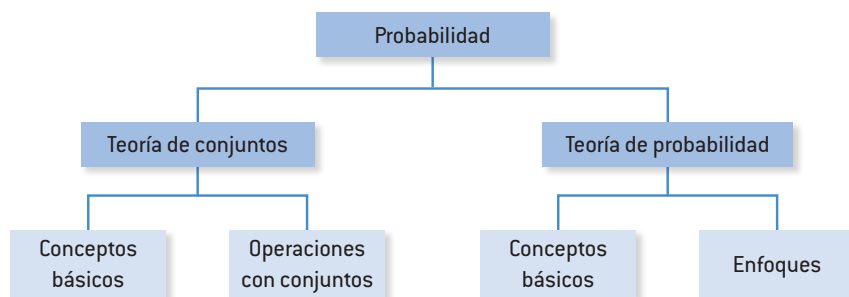
a) $\frac{3}{20}$	b) $\frac{1}{50}$	c) $\frac{2}{33}$
-------------------	-------------------	-------------------

Si no tuviste problemas o dudas con el examen anterior, ahora sí comienza el estudio del bloque 4.

Probabilidad

El término *probabilidad* es un concepto vago que se usa en la vida diaria para indicar cuán posible es que se presente u ocurra un evento en el futuro. Conocer “algo” o “mucho” del futuro nos ayuda a tomar decisiones. Por ello es importante estudiar la aplicación de la probabilidad y cómo medirla, lo mismo que cómo emplearla para hacer inferencias.

En este bloque presentaremos las ideas que sustentan la teoría de la probabilidad, para lo cual estudiaremos, entre otros temas, la teoría de conjuntos. En el esquema siguiente se muestran los temas que abordaremos.



Desempeños

Al terminar de estudiar este bloque y realizar todas las actividades propuestas habrás aprendido a:

- Identificar los elementos de un conjunto y sus operaciones.
 - Analizar y reconocer las operaciones de un conjunto como base para la probabilidad.
 - Comprender las características de experimento, espacio muestral, punto muestral y evento como elementos básicos en la aplicación de la probabilidad simple.
-

Teoría de conjuntos

Para lograr un desarrollo ordenado de la teoría de la probabilidad requerimos conocer los conceptos básicos de la teoría de conjuntos. Desde el punto de vista matemático, un *conjunto* es una colección bien definida, es decir, es una colección específica descrita con claridad. Los objetos o personas que constituyen un conjunto se llaman *elementos* o *miembros* del conjunto.

Se debe tener mucho cuidado al definir los elementos que pertenecen al conjunto. Siempre ha de buscarse que las definiciones no sean ambiguas.

Descripción de conjuntos

Hay tres formas de describir o definir un conjunto:

- *Por enumeración*: cuando se escribe o elabora una lista de los elementos que constituyen el conjunto.
- *Por comprensión*: cuando se proporciona una regla con la que se identifican los elementos del conjunto.
- *Por diagrama de Venn*: es un método gráfico para representar conjuntos y sus relaciones. Consta de un rectángulo para representar el conjunto universo (adelante definiremos este concepto), dentro del cual se trazan círculos para representar los conjuntos.

Cuando un elemento pertenece a un conjunto se utiliza la notación $x \in A$, esto es, "x pertenece al conjunto A". Si no es un elemento del conjunto entonces se escribe $x \notin A$.

Se utilizan letras mayúsculas para denotar a un conjunto.

Define los conjuntos siguientes por enumeración, por comprensión y mediante un diagrama de Venn:

Ejemplo 1

- a. Los días de la semana.
- b. Los números múltiplos de 6 entre 10 y 30, inclusive.

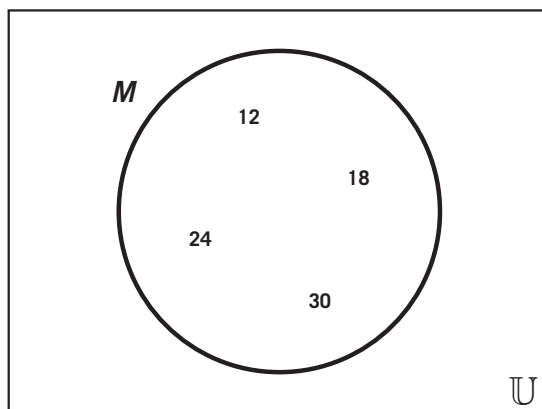
Solución

- a. Para el conjunto S , los días de la semana, tenemos:
 - Por comprensión: $S = \{\text{los días de la semana}\}$.
 - Por enumeración: $S = \{\text{lunes, martes, miércoles, jueves, viernes, sábado, domingo}\}$
 - Por diagrama de Venn:



- b. Para el conjunto M , los múltiplos de 6 entre 10 y 30, tenemos:
 - Por enumeración: $M = \{12, 18, 24, 30\}$

- ▶ Por comprensión: $M = \{x \in M \mid x = 6n, 10 \leq x \leq 30, n \in \mathbb{N}\}$, lo que se lee como “sea x un elemento del conjunto M , tal que x es múltiplo de 6 entre 10 y 30 inclusive, y n es un elemento del conjunto de los números naturales”.
- ▶ Por diagrama de Venn:



Conjunto universal

Se llama *conjunto universal* o simplemente *conjunto universo* el conjunto que contiene todos los elementos que interesan en una situación determinada. Se acostumbra denotarlo con una U .

En diagramas de Venn se utiliza un rectángulo para representar el conjunto universal; dentro de él se dibujan los conjuntos que lo forman, como se muestra en la figura 1.

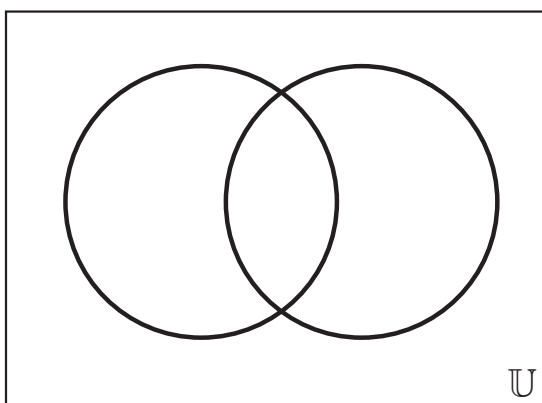


Figura 1 Conjunto universal o universo.

Ejemplo 2

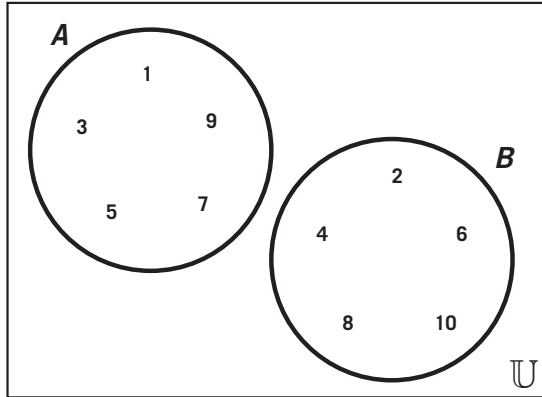
Supón que tenemos los dos conjuntos siguientes:

$$A = \{1, 3, 5, 7, 9\} \text{ y } B = \{2, 4, 6, 8, 10\}$$

Entonces, para este caso, el conjunto universo es:

$$U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

En diagrama de Venn:



Subconjunto

Sean dos conjuntos, A y B . Cuando todos los elementos de A están contenidos o pertenecen al conjunto B se dice que A es *subconjunto* de B . En notación matemática esto se representa con $A \subseteq B$.

Se dice que A es *subconjunto propio* de B , $A \subset B$, si B tiene por lo menos un elemento más que A (figura 2).

Observa que todo conjunto es subconjunto del conjunto universo.

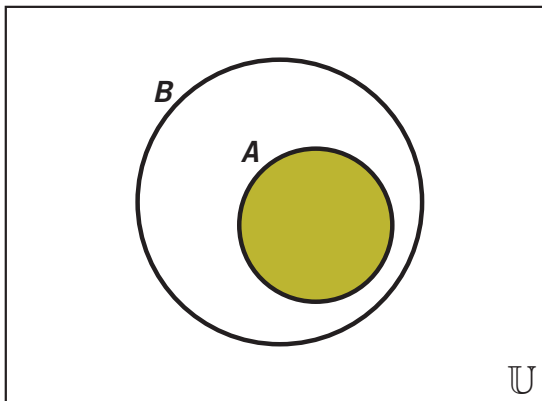
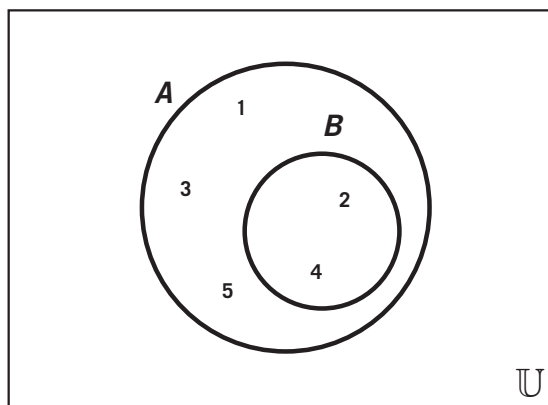


Figura 2 A es un subconjunto propio de B , es decir, $A \subset B$.

Sean los conjuntos $A = \{1, 2, 3, 4, 5\}$ y $B = \{2, 4\}$. Como todos los elementos de B pertenecen a A (de hecho, A tiene más elementos que B), entonces B es un subconjunto propio de A , esto es, $B \subset A$. El diagrama de Venn queda de la manera siguiente:

Ejemplo 3



Hay que señalar que el conjunto vacío es subconjunto de todo conjunto.

Conjunto vacío

El conjunto vacío o nulo es el que carece de elementos y se simboliza con $\emptyset$, o bien, mediante $\{ \}$.

Ejemplo 4

Define por enumeración el conjunto $E = \{\text{hombres vivos de más de 300 años de edad}\}$.

Solución

$E = \emptyset$, o bien, $E = \{ \}$

Cardinalidad de un conjunto

La cardinalidad de un conjunto A cualquiera es el número de elementos que contiene, distintos unos de otros, y se simboliza con $\#(A)$.

Ejemplo 5

Determina la cardinalidad de cada uno de los conjuntos siguientes:

- $A = \{\text{días de la semana}\}$
- $B = \{x, \text{ tal que } x \text{ es número primo y } 10 < x < 25\}$
- $C = \{\text{número de grandes premios de fórmula 1 para 2011}\}$
- $D = \{\text{mano, corazón, club, diamante, trébol}\}$
- $E = \{ \}$

Solución

- $\#(A) = 7$
- $\#(B) = 5$
- $\#(C) = 19$
- $\#(D) = 5$
- $\#(E) = 0$

Operaciones con conjuntos

A continuación definimos las cuatro operaciones básicas con conjuntos: unión, intersección, complemento y diferencia.

Unión

Sean dos conjuntos A y B cualesquiera. La *unión* de los conjuntos A y B es el conjunto de todos los elementos que están en A , en B o en ambos. Se simboliza con $A \cup B$. En diagrama de Venn se ve como se muestra en la figura 3.

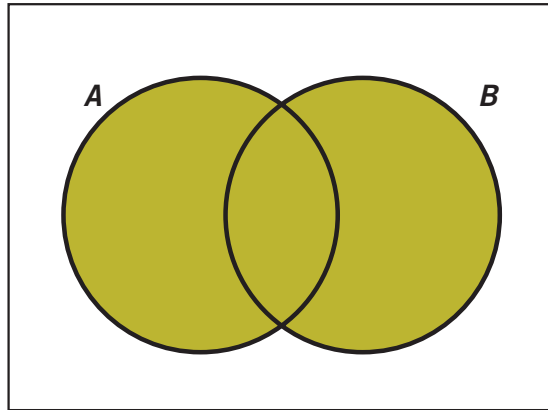


Figura 3 Unión de los conjuntos A y B .

Intersección

Sean dos conjuntos A y B cualesquiera. La *intersección*, de los conjuntos A y B $A \cap B$, es el conjunto de los elementos que están en A y también en B . En la figura 4 se muestra el diagrama de Venn para representar gráficamente la intersección de conjuntos.

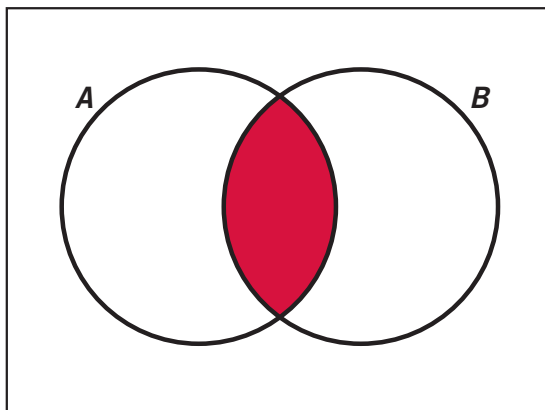


Figura 4 Intersección de los conjuntos A y B , $A \cap B$.

Cuando $A \cap B = \emptyset$, se dice que A y B son *conjuntos disjuntos* o mutuamente excluyentes (figura 5).

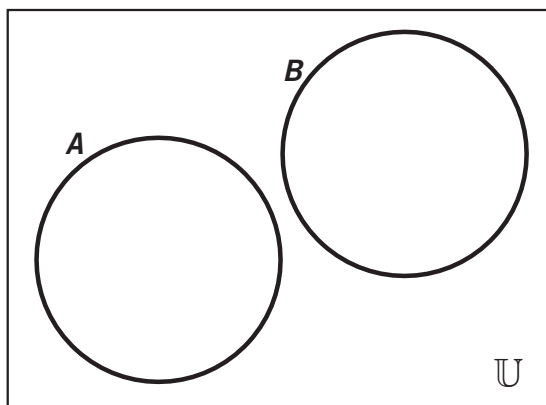


Figura 5 Conjuntos disjuntos o mutuamente excluyentes.

Complemento

El *complemento* del conjunto A es el conjunto de todos los elementos del universo que no están en A . En otras palabras, son todos los elementos que faltan a A para ser U . Entonces $A^c = U - A$. En la figura 6 se presenta gráficamente el complemento de conjuntos.

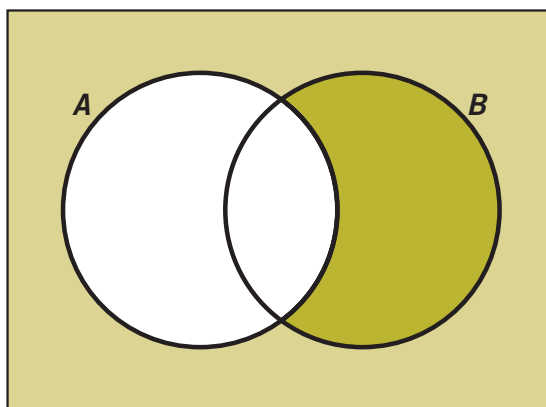


Figura 6 Complemento de conjuntos.

Diferencia

La *diferencia* de A y B , para cualesquiera conjuntos, es el conjunto de todos los elementos que están en A , pero no en B , es decir, son los elementos que exclusivamente pertenecen a A . Se simboliza como $A - B$ y el diagrama de Venn correspondiente se presenta en la figura 7.

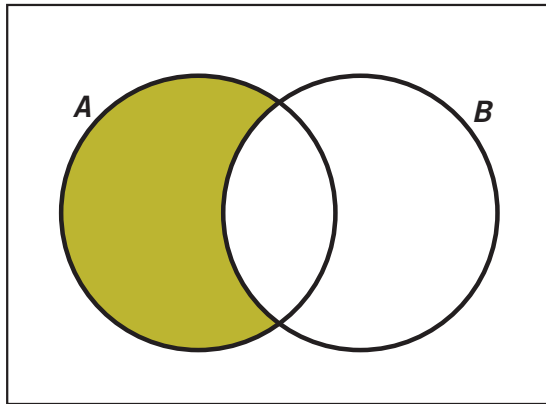


Figura 7 Diferencia de los conjuntos A y B , $A - B$.

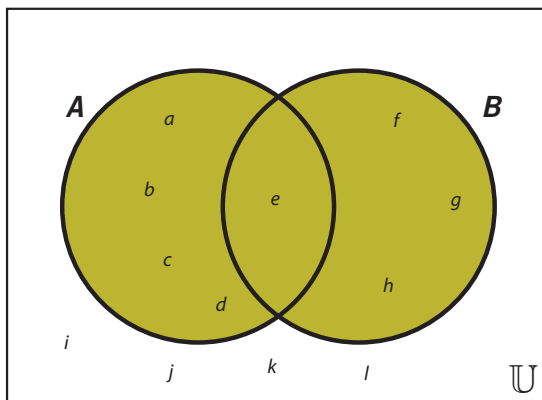
Sean los conjuntos $U = \{a, b, c, d, e, f, g, h, i, j, k, l\}$, $A = \{a, b, c, d, e\}$ y $B = \{e, f, g, h\}$. Determina los conjuntos siguientes; luego represéntalos tanto en lista como en diagrama de Venn y establece su cardinalidad.

Ejemplo 6

- | | |
|---------------|-------------------|
| a. $A \cup B$ | e. $(A \cup B)^c$ |
| b. $A \cap B$ | f. $(A \cap B)^c$ |
| c. B^c | g. $A^c \cup B^c$ |
| d. $B - A$ | h. $A^c \cap B^c$ |

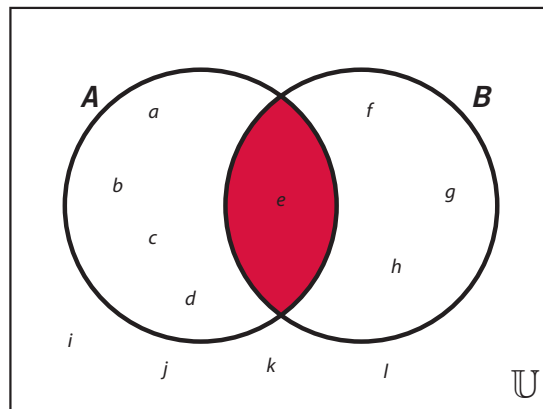
Solución

- a. $A \cup B = \{a, b, c, d, e, f, g, h\}$. En diagrama de Venn:



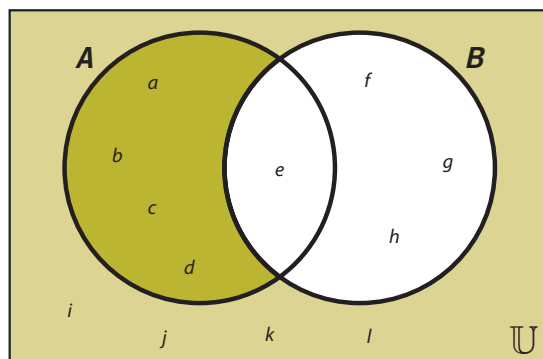
Por tanto, $\#(A \cup B) = 8$.

b. $A \cap B = \{e\}$. En diagrama de Venn:



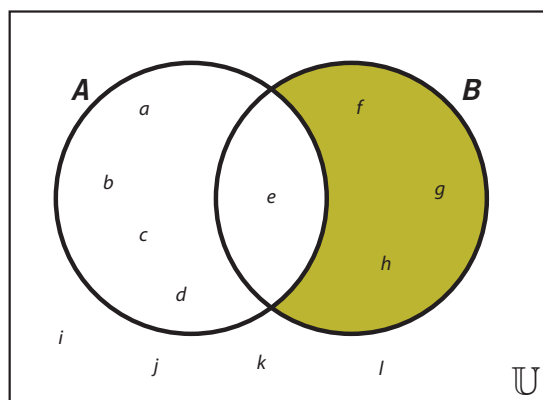
Por tanto, $\#(A \cap B) = 1$.

c. $B^c = U - B = \{a, b, c, d, i, j, k, l\}$. En diagrama de Venn:



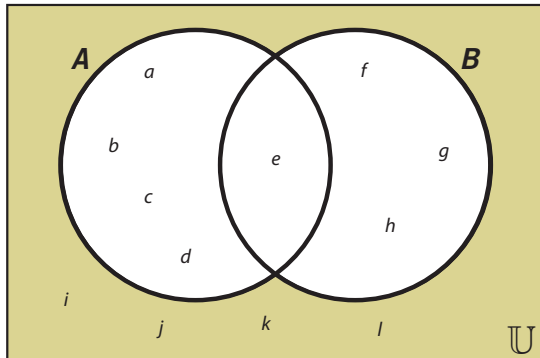
Por tanto, $\#(B^c) = 8$.

d. $B - A = \{f, g, h\}$. En diagrama de Venn:



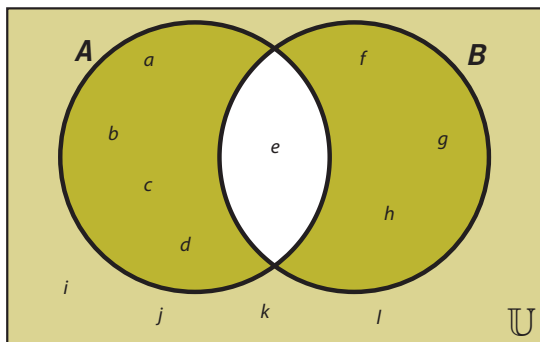
Por tanto, $\#(B - A) = 3$.

e. $(A \cup B)^c = U - (A \cup B) = \{i, j, k, l\}$. En diagrama de Venn:



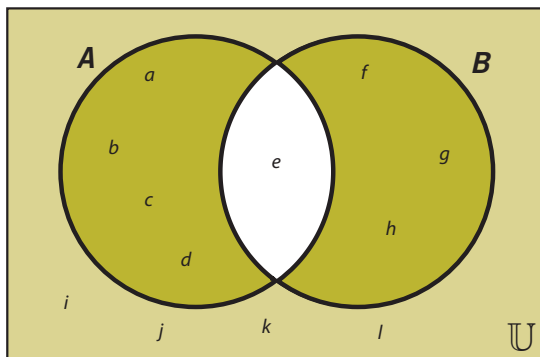
Por tanto, $\#[(A \cup B)^c] = 4$.

f. $(A \cap B)^c = U - (A \cap B) = \{a, b, c, d, f, g, h, i, j, k, l\}$. En diagrama de Venn:



Por tanto, $\#[(A \cap B)^c] = 11$.

g. $A^c \cup B^c = \{a, b, c, d, f, g, h, i, j, k, l\}$. En diagrama de Venn:



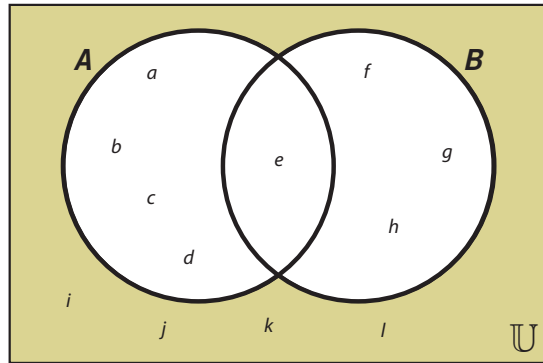
De estos últimos incisos se establece que:

$$(A \cap B)^c = A^c \cup B^c \text{ y}$$

$$(A \cap B)^c = A^c \cup B^c$$

Por tanto, $\#(A^c \cup B^c) = 11$.

h. $A^c \cap B^c = \{i, j, k, l\}$. En diagrama de Venn:



Por tanto, $\#(A^c \cap B^c) = 4$.

Ejemplo 7

Se realiza una encuesta entre 100 personas para conocer cuál es su destino turístico preferido: Acapulco o Cancún. De ellas, 45 respondieron que prefieren Cancún; 35 Acapulco y 25 que Cancún les gusta tanto como Acapulco. Determina:

- ¿Cuántas personas prefieren sólo Cancún?
- ¿Cuántas personas prefieren otros destinos que no sean Cancún o Acapulco?

Solución

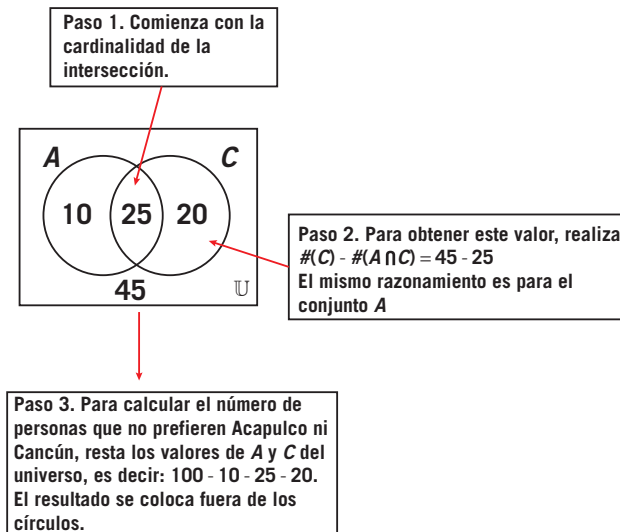
Primero determinamos los eventos:

- A: personas que prefieren Acapulco como destino turístico.
- C: personas que prefieren Cancún como destino turístico.

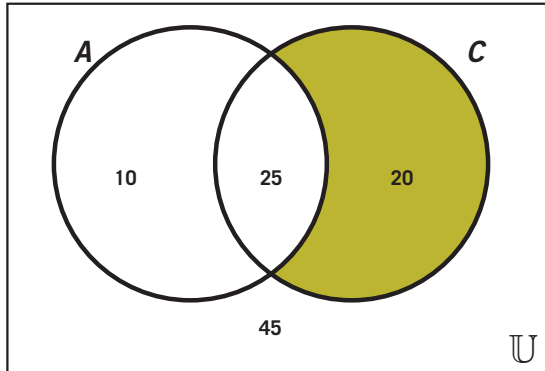
Establecemos la cardinalidad de cada conjunto:

- $\#(A) = 35$
- $\#(C) = 45$
- $\#(A \cap C) = 25$

Dibujamos el diagrama de Venn correspondiente:

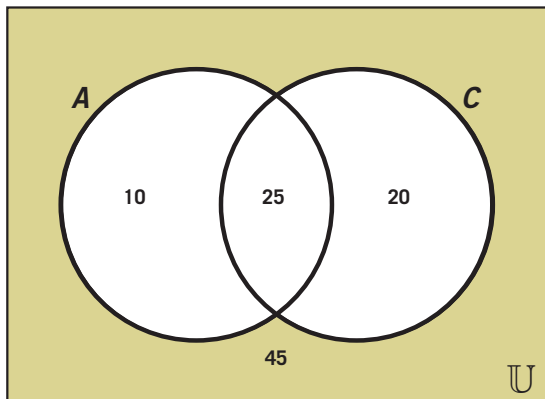


- a. El número de personas que prefieren únicamente Cancún se refiere a:



O bien, $\#(C - A) = \#(C) - \#(A \cap C) = 45 - 25 = 20$.

- b. ¿Cuántas personas prefieren otros destinos que no sean Cancún o Acapulco? En el diagrama siguiente se establece:



Es decir, $\#([A \cup C]^c) = \#(U) - \#(A \cup C) = 100 - 55 = 45$.

Se preguntó a 90 personas sobre sus deportes preferidos. Los resultados fueron:

- A 40 les gusta el fútbol americano
- A 39 el fútbol soccer
- 21 prefieren la fórmula 1
- A 13 les gusta el soccer y el americano
- 7 dijeron que les gusta el fútbol americano y la fórmula 1
- 8 prefieren el fútbol soccer y la fórmula 1
- A tres les gustan los tres deportes.

Determina:

- a. ¿Cuántas personas prefieren únicamente la fórmula 1?
- b. ¿A cuántas personas les gusta el fútbol americano y el soccer, pero no la fórmula 1?

Ejemplo 8

- c. ¿Cuántos prefieren la fórmula 1 y el futbol soccer, pero no el americano?
- d. ¿Cuántos prefieren deportes distintos del futbol americano, el soccer y la fórmula 1?

Solución

Determinamos los eventos:

- A : personas aficionadas al futbol americano.
- S : personas aficionadas al futbol soccer.
- F : personas aficionadas a la fórmula 1.

Identificamos y determinamos la cardinalidad de cada conjunto:

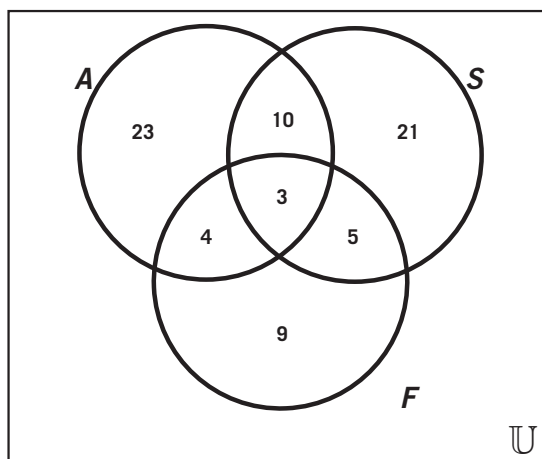
- $\#(A) = 40$
- $\#(S) = 39$
- $\#(F) = 21$

Entonces:

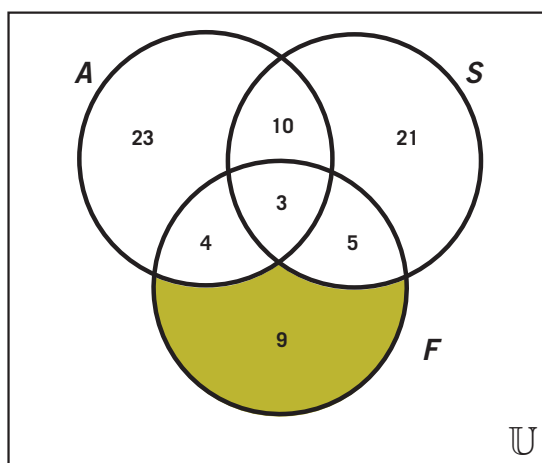
- $\#(A \cap S) = 13$
- $\#(A \cap F) = 7$
- $\#(F \cap S) = 8$
- $\#(A \cap F \cap S) = 3$

Dibujamos el diagrama de Venn correspondiente:

Recuerda empezar por la intersección de los tres conjuntos.

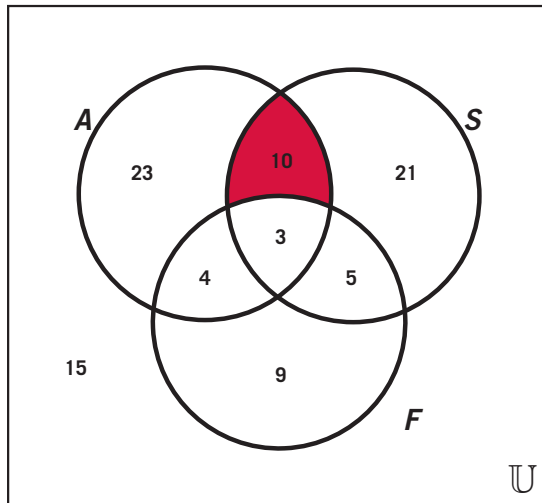


- a. El número de personas que prefieren únicamente la fórmula 1 son:



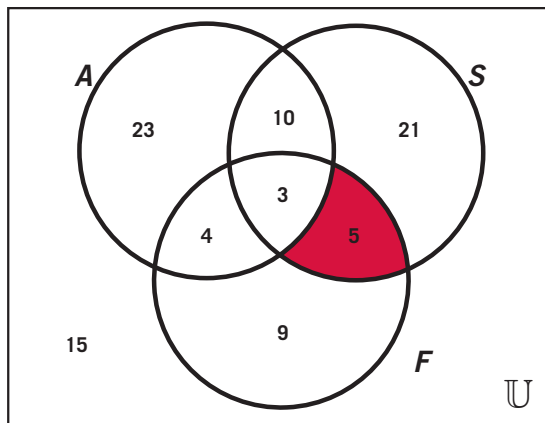
En notación: $\#(F - (A \cup S)) = 9$.

b. El número de personas que les gusta el futbol americano y el soccer, pero no la fórmula 1 son:



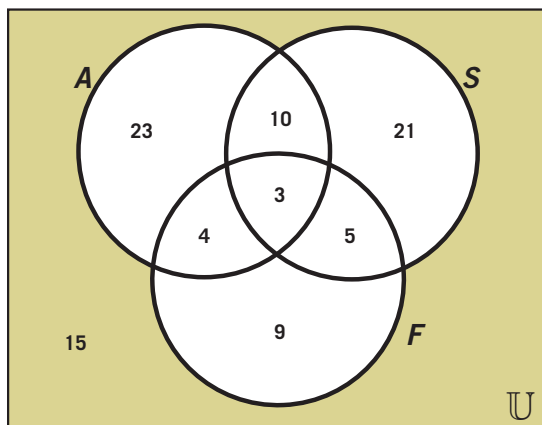
En notación: $\#[(A \cap S) - F] = 10$.

c. Los que prefieren la fórmula 1 y el futbol soccer, pero no el futbol americano son:



Es decir, $\#[(F \cap S) - A] = 5$.

d. Los aficionados a deportes distintos del futbol americano, el soccer y la fórmula 1 son:



O bien, $\#[(A \cup F \cup S)^c] = 15$.

Una vez que hemos revisado la teoría de conjuntos pasemos a la teoría de la probabilidad.

Probabilidad

La *probabilidad* es el valor entre cero y uno, ambos inclusive, que mide la factibilidad de que cierto resultado, al que llamaremos *evento*, ocurra (figura 8).

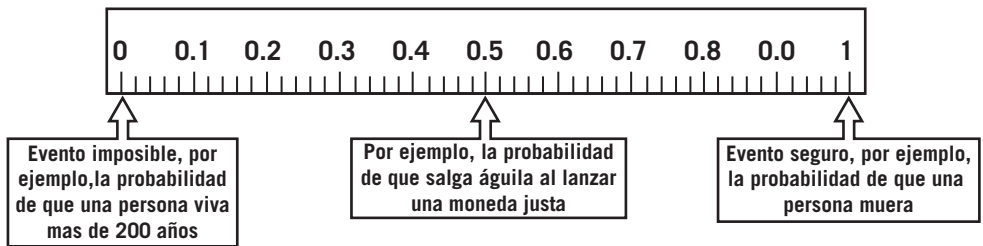


Figura 8 Probabilidad.

Conceptos básicos

Se definirán a continuación los conceptos más relevantes de la teoría de la probabilidad:

- **Experimento:** operación que consiste en observar los resultados en ciertas condiciones.
- **Evento:** conjunto de uno o más resultados de un experimento. Se puede hablar de *evento simple*, que es un resultado del espacio muestral con una sola característica. Un *evento conjunto* es, por su parte, un resultado del espacio muestral con dos o más características.
- **Espacio muestral (Ω):** conjunto de todos los posibles eventos o resultados que puedan ocurrir.
- **Punto muestral:** cada uno de los elementos del espacio muestral.

A continuación se ilustran los conceptos antes descritos.

Una probabilidad no puede ser negativa ni mayor que 1. Se puede expresar con decimales o en porcentaje.

Ejemplo 9

Experimento:	Lanzar un dado y observar el número que cae hacia arriba
Espacio muestral:	1, 2, 3, 4, 5, 6
Punto muestral:	Cada uno de los valores del espacio muestral
Evento simple:	Que sea un número par (una sola característica), es decir, {2, 4, 6}
Evento conjunto:	Que sea un número par mayor a 4 (dos características), es decir, {6}

Enfoques de la probabilidad

Hay dos enfoques para asignar probabilidades: el enfoque objetivo y el subjetivo. Se llama *enfoque objetivo* porque detrás de él existen fundamentos matemáticos (teóricos) para determinar la probabilidad de un evento. En contraparte, el *enfoque subjetivo* se basa en la experiencia o conocimiento del investigador.

El enfoque objetivo se divide a su vez en probabilidad clásica y probabilidad empírica (figura 9).



ENFOQUE CLÁSICO

Se basa en la idea de que los resultados de un experimento son igualmente posibles. Se calcula de la manera siguiente:

$$\text{Probabilidad de un evento} = \frac{\text{casos favorables}}{\text{casos totales}}$$

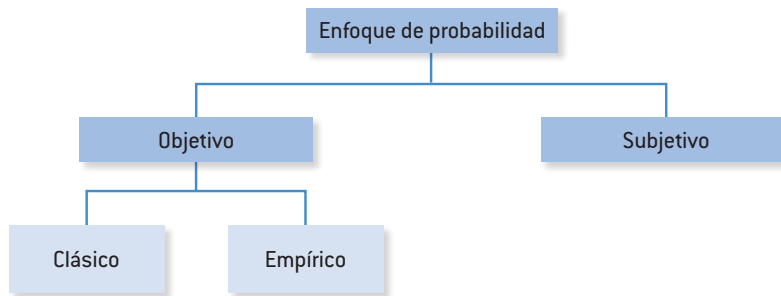


Figura 9.

En una caja hay 10 pelotas, de las cuales cuatro son azules. ¿Cuál es la probabilidad de que, sin ver, saques una pelota azul?

Ejemplo 10

Solución

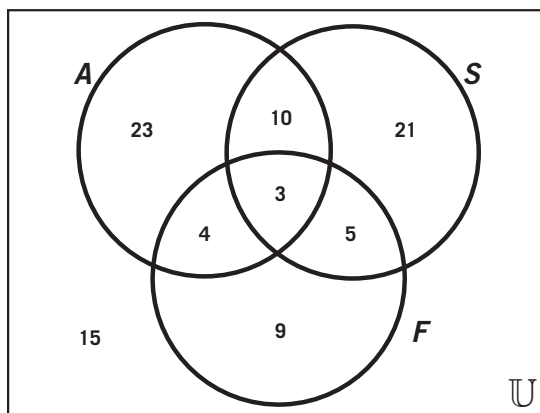
Sea A el evento de sacar una pelota azul de la caja. Como cada pelota tiene la misma probabilidad de ser elegida, utilizamos el enfoque clásico:

$$P(A) = \frac{\text{casos favorables}}{\text{casos totales}} = \frac{4}{10} = 0.4$$

Por tanto, la probabilidad de sacar una pelota azul es de 0.4.

Retomemos el ejemplo 8, sobre los deportes preferidos de un conjunto de personas. Los resultados fueron los siguientes:

Ejemplo 11



donde:

- A: personas aficionadas al futbol americano.
- S: personas aficionadas al futbol soccer.
- F: personas aficionadas a la fórmula 1.

A partir de esta información, calcula la probabilidad de que una persona elegida al azar:

- a. Prefiera únicamente la fórmula 1.
- b. Le guste el futbol americano y el soccer, pero no la fórmula 1.
- c. Prefiera la fórmula 1 y el futbol soccer, pero no el americano.
- d. Prefiera deportes distintos del futbol americano, el soccer y la fórmula 1.

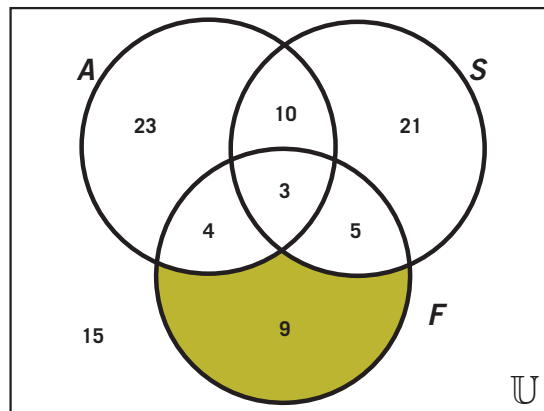
Solución

Utilizando el diagrama de Venn y aplicando el enfoque clásico de la probabilidad podremos determinar lo que se pide en cada inciso.

a.

$$P(F - (A \cup S)) = \frac{9}{90} = 0.10$$

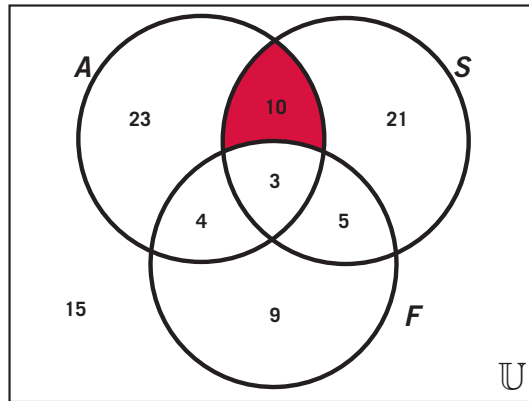
o bien, 10% de las personas prefieren únicamente la fórmula 1. En diagrama de Venn, como vimos en el ejemplo 8, esto es:



b.

$$P((A \cap S) - F) = \frac{10}{90} = 0.1111$$

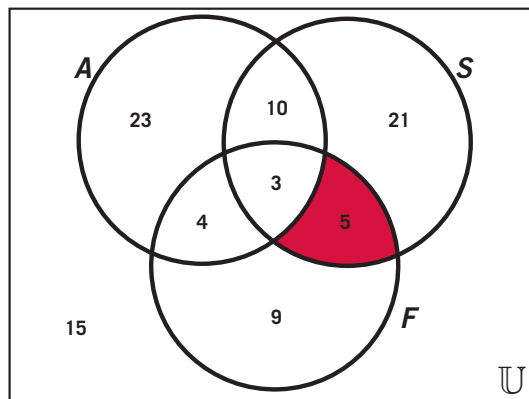
es decir, 11.11% gustan del futbol americano y el soccer, pero no de la fórmula 1. En diagrama de Venn:



c.

$$P((F \cap S) - A) = \frac{5}{90} = 0.0555$$

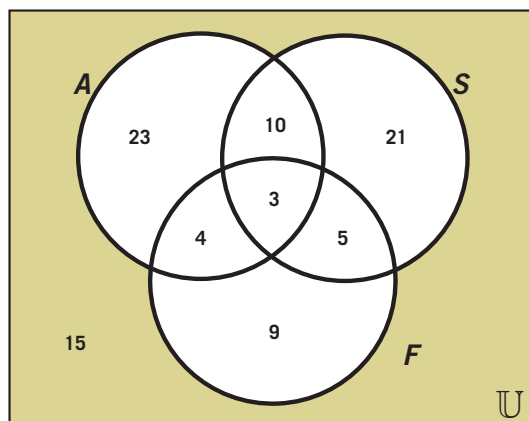
o bien, 5.55% prefiere la fórmula 1 y el futbol soccer, pero no el americano. En diagrama de Venn:



d.

$$P[(A \cup F \cup S)^c] = \frac{15}{90} = 0.1667$$

es decir, 16.67% prefiere deportes distintos del futbol americano, soccer y fórmula 1. En diagrama de Venn:



En muchas ocasiones la información relacionada con un problema se puede presentar mediante una tabla, a la que se le llama *tabla de contingencia*. Se trata de una tabla que se usa para clasificar el número de observaciones respecto a dos características o variables de interés.

Ejemplo 12

En una universidad 600 alumnos estudian en la escuela de negocios (economía, administración y relaciones internacionales, entre otras carreras), 550 estudian en la escuela de ingeniería y 350 en la de humanidades.

a. Con esta información, construye una tabla de contingencia.

Determina la probabilidad de que un estudiante, elegido al azar:

- b. Estudie en la escuela de ingeniería.
- c. Estudie en la escuela de humanidades.
- d. Estudie en la escuela de ingeniería o de negocios.

Solución

Los eventos de este problema son:

- > N : estudia en la escuela de negocios.
- > I : estudia en la escuela de ingeniería.
- > H : estudia en la escuela de humanidades.

a. Realizamos la tabla de contingencia:

Tipo de carrera	Evento	Número de estudiantes
Negocios	N	600
Ingeniería	I	550
Humanidades	H	350
Total		1500

b. La probabilidad de que una persona estudie en la escuela de ingeniería es:

$$P(I) = \frac{550}{1500} = 0.3667$$

o bien, 36.67%.

c. La probabilidad de que una persona estudie en la escuela de humanidades:

$$P(H) = \frac{350}{1500} = 0.2333$$

o bien, 23.33%.

d. Por último, la probabilidad de que una persona estudie en la escuela de ingeniería o de negocios es:

$$P(I \cup N) = P(I) + P(N) = \frac{550}{1500} + \frac{600}{1500} = \frac{1150}{1500} = 0.7667$$

o bien, 76.67%.

Enfoque empírico

Este enfoque se basa principalmente en frecuencias relativas, esto es, en el número de veces que ocurrió cierto evento en el pasado, y se calcula de esta manera:

$$\text{Probabilidad de un evento} = \frac{\# \text{ veces que ocurrió en el pasado}}{\# \text{ total de observaciones}}$$

Conforme el número de observaciones crece, el enfoque empírico se acerca al valor del enfoque clásico.

Se lanza una moneda varias veces y se cuenta el número de veces que cae sol. Conforme este experimento se repita más veces, llegaremos a que la probabilidad de que salga sol es de 0.5.

Ejemplo 13

Una persona lanza un dado “justo” 40 veces y anota cuántas veces salió cada número. Al terminar los lanzamientos, calcula cada una de las probabilidades como el número de veces que salió cada número dividido entre el número de lanzamientos realizados; llegó a los resultados siguientes:

Ejemplo 14

Número que salió	Número de veces	Probabilidad
1	8	$\frac{8}{40} = 0.20$
2	6	$\frac{6}{40} = 0.15$
3	10	$\frac{10}{40} = 0.25$
4	4	$\frac{6}{40} = 0.10$
5	6	$\frac{6}{40} = 0.15$
6	6	$\frac{6}{40} = 0.15$

Si se hicieran más lanzamientos, llegaríamos a que la probabilidad para cada cara del dado se aproximaría a $\frac{1}{6}$, es decir, 0.1667, resultado que coincidiría con el enfoque clásico.

Enfoque subjetivo

Es cuando una persona asigna la probabilidad a cierto evento con base en información disponible o en su experiencia.

Para determinar la probabilidad de que los Acereros de Pittsburgh lleguen al Supertazón el próximo año nos basamos en el desempeño del equipo en la presente campaña.

Ejemplo 15

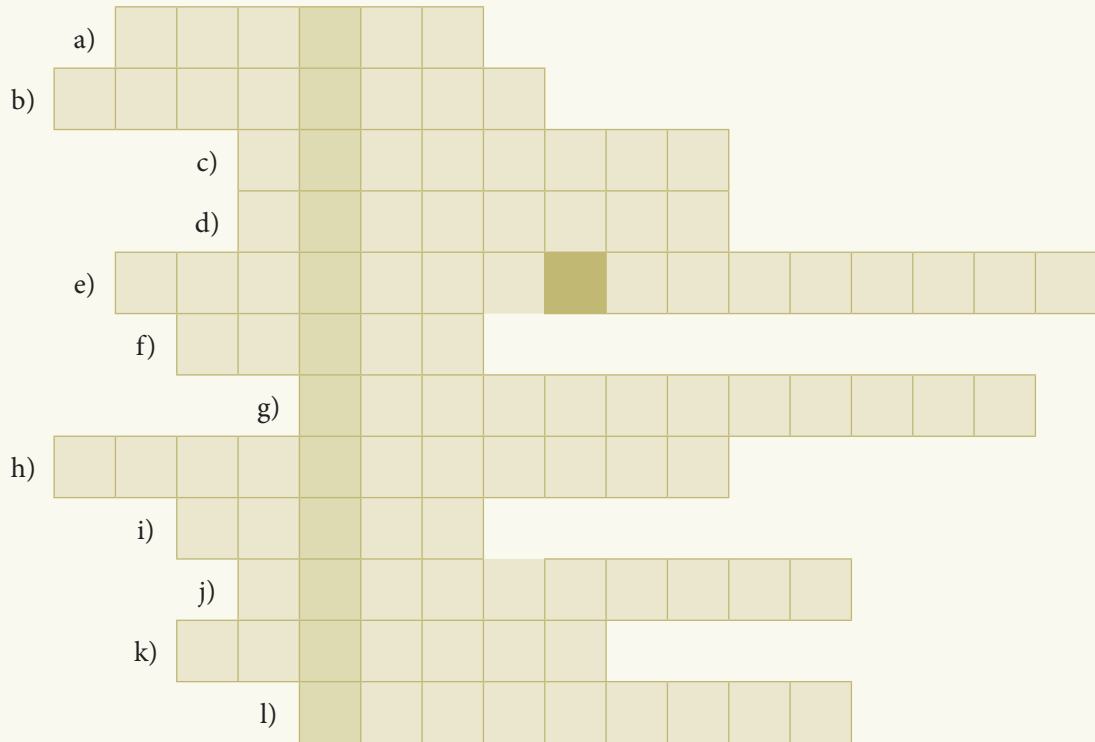
En muchas ocasiones, las empresas desean conocer la probabilidad de que un nuevo producto tenga éxito en el mercado. Para ello recurren a la experiencia o conocimiento del negocio de sus directivos o de sus vendedores. O bien, organizan grupos foco (*focus group*), en los que las personas expresan el grado de aceptación del producto. De esta manera, se utiliza el criterio de las personas para poder establecer la probabilidad de éxito o de fracaso.

Ejemplo 16

Ejercicios

I. TEORÍA Y ALGO MÁS

1. Completa el crucigrama siguiente y descubre la palabra secreta:



- a) Es el evento que sólo tiene una característica.
- b) Es la representación gráfica de la relación entre dos o más conjuntos, y fue desarrollado por Venn-Euler.
- c) Es el evento que tiene dos o más características.
- d) Es el enfoque de la probabilidad que está fundamentado matemáticamente.
- e) Es el conjunto de todos los posibles resultados de un experimento (dos palabras).
- f) Es otra forma de representar los posibles resultados de un experimento.
- g) Es el conjunto de los elementos que están en A y en B .
- h) Es el conjunto de los elementos que no están en A .
- i) Es el conjunto de los elementos que están en A , en B o en ambos conjuntos.
- j) Es la operación entre A y B para cualesquiera conjuntos, tal que todos los elementos están en A pero no en B .
- k) Enfoque probabilístico mediante el que se establece que la probabilidad es la proporción de los casos favorables respecto a los casos totales.
- l) Son los conjuntos que no tienen nada en común, es decir, su intersección resulte en un conjunto vacío.

II. PIENSO,... ¿Y LUEGO?

2. En el artículo “Futbolcifras”, escrito por Sergio Sarmiento y publicado en el diario *Reforma* el 10 de junio de 2010, se menciona:

México tiene un primer lugar en las tablas históricas de la Copa del mundo de fútbol: somos el país que acumula más derrotas. De 45 juegos, la selección nacional ha ganado solamente 11, empatado 12 y perdido 22...

Brasil es el equipo nacional más exitoso según datos acumulados de 22 130 juegos internacionales de 1980 a 2001 por el matemático Russell Gerard (http://www.staff.city.ac.uk/~sc397/football/aifr21_2.htm y Soccernomics) con un porcentaje ganador de 74.2% (los empates son considerados como media victoria). Le siguen Alemania (unida), Francia e Italia. Curiosamente, Iraq ocupa el quinto lugar con 69.2%, ayudado por la débil región en la que juega. México queda muy lejos de estos primeros lugares con un porcentaje ganador de 61.3% a pesar de jugar en la también fácil Concacaf. Con esta metodología, el porcentaje ganador de México en Copas del Mundo sería de apenas 35.5%.

De acuerdo con esta información, ¿cuál es la probabilidad de que la selección nacional gane un partido? ¿Y la probabilidad de que empate? ¿Consideras que México podría considerarse una potencia en un deporte como el fútbol soccer? ¿Qué se debería hacer para revertir estos resultados? ¿Qué hacen países como Brasil y Alemania para alcanzar el éxito?

III. PURO DEPORTE

3. Sólo para practicar. Sean los conjuntos:

- $U = \{\text{países de América}\}$
- $A = \{\text{Argentina, Chile, Paraguay, Uruguay}\}$
- $B = \{\text{Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, Panamá}\}$
- $C = \{\text{México, Estados Unidos, Canadá}\}$
- $D = \{\text{Venezuela, Ecuador, Bolivia, Perú, Colombia}\}$

Obtén una lista de los países que forman cada uno de los conjuntos siguientes:

- a) $A \cup B$
- b) $E - B$
- c) $B \cap E$
- d) $A \cup D$
- e) $E - [C \cup A]$

4. Sólo para practicar. Sean los conjuntos siguientes:

- $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$
- $A = \{\text{números pares}\}$
- $B = \{\text{números impares}\}$
- $C = \{\text{números primos}\}$
- $D = \{\text{números múltiplos de 3}\}$
- $E = \{\text{números múltiplos de 5}\}$

Determina:

- a) Los elementos de cada conjunto.
- b) $B \cup E$
- c) A^c
- d) $B \cap C$
- e) $B - D$
- f) $(A \cup C) - E$

5. *Sólo para practicar.* Sean A y B dos conjuntos cualesquiera. Traza los diagramas de Venn respectivos y sombrea la operación que se indica:

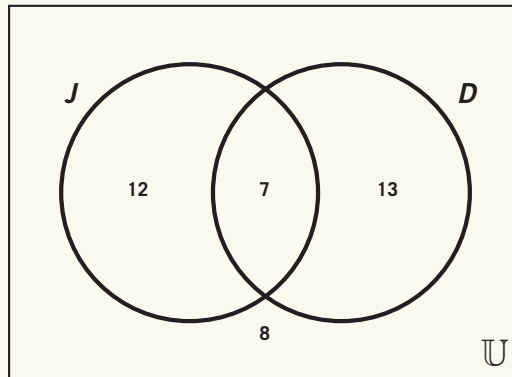
- | | |
|-----------------|-------------------|
| a) A^c | e) $(A \cup B)^c$ |
| b) $A^c \cap B$ | f) $(A \cap B)^c$ |
| c) $A^c \cup B$ | g) $A^c \cap B^c$ |
| d) $B - A$ | h) $A^c \cup B^c$ |

6. *Sólo para practicar.* Sean A , B y G tres conjuntos cualesquiera. Traza los diagramas de Venn respectivos y sombrea la operación que se te indica:

- | | |
|---------------|----------------------------|
| a) A | h) $G^c \cap B^c$ |
| b) $A \cup B$ | i) $A \cap G^c$ |
| c) $A \cup G$ | j) $(A \cup B) \cap G$ |
| d) $A \cap B$ | k) $(A - B) \cap G$ |
| e) $B \cap G$ | l) $(A^c \cap G) \cup B$ |
| f) $A - B$ | m) $(A^c \cap G^c) \cap B$ |
| g) $G - B$ | n) $(A \cup B \cup C)^c$ |

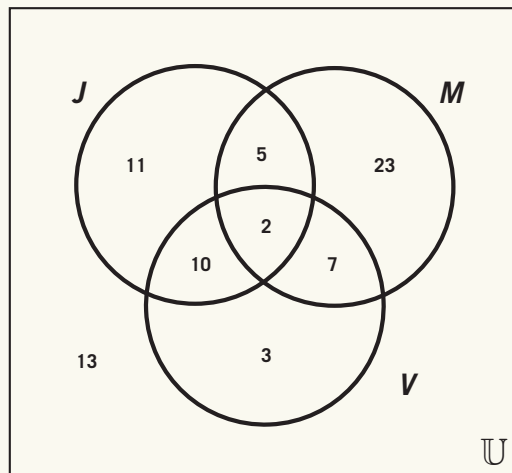
7. *Sólo para practicar.* Dado el diagrama de Venn mostrado enseguida, determina:

- a) $\#(D)$
 b) $\#(J^c)$
 c) $\#(J \cup D)$
 d) $\#(J - D)$
 e) $\#(J \cap D)$
 f) $\#(J \cup D)^c$
 g) $\#(J^c \cup D)$
 h) $\#(J \cap D^c)$



8. *Sólo para practicar.* Dado el diagrama de Venn mostrado abajo, determina:

- a) $\#(V) = 22$
 b) $\#(J^c) = 46$
 c) $\#(V \cup M) = 50$
 d) $\#(J - M) = 21$
 e) $\#(J \cap V) - M = 10$
 f) $\#(J \cup V \cup M)^c = 13$
 g) $\#[(J \cup V) \cap M] = 14$
 h) $\#[(V \cup M) - J] = 33$



9. *Sólo para practicar.* Sean A y B dos conjuntos tales que $\#U = 75$, $\#(A - B) = 10$, $\#(A \cap B) = 25$ y $\#(B - A) = 15$. Encuentra $\#(A^c \cap B^c)$.
10. *Sólo para practicar.* Sean A y B dos conjuntos tales que $\#U = 100$, $\#(A - B) = 50$, $\#(A^c \cap B^c) = 10$ y $\#(A \cap B) = 5$. Encuentra:
- $\#(B)$
 - $\#(A^c \cap B)$.
11. *Educación.* La Evaluación Nacional de Logro Académico en Centros Escolares (ENLACE) es una prueba del Sistema Educativo Nacional que se aplica a planteles públicos y privados del país. En educación básica se aplica a niñas y niños de tercero a sexto grados de primaria. De 120 niños de primaria que presentaron la ENLACE en 2010, los resultados fueron:
- 52 tuvieron nivel bueno o más en Español
 - 54 obtuvieron nivel bueno o más en Matemáticas
 - 40 tuvieron nivel bueno o más en Matemáticas y en Español

De acuerdo con esta información:

- Traza el diagrama de Venn respectivo, indicando los eventos por analizar.
 - Determina la probabilidad de que un estudiante elegido al azar haya obtenido un nivel bueno o más en Matemáticas, pero no en Español.
 - Determina la probabilidad de que un estudiante elegido al azar no haya obtenido un nivel bueno o más en Matemáticas ni en Español.
 - Con base en estos resultados, organiza con tus compañeros de clase una lluvia de ideas sobre cómo se podría mejorar el sistema educativo mexicano.
12. *Sociedad.* Una persona va al tianguis del Chopo y pregunta a 90 jóvenes sobre su gusto musical. Los resultados son los siguientes:
- 38 prefieren el *heavy metal*
 - 29 gustan del *punk*
 - 37 prefieren el *gothic rock*
 - 9 escuchan *heavy metal* y *punk*
 - 10 gustan del *punk* y *gothic rock*
 - 11 escuchan *heavy metal* y *gothic rock*
 - 3 gustan de estas tres corrientes.

Con estos datos:

- Traza el diagrama de Venn e indica los eventos por analizar.
- ¿Cuál es la probabilidad de que a una persona elegida al azar le guste el *gothic rock* y el *punk*, pero no el *heavy metal*?
- ¿Cuál es la probabilidad de que a una persona elegida al azar le guste únicamente el *gothic rock*?
- ¿Cuál es la probabilidad de que a una persona elegida al azar no le guste el *gothic* ni el *punk* ni el *heavy metal*?
- Una tribu urbana es un grupo de gente que se comporta de acuerdo con la ideología de una subcultura, que se origina y se desarrolla en el ambiente de una urbe o ciudad. Las tribus urbanas se caracterizan porque sus miembros muestran una estética similar, entre varios individuos de la misma tendencia. Suelen mostrar convicciones sociopolíticas o creencias religiosas o de carácter místico, según el movimiento o tribu urbana a la que pertenecen (<http://www.detrabusurbanas.com/tipos>). Junto con tus compañeros de clase, realicen una mesa redonda donde expongan las características e ideologías de los *metaleros*, los *punks* y los *góticos*. ¿Crees que la apariencia de una persona sea importante?

13. *Uso de la tecnología.* En la actualidad es común utilizar la tecnología para demostrar nuestros “afectos”. El pasado 14 febrero se preguntó a 150 personas cómo habían demostrado su “amor”. Los resultados fueron los siguientes:

- 50 personas enviaron un mensaje de texto con su teléfono celular
- 60 mandaron un mensaje por las redes sociales
- 50 lo hicieron mediante un mensaje de correo electrónico
- 15 personas enviaron un mensaje de texto por teléfono celular y por una red social
- 25 lo hicieron mediante una red social y correo electrónico
- 12 mandaron mensajes de texto y correo electrónico
- 5 enviaron mensajes de texto con el teléfono celular, por las redes sociales y por correo electrónico.

De acuerdo con esta información:

- a) Traza el diagrama de Venn correspondiente, indicando los eventos por analizar.
 - b) ¿Cuál es la probabilidad de que una persona elegida al azar haya enviado mensajes sólo por las redes sociales?
 - c) ¿Cuál es la probabilidad de que una persona elegida al azar haya mandado mensajes por su teléfono celular o mediante un correo electrónico, pero no por las redes sociales?
 - d) ¿Cuál es la probabilidad de que una persona elegida al azar no haya utilizado alguno de estos medios para enviar mensajes el 14 de febrero?
 - e) Forma un equipo y discute con tus compañeros si enviar mensajes por estos medios tecnológicos no es una forma de despersonalización de los sentimientos. Expongan sus ideas en la clase.
14. *Psicología.* Un rasgo humano es sentir celos, no sólo hacia la pareja sino también en el ámbito familiar y de trabajo. Los celos están constituidos por otras emociones como el enojo, el miedo y la inseguridad. En los celos se ven involucradas tres personas y surgen por el anhelo de no perder lo que se posee. A partir de esta información, se hizo una encuesta entre 150 personas y las siguientes fueron las actitudes de “celos” que han experimentado:

Actitud de la pareja	Número
Llama constantemente por teléfono	38
Ha provocado una escena de celos	34
Ha revisado sus mensajes	30
Lo o la ha seguido	27
Ha revisado sus cosas	21
Total	150

- a) ¿Cuál es la probabilidad de que una persona elegida al azar llame constantemente por teléfono?
 - b) ¿Cuál es la probabilidad de que por celos una persona elegida al azar haya seguido a su pareja?
 - c) ¿Cuál es la probabilidad de que una persona elegida al azar haya provocado una escena de celos o haya revisado sus mensajes?
15. *Responsabilidad social.* Uno de los graves problemas que enfrenta actualmente la sociedad es el alcoholismo. En fechas recientes se ha visto una tendencia a la alza en el consumo de bebidas alcohólicas entre los jóvenes. La embriaguez causa muchos accidentes de tránsito, como se muestra en la tabla siguiente:

Tipo de accidente	Presentó aliento alcohólico
Colisión con vehículo automotor	22 733
Colisión con peatón (atropellamiento)	934
Colisión con animal	135
Colisión con objeto fijo	7 533
Volcadura	1 164
Caída de pasajero	239
Salida del camino	1 316
Incendio	48
Colisión con ferrocarril	9
Colisión con motocicleta	1 705
Colisión con ciclista	587
Otro	326
Total	36 729

Fuente: INEGI. Estadísticas de accidentes de tránsito terrestre en zonas urbanas y suburbanas (2009).

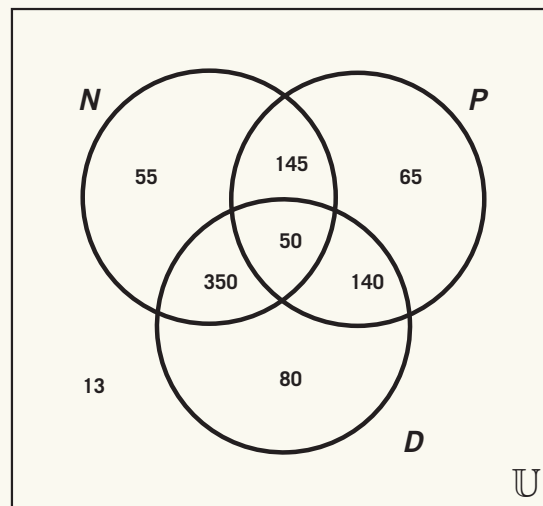
- ¿Cuál es la probabilidad de que una persona elegida al azar haya tenido una colisión con vehículo automotor?
- ¿Cuál es la probabilidad de que una persona elegida al azar haya tenido una volcadura?
- ¿Cuál es la probabilidad de que una persona elegida al azar haya tenido una colisión con un peatón?
- Investiga qué es la *responsabilidad social*. Junto con tus compañeros realicen una discusión respecto de qué consideran que podría hacerse para reducir la cantidad de accidentes y cuál es la responsabilidad de cada ciudadano en este aspecto.

Evaluación sumativa

Un escalón hacia arriba

Contesta los reactivos siguientes, que te permitirán saber si has aprendido lo que se esperaba.

- Sean A y B dos conjuntos tales que $\#U = 25$, $\#(A \cup B)^c = 3$, $\#(B - A) = 7$, $\#(A^c) = 10$. Encuentra:
 a) $\#(A)$ y b) $\#(A \cap B)$
- Se entrevistó a 885 jóvenes entre 18 y 29 años de edad sobre los programas de televisión que regularmente ven: noticiarios (N), series policiacas (P) y programas deportivos (D). Con la información obtenida se trazó el diagrama presentado abajo. Con base en los datos que se presentan ahí, responde:
 a) ¿Cuántos jóvenes ven cuando mucho dos programas de televisión?
 b) ¿Cuántas personas ven exactamente dos programas de televisión?



- Se preguntó a 50 personas por qué beben agua natural embotellada. Los resultados son los siguientes:
 - 27 personas lo hacen por gusto
 - 28 por seguir una dieta
 - 30 por facilidad para comprar la botella
 - 13 por gusto y por dieta
 - 17 por dieta y por facilidad
 - 19 por gusto y por facilidad
 - 10 por gusto, por facilidad y por seguir una dieta

A partir de estos datos:

- a) Traza el diagrama de Venn e indica los eventos por estudiar.
 - b) ¿Cuál es la probabilidad de que una persona consuma agua embotellada por gusto o por seguir una dieta, pero no por facilidad?
 - c) ¿Cuál es la probabilidad de que no beba agua embotellada por las razones antes descritas?
4. Se hizo una encuesta telefónica en la que se preguntó a las personas para qué utilizan internet. Los resultados fueron:

Tipo de uso	Núm. de usuarios
Para obtener información	12
Para comunicarse	25
Para apoyar la educación/capacitación	14
Para entretenimiento	30
Para operaciones bancarias	9
Total	90

Determina la probabilidad de que una persona elegida al azar utilice internet para:

- a) Obtener información o para realizar operaciones bancarias.
- b) Comunicarse o para entretenimiento.

Proyecto 4

VIOLENCIA EN EL NOVIAZGO

Según la Organización Mundial de la Salud (oms), sólo tres de cada 10 adolescentes denuncian que sufren violencia en el noviazgo. El problema de que no lo denuncien es que no se sabe identificarla. En otras palabras, es una situación que se mantiene en silencio porque la gran mayoría considera que son conductas normales, que no se intensificaron y, por tanto, no se atreven a denunciar, o bien, porque muchas personas creen que la violencia es sólo insultos o golpes. Pero esto no es así.

La violencia se define como todo acto o conducta que se hace para ofender o humillar a otra persona. Las etapas son las siguientes:

- *Psicológica*: hay gritos o insultos, conductas de desvaloración, amenazas, chantajes, celos.
- *Después la violencia física*: hay arañazos, empujones, jalones, pellizcos, bofetadas, golpes, patadas, entre otros.
- *La violencia sexual*: toqueteos que se hacen sin el consentimiento de la otra persona.

De acuerdo con diversos estudios sobre la violencia en el noviazgo se establece que la causa que la origina es la idiosincrasia machista y patriarcal que existe en nuestra sociedad, en la que el hombre es el “fuerte”, el que provee de recursos al núcleo familiar, en tanto que la mujer se dedica únicamente a labores del hogar. Es decir, se trata fundamentalmente de un problema de educación.

Supón que te han contratado para elaborar un análisis estadístico sobre la violencia en el noviazgo y que te entregan una encuesta realizada a 85 personas, tanto hombres como mujeres, con edades entre los 15 y los 24 años (archivo:[violencia.xls](#)). Representa la información por tipo de violencia, además de algunos de los datos siguientes:

- Porcentaje de personas que sufren violencia únicamente de manera física.
- Proporción de personas que sufren violencia psicológica.
- Fracción de personas que han sufrido violencia psicológica, física y sexual.

Escribe un informe de las conclusiones que obtienes, las cuales han de responder a las preguntas siguientes:

- ¿Has vivido esta situación? ¿Qué hiciste o has hecho para salir de este problema?
- ¿Qué consecuencias trae consigo esta situación cuando no se denuncia?
- ¿Qué beneficios trae denunciarla?
- ¿Qué propones para crear conciencia en las personas sobre este problema?
- ¿Consideras que las campañas de prevención son eficientes?

Para que tengas una visión más amplia sobre este tema mira el video que se presenta en esta dirección:

http://www.sep.gob.mx/wb/sep1/violencia_en_el_noviazgo



Competencias

- Argumenta el uso del árbol de probabilidad en la resolución de un problema, con métodos numéricos, gráficos, analíticos variacionales, mediante representaciones lingüísticas, matemáticas o gráficas y el uso de las tecnologías de la información y la comunicación
- Identifica las ideas clave en un texto sobre las técnicas de conteo e infiere conclusiones a partir de ellas
- Estructura ideas y argumentos de manera clara, coherente y sintética sobre cómo se construye un árbol de probabilidad, interpretando tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos
- Propone maneras de resolver un problema en equipo, definiendo un curso de acción con pasos específicos, asumiendo una actitud constructiva y aportando puntos de vista con apertura, considerando los de otras personas de manera reflexiva
- Formula y resuelve problemas matemáticos aplicando diferentes enfoques

5. Técnicas de conteo

Contenido

- Árbol de probabilidad
- Técnicas de conteo y su aplicación

Por cierto...

En una bellísima colina de la Selva Negra alemana se alza majestuoso el Instituto de Matemáticas de Oberwolfach, creado por el inminente J. Süss y donde 51 semanas al año se celebra, cada semana, un congreso de matemáticas por invitación de 60 personas.

Como las servilletas de cada uno se ponen en una bolsita de plástico con el nombre y dichas bolsitas son distribuidas al azar (a la comida y a la cena) en las diferentes mesas, pronto nació el “problema de Oberwolfach” en relación con calcular la probabilidad de sentarse al lado de alguien una vez o varias. ¡Los matemáticos juntos no pueden ni comer tranquilos! (Alsina, 2008).

Ubícate

Antes de comenzar con el estudio de las técnicas de conteo, contesta este examen para determinar si tienes los conocimientos necesarios para comprender cabalmente este bloque.

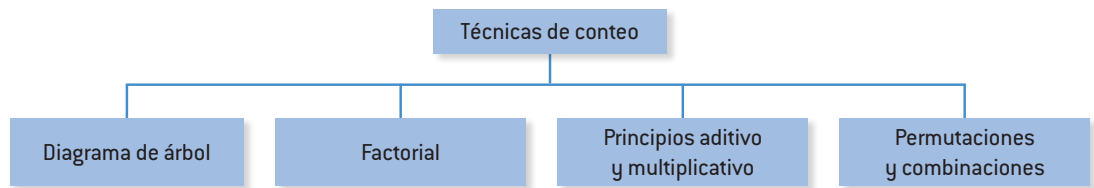
1. Define:
 - a) Espacio muestral
 - b) Evento
 - c) Probabilidad
 - d) Enfoque clásico de la probabilidad
2. Si se lanzan dos dados “justos” y de distinto color, determina:
 - a) El espacio muestral de este experimento
 - b) ¿Cuál es la probabilidad de que sus caras superiores sumen 5?
3. Hay tres lugares desocupados en la fila H de la sala 3 del cine. A, J y F no saben en qué orden acomodarse.
 - a) Enumera las posibilidades que tienen para sentarse.
 - b) ¿Cuál es la probabilidad de que A y J queden juntos?

Si no tuviste problemas o dudas con el examen anterior, ahora sí comienza el estudio del bloque 5.

Introducción

En el capítulo anterior comenzamos a estudiar la teoría de la probabilidad. Como aprendiste, de acuerdo con el enfoque clásico, se calcula como el número de casos favorables entre el número de casos totales. Muchas veces es fácil determinar cuántos casos favorables y cuántos casos totales hay en un experimento; por ejemplo, la probabilidad de que al tirar un dado el número que aparezca en la cara superior sea par resulta de dividir el total de números pares que hay en el dado, 3, entre el número total de lados que tiene el dado, que es 6, por lo que la probabilidad es $\frac{3}{6} = \frac{1}{2}$.

Pero en muchos otros experimentos no resulta tan sencillo enumerar los casos totales. Y debemos ser prácticos, así que en este capítulo revisaremos las técnicas de conteo, que nos permitirán calcular de forma más fácil el número de casos favorables y de casos totales en cierto experimento. En el esquema siguiente se muestran los temas que abordaremos.



Desempeños

Al terminar de estudiar este bloque y realizar todas las actividades propuestas habrás aprendido a:

- Sintetiza los elementos de la probabilidad conjunta para resolver problemas de su vida cotidiana.
- Analiza las características de una representación de forma tabular, gráfica y funciones de probabilidad mediante gráficas de árbol.
- Describe las características de los eventos mutuamente excluyentes e independientes para su aplicación en la solución de problemas.
- Aplica las características de la probabilidad en los diagramas de árbol para la resolución de problemas.
- Explica las condiciones del cálculo de la probabilidad analizando el teorema de Bayes dentro de la solución de problemas.

Diagrama de árbol

Un diagrama de árbol sirve para representar todos los eventos posibles de un experimento. Se establece una raíz y a su derecha se dibujan cada una de las ramas o trayectorias referentes a cada uno de los eventos.

Un diagrama de árbol permite visualizar eventos simples y eventos conjuntos o compuestos. Al final de cada rama se acostumbra escribir la cardinalidad de cada evento. Si se necesita calcular alguna probabilidad, se recurre al enfoque clásico que aprendiste en el bloque anterior *casos favorables/casos totales*.

En una universidad 600 alumnos estudian en la escuela de negocios (economía, administración y relaciones internacionales, entre otras carreras), 550 estudian en la escuela de ingeniería y 350 en la de humanidades.

Ejemplo 1

- a. Con esta información, traza un diagrama de árbol.

Determina la probabilidad de que un estudiante, elegido al azar, estudie en la escuela de:

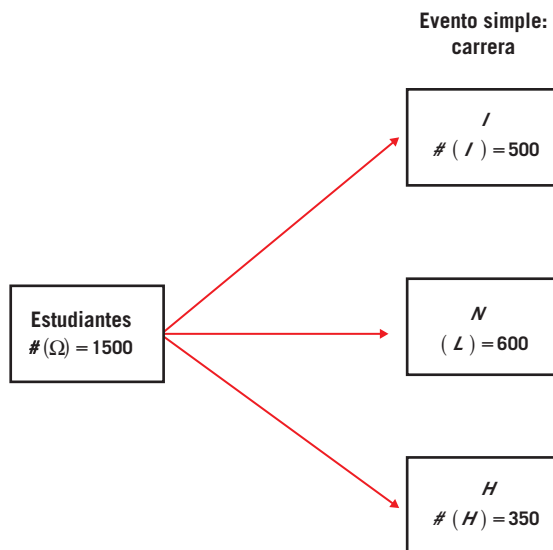
- b. Ingeniería.
- c. Humanidades.
- d. Ingeniería o negocios.

Solución

Los eventos de este problema son:

- N : estudia en la escuela de negocios.
- I : estudia en la escuela de ingeniería.
- H : estudia en la escuela de humanidades.

- a. Trazamos el diagrama de árbol:



Establecemos la raíz que, en este caso, son los estudiantes. La primera rama se refiere a la clasificación de los eventos, es decir, la escuela donde estudian.

- b. La probabilidad de que una persona estudie en la escuela de ingeniería es:

$$P(I) = \frac{550}{1500} = 0.3667$$

o bien, 36.67%.

c. La probabilidad de que una persona estudie en la escuela de humanidades:

$$P(H) = \frac{350}{1500} = 0.2333$$

o bien, 23.33%.

d. Por último, la probabilidad de que una persona estudie en la escuela de ingeniería o de negocios es:

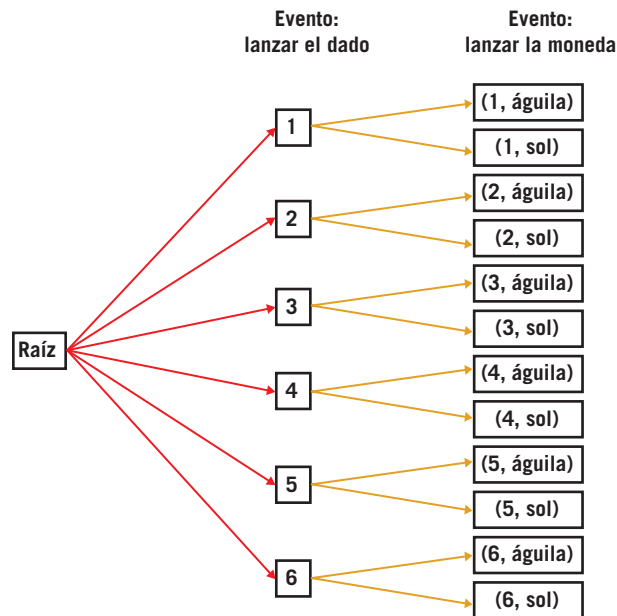
$$P(I \cup N) = P(I) + P(N) = \frac{550}{1500} = \frac{600}{1500} = \frac{1150}{1500} = 0.7667$$

o bien, 76.67%.

Ejemplo 2

Se tiene un dado y una moneda. Se lanza el dado y se anota el número que queda hacia arriba. Posteriormente se lanza la moneda y se anota el resultado. Determina el espacio muestral para este experimento mediante un diagrama de árbol.

Solución



es decir, el espacio muestral es

$$\Omega = \{(1, \text{águila}), (1, \text{sol}), (2, \text{águila}), (2, \text{sol}), (3, \text{águila}), (3, \text{sol}), (4, \text{águila}), (4, \text{sol}), (5, \text{águila}), (5, \text{sol}), (6, \text{águila}), (6, \text{sol})\}$$

y su cardinalidad es 12.

Así, si se desea, por ejemplo, calcular la probabilidad de que salga 6 y águila, se aplica el enfoque clásico de la probabilidad:

$$\frac{\text{casos favorables}}{\text{casos totales}} = \frac{1}{12}$$

Para determinar la cardinalidad del espacio muestral o de cualquier evento sin hacer una lista se emplean las técnicas de conteo.

Técnicas de conteo

Las técnicas de conteo sirven para contar los casos posibles de un conjunto; permiten reducir cálculos cuando no es tan sencillo enumerar los elementos de un conjunto. Por tanto, es necesario comprender muy bien en qué casos y cómo aplicar esos métodos. Para ello, estudiaremos la notación factorial, los principios multiplicativo y aditivos las permutaciones y las combinaciones.

Notación factorial

La notación factorial se refiere al uso del símbolo ! para representar el producto de números enteros positivos de la manera siguiente:

$$n! = n(n-1) \cdot (n-2) \dots 3 \cdot 2 \cdot 1$$

Por convención, $0! = 1$

Calcula:

- a. $3!$
- b. $5!$
- c. $10!$

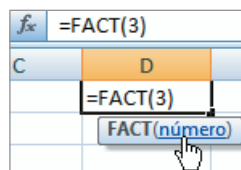
Ejemplo 3

Solución

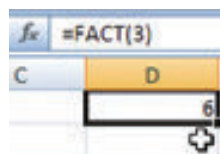
Desarrollando la notación factorial:

- a. $3! = 3 \cdot 2 \cdot 1 = 6$
- b. $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$
- c. $10! = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 3\,628\,800$

En Excel se utiliza la función `=FACT()`, indicando entre los paréntesis el número o factor como se ilustra aquí:



Se oprime a continuación la tecla Intro para ver el resultado.



Principio multiplicativo

Si una operación puede realizarse en m formas y una segunda puede hacerse en n formas, entonces las dos operaciones pueden realizarse juntas en $m \cdot n$ formas.

Ejemplo 4

Retomemos el ejemplo 2, en el que se lanza el dado y posteriormente la moneda. ¿Cuántos resultados son posibles?

Solución

Si revisamos el diagrama de árbol de la solución del ejemplo 2 veremos que hay 12 resultados posibles. Al aplicar el principio multiplicativo obtenemos:

$$\text{número de caras del dado} \times \text{número de caras de la moneda} = 6 \times 2 = 12$$

Ejemplo 5

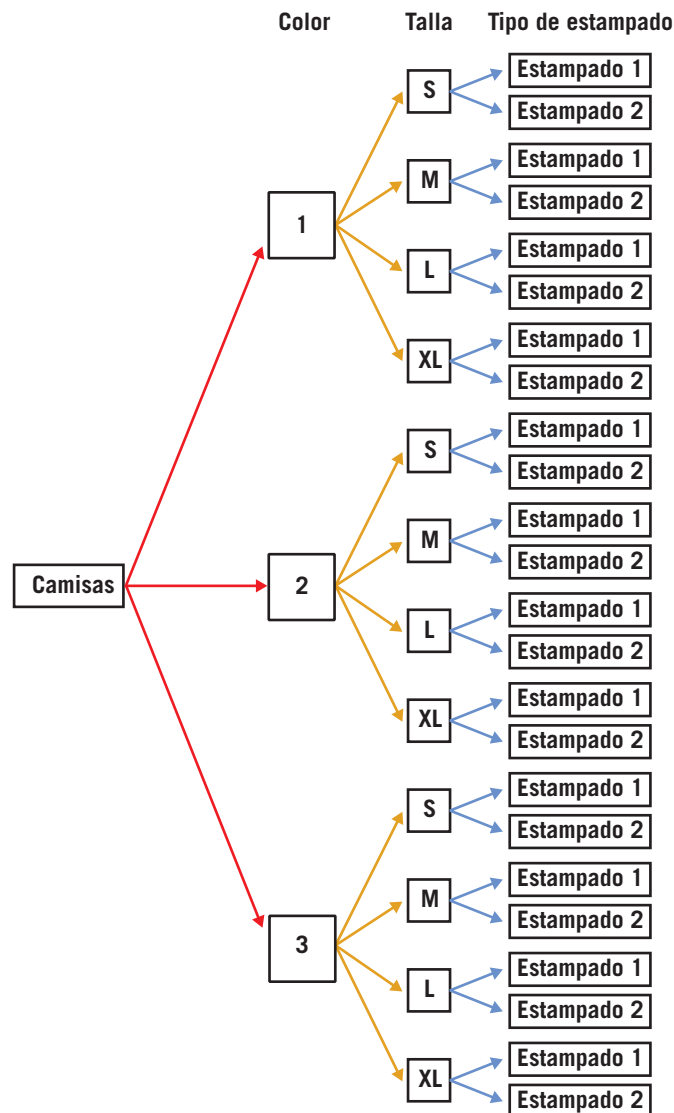
Supón que en una tienda de ropa se ofrecen camisas en tres colores, los cuales están disponibles en cualesquiera de cuatro tallas, con dos tipos de estampado. ¿Cuántas camisas diferentes tienen?

Solución

De acuerdo con el principio multiplicativo, esta tienda puede ofrecer una gama de

$$\{3\}\{4\}\{2\} = 24$$

Camisas diferentes. El diagrama de árbol es de esta manera:



Principio aditivo

Si una operación puede realizarse en m formas y otra en n y ambas no pueden realizarse juntas, entonces el número total de formas en las que se pueden realizar es $m + n$.

Supón que planeas un viaje y debes decidir entre trasladarte por autobús o avión. Si hay tres rutas para el autobús y dos para el avión, ¿cuántas rutas en total tienes disponibles?

Ejemplo 6

Solución

Como no se puede viajar en autobús y avión al mismo tiempo, entonces se usa el principio aditivo, por lo que tienes

$$3 + 2 = 5$$

cinco rutas disponibles para el viaje.

Permutaciones

Una *permutación* es un arreglo de todos los elementos de un conjunto, o de una parte de ellos, en el que importa el orden.

TEOREMA 1. Sean $n, r \in \mathbb{N}$ no negativos, con $r \leq n$, entonces:

- El número de permutaciones de n objetos distintos es $n!$
- El número de permutaciones de n objetos distintos, tomando r a la vez, es

$${}_nP_r = \frac{n!}{(n-r)!}$$

¿Cuántas palabras distintas de seis letras se pueden formar con las letras A, B, C, D, E y F ?

Ejemplo 7

Solución

Como nos importa el orden, ya que $ABCDEF \neq FEDCBA$, entonces utilizamos permutaciones, es decir:

$$6! = 720$$

Podemos formar 720 palabras con esas letras, independientemente de si tienen significado o no.

¿Cuántas palabras de cuatro letras pueden formarse con las letras A, B, C, D, E y F ?

Ejemplo 8

Solución

Como nos importa el orden, entonces utilizamos permutaciones tomando $r = 4$, de modo que sustituyendo valores en la fórmula de permutaciones se obtiene:

$${}_6P_4 = \frac{6!}{(6-4)!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = \frac{720}{2} = 360$$

esto es, pueden formarse 360 palabras de cuatro letras.

Si queremos resolver este ejemplo mediante una hoja de cálculo como Excel hemos de usar la función =PERMUTACIONES(), con los valores correspondientes de n y r , respectivamente:

fx =PERMUTACIONES(6,4)			
C	D	E	F
	=PERMUTACIONES(6,4)		
	PERMUTACIONES(número, tamaño)		

Se oprime Intro y listo, obtenemos el resultado.

Ejemplo 9

¿Cuántas palabras distintas pueden formarse con las letras A, B, C, D, E, F, si es necesario que D, E y F estén juntas?

Solución

Como se pide que D, E y F estén juntas, entonces consideramos estas letras como un solo elemento:

A B C (D E F)

por lo que contamos con cuatro elementos que debemos “mover”, es decir, tenemos 4!

Pero D, E y F pueden estar en distinto orden y aun así seguir juntas, lo que equivale a 3!

Como estas dos operaciones se pueden llevar a cabo al mismo tiempo, aplicamos el principio multiplicativo:

$$4! \times 3! = 144 \text{ palabras}$$

Podemos formar 144 palabras con las condiciones descritas.

Combinaciones

Dado un conjunto de n objetos distintos, cualquier subconjunto no ordenado de tamaño r de esos objetos se llama *combinación*.

TEOREMA 2. Sean $n, r \in N$ no negativos, con $r \leq n$; entonces el número de formas de elegir r elementos de un conjunto de n objetos distintos, *sin importar el orden*, es:

El término $\binom{n}{r}$ se llama coeficiente *binomial*.

$${}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Ejemplo 10

En un grupo de 30 alumnos se quiere formar equipos de cuatro personas. ¿Cuántos equipos pueden formarse?

Solución

Como no importa el orden, entonces calculamos las combinaciones de $n = 30$ y $r = 4$:

$${}_{30}C_4 = \binom{30}{4} = \frac{30!}{4!(30-4)!} = 27\,405$$

Pueden formarse 27 405 equipos. Con Excel se usa la función =COMBINAT(), en la que primero se anota el valor n y después el de r :

fx =COMBINAT(30,4)			
C	D	E	
	=COMBINAT(30,4)		
	COMBINAT(número, tamaño)		

Se oprime Intro y se obtiene el resultado.

¿Cuántas palabras de cuatro letras se pueden formar con las letras A, B, C, D, E, F, sin importar el orden de éstas?

Ejemplo 11

Solución

Como no importa el orden, entonces utilizamos combinaciones con $r = 4$, es decir,

$${}_6C_4 = \frac{6!}{4!(6-4)!} = \frac{720}{48} = 15$$

Pueden formarse 15 palabras distintas de cuatro letras.

En combinaciones, tener *ABCD* es exactamente lo mismo que tener *DCBA*, por eso el número de combinaciones es menor que el de las permutaciones.

¿Cuántas palabras de cuatro letras se pueden formar con las letras A, B, C, D, E, F, sin importar el orden de éstas y queremos que cada palabra tenga una vocal y tres consonantes?

Ejemplo 12

Solución

En este caso, la condición establece que la palabra que hemos de formar debe tener una vocal y tres consonantes; es decir:

- De dos vocales disponibles debemos elegir una:

$${}_2C_1 = \frac{2!}{1!(2-1)!} = 2$$

- De cuatro consonantes disponibles elegimos tres:

$${}_4C_3 = \frac{4!}{3!(4-3)!} = 4$$

Como se pueden elegir vocales y consonantes al mismo tiempo, utilizamos el principio multiplicativo:

$${}_2C_1 \times {}_4C_3 = 2 \times 4 = 8$$

Podemos formar ocho palabras con las letras y las condiciones indicadas.

Supón que tienes cinco revistas: tres sobre chismes del espectáculo y dos de automóviles.

- ¿Cuál es la probabilidad de que al elegir tres revistas al azar, dos sean sobre chismes y una sobre autos?

Ejemplo 13

- b. ¿Cuál es la probabilidad de que al elegir tres revistas al azar, al menos una sea sobre chismes?

Solución

- a. Primero calculamos el número de casos totales, es decir, el número de formas que hay para elegir tres revistas, sin importar de qué tema sean:

$${}_5C_3 = 10$$

Después determinamos el número de casos favorables, o sea, la cantidad de formas con que contamos para elegir dos revistas de chismes y una de autos:

$${}_3C_2 \cdot {}_2C_1 = 3$$

Por último, hallamos la probabilidad:

$$P(\text{dos revistas de chismes y una de autos}) = \frac{3}{10} = 0.3$$

Tenemos una probabilidad de 30% de que al elegir al azar tres revistas, dos sean sobre chismes y una de autos.

- b. Como se pide *al menos* sacar una revista de chismes, entonces las posibilidades son de que sean una, dos o tres revistas de chismes, es decir,

- Una revista de chismes y dos de autos:

$${}_3C_1 \cdot {}_2C_2 = 3$$

- Dos revistas de chismes y una de autos:

$${}_3C_2 \cdot {}_2C_1 = 3$$

- Tres revistas de chismes y ninguna de autos

$${}_3C_3 \cdot {}_2C_0 = 1$$

Como los eventos anteriores no pueden presentarse al mismo tiempo, entonces se aplica el principio aditivo:

$$3 + 3 + 1 = 7$$

Por último, calculamos la probabilidad que se nos pide:

$$P(\text{al menos una revista de chismes}) = \frac{7}{10} = 0.7$$

es decir, tenemos una probabilidad de 70% de que al tomar tres revistas, al menos una sea de chismes.

Ejercicios

I. TEORÍA Y ALGO MÁS

1. Ésta es una versión “estadística” del juego “Tripas de gato”, donde debes unir la operación con el resultado que le corresponda:

${}_5P_3$	$6!$	0.25
0.1667	${}_3C_1 * {}_2C_1 / {}_5C_2$	$3,628,800$
10	840	
${}_{10}C_2$	15	${}_9P_6 / {}_9P_9$
45	${}_6P_3 * {}_4P_3$	${}_5C_3$
${}_5C_4 * {}_3C_1$	2.880	60
$10!$	$15\ 504$	${}_{20}C_5$
${}_7P_4$	0.6	720
	$3! / 4!$	

II. PIENSO,... ¿Y LUEGO?

2. En el libro *El diablo de los números* (Enzensberger, 1998) se hace alusión a los números triangulares y a un “viejo triángulo numérico”:

Tocó la pizarra con su vara, y ahí estaba el triángulo, en todo su esplendor y a cuatro colores.

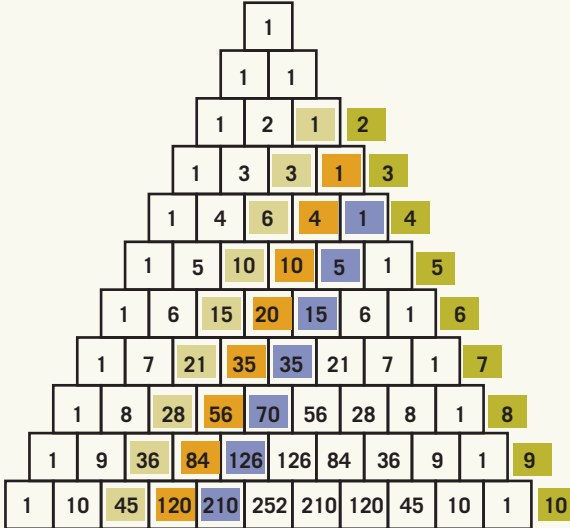
Más cómodo imposible —dijo el viejo diablo de los números—. Al estrechar las manos, simplemente cuentas los cubos verdes de arriba abajo: con dos personas un apretón de manos, con tres personas tres, con once personas 55.

Para nuestra cuadrilla de limpieza necesitas los cubos rojos. Vuelves a contar de arriba abajo. Empiezas con tres personas, con ellas no hay más que una posibilidad. Si puedes elegir cuatro personas dispones de cuatro combinaciones, con cinco personas ya son diez. ¿Y qué pasa cuando están los once alumnos?

Entonces son 165 —respondió Robert—. Es realmente sencillo. Este triángulo numérico es casi tan bueno como una calculadora. Pero ¿para qué sirven los cubos amarillos?

—Oh —dijo el anciano—, ya sabes que yo no me doy fácilmente por satisfecho. Nosotros, los diablos de los números, siempre llevamos todo hasta el extremo. ¿Qué harás si las tres personas que tienes no son suficientes para el trabajo? Tendrás que escoger cuatro y la fila amarilla te dirá cuántas posibilidades hay, por ejemplo, para elegir un cuarteto a partir de ocho personas.

—Setenta —dijo Robert...



Forma un equipo de cuatro personas y luego:

- Investiga cuáles son los números triangulares.
- Investiga qué es y cómo se construye el llamado “Triángulo de Pascal”.
- ¿Cuáles son las similitudes que hay entre los coeficientes binomiales y las combinaciones? Discute los resultados con tus compañeros de equipo y expongan ante el grupo sus conclusiones.

PURO DEPORTE

3. *Sólo para practicar.* Se lanza una moneda “justa” y se anota “águila” o “sol”. Posteriormente se lanza la moneda nuevamente anotando la figura que salió. Construye el diagrama de árbol del espacio muestral de este experimento.
4. *Sólo para practicar.* Se lanza una moneda “justa”. Si sale “águila”, se lanza la moneda nuevamente. Si sale “sol”, se lanza un dado “justo”. Construye el diagrama de árbol del espacio muestral de este experimento.
5. *Sólo para practicar.* Se tiene una baraja inglesa de 52 cartas, divididas en cuatro palos: picas, corazones, diamantes y tréboles. Se saca una carta de dicha baraja y se anota si es un número o una letra, así como su color. Construye el diagrama de árbol del espacio muestral de este experimento, donde se debe mostrar la cardinalidad de cada evento.
6. *Sólo para practicar.* Calcula:
 - a) $9!$
 - b) $3! \times 5!$
 - c) ${}_{15}C_5$
 - d) ${}_{12}P_7$
 - e) ${}_5C_0 + {}_5C_1 + {}_5C_2$
 - f) ${}_6P_4 + {}_6P_5 + {}_6P_6$
 - g) ${}_4C_0 \times {}_3C_2 + {}_4C_1 \times {}_3C_1$

7. *Para pensar un rato.* Calcula:

- a) ${}_0C_0$
- b) ${}_1C_0 + {}_1C_1$
- c) ${}_2C_0 + {}_2C_1 + {}_2C_2$
- d) ${}_3C_0 + {}_3C_1 + {}_3C_2 + {}_3C_3$
- e) ${}_4C_0 + {}_4C_1 + {}_4C_2 + {}_4C_3 + {}_4C_4$
- f) ${}_5C_0 + {}_5C_1 + {}_5C_2 + {}_5C_3 + {}_5C_4 + {}_5C_5$
- g) Con base en los incisos anteriores, deduce el resultado de
 ${}_nC_0 + {}_nC_1 + {}_nC_2 + \dots + {}_nC_{n-2} + {}_nC_{n-1} + {}_nC_n$

8. *Sólo para practicar.* Se tienen los números 0, 3, 4, 6, 7 y 9. Si no se permiten repeticiones:

- a) ¿Cuántos números de tres dígitos se pueden formar?
- b) ¿Cuántos de éstos son menores que 400?
- c) ¿Cuántos son pares?
- d) ¿Cuántos son impares?
- e) ¿Cuántos son múltiplos de 5?

9. *Sólo para practicar.* Los números telefónicos en la ciudad de Cuernavaca, Morelos, constan de siete dígitos.

- a) ¿Cuántas líneas telefónicas pueden crearse en esa ciudad? Recuerda que no pueden comenzar con cero.
- b) Si los tres primeros dígitos representan una zona de Cuernavaca, ¿cuántas líneas telefónicas pertenecen a la zona 326?

10. *Sólo para practicar.* Se tiene la palabra MURCIÉLAGO. Si no se permiten repeticiones al usar las letras que la forman, determina:

- a) El número de palabras que puedes crear.
- b) El número de palabras con cinco letras.
- c) La probabilidad de que una palabra de cinco letras termine en vocal.

11. *Parque vehicular.* La Ciudad de México tiene placas de automóviles con tres números seguidos por tres letras, es decir, son placas del tipo 999-AAA. Sin embargo, no pueden empezar con 0 (cero) ni se utilizan las letras I, O, Q y Ñ.

- a) ¿Cuántas placas diferentes son posibles?
- b) ¿Cuántas placas diferentes terminan en vocal?
- c) ¿Cuál es la probabilidad de que una placa, elegida al azar, sea impar?
- d) De acuerdo con el resultado obtenido en el inciso (a), abre un debate acerca de las consecuencias que traería consigo tener ese número de vehículos circulando en la ciudad. Además, realicen propuestas para mejorar los problemas de circulación vial que se viven en las grandes ciudades.

12. *Servicios restauranteros.* Una cadena de hamburguesas anuncia que se ofrecen 512 maneras de pedir una hamburguesa. Un cliente puede elegir cualquiera de estos ingredientes: catsup, mostaza, mayonesa, aderezo de la casa, cebolla, pepinillos, jitomate, tocino y lechuga. ¿Crees que lo que afirma la empresa sea cierto? Justifica tu respuesta.

- 13. Cultura.** Una persona tiene un libro de cada uno de los autores siguientes: Mario Vargas Llosa, Alberto Ruy Sánchez, Gabriel García Márquez, Günther Grass, Patrick Süskind, Heinrich Böll, Arturo Pérez-Reverte y Camilo José Cela. Como aún no los ha leído, quiere ponerlos en una repisa cerca de su cama. De acuerdo con esto, determina:
- ¿De cuántas formas puede acomodar los libros sobre la repisa?
 - ¿De cuántas maneras puede ordenarlos si quiere que los libros de autores latinoamericanos estén juntos?
 - ¿De cuántas formas puede acomodarlos, si los autores latinoamericanos deben estar juntos, así como los autores alemanes y los españoles también?
 - Determina las probabilidades correspondientes a los incisos (b) y (c).
 - Lee esta frase, de John Ernest Steinbeck: “Por el grosor del polvo en los libros de una biblioteca pública puede medirse la cultura de un pueblo”.
Organiza con tus compañeros de clase una mesa redonda y expongan sus puntos de vista sobre la lectura y el desarrollo de un pueblo. ¿A qué conclusiones llegan?
- 14. Finanzas.** Se va a formar un portafolio de inversión de cuatro instrumentos financieros a partir de cinco de renta fija y seis de renta variable. ¿De cuántas maneras podría formarse el portafolio si:
- No hay restricciones?
 - Debe haber dos instrumentos de renta fija y dos de renta variable?
 - No debe haber ningún instrumento de renta fija?
 - Debe haber tres instrumentos de renta fija?
 - Calcula las probabilidades para los incisos anteriores.
 - Investiga qué es un instrumento de renta fija y otro de renta variable.
 - Reflexiona sobre la importancia del ahorro y comparte con tus compañeros de clase las conclusiones a las que llegaste.
- 15. Fútbol.** Tres parejas (tres hombres y tres mujeres) compran boletos para ir a un partido de los Pumas. Cuando llegan al estadio de Ciudad Universitaria les surge la duda de cómo sentarse en la fila. ¿De cuántas maneras pueden hacerlo si:
- No hay restricciones?
 - Los hombres deben sentarse juntos?
 - Deben sentarse por parejas?
 - ¿Cuál es la probabilidad de cada uno de los eventos descritos en los incisos anteriores?
 - ¿Crees que el fútbol es un deporte sólo para hombres? Discútelo con tus compañeros de clase.
- 16. Trabajo en equipo.** En un grupo de 25 alumnos hay 13 chicas y 12 chicos. Se desean formar equipos de trabajo para realizar una investigación:
- ¿Cuántos equipos de cinco personas pueden formarse?
 - ¿Cuántos equipos pueden formarse si debe haber una chica?
 - ¿Cuántos equipos pueden formarse si debe haber dos chicas?
 - Determina las probabilidades correspondientes a los incisos anteriores.
 - ¿Qué ventajas o desventajas hay al trabajar en equipo? ¿Qué habilidades se desarrollan de esta manera? Coméntalo con tus compañeros de equipo y expongan sus ideas en clase.

17. *Control de calidad.* Se tiene una caja con 24 latas y se sabe que dos están contaminadas. Se van a seleccionar tres latas al azar para someterlas a una prueba de control de calidad, es decir, para medir los estándares de calidad de la empresa.
- a) ¿Cuántas combinaciones de tres latas pueden hacerse?
 - b) ¿Cuál es la probabilidad de que se seleccione una lata contaminada para la prueba?
 - c) ¿Cuál es la probabilidad de que se seleccione al menos una lata contaminada para la prueba?
 - d) ¿Y la probabilidad de que no se elijan latas contaminadas para la prueba?
 - e) Forma un equipo e investiga qué es la filosofía de la calidad según W. Edwards Deming. Con el resultado, elaboren un mapa mental sobre las ideas principales. Junto con tus compañeros de clase, realicen una mesa redonda, en la que han de exponer la importancia de llevar a cabo controles de calidad en cualquier empresa y los beneficios que implican tanto al consumidor como a la empresa y al país.

Evaluación sumativa

Un escalón hacia arriba

Contesta los reactivos siguientes, que te permitirán saber si has aprendido lo que se esperaba.

1. Obtén el resultado de:

a) $6!$

b) $\frac{3!}{5!}$

c) ${}_{25}C_3$

d) ${}_6P_2$

e) ${}_7C_4 \times {}_3C_2$

f) $\frac{{}_5P_3}{{}_5P_5}$

2. De una caja que contiene 10 botellas de refresco, tres tienen la taparroscas rota. Si se eligen cuatro botellas al azar:

a) ¿Cuál es la probabilidad de que haya dos con la taparroscas rota?

b) ¿Cuál es la probabilidad de que haya cuando mucho dos con la taparroscas rota?

3. En una fiesta hay cuatro alemanes, tres italianos y dos franceses. Supón que hay sillas acomodadas en fila y determina la probabilidad de que los alemanes:

a) Se sienten juntos.

b) No se sienten juntos.

Proyecto 5

RUTA ÓPTIMA

La empresa JDA brinda servicios integrados y soluciones basadas en las necesidades de sus clientes para la transportación de correspondencia y paquetería en todo el país. JDA es una empresa mexicana que empezó a operar en 2007, pero ha enfrentado algunos problemas de logística, lo que ha repercutido en el aumento de sus costos de operación.

En los últimos dos años se ha dado cuenta que las rutas terrestres que salen de la Ciudad de México y que han implicado mayores costos son:

- Cuernavaca, Morelos
- Puebla, Puebla
- Querétaro, Querétaro

Supón que has sido contratado para optimizar las rutas destinadas a la empresa JDA, esto es, para:

- Determinar todas las rutas posibles para la entrega de paquetes a esas tres ciudades.
- Calcular la ruta que minimice costos por concepto de peaje y de combustible.

Escribe un informe breve con la ruta que más conviene a la empresa JDA; justifica la decisión.

Aspectos para tener en cuenta al escribir el informe:

- Indica la técnica que utilizaste para determinar las rutas posibles.
- Muestra en un mapa dichas rutas.
- Para el cálculo de las tarifas de peaje y gasto aproximado de combustible puedes consultar la página: http://aplicaciones4.sct.gob.mx/sibuac_internet/ControllerUI?action=cmdEscogeRuta
- Explica por qué es importante para una empresa minimizar costos y qué impacto tiene para el cliente.

En los últimos años han crecido empresas privadas de servicios de paquetería en nuestro país, lo que ha implicado que Correos de México esté reduciendo su participación en el mercado. ¿Qué se podría hacer para reactivar Correos de México?



Competencias

- Expresa ideas y conceptos mediante representaciones lingüísticas y matemáticas relacionadas con eventos mutuamente excluyentes e independientes, siguiendo instrucciones y procedimientos de manera reflexiva; además, ordena información de acuerdo con categorías y jerarquías relacionadas con la probabilidad conjunta
- Sintetiza evidencias obtenidas mediante la experimentación para producir conclusiones y formular nuevas preguntas relacionadas con la probabilidad condicional
- Estructura ideas y argumentos sobre el teorema de Bayes de manera clara, coherente y sintética
- Define metas y da seguimiento a sus procesos de construcción de conocimiento sobre los eventos mutuamente excluyentes e independientes
- Propone maneras de resolver un problema o desarrollar un proyecto en equipo, definiendo un curso de acción con pasos específicos, asumiendo una actitud constructiva, congruente con los conocimientos y habilidades con los que cuenta dentro de distintos equipos de trabajo

6. Probabilidad conjunta

Contenido

- Eventos mutuamente excluyentes
- Eventos independientes
- Probabilidad condicional
- Teorema de Bayes

El primer trabajo importante en estadística en el sentido moderno lo llevó a cabo un insólito investigador: un tendero de Londres del siglo xvii. John Graunt (1620-1674) estudió latín y francés por su cuenta y empezó a interesarse por las *Listas de mortalidad* (cifras semanales de los fallecimientos, parroquia por parroquia) publicadas en Londres desde 1604. El proceso de emisión de estos informes surgió principalmente con el fin de disponer de una señal de alarma rápida ante devastadoras epidemias.

Graunt fue también pionero en la construcción de una distribución de edades o “Tablas de vida”¹ de la población viva a partir de las cifras de muertes y sus causas, cuya trascendencia política fue considerable (Livio, 2009).

Ubícate

Antes de comenzar con el estudio de la probabilidad conjunta, contesta el examen siguiente para determinar si tienes los conocimientos necesarios para comprender cabalmente este bloque.

1. Se sabe que $\#(A - B) = 35$, $\#(A^c \cap B) = 15$ y $\#(A^c) = 60$. Con base en estos datos, determina:
 - a) $\#(A \cap B)$
 - b) $\#(B)$
2. En la tabla que sigue se muestra el número de personas entre 12 y 65 años clasificadas según sus hábitos de consumo de tabaco, de acuerdo con la Encuesta Nacional de Adicciones

Patrones de consumo	Total
Fumador	16 371 601
Ex fumador	12 155 232
No fumador	41 240 234
Total	69 767 067

A partir de estos datos, determina:

- a) Si se elige una persona al azar, ¿cuál es la probabilidad de que sea exfumador?
- b) ¿Cuál es la proporción de fumadores?
- c) Representa la información mediante un árbol de probabilidad.

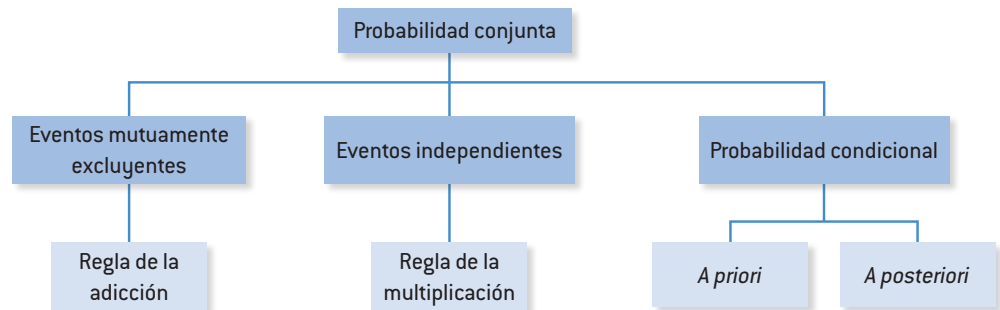
¹ Las tablas de vida son la base para el cálculo de los seguros de vida. En ellas se muestran las probabilidades de vivir según la edad.

3. En un grupo de 50 personas se sabe que 30 tienen i-Phone, 25 Blackberry y 10 ambas marcas de teléfonos. Con base en estos datos, haz o contesta lo que se indica:
- Traza el diagrama de Venn respectivo.
 - ¿Cuál es la probabilidad de que una persona escogida al azar tenga i-Phone o Blackberry?
 - ¿Cuál es la probabilidad de que una persona escogida al azar tenga i-Phone, pero no Blackberry?

Si no tuviste problemas o dudas con el examen anterior, ahora sí comienza el estudio del bloque 6.

Probabilidad conjunta

En el bloque anterior aprendiste a calcular las probabilidades simples (de una sola característica). En este bloque proseguiremos con el estudio de la teoría de probabilidad, pero ahora lo haremos para eventos conjuntos, es decir, aquellos que contemplan dos o más características. En el esquema siguiente se muestran los temas que abordaremos.



Desempeños

Al terminar de estudiar este bloque y realizar todas las actividades propuestas habrás aprendido a:

- Sintetizar los elementos de la probabilidad conjunta para resolver problemas de tu vida cotidiana.
- Analizar las características de una representación de forma tabular, gráfica y funciones de probabilidad mediante gráficas de árbol.
- Describir las características de los eventos mutuamente excluyentes e independientes para su aplicación en la resolución de problemas.
- Aplicar las características de la probabilidad en los diagramas de árbol para la resolución de problemas.
- Explicar las condiciones del cálculo de la probabilidad analizando el teorema de Bayes dentro de la solución de problemas.

Composición de eventos

Como recordarás, la *probabilidad* es el valor entre cero y uno, ambos inclusive, que mide la viabilidad de que cierto resultado o evento ocurra. Por supuesto, para el cálculo de esa probabilidad seguiremos utilizando el enfoque clásico:

$$\text{Probabilidad} = \frac{\text{casos favorables}}{\text{casos totales}}$$

La composición de eventos se presenta cuando se realizan operaciones para calcular la probabilidad de que ocurran dos o más eventos.

Definición y representación de eventos

Para determinar esa probabilidad retomaremos lo aprendido acerca de la teoría de conjuntos.

Probabilidad de la unión

Sean dos eventos A y B . La probabilidad de la unión de estos eventos, $P(A \cup B)$, se define como la probabilidad de que ocurra A , la probabilidad de que ocurra B , o bien, la probabilidad de que ocurran ambos eventos, es decir,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

lo cual, representado en diagrama de Venn, se grafica como en la figura 1.

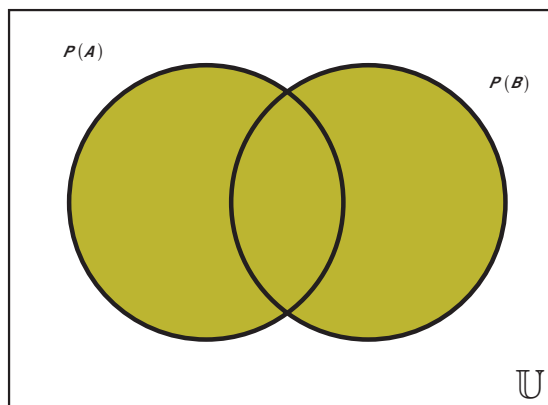


Figura 1 Diagrama de Venn de la probabilidad de la unión de dos eventos.

Como puedes observar en la figura 1, en la teoría de conjuntos A y B reciben el nombre de *conjuntos*, en tanto que en la teoría de la probabilidad (tema del bloque anterior y del presente) se llaman *eventos*.

Probabilidad de la intersección

Sean dos eventos A y B . La probabilidad de que ocurra la intersección de estos eventos, $P(A \cap B)$, se define como la probabilidad de que ocurra A y de que se presente B , lo cual, representado en diagrama de Venn, queda como en la figura 2.

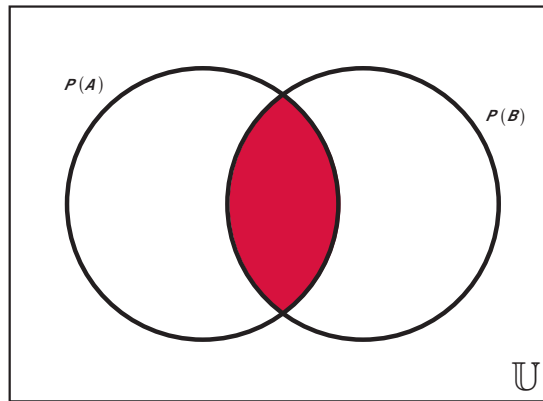


Figura 2 Diagrama de Venn de la probabilidad de que se presente la intersección de dos eventos.

Probabilidad del complemento

Sea el evento A . La regla del complemento establece que la probabilidad de que ocurra A más la probabilidad de que no ocurra A , $P(A^c)$, es igual a 1, es decir:

$$P(A) + P(A^c) = 1$$

En la figura 3 se muestra el diagrama de Venn correspondiente.

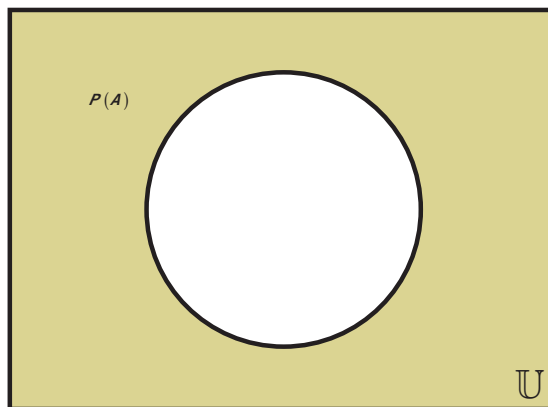


Figura 3 Diagrama de Venn de la probabilidad de que se presente el complemento de un evento.

Probabilidad de eventos disjuntos o mutuamente excluyentes

Sean dos eventos A y B , tales que son mutuamente excluyentes. La probabilidad de que ocurran los dos al mismo tiempo es cero, esto es

$$P(A \cap B) = 0$$

En la figura 4 se muestra el diagrama de Venn respectivo.

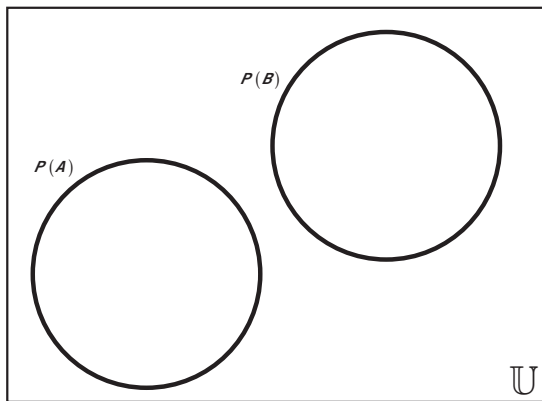


Figura 4 Diagrama de Venn de la probabilidad de eventos mutuamente excluyentes o disjuntos.

Probabilidad de la diferencia

Sean dos eventos A y B . La probabilidad de que ocurra el evento A , pero que no ocurra el evento B se calcula como:

$$P(A - B) = P(A) - P(A \cap B)$$

Gráficamente, la probabilidad de la diferencia se muestra en la figura 5.

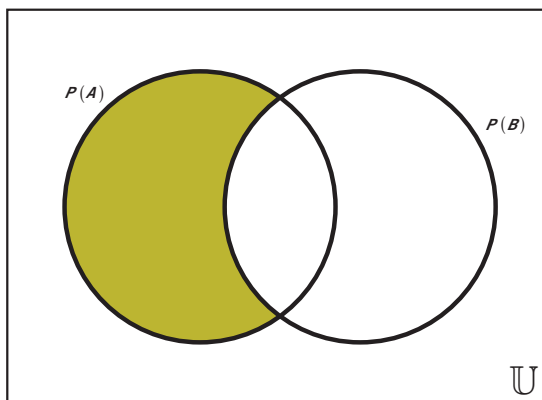


Figura 5 Diagrama de Venn de la probabilidad de que ocurra la diferencia de dos eventos.

En la figura 5, observa que

$$P(A - B) \neq P(B - A)$$

Supón que tienes dos eventos A y B y sabes que:

$$P(A) = 0.4, \quad P(B) = 0.3 \quad \text{y} \quad P(A \cap B) = 1$$

Con base en estos datos, calcula:

- $P(A \cup B)$
- $P(B - A)$
- $P(B^c)$

Ejemplo 1

Solución

Al aplicar las fórmulas que hemos visto se obtiene:

- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.3 - 0.1 = 0.6$
- $P(B - A) = P(B) - P(A \cap B) = 0.3 - 0.1 = 0.2$
- En este caso, de la regla del complemento despejamos $P(B^c)$, de modo que tenemos:

$$\begin{aligned} P(B) + P(B^c) &= 1 \\ P(B^c) &= 1 - P(B) \\ P(B^c) &= 1 - 0.3 = 0.7 \end{aligned}$$

Regla de la adición para eventos no excluyentes

Sean dos eventos A y B . La probabilidad de la unión de estos eventos, $P(A \cup B)$, se define como la probabilidad de que ocurra A , de que ocurra B , o de que ocurran ambos eventos, es decir,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

En la figura 6 se muestra el diagrama de Venn de esta probabilidad.

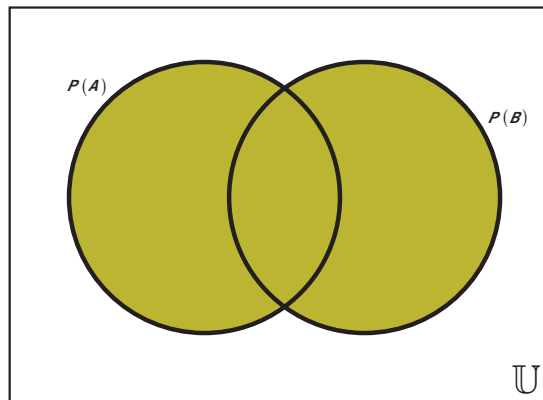


Figura 6 Diagrama de Venn de la adición para eventos no excluyentes.

Ejemplo 2

En cierta escuela se hace un estudio sobre el porcentaje de aprobación en dos materias “coco”: Física y Matemáticas. El estudio muestra que la probabilidad de que un estudiante apruebe Matemáticas es de 0.6, de que apruebe Física es de 0.55 y de que apruebe ambas es de 0.35. ¿Cuál es la probabilidad de que un alumno apruebe Matemáticas o Física?

Solución

Sean los eventos:

- M : aprobar Matemáticas
- F : aprobar Física

cuyas probabilidades son:

$$P(M) = 0.6, \quad P(F) = 0.55 \quad \text{y} \quad P(M \cap F) = 0.35$$

Por consiguiente, la probabilidad de que un alumno apruebe Matemáticas o Física es:

$$P(M \cup F) = P(M) + P(F) - P(M \cap F) = 0.6 + 0.55 - 0.35 = 0.8$$

Regla de la adición para eventos disjuntos o mutuamente excluyentes

Sean dos eventos A y B , tales que son mutuamente excluyentes, es decir, la probabilidad de que ocurran los dos al mismo tiempo es cero, $P(A \cap B) = 0$. Por tanto, la probabilidad de la unión se calcula como:

$$P(A \cup B) = P(A) + P(B)$$

En la figura 7 se muestra el diagrama de Venn correspondiente.

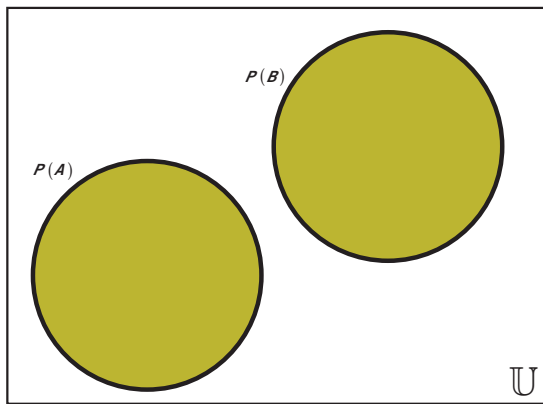


Figura 7 Diagrama de Venn de la adición de eventos mutuamente excluyentes o disjuntos.

En el área de recursos humanos de una empresa se han recibido 500 solicitudes de empleo. Se sabe que 325 son de mujeres y el resto de hombres. ¿Cuál es la probabilidad de que al tomar una solicitud al azar provenga de un hombre o de una mujer?

Ejemplo 3

Solución

Sean los eventos:

- M : solicitante mujer
- H : solicitante hombre

cuyas probabilidades son:

$$P(M) = \frac{325}{500} = 0.65, \quad P(H) = \frac{175}{500} = 0.35 \quad \text{y} \quad P(M \cap H) = 0$$

Entonces, para calcular la probabilidad de que al azar se tome la solicitud proveniente un hombre o de una mujer es:

$$P(M \cup H) = P(M) + P(H) = 0.65 + 0.35 = 1$$

es decir, es la probabilidad del evento seguro.

Independencia de eventos

Dos eventos son *independientes* si el que ocurra uno de ellos no influye en la probabilidad de que ocurra el otro, es decir,

$$P(A \cap B) = P(A)P(B)$$

De otra manera, los eventos son *dependientes*.

Ejemplo 4

Se tiene una baraja y se toma una carta. Sean los eventos:

- A : se observa un as
- B : se observa una carta negra

y las probabilidades asociadas a cada evento son:

$$P(A) = \frac{1}{13}, \quad P(B) = \frac{1}{2} \quad \text{y} \quad P(A \cap B) = \frac{1}{13}$$

¿Los eventos A y B son independientes?

Solución

Se verifica que se cumpla $P(A \cap B) = P(A)P(B)$

$$\frac{1}{13} = P(A \cap B) \neq P(A)P(B) = \left(\frac{1}{13}\right)\left(\frac{1}{2}\right) = \frac{1}{26}$$

Por tanto, A y B son eventos *dependientes*.

Visualización de eventos: tablas de contingencia y árboles de probabilidad

Para poder visualizar un problema y hacer los cálculos de las probabilidades de una forma más sencilla, además de los diagramas de Venn pueden usarse las tablas de contingencia y los árboles de probabilidad.

Tabla de contingencia

La de contingencia es una tabla que se emplea para clasificar el número de observaciones respecto a dos características o variables de interés.

Ejemplo 5

Una universidad tiene una población de 1500 alumnos. Se sabe que 850 estudian alguna licenciatura (economía, administración o relaciones internacionales, entre otras) y 650 alguna ingeniería. De los que estudian ingeniería, 350 son mujeres, en tanto que 600 mujeres son las que estudian alguna licenciatura. Elabora una tabla de contingencia para resumir esta información.

Solución

Identificamos los eventos de este problema:

- M : estudiante mujer
- H : estudiante hombre
- L : estudia alguna licenciatura
- I : estudia alguna ingeniería

Elaboramos una tabla:

Colocamos un evento por columna o por fila, de acuerdo con su clasificación; en este caso, por género o por carrera

Género/Carrera	L	I	Total
M			
H			
Total			

Además, se ponen los totales por fila y por columna

Identificamos ahora la cardinalidad de cada evento:

$$\#(L) = 850, \quad \#(I) = 650, \quad \#(L \cap M) = 600 \quad \text{y} \quad \#(I \cap M) = 350$$

Ya con esto, comenzamos a llenar la tabla de contingencia:

Cada una de estas celdas es la intersección de los eventos, es decir, donde se cumplen las dos características

Género/Carrera	L	I	Total
M	600	350	950
H	250	300	550
Total	850	650	1500

Los datos faltantes se obtienen con una simple resta

Los eventos simples son los totales por columna o por fila

Los totales por fila y por columna deben dar la cardinalidad del espacio muestral

De esta manera, podemos calcular la probabilidad, por ejemplo, de que un hombre estudie ingeniería:

$$P(H \cap I) = \frac{300}{1500} = 0.2, \text{ o bien, } 20\%$$

Árbol de probabilidad

Un árbol de probabilidad es un diagrama que se emplea para representar, igual que la tabla de contingencia, todos los resultados posibles de un experimento estadístico. Para elaborarlo hay que establecer una raíz y las ramas se refieren a cada uno de los eventos.

Retomemos el ejemplo anterior, en el que se presenta el caso de una universidad que tiene una población de 1500 alumnos. Se sabe que 850 estudian alguna licenciatura (economía, administración o relaciones internacionales, entre otras), y 650 alguna ingeniería. De los que estudian ingeniería, 350 son mujeres, en tanto que 600 mujeres son las que estudian alguna licenciatura. Elabora un diagrama de árbol para representar la información anterior.

Ejemplo 6

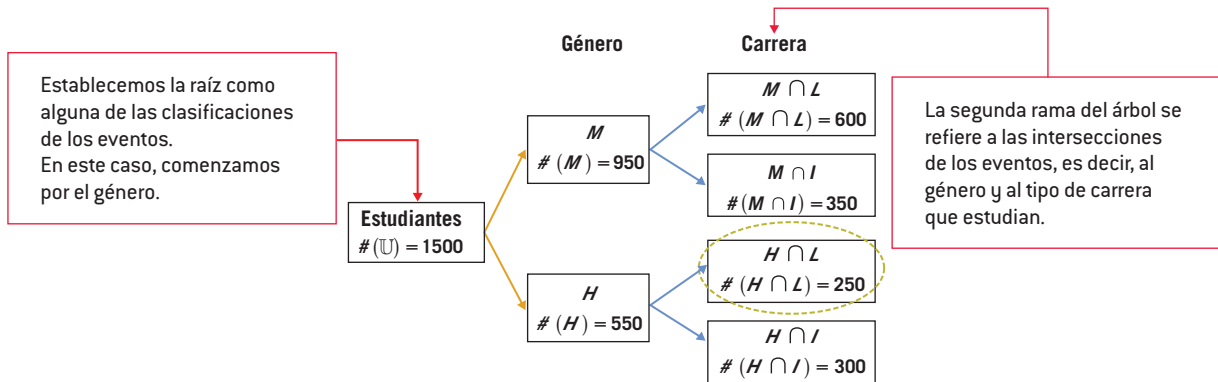
Solución

Los eventos de este problema son:

- *M*: estudiante mujer
- *H*: estudiante hombre

- L : estudia alguna licenciatura
- I : estudia alguna ingeniería

Elaboramos el diagrama de árbol:



Así calculamos, por ejemplo, la probabilidad de que un hombre estudie alguna licenciatura:

$$P(H \cap L) = \frac{250}{1500} = 0.1667, \text{ o bien, } 16.67\%$$

Probabilidad condicional

A menudo sucede que un evento influye en que se presente u ocurra otro. Esto recibe el nombre de *probabilidad condicional*, que se define como la probabilidad de que el evento A ocurra, *puesto que* ya se presentó el evento B . La probabilidad condicional se calcula como sigue:

Observa que

$$P(A|B) \neq P(B|A)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

donde $P(B) > 0$.

A este tipo de probabilidad condicional se le conoce como *a priori*, ya que sucede un evento y se quiere conocer la probabilidad de su efecto.

Cuando se determina una probabilidad condicional el espacio muestral se “reduce”, es decir, como ocurrió el evento B , entonces el “nuevo” espacio muestral tiene la característica del evento B . El diagrama de Venn correspondiente se presenta en la figura 8.

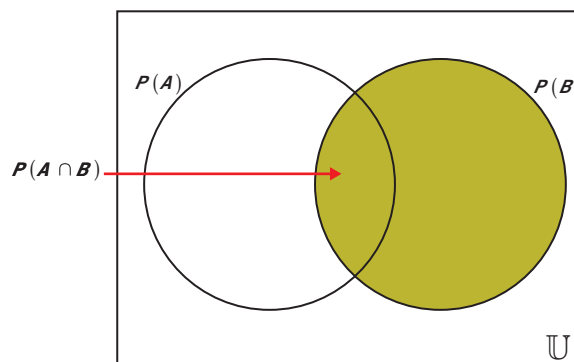


Figura 8 Diagrama de Venn de la probabilidad condicional.

A partir de la fórmula para calcular la probabilidad condicional se deduce que:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B)P(B)$$

A este resultado se le conoce como la *ley de la multiplicación*.

También se puede comprobar la independencia de eventos (recuerda que hay independencia de eventos cuando el que ocurra uno de ellos no influye en que se presente otro), siempre que se cumpla que:

$$P(A|B) = P(A)$$

o bien

$$P(B|A) = P(B)$$

Retomemos el ejemplo de la universidad, cuya población es de 1500 estudiantes, de los cuales 850 estudian alguna licenciatura y 650 alguna ingeniería. Determina la probabilidad de que un alumno elegido al azar:

Ejemplo 7

- Estudie ingeniería, puesto que es hombre.
- Si estudia alguna licenciatura, entonces es mujer.

Solución

- Probabilidad de que un alumno elegido al azar estudia ingeniería, puesto que es hombre:

$$P(I|H) = \frac{P(I \cap H)}{P(H)}$$

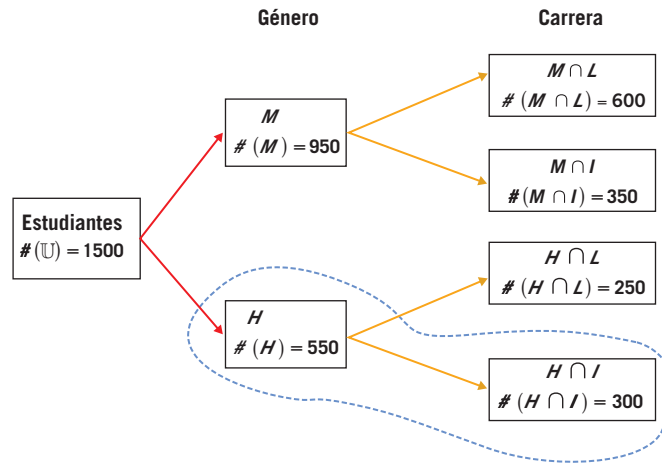
La tabla de contingencia respectiva es la siguiente:

Género/Carrera	L	I	Total
M	600	350	950
H	250	300	550
Total	850	650	1500

De los que son hombres, se identifica cuántos estudian ingeniería

Como el evento que ocurre primero (la condición) es ser hombre, el espacio muestral se reduce únicamente al evento H

Y el árbol de probabilidad es:



De esta manera:

$$P(I|H) = \frac{P(I \cap H)}{P(H)} = \frac{300}{550} = 0.5454$$

De modo que la probabilidad de que un alumno elegido al azar estudie ingeniería, puesto que es hombre, es de 54.54%.

b. La probabilidad de que un alumno elegido al azar estudie licenciatura sea mujer es decir

$$P(M|L) = \frac{P(M \cap L)}{P(L)}$$

Género/Carrera	L	I	Total
M	600	350	950
H	250	300	550
Total	850	650	1500

Como el evento que ocurre primero (la condición) es ser estudiante de licenciatura, el espacio muestral se reduce únicamente al evento L

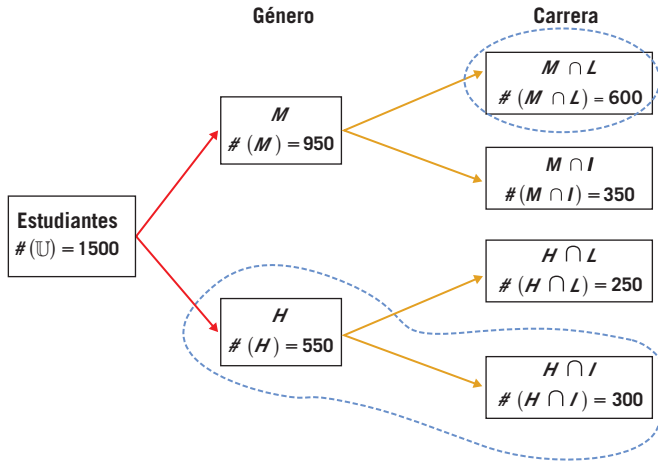
En el caso del árbol de probabilidad, como no lo tenemos por tipo de carrera, entonces podemos hacer dos cosas:

- Elaborar un nuevo árbol de probabilidad, en el que la primera rama sea el tipo de carrera.
- Los que estudian licenciatura pueden calcularse como:

$$P(L) = P(M \cap L) + P(H \cap L) = \frac{600}{1500} + \frac{250}{1500} = \frac{850}{1500} = 0.5666$$

pues consideramos tanto a los hombres como a las mujeres que estudian una licenciatura.

El árbol de probabilidad correspondiente es:



De esta manera:

$$P(M|L) = \frac{P(M \cap L)}{P(L)} = \frac{600}{850} = 0.7058$$

es decir, la probabilidad de que un alumno elegido al azar estudie licenciatura y sea mujer es de 70.58%.

Teorema de Bayes

Thomas Bayes (Inglaterra 1702-1761) fue un teólogo protestante que dedicó gran parte de su vida a demostrar la existencia de dios para quienes no creían. En su afán por probar mediante métodos matemáticos que dios es la única causa de todo lo que existe dio con una fórmula que hoy en día resulta muy útil y que se conoce como *teorema de Bayes*.

El teorema de Bayes sirve para calcular la probabilidad condicional *a posteriori*. En otras palabras, sirve para determinar la probabilidad de una de las causas, puesto que ya se observó el efecto, y se calcula de la manera siguiente:

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^n P(B_i)P(A|B_i)}$$

donde B_j es la causa j y A es el efecto, y B_i es la causa i , donde $i = 1, 2, 3, \dots, n$.

Utilizando la ley de la multiplicación tenemos que:

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^n P(B_i)P(A|B_i)} = \frac{P(A \cap B_j)}{\sum_{i=1}^n P(A \cap B_i)} = \frac{P(A \cap B_j)}{P(A)}$$

De esta manera llegamos a la definición de probabilidad condicional. Es importante señalar que el denominador del teorema de Bayes se conoce como *probabilidad total*. Este teorema tiene las características siguientes:

- Las causas (B_i) son mutuamente excluyentes.
- La suma de las probabilidades asociadas a las causas es 1, es decir, $\sum_{i=1}^n P(B_i) = 1$

La fórmula del teorema de Bayes muchas veces impresiona, pero en realidad es fácil de usar, como se advierte en el ejemplo que sigue.

Ejemplo 8

En México, las ventas de automóviles se distribuyen de la manera siguiente:

- 20% son Nissan
- 14% son Volkswagen
- El resto es de otras marcas

Además, cuando el auto es Nissan, tiene una probabilidad de 73% de ser robado; si es Volkswagen, tiene una probabilidad de 9% de ser robado, y 18% de ser robado si es de otras marcas.

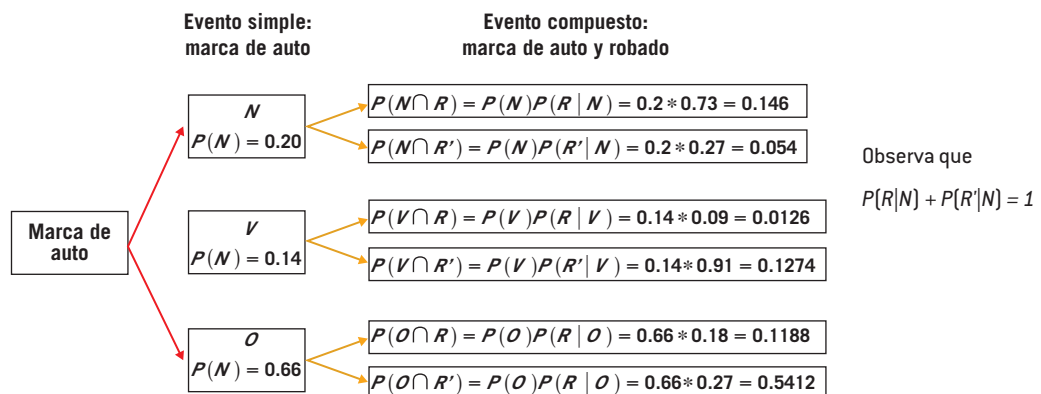
- ¿Cuál es la probabilidad de que se roben un auto?
- Si se han robado un auto, ¿cuál es la probabilidad de que sea un Volkswagen?

Solución

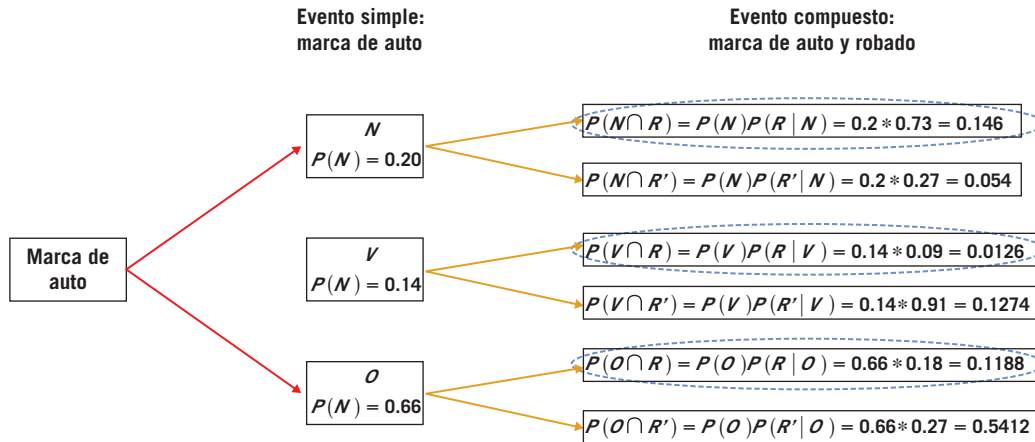
Determinamos los eventos y sus respectivas probabilidades, así como el árbol de probabilidad correspondiente:

- N: auto de la marca Nissan $P(N) = 0.20$ $P(R|N) = 0.73$
- V: auto de la marca Volkswagen $P(V) = 0.14$ $P(R|V) = 0.09$
- O: otras marcas de autos $P(O) = 0.66$ $P(R|O) = 0.18$
- R: robo, en este caso es el efecto.

El árbol de probabilidad queda de la siguiente manera:



- En este inciso debemos calcular la probabilidad total, que se refiere a los autos robados, sin importar la marca, que queda como se muestra en este árbol de probabilidad:



por ende:

$$\begin{aligned}
 P(R) &= P(N)P(R|N) + P(V)P(R|V) + P(O)P(R|O) \\
 P(R) &= (0.20)(0.73) + (0.14)(0.09) + (0.66)(0.18) \\
 P(R) &= 0.2774
 \end{aligned}$$

b. Por el teorema de Bayes,

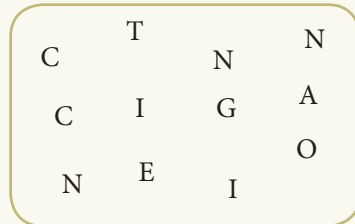
$$P(V|R) = \frac{P(V)P(R|V)}{P(R)} = \frac{(0.14)(0.09)}{0.2774} = \frac{0.0126}{0.2774} = 0.5263$$

Así, la probabilidad de que si un auto es robado, sea Volkswagen es de 52.63%.

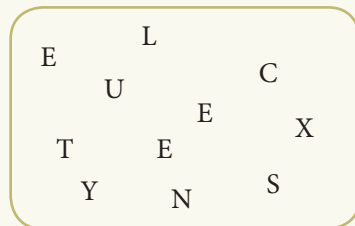
Ejercicios

I. TEORÍA Y ALGO MÁS

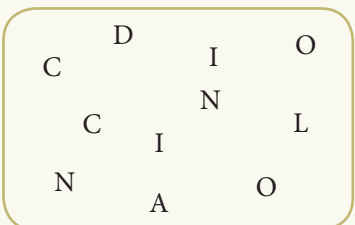
1. Ordena las letras hasta encontrar los conceptos presentados en este bloque; puedes ayudarte de las definiciones que aparecen debajo de cada cuadro.



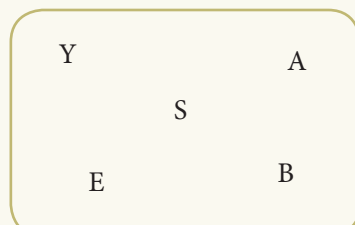
- a) Es una tabla que se utiliza para clasificar el número de observaciones con respecto a dos características o variables de interés.



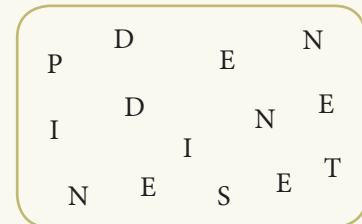
- c) Cuando la probabilidad de que ocurran dos eventos al mismo tiempo es cero, se dice que son eventos mutuamente...



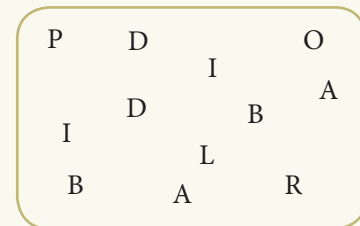
- e) Es la probabilidad de que el evento A ocurra, *dado que* ya ocurrió el evento B. Se le conoce como probabilidad *a priori*.



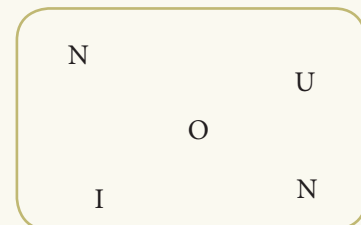
- g) Teorema que sirve para calcular la probabilidad condicional *a posteriori*.



- b) Si la ocurrencia de un evento no influye en la probabilidad de la ocurrencia de otro se dice que son eventos...



- d) Es el valor entre cero y uno, inclusive, que mide la viabilidad de que un cierto resultado (evento) ocurra.



- f) Se define como la probabilidad de que ocurra A o la probabilidad de que ocurra B o que ocurran ambos eventos.

II. PIENSO,... ¿Y LUEGO?

2. *Problema de Monty Hall.* “Marilyn describía al concursante en un juego que tiene que elegir una entre tres puertas. Detrás de una de ellas se encuentra el premio importante —un coche— y detrás de las otras dos hay sendas cabras. Las probabilidades de que el automóvil esté detrás de cualquier puerta dada son iguales: una entre tres. Pero una vez que el concursante ha elegido una puerta, el presentador (que sabe dónde está el automóvil) señala otra puerta, diferente de la escogida por el concursante, y la abre para mostrar que hay una cabra detrás.

“El presentador siempre puede hacer esto: puesto que hay sólo un automóvil, al menos una de las dos puertas no elegidas esconde una cabra. A veces las dos lo hacen (cuando el concursante ha escogido la puerta que esconde el automóvil)” (Stewart, 2001).

Discute con tus compañeros si consideran que mejora la probabilidad que tiene el concursante de ganar el automóvil si cambia de opinión y escoge la puerta que no eligió al principio y que el presentador no ha abierto.

III. PURO DEPORTE

3. *Sólo para practicar.* Sean A y B dos eventos tales que

$$P(A - B) = 0.1, \quad P(A \cap B) = 0.25 \quad \text{y} \quad P(B - A) = 0.15$$

Encuentra $P(A^c \cap B^c)$.

4. *Sólo para practicar.* Sean A y B dos eventos tales que

$$P(A - B) = 0.5, \quad P(A^c \cap B^c) = 0.1 \quad \text{y} \quad P(A^c) = 0.3$$

Encuentra **a.** $P(B)$ y **b.** $P(A^c \cap B)$

5. *Sólo para practicar.* Supón que se estableció la tabla de contingencia siguiente:

	B	B^c
A	10	20
A^c	40	30

- a)** Con esta información, construye el árbol de probabilidad correspondiente.
b) Representa la tabla de contingencia con diagramas de Venn.
c) ¿Cuál es la probabilidad del evento:

- | | |
|------------------|---------------------|
| i. A ? | vi. A^c y B ? |
| ii. B ? | vii. A o B ? |
| iii. A^c ? | viii. A o B^c ? |
| iv. A y B ? | ix. A^c o B^c ? |
| v. A y B^c ? | x. $B - A$? |

6. *Sólo para practicar.* Dado el árbol de probabilidad siguiente:

- a)** Construye la tabla de contingencia respectiva.
b) ¿Cuál es la probabilidad de que suceda:
- X ?
 - A o X ?
 - B o X^c ?
 - C , dado X ?
 - X , dado C ?

7. **Educación.** De acuerdo con información de la Secretaría de Educación Pública, el número de alumnos inscritos en educación básica en el ciclo 2009-2010 en todo el país fue:

Matrícula escolar en educación básica por sexo
Ciclo escolar 2009-2010

Sexo	Preescolar	Primaria	Secundaria	Total
Hombres	2 327 725	7 593 412	3 083 130	13 004 267
Mujeres	2 280 530	7 267 292	3 044 772	12 592 594
Total	4 608 255	14 860 704	6 127 902	25 596 861

Fuente: SEP, Sistema educativo de los Estados Unidos Mexicanos. Principales cifras, ejercicio escolar 2009/2010, en <http://www.sep.gob.mx> <consulta hecha el 8 de junio de 2011>.

Con esta información, determina la probabilidad de que un alumno elegido al azar:

- Sea mujer.
 - Curse preescolar o primaria.
 - Sea hombre y esté en preescolar.
 - Sea mujer y curse nivel secundaria.
 - Sea hombre o curse nivel primaria.
 - Sea hombre, puesto que estudia en el nivel secundaria.
 - Estudie secundaria, si es mujer.
 - ¿Son eventos independientes ser hombre y estudiar el nivel secundaria?
 - Uno de los graves problemas que hay en el país es la deserción de alumnos del sistema educativo. Junto con tus compañeros de equipo, investiga cuáles son las causas para que una persona deje de estudiar y luego comenten algunas acciones que podrían emprenderse para evitarlo.
8. **Turismo.** El aeropuerto internacional de la Ciudad de México (AICM), por el número de operaciones aéreas, es el más transitado de América Latina. A continuación se presenta el número de pasajeros que llegaron al país durante 2011:

Pasajeros por terminal (llegadas durante 2011)

	Nacional	Internacional	Total
Terminal 1	4 098 215	2 109 546	6 207 761
Terminal 2	4 601 073	2 347 327	6 948 400
Total	8 699 288	4 456 873	13 156 161

Fuente: <http://www.aicm.com.mx>

A partir de estos datos, responde lo siguiente:

- ¿Qué proporción de pasajeros arribaron al AICM por la terminal 1? ¿Y por la terminal 2?
- ¿Qué porcentaje son pasajeros nacionales? ¿Y pasajeros internacionales?
- Si una persona, elegida al azar, dice ser pasajero nacional, ¿cuál es la probabilidad de que haya llegado a la terminal 1?
- Cuando una persona arriba a la terminal 2, ¿cuál es la probabilidad de que sea internacional?
- ¿El arribar a la terminal 1 y ser pasajero internacional son eventos independientes? Explica tu respuesta.
- El AICM es insuficiente para atender la demanda de operaciones aéreas y convertirse en la terminal aérea más importante de América Latina. Hace algunos años se trató de construir un aeropuerto nuevo, pero hubo muchas dificultades para ello. Discute con tus compañeros tanto los beneficios como los perjuicios de “sacar” el aeropuerto referido de la Ciudad de México.

9. *Comportamiento del consumidor.* Se seleccionó una muestra de 1004 encuestados en el área metropolitana para determinar cierta información acerca del comportamiento de los consumidores. Entre las preguntas estaba: “¿Disfruta usted comiendo chocolates?”

De 527 hombres que respondieron a la encuesta, 302 contestaron que sí; de 477 mujeres, 452 respondieron que sí.

- Elabora una tabla de contingencia para evaluar las probabilidades.
- Da un ejemplo de evento simple.
- Da un ejemplo de evento conjunto.
- ¿Cuál es el complemento de “disfrutar comiendo chocolates”? ¿Cuál es la probabilidad de que un encuestado elegido al azar sea hombre?
- Disfrute comiendo chocolates?
- Sea mujer y disfrute comiendo chocolates?
- Sea hombre y no disfrute comiendo chocolates?
- Sea mujer o disfrute comiendo chocolates?
- Sea hombre o no disfrute comiendo chocolates?
- Sea hombre o mujer?
- Sea mujer, puesto que disfruta comiendo chocolates?
- ¿Son eventos independientes ser mujer y disfrutar comiendo chocolates?

10. *Problemas sociales.* En el artículo “El vals de la desilusión”, publicado por Roberto Zamarripa en el diario *Reforma* (21 de noviembre de 2011) se presenta la información siguiente: “México tiene 36.2 millones de habitantes entre 12 y 29 años de edad. El equivalente al número de electores en España. Que sólo estudien hay 14 millones de muchachos. Que sólo trabajen hay 10 millones 300 mil. Que estudien y trabajen hay 3 millones 962 mil”.

Con base en esta información:

- Elabora una tabla de contingencia.
- ¿Cuál es la probabilidad de que una persona entre 12 y 29 años:
- No estudie?
 - No trabaje?
 - Sea *nini*, es decir, *ni* estudie *ni* trabaje?
 - Que no trabaje, puesto que no estudia?
 - Que estudie, puesto que trabaja?
 - ¿Son eventos independientes trabajar y estudiar?
 - Discute con tus compañeros sobre el futuro que espera a los *ninis*, así como el problema social que representa para el país este fenómeno social. ¿Qué proponen ustedes para crear una campaña para que los jóvenes cobren conciencia de que deben ser productivos?

11. *Sólo para practicar.* Supón que despiertas en la noche con un terrible dolor y te diriges al baño sin encender la luz. En la oscuridad, alcanzas uno de tres frascos de pastillas, lo destapas e ingieres una pastilla. Una hora más tarde te pones muy enfermo y recuerdas que en dos frascos hay aspirinas y en el otro veneno. Supón que tienes un libro médico y dice que 80% de los individuos normales muestran los síntomas que presentas después de ingerir veneno y que únicamente 5% tiene los síntomas después de tomar aspirinas.

- Realiza el árbol de probabilidad correspondiente.
- ¿Cuál es la probabilidad de que presentes síntomas de envenenamiento?
- Si cada frasco tiene la misma probabilidad de ser escogido en la oscuridad, ¿cuál es la probabilidad de morir envenenado, puesto que presentas los síntomas?²

² Problema del curso de Cálculo de probabilidades I, impartido por Norma Alicia Rosas.

- 12. Industria.** Una empresa dedicada a la fabricación de bolsas de plástico tiene tres máquinas (A, B y C) para ello. La máquina A realiza 25% de la producción de bolsas; la máquina B, 35%, y la máquina C, 40%. Se ha determinado que cuando las bolsas son hechas por la máquina A, 5% tiene defectos; de la máquina B, 10% los tiene, y de la máquina C, 3%. De acuerdo con estos datos:

- Realiza el diagrama de árbol correspondiente.
- ¿Cuál es la probabilidad de que una bolsa, elegida al azar, esté defectuosa?
- Si una bolsa está defectuosa, ¿cuál es la probabilidad de que haya sido producida por la máquina A? ¿Por la máquina B? ¿Y por la máquina C?
- En el artículo “Biodegradables”, publicado por Sergio Sarmiento el 20 de agosto de 2010 en el diario *Reforma*:

Las bolsas de plástico con las que se empaacan los productos en tiendas y supermercados representan menos de 1% de la basura. Pero vamos a suponer que, efectivamente, las autoridades logren que 100% sea remplazado por otras bolsas de plástico “biodegradable”. ¿Ayudaría esto a resolver 1% de los problemas del ambiente? Para nada. El éxito agravaría los problemas ambientales.

Las bolsas “biodegradables” sólo sirven para promover la imagen ecologista de los políticos y aliviar las culpas de las buenas conciencias. Su impacto sobre el ambiente puede ser peor que el polietileno.

Organiza una mesa redonda para discutir los problemas ambientales que existen en el lugar donde vives y qué medidas consideras que podría emprender el gobierno para ayudar a mejorar el ambiente.

- 13. Democracia.** En México, se realizó un estudio de las preferencias electorales y se obtuvieron los resultados siguientes:

- 45% se declararon “rojos”
- 35%, “azules”
- 20%, “amarillos”

Se sabe que 85% de los “rojos” realmente votará; 70% de los azules lo hará, y 58% de los amarillos irá a votar.

- Elabora el árbol de probabilidad respectivo.
- ¿Cuál es la probabilidad de que una persona, independientemente de su preferencia, realmente votará?
- Si una persona realmente vota, ¿cuál es la probabilidad de que fuera por el partido “azul”?
- Investiga qué es el Instituto Federal Electoral (IFE) y cuáles son sus funciones y escribe un informe con la información que obtengas. Preséntalo a tus compañeros, explicando por qué los anuncios del IFE afirman: “Lo que hace grande a un país es la participación de su gente”.

- 14. Salud.** El cáncer de pulmón consiste en un crecimiento anormal de las células de ese órgano. La enfermedad se produce habitualmente en las paredes internas de los bronquios, que al crecer puede obstruir el paso del aire y alterar la respiración. Por ello, produce generalmente falta de aire, ahogo y fatiga. El hábito de fumar tabaco es la causa principal del cáncer de pulmón, tanto en hombres como en mujeres (<http://www.elmundo.es/elmundosalud/especiales/cancer/pulmon5.html>). Se sabe que 40% de las personas fuma o ha fumado alguna vez en su vida. Además, se sabe que 90% de quienes fuman o han fumado presentan cáncer de pulmón y que 25% de las personas no fumadoras pueden desarrollar cáncer de pulmón. De acuerdo con esta información:

- a) Elabora el árbol de probabilidad respectivo.
 - b) Calcula la probabilidad de que una persona presente cáncer de pulmón, sea o no fumadora.
 - c) Si una persona desarrolla cáncer de pulmón, ¿cuál es la probabilidad de que haya sido fumadora?
 - d) ¿Cuál es la probabilidad de que una persona sea fumadora “pasiva” y desarrolle cáncer de pulmón?
 - e) ¿Crees que las campañas emprendidas en el país para alentar a la gente a que deje de fumar han sido eficaces? Explica tu respuesta. ¿Qué propondrías para lograr una influencia mayor en la población a este respecto y que un número más grande de personas dejara de fumar?
- 15. Sexualidad.** México registra altos porcentajes de embarazos en adolescentes. Entre las causas se encuentran la insuficiente educación sexual, la falta de disponibilidad de métodos anticonceptivos, la poca preparación en materia sexual de maestros y gobiernos locales que obstaculizan políticas públicas en la materia. De acuerdo con estudios realizados, 72.5% de las mujeres utilizan anticonceptivos. Se sabe que 23% de las mujeres que no emplearon algún método anticonceptivo se embarazan; en tanto que 14% de las mujeres que sí lo usaron resultaron embarazadas. Con base en estos datos:
- a) Elabora el árbol de probabilidad correspondiente.
 - b) ¿Qué porcentaje existe de embarazos?
 - c) Si una persona se embaraza, ¿cuál es la probabilidad de que haya sido por no utilizar algún método anticonceptivo?
 - d) Si una persona se embaraza, ¿cuál es la probabilidad de que haya sido así pese a haber utilizando algún método anticonceptivo?
 - e) Forma un equipo e investiga las ventajas y las desventajas de los métodos anticonceptivos que hay en el mercado. De acuerdo con los resultados que obtuvieron anteriormente, diseñen un cartel para evitar los embarazos no deseados.

Evaluación sumativa

Un escalón hacia arriba

Contesta los reactivos siguientes que te permitirán determinar lo que has aprendido. ¡Adelante!

1. Se sabe que

$$P(A - B) = 0.35, \quad P(A^c \cap B) = 0.15 \quad \text{y} \quad P(A^c) = 0.60$$

Con base en estos datos, determina:

- $P(A \cap B)$
 - $P(B)$
 - $P(A \cup B)$
2. En un estudio sobre tarjetas bancarias se analiza con qué frecuencia usan sus tarjetas de crédito o débito los consumidores jóvenes entre 18 y 24 años de edad. Sean los eventos:
- E : persona entre 18 y 24 años de edad
 - T : uso de tarjeta de crédito o débito

Los resultados del estudio son los siguientes:

	T	T^c	Total
E	7%	6%	13%
E^c	30%	57%	87%
Total	37%	63%	100%

- Si un consumidor usa su tarjeta, ¿cuál es la probabilidad de que sea mayor de 24 años?
 - ¿Los eventos E y T son independientes? Explica tu respuesta.
3. Una agencia distribuidora de automóviles tiene tres vendedores: Él señor Pérez (P) vende 20% del total de las unidades vendidas al mes, el señor Hernández (H), 30% y el señor Fernández (F), 50%. El 1% de los autos que P vende es del tipo de lujo; el 4% de los que vende H es de lujo y de los que vende F 2% es de lujo.
- Con base en estos datos:
- Elabora el árbol de probabilidad respectivo.
 - Si se vende un auto de lujo, ¿cuál es la probabilidad de que el vendedor haya sido H ?
 - ¿Cuál es la probabilidad de que se venda un auto de lujo?

Proyecto 6

SOBREPESO Y OBESIDAD

La obesidad y el sobrepeso se definen como una acumulación anormal o excesiva de grasa que puede resultar nociva para la salud. El índice de masa corporal (IMC) —el peso en kilogramos dividido entre el cuadrado de la talla en metros (kg/m^2)— es una indicación simple de la relación entre el peso y la talla que se utiliza frecuentemente para identificar el sobrepeso y la obesidad en los adultos, tanto en el plano individual como poblacional.

El IMC constituye la medida poblacional más útil del sobrepeso y la obesidad, pues la forma de calcularlo no varía en función del sexo ni de la edad en la población adulta. No obstante, debe considerarse como una guía aproximativa, pues puede no corresponder al mismo grado de obesidad en diferentes individuos.

La Organización Mundial de la Salud (OMS) define el sobrepeso como un IMC igual o superior a 25, y la obesidad como un IMC igual o superior a 30. Estos umbrales sirven de referencia para las evaluaciones individuales, pero hay pruebas de que el riesgo de enfermedades crónicas en la población aumenta progresivamente a partir de un IMC de 21.

Para mayor información consulta la página: <http://www.who.int/mediacentre/factsheets/fs311/es/index.html>

Supón que has sido contratado para llevar a cabo un estudio para la Secretaría de Salud de la población donde vives acerca del sobrepeso y la obesidad entre jóvenes de 15 a 20 años. Para ello:

- Realiza 15 encuestas a hombres y 15 a mujeres entre 15 y 20 años, en la que preguntarás su peso (en kilogramos, kg) y su estatura (en metros, m).
- Con la información obtenida, calcula el IMC correspondiente.
- Representa la información obtenida mediante una tabla de contingencia, un árbol de probabilidad o, mejor aún, con ambos. (*Sugerencia:* clasifica la información por género y por IMC, de acuerdo con la OMS.)
- Calcula porcentajes que consideres importantes para mostrar la problemática de sobrepeso y obesidad que hay entre jóvenes de 15 a 20 años.
- Escribe un informe breve con las conclusiones a las que llegues. En el informe has de contemplar estos aspectos:
 - a) Las causas de la obesidad y el sobrepeso.
 - b) Las repercusiones frecuentes del sobrepeso y la obesidad en la salud.
 - c) Cómo reducir la obesidad y sobrepeso.
 - d) Los beneficios/prejuicios sociales, de salud y económicos que acarrearán los problemas de obesidad y sobrepeso.



Competencias

- Expresa ideas y conceptos mediante representaciones lingüísticas y matemáticas relacionadas con las variables aleatorias y sus representaciones en distribuciones de probabilidad
- Sigue procedimientos de manera reflexiva, ordena información de acuerdo con categorías, jerarquías y las relaciona con base en las distribuciones de probabilidad, para las variables aleatorias discretas
- Estructura ideas y argumentos de manera clara, coherente y sintetiza los resultados de las distribuciones de probabilidad de variables aleatorias discretas manejando las tecnologías de la información para el análisis de resultados obtenidos en la representación tabular, gráfica y funciones de probabilidad
- Define metas y da seguimiento a sus procesos de construcción de conocimiento sobre las distribuciones de probabilidad de variables aleatorias discretas
- Propone maneras de resolver un problema o desarrollar un proyecto en equipo, definiendo un curso de acción con pasos específicos, asumiendo una actitud constructiva, congruente con los conocimientos y habilidades con los que cuenta dentro de distintos equipos de trabajo

7. Distribución de probabilidad de variables aleatorias discretas

Contenido

- Variables aleatorias discretas

Por cierto...

Blaise Pascal (Francia, 1623-1662) empezó sus investigaciones geométricas a la edad de 12 años y a los 16 ya había descubierto numerosos teoremas sobre proyecciones; a los 23 inventó la “pascalina”, la primera calculadora de la historia... Muchos resultados que hoy usamos llevan su nombre (el teorema de Pascal, el triángulo de Pascal...), el físico francés tuvo una influencia directa en el desarrollo de la matemática de su tiempo y nos dejó una muy sabia (y a veces desconcertante) colección de pensamientos.

Pero también se equivocó: “La excitación que un jugador siente cuando hace una apuesta es igual a la cantidad que puede ganar multiplicada por la probabilidad de ganarla”.

De ser así, la excitación en una partida de póquer apostando pequeñas cantidades sería muy baja... y en juegos sin apuestas la excitación sería siempre nula. A Pascal le faltó sumar a lo anterior la cantidad que se puede perder multiplicada por la probabilidad de perder. Así habría obtenido la verdadera esperanza matemática del apostador. (Alsina, 2008)

Ubícate

Antes de comenzar con el estudio del tema de este bloque, contesta este examen para revisar si tienes los conocimientos necesarios para comprenderlo cabalmente.

1. Calcula:

a) $5!$

b) ${}_7C_4$

2. Éste es el número de empleados por departamento en cierta empresa:

Departamento	A	B	C	D
Empleados	25	12	14	33

Si un empleado es elegido al azar, calcula la probabilidad de que:

a) Pertenezca al departamento A. b) Trabaje en el departamento B o C.

3. Tienes 5 plumones azules y 4 rojos. Si tomas 3 al azar, calcula la probabilidad de que:

a) Dos sean azules.

b) Cuando mucho uno sea azul.

4. Sean los datos:

0 -1 3 4 4 5 0

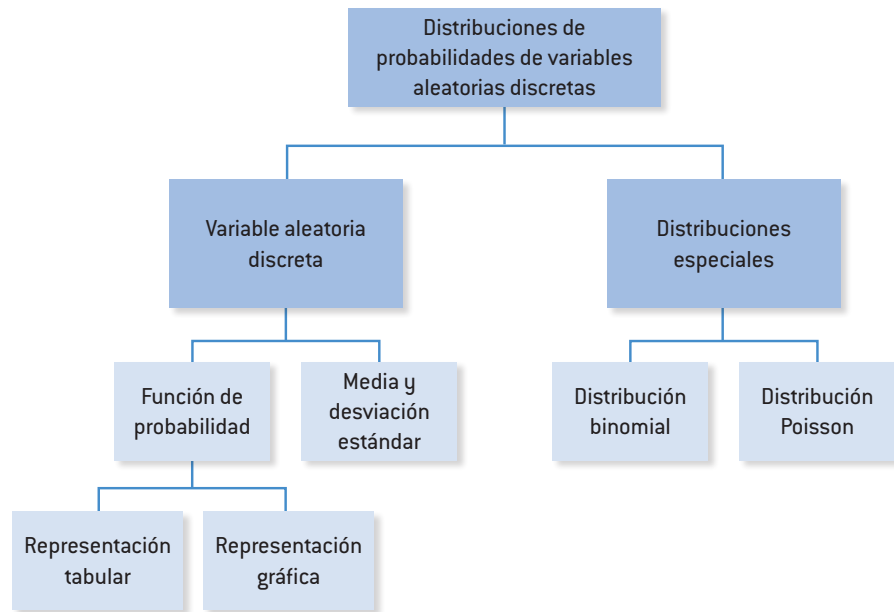
a) Calcula la media poblacional.

b) Calcula la varianza y la desviación estándar poblacionales.

Si no tuviste problemas o dudas con el examen anterior, ahora sí comienza con el estudio del bloque 7.

Distribución de probabilidades de variables aleatorias discretas

En este bloque proseguiremos el estudio de la probabilidad que empezamos en el anterior, pero ahora aprenderemos conceptos como el de variable aleatoria y distribuciones de probabilidad, así como a calcular la esperanza matemática, entre otras cosas. En el esquema siguiente se muestran los temas que abordaremos.



Desempeños

Al terminar de estudiar este bloque y realizar todas las actividades propuestas habrás aprendido a:

- Identificar y analizar los elementos de una variable aleatoria discreta para su aplicación al resolver problemas de diversos contextos.
 - Explicar las características de una representación de forma tabular, gráfica y funciones de probabilidad mediante la media y la desviación estándar, que te permita interpretar soluciones a problemas de tu entorno para tu autoevaluación.
-

Variable aleatoria discreta

Una *variable aleatoria* es un medio para describir los resultados del espacio muestral empleando números, es decir, es una función¹ que asigna valores numéricos a cada uno de los eventos del espacio muestral. Puede ser discreta o continua y se acostumbra representarla con una X . Cabe señalar que estudiaremos las variables aleatorias continuas en el próximo bloque; en éste centraremos el análisis en las variables aleatorias discretas.

Como ejemplos de variables aleatorias tenemos:

- Número de veces que sale 6 al lanzar un dado dos veces.
- El tiempo que requieres para hablar por teléfono celular con tu mejor amig@: los valores serían de 0 a k , donde k es el tiempo que tengas disponible, de que dure la batería o de que tu amig@ resista.

Las *variables aleatorias discretas* son aquellas en que el número de valores posibles que toman es finito o infinito, pero que puede numerarse.

En la tabla siguiente se muestran algunos ejemplos de variables aleatorias discretas:

Ejemplo 1

Experimento	Variable aleatoria (X)	Valores posibles de X
Lanzar una moneda cuatro veces	Contar el número de águilas que salen	0, 1, 2, 3, 4
Realizar el control de calidad de 10 teléfonos celulares	Contar el número de teléfonos celulares que presentan algún defecto	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Jugar un partido de fútbol	Contar el número de goles que anota el equipo contrario	0, 1, 2, 3, 4, ...
Concierto de Rammstein en el Palacio de los Deportes	Contar el número de asistentes	0, 1, 2, 3, ..., 27 999, 28 000

Distribución de probabilidad

Una vez que determinamos los valores de la variable aleatoria, les asignamos una probabilidad, y a esto se le llama *distribución de probabilidad*.

Las propiedades que debe cumplir una distribución de probabilidad son:

- a. $0 \leq P(X_i) \leq 1$
- b. $\sum_{i=1}^n P(X_i) = 1$

donde X_i es el valor de la variable aleatoria y $P(X_i)$ es la probabilidad asociada al valor X_i .

Una distribución de probabilidad puede representar:

- En forma tabular: es una tabla donde se muestran en una columna todos los posibles valores de la variable aleatoria y en otra columna las probabilidades asociadas a cada valor.
- En forma gráfica: es una gráfica de barras cuyas alturas corresponden a la probabilidad de cada valor de la variable aleatoria.
- En forma de función: cuando la distribución de probabilidad se define mediante una fórmula. A esta expresión se le llama *función de probabilidad*.

¹ Recuerda que una función es una relación entre dos conjuntos, el dominio y el rango, con la característica de que a un elemento del dominio le corresponde uno solo del rango.

Ejemplo 2

Se lanzan dos monedas “justas” y se cuenta el número de “soles”. Determina la función de probabilidad para este experimento y traza la gráfica correspondiente.

Solución

Sea S : sol y A : águila.

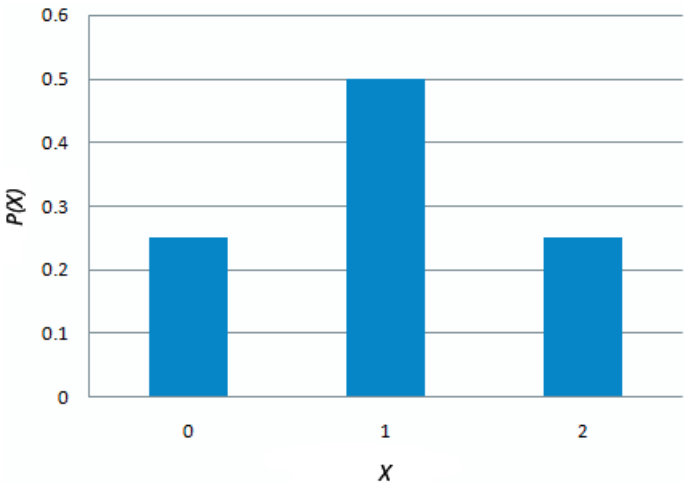
El espacio muestral asociado a este experimento es

$$\{ \{A, A\}, \{S, A\}, \{A, S\} \{S, S\} \}$$

► Distribución de probabilidad y representación gráfica:

Resultados posibles	X	$P(X)$
$\{A, A\}$	0	$\frac{1}{4} = 0.25$
$\{S, A\}, \{A, S\}$	1	$\frac{2}{4} = 0.50$
$\{S, S\}$	2	$\frac{1}{4} = 0.25$

Observa que cada una de las probabilidades están entre 0 y 1 y que si sumas todas las probabilidades, el resultados es 1.



Ejemplo 3

Dada la tabla siguiente:

X	$P(X=x)$
0	0.20
1	0.25
2	0.15
3	0.40

- a. Comprueba si es una distribución de probabilidad.
- b. Calcula $P(X=1)$
- c. Calcula $P(X<2)$
- d. Calcula $P(X\geq 2)$

Solución

- a. Para saber si es una distribución de probabilidad debemos comprobar que se cumplen las propiedades:

- Cada una de las probabilidades están entre cero y uno.
- Si sumamos las probabilidades $(0.20 + 0.25 + 0.15 + 0.40)$, el resultado es 1.

Por tanto, sí es una distribución de probabilidad.

- b. Buscamos en la tabla: $P\{X = 1\} = 0.25$
- c. Para calcular $P\{X < 2\}$, primero determinamos los valores de X que son menores a 2, es decir, $X = 0, 1$. Como son eventos independientes, sumamos las probabilidades asociadas como se muestra a continuación:

$$P\{X < 2\} = P\{X = 0\} + P\{X = 1\} = 0.20 + 0.25 = 0.45$$

Este cálculo recibe el nombre de *probabilidad acumulada*.

- d. Para calcular $P\{X \geq 2\}$ disponemos de dos formas:

- i. Determinamos los valores de X mayores o iguales a 2 y sumamos sus probabilidades:

$$P\{X \geq 2\} = P\{X = 2\} + P\{X = 3\} = 0.15 + 0.40 = 0.55$$

- ii. Hacemos un “truco” con el concepto de complemento:

$$P\{X \geq 2\} = 1 - P\{X < 2\} = 1 - 0.45 = 0.55$$

Observa que
 $P\{X \geq 2\} + P\{X < 2\} = 1$.
 O bien,
 $P\{X > 2\} + P\{X \leq 2\} = 1$.

Se lanza un dado “justo” y se apunta el número que se observa en la cara superior. De acuerdo con esto:

Ejemplo 4

- a. Determina la distribución de probabilidad.
- b. Calcula la probabilidad de que el número sea un número par.
- c. Calcula la probabilidad de que el número sea menor que 3.

Solución

Como es un dado “justo”, cada una de sus caras tiene la misma probabilidad de “salir”. Entonces, X es el número que aparece en la cara superior, donde $X = 1, 2, 3, 4, 5, 6$.

- a. La distribución de probabilidad queda de esta forma:

X	$P\{X = x\}$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

b. La probabilidad de que sea un número par se calcula así:

$$P(X \text{ sea par}) = P(X = 2 \text{ o } 4 \text{ o } 6) = P(X = 2) + P(X = 4) + P(X = 6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 0.5$$

c. La probabilidad de que el número sea menor que 3 se determina de esta manera:

$$P(X < 3) = P(X = 1 \text{ o } 2) = P(X = 1) + P(X = 2) = \frac{1}{6} + \frac{1}{6} = 0.3333$$

Media y desviación estándar

En lenguaje “probabilístico” la media se conoce como *el momento 1* y la varianza como *el momento 2* de una variable aleatoria. Cabe recordar que también son características de una distribución.

Como recordarás, en el bloque 3, “Medidas de tendencia central y de dispersión” estudiamos las medidas de tendencia central, como la media, y las medidas de dispersión, como la varianza de un conjunto de datos. Entonces, al hablar de una distribución de probabilidad definiremos la media y la desviación estándar como se explica en los apartados siguientes.

Media o valor esperado

La media o valor esperado es una medida de localización central de la variable aleatoria. Se calcula de la manera siguiente:

$$\mu = E(X) = \sum_{i=1}^n X_i P(X_i)$$

es decir, es un promedio ponderado de los valores que toma la variable aleatoria (X_i), donde las ponderaciones (o pesos) son las probabilidades [$P(X_i)$].

Varianza y desviación estándar

La varianza y la desviación estándar son medidas de dispersión o variabilidad de la variable aleatoria. La primera se calcula de la siguiente manera:

$$\sigma^2 = Var(X) = \sum_{i=1}^n P(X_i)(X_i - \mu)^2$$

En otras palabras, es el promedio ponderado de las distancias al cuadrado de cada valor de la variable aleatoria respecto a su media. Esto indica cuán alejados están los valores en relación con la media. Se eleva al cuadrado para hacer más evidente esa distancia (variabilidad).

También puede calcularse así:

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2$$

Como el resultado de la varianza se expresa en unidades al cuadrado, determinamos la desviación estándar calculando sólo la raíz cuadrada de la varianza.

Ejemplo 5

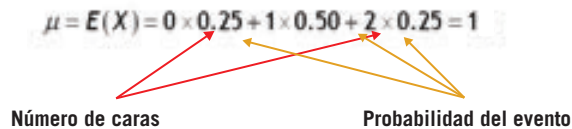
Para seguir con el ejemplo 2, calcula la media y la desviación estándar del número de soles que pueden caer cuando se lanzan dos monedas.

Solución

La distribución de probabilidad para este ejemplo es:

X	$P(X)$
0	$\frac{1}{4} = 0.25$
1	$\frac{2}{4} = 0.50$
2	$\frac{1}{4} = 0.25$

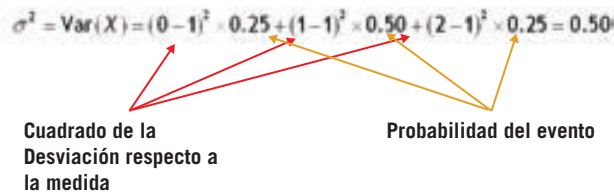
Con base en esta información, calculamos la media y la desviación estándar. Primero la media:

$$\mu = E(X) = 0 \times 0.25 + 1 \times 0.50 + 2 \times 0.25 = 1$$


Número de caras

Probabilidad del evento

Ahora la varianza:

$$\sigma^2 = \text{Var}(X) = (0-1)^2 \times 0.25 + (1-1)^2 \times 0.50 + (2-1)^2 \times 0.25 = 0.50$$


Cuadrado de la Desviación respecto a la medida

Probabilidad del evento

O bien, utilizamos otra fórmula para la varianza donde se calcula la esperanza de las X al cuadrado:

$$E(X^2) = \sum X_i^2 P(X_i) = 0^2 \times 0.25 + 1^2 \times 0.5 + 2^2 \times 0.25 = 1.5$$

De esta manera,

$$\sigma^2 = \text{Var}(X) = E(X^2) - [E(X)]^2 = 1.5 - (1)^2 = 0.5$$

Por tanto, la desviación estándar es

$$\sigma = \sqrt{0.50} = 0.7071$$

A continuación revisaremos dos distribuciones especiales de probabilidad; se les llama *especiales* porque son de las más comunes en la vida real.

Distribución binomial

Antes de definir la distribución binomial debemos explicar lo que se conoce como *ensayo de Bernoulli*. Un ensayo de Bernoulli es aquel que sólo tiene dos resultados: éxito-fracaso, falso-verdadero, sí-no, 0-1, cumple-no cumple.

Por ejemplo, lanzar una moneda para saber qué caerá es un ensayo de Bernoulli, ya que los únicos resultados posibles son “águila” o “sol”. Otro ejemplo es el control de calidad de cierto producto, pues el artículo cumple o no las especificaciones, es decir, sólo tiene dos resultados posibles.

Ahora bien, hablamos de una *distribución binomial* cuando se satisfacen las características siguientes:

- n ensayos de Bernoulli, es decir, se repite muchas veces un ensayo que tiene sólo dos respuestas.
- Los ensayos son independientes, esto es, el resultado de un ensayo no implica el resultado de otro.
- La probabilidad de cada ensayo es constante. Aquí se habla de probabilidad de éxito, y se entiende por éxito lo que queremos analizar.

Función de probabilidad

La función de probabilidad para una variable aleatoria X que se distribuye como una binomial es:

$$P(X = x) = {}_nC_x p^x q^{n-x}$$

donde $P(X = x)$ es la probabilidad de x éxitos dados n y p .

X es el número de éxitos, con $X = 0, 1, 2, \dots, n$.

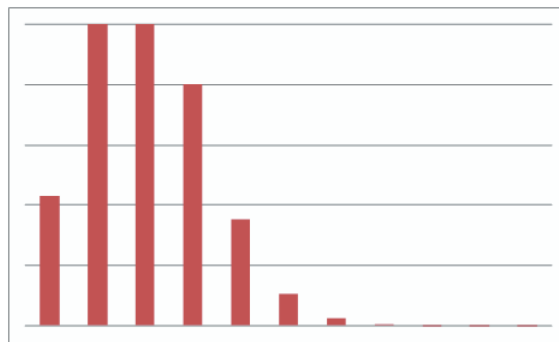
p es la probabilidad de éxito.

q es la probabilidad de fracaso y se calcula como $1 - p$.

n es el número de ensayos de Bernoulli.

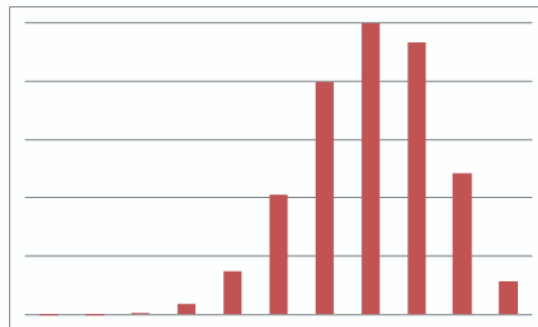
La gráfica de una distribución binomial, por ser una función discreta, se compone de columnas, donde la altura representa la probabilidad asociada a cada valor de X . Pueden presentar sesgo o ser simétricas, según el valor de p :

- Si $p < 0.5$, la gráfica está sesgada a la derecha:



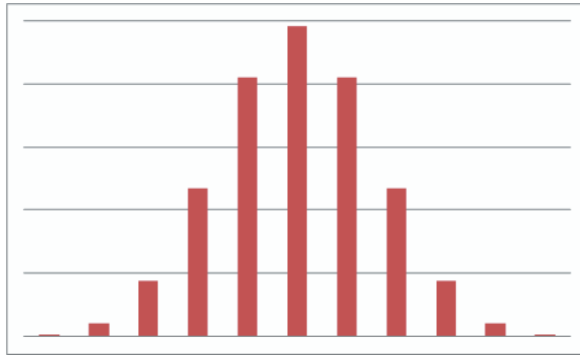
$p < 0.5$, por lo que hay sesgo a la derecha.

- Si $p > 0.5$, la gráfica está sesgada a la izquierda:



$p > 0.5$, por lo que hay sesgo a la izquierda.

- Si $p = 0.5$, la gráfica es simétrica, es decir, no presenta sesgo:



$p = 0.5$, por lo que no hay sesgo y la gráfica es simétrica.

Se sabe que X es una variable aleatoria que sigue una distribución binomial. Con $n = 5$ y $p = 0.4$, determina:

- $P(X = 1)$
- $P(X < 2)$
- $P(X > 2)$

Ejemplo 6

Solución

- Aplicando la función de distribución binomial:

$$P(X = 1) = {}_5C_1 (0.4)^1 (0.6)^{5-1} = 0.2592$$

Para realizar este cálculo en Excel, se utiliza la función `DISTR.BINOM()` de la forma siguiente:

INTERSECCION.EJE					
	A	B	C	D	E
1	=DISTR.BINOM(1,5,0.4,FALSO)				
2	DISTR.BINOM(núm_éxito, ensayos, prob_éxito, acumulado)				

donde *núm_éxito* es el valor de X , del cual nos interesa calcular la probabilidad.
ensayos es el valor de n .
prob_éxito es la probabilidad de éxito p .
acumulado se anota "verdadero" si se quiere obtener la probabilidad acumulada, es decir, $P(X \leq x)$, o "falso" si sólo se desea calcular $P(X = x)$.

-

$$\begin{aligned}
 P(X < 2) &= P(X = 0) + P(X = 1) \\
 &= {}_5C_0 (0.4)^0 (0.6)^{5-0} + {}_5C_1 (0.4)^1 (0.6)^{5-1} \\
 &= 0.0778 + 0.2592 = 0.3370
 \end{aligned}$$

En Excel, con la función `DISTR.BINOM()` tenemos:

Observa que $P(X < 2)$ puede calcularse de dos formas:

- Como $P(X = 0) + P(X = 1)$
- O bien, $P(X \leq 1)$. Es aquí cuando podemos utilizar la opción Acumulado en la función de Excel.

```
=DISTR.BINOM(1,5,0.4,VERDADERO)
```

```
DISTR.BINOM(núm_éxito, ensayos, prob_éxito, acumulado)
```

c. Se puede calcular de manera directa:

$$\begin{aligned} P(X > 2) &= P(X = 3) + P(X = 4) + P(X = 5) \\ &= {}_5C_3(0.4)^3(0.6)^{5-3} + {}_5C_4(0.4)^4(0.6)^{5-4} + {}_5C_5(0.4)^5(0.6)^{5-5} \\ &= 0.2304 + 0.0768 + 0.0102 = 0.3174 \end{aligned}$$

O bien, con el complemento:

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \\ &= 1 - [{}_5C_0(0.4)^0(0.6)^{5-0} + {}_5C_1(0.4)^1(0.6)^{5-1} + {}_5C_2(0.4)^2(0.6)^{5-2}] \\ &= 1 - (0.0778 + 0.2592 + 0.3456) = 1 - 0.6826 = 0.3174 \end{aligned}$$

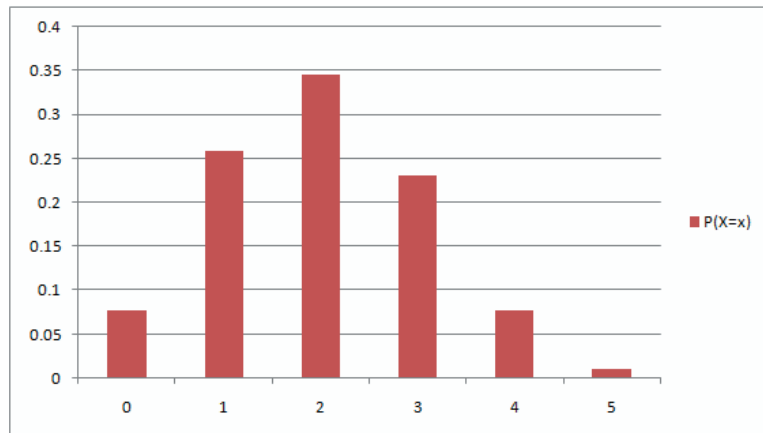
Por último, para Excel:

```
=1-DISTR.BINOM(1,5,0.4,VERDADERO)
```

```
DISTR.BINOM(núm_éxito, ensayos, prob_éxito, acumulado)
```

Se presentan a continuación tanto la distribución de probabilidad como la gráfica correspondiente a este problema:

X	$P(X = x)$
0	0.0778
1	0.2592
2	0.3456
3	0.2304
4	0.0768
5	0.0102



Ejemplo 7

El equipo Pumas de la UNAM fue líder general del Torneo de clausura 2011 de fútbol hasta la jornada 16. En esa jornada, Pumas enfrentó al Necaxa y, al ver sus estadísticas, tenemos que de ocho tiros al arco, anotó un gol. De acuerdo con esto, si en el partido siguiente dispara seis veces a la portería:

- ¿Cuál es la probabilidad de que Pumas anote dos goles?
- ¿Cuál es la probabilidad de que anote máximo dos goles?

Solución

Este problema es binomial, ya que al tirar a la portería tenemos sólo dos posibilidades: es gol o no. Sean entonces:

- X : la variable aleatoria del número de goles anotados
- p : la probabilidad de anotar un gol; en este caso, $p = \frac{1}{8}$.
- n : número de ensayos; aquí, $n = 6$

Así, tenemos lo siguiente.

$$\text{a. } P(X=2) = {}_6C_2 [0.125]^2 [0.875]^{6-2} = 0.1374$$

b.

$$\begin{aligned} P(X \leq 2) &= P(X=0) + P(X=1) + P(X=2) \\ &= {}_6C_0 [0.125]^0 [0.875]^{6-0} + {}_6C_1 [0.125]^1 [0.875]^{6-1} + {}_6C_2 [0.125]^2 [0.875]^{6-2} \\ &= 0.4488 + 0.3847 + 0.1374 = 0.9709 \end{aligned}$$

Media y desviación estándar

Mediante las definiciones de media y desviación estándar, tenemos los conceptos siguientes.

MEDIA O VALOR ESPERADO

$$\mu = E(X) = \sum_{i=1}^n X_i P(X_i) = \sum_{i=1}^n X_i \left({}_n C_{x_i} p^{x_i} q^{n-x_i} \right) = np$$

Dicho con palabras, el valor esperado de una distribución binomial es el número de ensayos por la probabilidad de éxito.

VARIANZA Y DESVIACIÓN ESTÁNDAR

$$\sigma^2 = Var(X) = \sum_{i=1}^n P(X_i) (x_i - \mu)^2 = \sum_{i=1}^n \left({}_n C_{x_i} p^{x_i} q^{n-x_i} \right) (X_i - \mu)^2 = npq$$

Entonces, la desviación estándar de una distribución binomial se calcula como la raíz cuadrada del número de ensayos por la probabilidad de éxito por la probabilidad de fracaso.

Se sabe que X es una variable aleatoria que sigue una distribución binomial, con $n = 5$ y $p = 0.4$. Determina su media y desviación estándar.

Ejemplo 8

Solución

- Media: $= np = 5(0.4) = 2$

$$\text{Desviación estándar: } \sigma = \sqrt{npq} = \sqrt{5(0.4)(0.6)} = 1.0954$$

Distribución de Poisson

La distribución de Poisson es otra de las distribuciones “especiales”, es decir, de las más comunes en la vida real. Sirve para representar el número de eventos de poca frecuencia (a veces se les llama *eventos raros*) que ocurren en el tiempo o en el espacio.

Los siguientes son ejemplos de distribución de Poisson:

- El número de mensajes recibidos en un teléfono celular en una hora.

- El número de automóviles que llega a un estacionamiento entre las 7 y las 9 de la mañana.
- El número de faltas ortográficas por capítulo.
- El número de manchas por metro cuadrado de tela.

En una distribución de Poisson necesitamos el número promedio de eventos que ocurren en un intervalo de tiempo o espacio.

Función de probabilidad

La función de probabilidad de una distribución de Poisson es:

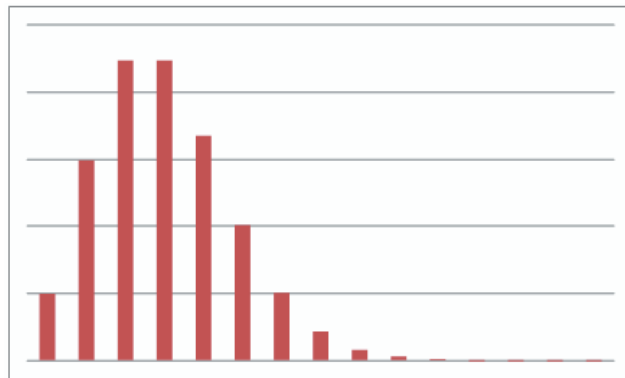
$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

donde $P(X = x)$ es la probabilidad de x ocurrencias, dado λ , y $x = 0, 1, 2, 3, 4, \dots$

λ es el número promedio de eventos que ocurren por unidad de tiempo o de espacio.

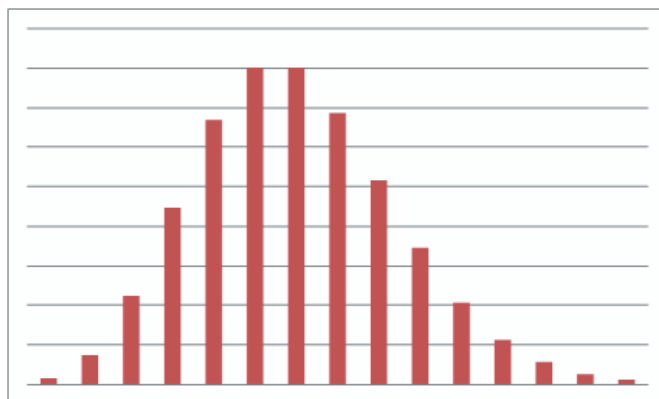
Igual que la distribución binomial, la gráfica de una distribución de Poisson se compone de columnas, cuya altura representa la probabilidad asociada a cada valor de X . Pueden presentar sesgo o ser simétricas, según el valor de λ :

- Si $\lambda < 6$, la gráfica está sesgada a la derecha:



Como $\lambda < 6$, la gráfica muestra sesgo a la derecha.

- Si λ es cercana a 6, la gráfica es casi simétrica:



Puesto que λ es cercana a 6, la gráfica es casi simétrica.

Se sabe que X es una variable aleatoria que sigue una distribución de Poisson con $\lambda = 3$. Determina:

Ejemplo 9

- $P(X = 2)$
- $P(X < 3)$
- $P(X > 3)$

Solución

- De acuerdo con la función de distribución de Poisson:

$$P(X = 2) = \frac{e^{-3} 3^2}{2!} = 0.2240$$

Para calcular esta probabilidad en Excel se usa la función `POISSON()`:

INTERSECCION.EJE					
	A	B	C	D	E
1	=POISSON(2,3,FALSO)				
2	POISSON(x, media, acumulado)				

donde X es el valor del cual nos interesa calcular la probabilidad.
media es el valor de λ , es decir, del número promedio de eventos por unidad de tiempo o espacio.
acumulado se introduce “verdadero” si se quiere obtener la probabilidad acumulada, es decir, $P(X \leq x)$; o “falso” si sólo se desea calcular $P(X = x)$.

En el caso de una distribución de Poisson, cuando tengamos $P(X > x)$ o $P(X \geq x)$ debemos aplicar siempre el “truco” del complemento. Esto es porque no tenemos un “tope” superior como en el caso de una binomial (n). En teoría, el número de eventos por unidad de tiempo o de espacio puede ser infinito.

-

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= \frac{e^{-3} 3^0}{0!} + \frac{e^{-3} 3^1}{1!} + \frac{e^{-3} 3^2}{2!} = 0.0498 + 0.1494 + 0.2240 = 0.4232$$

O bien, en Excel:

=POISSON(2,\$E\$2,VERDADERO)
POISSON(x, media, acumulado)

-

$$P(X > 3) = 1 - P(X \leq 3)$$

$$= 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]$$

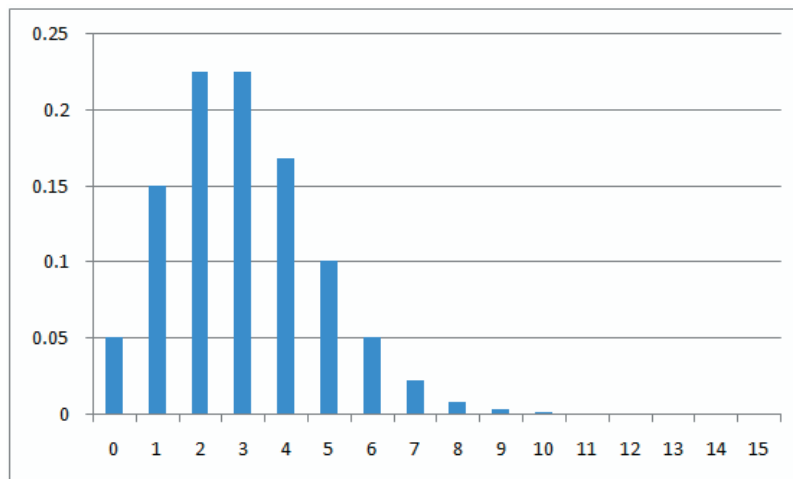
$$= 1 - \left[\frac{e^{-3} 3^0}{0!} + \frac{e^{-3} 3^1}{1!} + \frac{e^{-3} 3^2}{2!} + \frac{e^{-3} 3^3}{3!} \right]$$

$$= 1 - [0.0498 + 0.1494 + 0.2240 + 0.2240]$$

$$= 0.3528$$

X	$P(X)$
0	0.0498
1	0.1494
2	0.2240
3	0.2240
4	0.1680
5	0.1008
6	0.0504
7	0.0216
8	0.0081
9	0.0027
10	0.0008
11	0.0002
12	0.0001
13	0.0000
14	0.0000
15	0.0000

Se presentan a continuación la distribución de probabilidad y la gráfica de este problema:



Ejemplo 10

Volvamos con los Pumas de la UNAM. En el Torneo de clausura 2011, hasta la jornada 16, en promedio, anotaron 1.69 goles por cada 90 minutos. De acuerdo con esto:

- ¿Cuál es la probabilidad de que Pumas anote dos goles en los próximos 90 minutos?
- ¿Cuál es la probabilidad de que anote cuando mucho un gol durante el primer tiempo?

Solución

Se trata de un caso típico de distribución de Poisson, ya que tenemos la información del número promedio de eventos por unidad de tiempo, que es de 90 minutos.

a. Sean:

- X : la variable aleatoria del número de goles anotados en 90 minutos
- $\lambda = 1.69$ goles en promedio por 90 minutos

Sustituyendo valores queda:

$$P(X = 2) = \frac{e^{-1.69} 1.69^2}{2!} = 0.2635$$

b. Para resolver este inciso debemos observar que la unidad de tiempo cambió. Para eso calculamos la λ correspondiente. Sean

- X : la variable aleatoria del número de goles anotados en 45 minutos
- λ : 0.845 goles en promedio por 45 minutos

Sustituyendo,

$$P(X \leq 1) = \frac{e^{-0.845} 0.845^0}{0!} + \frac{e^{-0.845} 0.845^1}{1!} = 0.4296 + 0.3630 = 0.7925$$

Para calcular esta nueva λ sólo debemos aplicar una regla de tres:

$$\begin{array}{l} 1.69 \text{ goles} - 90 \text{ min} \\ \lambda - 45 \end{array}$$

Y listo. Esto es porque se supone que los eventos, durante la unidad de tiempo dada, se distribuyen de manera uniforme.

Media y desviación estándar

Con las definiciones de media y desviación estándar tenemos los conceptos siguientes.

MEDIA O VALOR ESPERADO

$$\mu = E(X) = \sum_{i=1}^n X_i P(X_i) = \sum_{i=1}^n X_i \left(\frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \right) = \lambda$$

Dicho con palabras, el valor esperado de una distribución de Poisson es el número promedio de eventos por unidad de tiempo o espacio.

VARIANZA Y DESVIACIÓN ESTÁNDAR

$$\sigma^2 = Var(X) = \sum_{i=1}^n (X_i)(X_i - \mu)^2 = \sum_{i=1}^n \left(\frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \right) = (X_i - \mu)^2 = \lambda$$

Entonces, la desviación estándar de una distribución de Poisson se calcula como la raíz cuadrada del número promedio de eventos por unidad de tiempo o espacio.

La distribución de Poisson tiene esa particularidad: su media y su varianza son iguales.

Se sabe que X es una variable aleatoria que sigue una distribución de Poisson con $\lambda = 3$. Determina su media y su desviación estándar.

Ejemplo 11

Solución

- Media: $\mu = \lambda = 3$
- Desviación estándar: $\sigma = \sqrt{\lambda} = \sqrt{3} = 1.7321$

Aproximación de la distribución binomial a la distribución de Poisson

Sea una variable aleatoria que se distribuye como una binomial y si se cumplen las tres condiciones siguientes:

1. n es bastante grande ($n \rightarrow \infty$)
2. p es pequeña ($p \rightarrow 0$)
3. $np < 7$

entonces la variable aleatoria puede aproximarse a una distribución de Poisson con $\lambda = np$.

Se utiliza esta aproximación pues resulta más sencillo trabajar mediante Poisson que con una binomial.

Sea X una variable aleatoria distribuida como binomial, con $n = 600$ y $p = 0.002$. Con esta información construye la distribución de probabilidad de la binomial y la de Poisson.

Ejemplo 12

Solución

Como n es grande, p pequeña y $np = 1.2 < 7$, se puede trabajar como una distribución de Poisson. Entonces,

Cuanto más grande sea n y más pequeño p , la aproximación será mucho mejor.

	Distr. binomial	Distr. Poisson
X	$P(X=x)$	$P(X=x)$
0	0.3008	0.3012
1	0.3617	0.3614
2	0.2171	0.2169
3	0.0867	0.0867
4	0.0259	0.0260
5	0.0062	0.0062
6	0.0012	0.0012
7	0.0002	0.0002
8	0.0000	0.0000
9	0.0000	0.0000
10	0.0000	0.0000
$\vdots$		
599	0.0000	0.0000
600	0.0000	0.0000

Ejercicios

I. TEORÍA Y ALGO MÁS

1. Contesta verdadero o falso según corresponda en los enunciados siguientes.
 - a) Una variable aleatoria es cuando se le asigna un número a cada evento del espacio muestral.
 - b) Una variable aleatoria no puede ser discreta.
 - c) Una distribución de probabilidad es una tabla donde se muestran los valores de la variable aleatoria, así como las probabilidades respectivas.
 - d) Las propiedades de una distribución discreta son que la probabilidad de cada valor de la variable aleatoria toma valores entre 0 y 1, y que la suma de todas ellas siempre es 1.
 - e) Un ensayo de Bernoulli tiene al menos dos resultados posibles.
 - f) Una distribución binomial surge, en principio, de repetir n veces un ensayo de Bernoulli.
 - g) Si la probabilidad de éxito es pequeña o el número de ensayos es muy grande, entonces una variable con distribución binomial se puede modelar como una de Poisson.
 - h) A p se le conoce como *probabilidad de éxito*, ya que se refiere a la ocurrencia de eventos positivos o agradables.
 - i) Una distribución de Poisson sirve para determinar la ocurrencia de eventos en un tiempo o espacio dado.
 - j) En una distribución de Poisson, la media y la desviación estándar tienen el mismo valor.

II. PIENSO, ... ¿Y LUEGO?

2. El *sexting* es una forma de relación virtual entre adolescentes y jóvenes por medio de redes sociales de internet. Este fenómeno consiste en que los(as) chicos(as) se toman fotografías o videos con teléfono celular o cámara, semi o totalmente desnudos(as), en poses atractivas, para enviarlas por mensaje de texto o correo electrónico a contactos personales.

La comunidad *flogger* (usuarios de las redes sociales) define el *sexting* como un recurso para tener popularidad.

Según indicó “una encuesta en internet realizada para The Associated Press y MTV por la firma Knowledge Networks, en septiembre de 2009 24% de los adolescentes de 14 a 17 años en Estados Unidos había estado involucrado en algún tipo de *sexting*. Un sondeo telefónico, celebrado en diciembre de 2009, del Proyecto de Internet y Vida Estadounidense del Centro de Investigación Pew, arrojó que 5% de los estadounidenses de 14 a 17 años había enviado fotos o videos desnudos o casi desnudos”. (Hoffman, 2011).

- a) ¿Crees que esta situación es similar en México? ¿Qué consecuencias en el corto y en el largo plazos genera enviar fotos “sugereentes” por alguna red social?
- b) Si estuvieras en una fiesta a la que asisten 30 chicos y chicas, ¿cuántos de ellos crees que hayan enviado una foto donde aparezcan desnudos o semidesnudos? ¿Cuál sería la probabilidad de que cuando mucho tres hayan mandado una foto de este tipo?

III. PURO DEPORTE

3. *Sólo para practicar*. Se lanzan dos dados “justos” y se suman los valores que aparecen en la cara superior.
 - a) Indica el espacio muestral para este experimento.
 - b) Construye la tabla de distribución correspondiente y traza su gráfica.

4. *Sólo para practicar.* A continuación se presentan tres supuestas distribuciones de probabilidad:

(i)		(ii)		(iii)	
X	P(X)	X	P(X)	X	P(X)
0	0.2	0	0.3	0	0.2
1	0.1	1	0.4	1	0.3
2	-0.3	2	0.1	2	0.1
3	0.4	3	0.3	3	0.4

- a) Para cada una de ellas, explica si satisfacen las propiedades de toda distribución de probabilidad discreta.
- b) De la que realmente es una distribución de probabilidad, calcula la media y la desviación estándar.
5. *Negocios.* A continuación se presentan las utilidades estimadas del próximo año para una empresa de la industria del vestido:

X (en mill. de pesos)	-6	-3	0	5	10
P(X)	0.15	0.25	0.3	?	?

- a) Completa la tabla considerando que la probabilidad de obtener una utilidad de \$10 000 000 es el doble de la probabilidad de ganar \$5 000 000.
- b) ¿Cuál es la probabilidad de que la empresa sea rentable?
- c) ¿Cuál es la ganancia esperada en el primer año?
6. *Negocios.* El Café de la Sirena ha visto que 22% de los pedidos son frapuccinos, 25% capuccinos, 15% latte, 25% moka y 13% té. Los precios de las bebidas en tamaño grande son \$36, \$30, \$30, \$35 y \$34, respectivamente.
- a) Determina la distribución de probabilidad para la venta de bebidas en el Café de la Sirena.
- b) Calcula la cantidad promedio que se cobra por una bebida.
- c) ¿Cuál es la varianza y la desviación estándar de la cantidad que se cobra por una bebida?
7. *Seguridad.* Según un anuncio comercial de colchones, 30% de los accidentes automovilísticos se deben a la falta de sueño. Supón que en un día cualquiera ocurren 10 accidentes automovilísticos.
- a) Determina la distribución de probabilidad y traza la gráfica correspondiente.
- b) ¿Qué tipo de sesgo presenta la distribución?
- c) Calcula la probabilidad de que cinco de los accidentes se hayan debido a la falta de sueño.
- d) Calcula la probabilidad de que al menos dos accidentes hayan sido ocasionados por la falta de sueño.
- e) ¿Cuántos accidentes esperarías que fueran provocados por falta de sueño?
- f) Investiga qué otras consecuencias fatales tiene la falta de sueño. Junto con tus compañeros de clase expongan las posibles medidas para prevenir accidentes fatales.
8. *Redes sociales.* En México, 6% de los 30 millones de usuarios de internet que hay utilizan *Twitter*. (Torres, 2010). Si conoces a 15 personas que sean asiduas a las redes sociales:
- a) ¿Cuál es la probabilidad de que 9 o más utilicen *Twitter*?
- b) ¿Cuál es la probabilidad de que máximo dos usen esa red?
- c) Organiza con tus compañeros de clase una mesa redonda y expongan cuáles son las ventajas y desventajas de usar *Twitter*. Además, hagan una lista sobre los temas que comunican a través de dicha red.

9. *Seguros.* De acuerdo con la Asociación Mexicana de Instituciones de Seguros (AMIS), en 2010 sólo 47% de los automóviles contaban con un seguro. Si se realizara una revisión a 18 automóviles:
 - a) ¿Cómo sería su distribución? Determina la distribución de probabilidad, traza la gráfica correspondiente e indica el tipo de sesgo que presenta.
 - b) ¿Cuántos autos esperas que cuenten con seguro?
 - c) ¿Cuál es la probabilidad de que 13 o más autos tengan seguro?
 - d) Investiga cuáles son las coberturas principales que tiene un seguro para automóviles. Realiza una discusión con tus compañeros sobre la importancia de contar con seguro.
10. *Sexualidad.* Con base en la Encuesta Nacional de Juventud 2005, 90% de los adolescentes entre 15 y 19 años afirmó que tuvo su primera relación sexual antes de los 18 años. Si en un grupo hay 20 adolescentes entre 15 y 19 años, entonces:
 - a) ¿Cuántos esperarías que contestaran que tuvieron su primera relación sexual antes de los 18 años?
 - b) ¿Cuál es la probabilidad de que 12 de ellos digan que tuvieron su primera relación sexual antes de los 18 años?
 - c) ¿Cuál es la probabilidad de que más de 17 digan que tuvieron su primera relación sexual antes de los 18 años?
 - d) Con tus compañeros de clase lleven a cabo una discusión sobre:
 - Los diferentes métodos anticonceptivos que hay en el mercado.
 - Las enfermedades de transmisión sexual (ETS).
 - Las consecuencias de no utilizar algún método anticonceptivo.
11. *Servicios.* De acuerdo con los datos proporcionados por la Secretaría de Salud en 2010, 34% de la población utiliza los servicios de salud públicos federales o estatales. Se quiere realizar una encuesta a 50 personas sobre la calidad de los servicios de salud. Encuentra:
 - a) ¿Qué proporción de las personas esperas que asistan a los servicios de salud públicos?
 - b) ¿Qué probabilidad existe de que 30 personas utilicen los servicios que brinda el gobierno?
 - c) ¿Qué probabilidad hay de que cuando mucho 10 utilicen estos servicios?
 - d) ¿Alguna vez has recurrido a los servicios públicos de salud? Con tus compañeros de clase comenten sus experiencias y realicen una lluvia de ideas para mejorar los servicios de salud.
12. *Psicología.* Según encuestas realizadas, uno de cada cuatro adolescentes estaría dispuesto a someterse a una cirugía plástica para parecerse a alguna personalidad como David Beckham, Usher o Justin Bieber. Supón que 15 personas asisten a una reunión,
 - a) ¿Cuál es la probabilidad de que al menos 10 de ellas quisieran parecerse a alguna personalidad?
 - b) ¿Cuál es la probabilidad de que exactamente cinco quisieran practicarse una cirugía plástica?
 - c) Según estudios psicológicos, esta situación es muy común entre los adolescentes, pues refleja un problema de autoestima y de creer que son valiosos sólo por parecerse a alguien famoso. ¿Consideras que pareciéndose a otra persona uno se pueda sentir contento consigo mismo? ¿Estarías dispuesto a practicar una cirugía estética para parecerte a tu ídolo? Discute con tus compañeros sobre este tema.
13. *Economía.* El Mar de Cortés, en México, es la zona donde se localiza el mayor número de granjas dedicadas al cultivo de ostras perleras. Según los productores, 3.6 de cada cien ostras pueden llegar a tener un buen valor comercial. Si tenemos 150 ostras,
 - a) Determina la probabilidad de que 12 de ellas tengan perlas de buen valor comercial.
 - b) Encuentra la probabilidad de que cuando mucho dos tengan perlas de buen valor comercial.

- c) Consulta la página <http://perlas.com.mx/es/> y con tus compañeros conversa sobre la importancia de apoyar esta industria y cuál podría ser la repercusión económica y social que tendría tanto en el nivel regional como nacional.
- 14. Salud.** De acuerdo con la Encuesta de Consumo de Alcohol, Tabaco y otras Drogas en la Ciudad de México (2009), 25.3% de las personas consumen marihuana, drogas inhalables o cocaína. Si tenemos un grupo de 25 personas,
- ¿Cuál es la probabilidad de que 10 de ellas consuman alguna de estas drogas?
 - ¿Cuál es la probabilidad de que al menos 20 de ellas consuman alguna de estas drogas?
 - De las siguientes 50 personas, ¿cuántas esperarías que consumieran alguna de estas drogas?
 - Actualmente, el problema de consumo de drogas se presenta en todos los estratos sociales. Por lo general se utilizan para “estar en ambiente”, o bien, para “evadir la realidad”, entre otras muchas razones. ¿Sabes las consecuencias que el consumo de drogas trae consigo? ¿Crees que las campañas para vivir sin drogas han sido eficaces? Comenta con tus compañeros qué acciones se pueden llevar a cabo para prevenir su consumo.
- 15. Demografía.** De acuerdo con cifras del Censo Poblacional 2010, en México hay, en promedio, 3.9 habitantes por vivienda. Con esta información:
- Construye la distribución de probabilidad, así como la gráfica correspondiente y descríbela, indicando qué tipo de sesgo presenta.
 - Calcula la probabilidad de que una vivienda sea habitada por cinco personas.
 - Encuentra la probabilidad de que en un hogar vivan más de tres personas.
 - En la actualidad existe la tendencia a que las familias sean pequeñas, es decir, que estén constituidas por los padres y dos hijos. ¿Qué ventajas y desventajas tiene una familia así en comparación con una familia “grande”? Discute con tus compañeros sobre este tema.
- 16. Turismo.** Al término de la Semana Santa de 2011, en la caseta de la autopista a Cuernavaca ingresaban 75 autos en promedio por minuto. Con este dato, calcula la probabilidad de que en un minuto dado, ingresaran:
- 80 autos.
 - 50 autos.
 - de que en 15 minutos, ingresaran 800 autos.
 - El sector turístico se ha visto afectado actualmente por la ola de violencia que vive México. Sin embargo, es un sector que representa importantes ingresos para nuestro país. ¿Qué propondrías para mejorar este sector? ¿Cómo podría llegar a ser México el principal destino turístico del mundo? Organiza una mesa redonda con tus compañeros y redacten sus conclusiones.
- 17. Pasatiempos.** Abel Pérez tiene como pasatiempo armar rompecabezas. Puede pasar horas escuchando música mientras coloca piezas. El rompecabezas que actualmente arma es de 5000 piezas, y en un día coloca en promedio 100 piezas.
- Determina la probabilidad de que en un día coloque entre 100 y 110 piezas, inclusive.
 - Calcula la probabilidad de que en un día en que esté muy inspirado coloque 150 piezas.
 - En un día cualquiera, Abel no está inspirado. ¿Cuál es la probabilidad de que coloque 45 piezas o menos?
 - Pregunta a tus compañeros de clase qué pasatiempo tienen. En grupo, comenten cuáles son las ventajas de tener un pasatiempo y cómo influyen en su calidad de vida.
- 18. Inseguridad.** De acuerdo con las estadísticas de criminalidad, en promedio se roba un auto cada tres minutos. Determina:
- La distribución de probabilidad y la gráfica correspondiente. Describe de manera breve la distribución.

- b) La probabilidad de que en los siguientes tres minutos se roben dos autos.
 - c) La probabilidad de que en los siguientes tres minutos se roben dos autos o menos.
 - d) La probabilidad de que se roben seis autos en los siguientes 15 minutos.
 - e) Uno de los factores que afectan los costos de los seguros de automóviles es el número de robos (siniestros) que se cometen. ¿Qué se podría hacer para evitar el robo de autos? Comenta con tus compañeros y expongan sus propuestas en clase.
- 19. Economía.** Una cadena de supermercados, cuyo nombre empieza con W, abrió 267 tiendas en México durante 2010. Con esta información,
- a) Determina λ para el número promedio de tiendas abiertas en un día.
 - b) ¿Cuál es la probabilidad de que en un día se abran dos tiendas?
 - c) ¿Cuál es la probabilidad de que en una semana se inauguren seis tiendas?
 - d) ¿Qué beneficios económicos trae la inversión de empresas trasnacionales como W? En la actualidad hay una discusión respecto a qué conviene más a las personas: los mercados tradicionales o los supermercados. Junto con tus compañeros, determinen las ventajas y desventajas de cada uno de ellos.
- 20. Deportes.** En un gimnasio sólo para mujeres se realizan ejercicios aeróbicos y ejercicios en aparatos. El circuito está constituido por nueve aparatos tonificadores y ocho estaciones para los ejercicios aeróbicos. En un día cualquiera, en hora pico (entre las 19 y 20 horas), llegan en promedio siete mujeres. En ese horario:
- a) ¿Cuál es la probabilidad de que lleguen menos de cinco mujeres?
 - b) ¿Cuál es la probabilidad de que estén ocupados todos los aparatos?
 - c) Y en un día extraordinario, ¿cuál es la probabilidad de que tanto los aparatos como las estaciones estén ocupados?
 - d) ¿Por qué es importante practicar algún deporte? ¿Qué beneficios trae su práctica? ¿Haces algún deporte? ¿A qué conclusiones llegas? Realiza una breve exposición con tus compañeros.
- 21. Aeronáutica.** Los aviones son los transportes más seguros que hay. Según los datos de las compañías aéreas que más accidentes han tenido, la probabilidad de sufrir un accidente es de 0.0002%. Si en un aeropuerto se registran en una semana determinada 3500 operaciones aéreas,
- a) ¿Aplicarías una distribución binomial o una de Poisson? Justifica tu respuesta.
 - b) ¿Cuál es la probabilidad de que haya cuando mucho un accidente aéreo?
 - c) ¿Cuáles son las implicaciones de que un aeropuerto esté dentro de la ciudad? Discute con tus compañeros.
- 22. Servicios.** Se calcula que 0.3% de quienes llaman a los servicios de atención a clientes de un banco recibirá una señal de línea ocupada. Si se reciben 1100 llamadas, entonces:
- a) ¿Aplicarías una distribución binomial o una de Poisson? Justifica tu respuesta.
 - b) ¿Cuál es la probabilidad de que por lo menos cinco personas hayan recibido una señal de línea ocupada?
 - c) ¿Qué significa “dar un buen servicio”? ¿Crees que en México tenemos actitud de servicio? Comenta con tus compañeros el tema y expongan sus conclusiones.

Evaluación sumativa

Un escalón hacia arriba

Contesta los reactivos siguientes, lo que te permitirá establecer si has comprendido cabalmente lo expuesto en este bloque o debes repasar algún tema. ¡Adelante!

1. Hay cuatro tarjetas numeradas de 0 a 3. Se eligen dos al azar y se suman los números que aparecen.
 - a) Determina el espacio muestral para este experimento.
 - b) Construye la distribución de probabilidad.
 - c) Calcula la media.
 - d) Calcula la desviación estándar.
2. Durante el Mundial 2010 de fútbol en Sudáfrica, *die Mannschaft* (la selección alemana) anotó 16 goles en siete partidos que jugó. Calcula:
 - a) El promedio de goles por partido.
 - b) La probabilidad de que anote más de dos goles en el próximo partido.
3. Según la empresa Nielsen, una de cada 10 personas en el mundo utiliza internet para compras por ese medio. Supón que se les pregunta a seis personas si realizan compras por internet y calcula:
 - a) La probabilidad de que tres personas usen internet para realizar sus compras.
 - b) La probabilidad de que al menos dos personas usen este medio para realizar sus compras.
4. Supón que 0.1% de las teclas de los nuevos teléfonos inteligentes están defectuosas. Para una muestra aleatoria de 1200 teclas,
 - a) ¿Utilizarías una distribución binomial o una de Poisson? ¿Por qué?
 - b) Encuentra la probabilidad de que ninguna de las teclas esté defectuosa.
 - c) Halla la probabilidad de que tres o más teclas estén defectuosas.

Proyecto 7

HÁBITOS CULTURALES

En 1982, la UNESCO declaró que “la cultura da al hombre la capacidad de reflexionar sobre sí mismo. Es ella la que hace de nosotros seres específicamente humanos, racionales, críticos y éticamente comprometidos. A través de ella discernimos los valores y efectuamos opciones. A través de ella, el hombre se expresa, toma conciencia de sí mismo, se reconoce como un proyecto inacabado, pone en cuestión sus propias realizaciones, busca incansablemente nuevas significaciones, y crea obras que lo trascienden”.

La cultura debe ser uno de los principales objetivos de cualquier gobierno. Sin embargo, la Encuesta Nacional de Hábitos, Prácticas y Consumos Culturales de CONACULTA, que se aplicó a 32 000 mexicanos, arroja que:

- 25% nunca ha ido al cine.
- 44% jamás ha visitado una zona arqueológica.
- 43% no ha entrado a un museo ni a una biblioteca.

Algunos promotores y trabajadores de la cultura han criticado estos resultados, pues, en su opinión, CONACULTA ha tenido una visión elitista de la cultura, ya que sólo se centra en las bellas artes.

El cine se considera arte porque es una forma de narrar historias o acontecimientos. De hecho, se le llama *el séptimo arte*. En México hay 4818 salas de exhibición, agrupadas en 519 complejos cinematográficos (*Anuario IMCINE 2010*).

Supón que te han contratado para elaborar un nuevo estudio para confirmar o refutar los resultados de la Encuesta Nacional de Hábitos, Prácticas y Consumos Culturales de CONACULTA. Para ello:

- Elabora un cuestionario para obtener información sobre si las personas a las que encuestas asisten al cine, cuántas veces al mes va al cine y si ven películas mexicanas o extranjeras.
- Aplica el cuestionario a 30 personas y captura la información.
- Con base en las respuestas, calcula la proporción de personas que van al cine, el número promedio de veces al mes que asiste una persona a las salas cinematográficas, así como la proporción de personas que ven películas mexicanas.
- Determina las distribuciones de probabilidad para cada reactivo y describe las distribuciones, indicando, por ejemplo, la probabilidad de que al menos 10 personas vayan al cine, o de que al menos asistan al cine dos veces al mes, o bien, de que al menos cinco personas vean cine mexicano.

Escribe un breve informe de las conclusiones a las que llegas.

Aspectos que has de contemplar para el informe:

- Compara los resultados que obtuviste de tu muestra con los resultados de la Encuesta Nacional.
- ¿Qué consecuencias económicas, culturales, sociales existen si las personas no asisten al cine?
- ¿Qué beneficios y prejuicios trae ver películas mexicanas? ¿Y extranjeras?
- Uno de los graves problemas por los que hay una baja asistencia a las salas cinematográficas es la piratería. ¿Qué propones para que haya mayor concurrencia al cine? Es decir, ¿qué propones para evitar la piratería?
- ¿Consideras que las campañas contra la piratería han dado buenos resultados? Explica tu respuesta.



Competencias

- Expresa ideas y conceptos mediante representaciones lingüísticas y matemáticas relacionadas con las variables aleatorias y sus representaciones en distribuciones de probabilidad
- Sigue procedimientos de manera reflexiva, ordena información de acuerdo con categorías y jerarquías y las relaciona con base en las distribuciones de probabilidad, para las variables aleatorias continuas
- Estructura ideas y argumentos de manera clara, coherente y sintetiza los resultados de las distribuciones de probabilidad de variables aleatorias continuas manejando las tecnologías de la información para el análisis de resultados obtenidos en la representación tabular, gráfica y funciones de probabilidad
- Define metas y da seguimiento a sus procesos de construcción de conocimiento sobre las distribuciones de probabilidad de variables aleatorias continuas
- Propone maneras de resolver un problema o desarrollar un proyecto en equipo, definiendo un curso de acción con pasos específicos, asumiendo una actitud constructiva, congruente con los conocimientos y habilidades con los que cuenta dentro de distintos equipos de trabajo

8. Distribución de probabilidad de variables aleatorias continuas

Contenido

- Variables aleatorias continuas

... Quetelet puso en marcha un ambicioso proyecto de recopilación de miles de medidas relacionadas con el cuerpo humano... y representó gráficamente la frecuencia de aparición de cada rasgo humano... Para su sorpresa, Quetelet descubrió que todas las características humanas siguen lo que ahora se denomina una *distribución de frecuencias normal* (o *gaussiana*, por el nombre del “príncipe de la matemática” Carl Friedrich Gauss, aunque no está demasiado justificado el porqué de esta denominación), con forma de campana...

La curva no era desconocida para Quetelet; los matemáticos y físicos la conocían desde mediados del siglo XVIII, y Quetelet estaba familiarizado con ella por su trabajo en astronomía; lo asombroso fue la asociación de esta curva con características humanas. Anteriormente, se la solía denominar *curva de error*, porque solía aparecer en cualquier tipo de errores de medida (Livio, 2009).

Ubícate

Antes de comenzar con el estudio de la probabilidad de variables aleatorias continuas, contesta el examen siguiente para que establezcas si tienes los conocimientos necesarios para comprender cabalmente los temas expuestos.

1. Integra:

$$\int_0^2 (x + 1)dx$$

2. Dada la muestra siguiente, calcula lo que se pide:

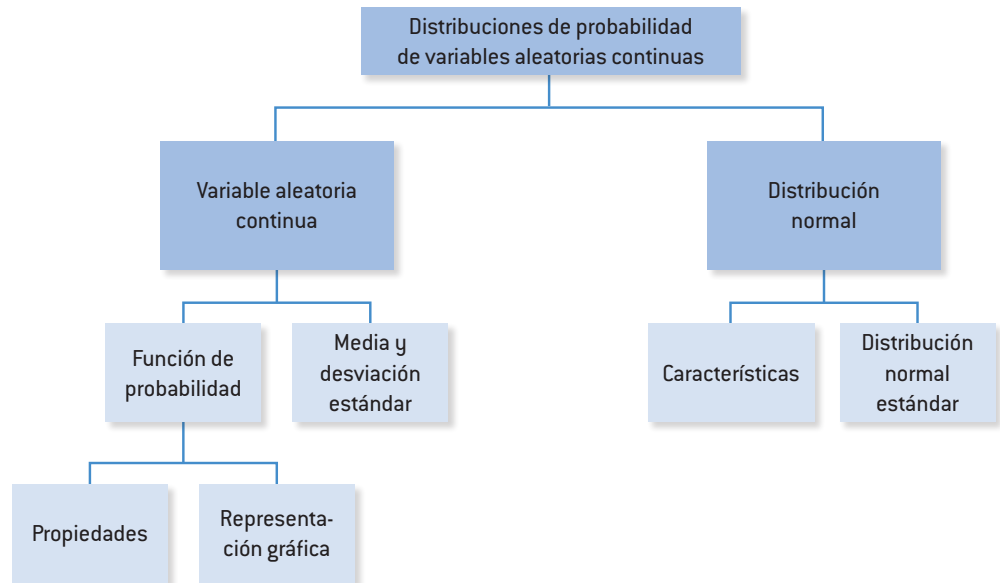
-7	11	9	-9	6	12	-9	8
----	----	---	----	---	----	----	---

- a) Media, mediana y moda
 - b) Varianza y desviación estándar
3. Compara los valores de la media, la mediana y la moda siguientes e indica si las distribuciones son simétricas, sesgadas a la derecha o sesgadas a la izquierda:
- a) Media = 5, mediana = 4.99, moda = 5.01
 - b) Media = 7.5, mediana = 9.0, moda = 9.5
 - c) Media = 3, mediana = 2.2, moda = 0.1

Si no tuviste problemas o dudas con el examen anterior, ahora sí comienza el estudio del bloque 8.

Función de densidad

En el bloque anterior aprendiste qué es y cómo usar la distribución de probabilidad de variables aleatorias discretas. En este bloque proseguiremos con el estudio de la distribución de probabilidad, pero ahora nos centraremos en las variables aleatorias continuas. En el esquema siguiente se muestran los temas que abordaremos.



Desempeños

Al terminar de estudiar este bloque y realizar todas las actividades propuestas habrás aprendido a:

- Identificar y analizar los elementos de una variable aleatoria continua para aplicarlos en la resolución de problemas de diversos contextos.
- Explicar las características de una representación de forma tabular, gráfica y funciones de probabilidad mediante la media y la desviación estándar, que te permita interpretar soluciones a problemas de tu entorno.
- Identificar el área bajo la curva normal estandarizada a partir de la distribución de probabilidad normal.

Variable aleatoria continua

Como aprendiste en el bloque anterior, una *variable aleatoria* es un medio para describir los resultados del espacio muestral mediante la asignación de valores. En el caso de una variable aleatoria continua, los valores numéricos provienen de un intervalo continuo, es decir, no son valores específicos, sino que puede ser cualquier valor entre dos números, a y b .

En otras palabras, con variables aleatorias continuas la probabilidad de que la variable aleatoria X sea igual a un valor es cero, $P(X = x) = 0$, ya que es imposible que X tenga exactamente ese valor.

En la tabla siguiente se muestran algunos ejemplos de variables aleatorias continuas:

Ejemplo 1

Experimento	Variable aleatoria (X)	Valores posibles de X
Tomar agua durante un día	Contar la cantidad de agua que se bebe	0 a k (una persona puede beber 330 ml, 2 l, 2.5 l, etc.)
Manejar un auto	Determinar la distancia recorrida en una semana	0 a k (una persona puede manejar 60.5 km, 0.14 km, 301 km, 567.3 km)
Realizar una llamada telefónica	Determinar el tiempo (exacto) que dura la llamada telefónica	0 a k (la llamada puede durar 1 s, 1 min, 1.7 min, 5.3 min, etc.)
Determinar la altura media de los estudiantes de una escuela	Medir la altura de los estudiantes de una escuela	0 a 2.5 m

Función de densidad

Se llama *función de densidad* a la distribución de probabilidad de una variable aleatoria continua. Para poder calcular las probabilidades, debemos usar cálculo integral, pues el área bajo una curva es, en estadística, una probabilidad.

Las propiedades que debe cumplir una función de densidad son las siguientes:

- $f(x) \geq 0$, para toda $x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(x) dx = 1$
- $P(a < x < b) = \int_a^b f(x) dx$

donde x es el valor de la variable aleatoria y $f(x)$ la probabilidad asociada con el valor x .

Una función de densidad puede representarse de dos formas:

- *Gráfica*: es una gráfica de línea, cuya área bajo la curva es la probabilidad entre dos valores de la variable aleatoria.
- *Como función*: cuando la distribución de densidad se define mediante una fórmula.

Media y desviación estándar de una variable continua

En el bloque anterior hicimos la observación de que la media y la varianza (momentos 1 y 2 de una función de probabilidad) son las dos principales características de una función de probabilidad, ya que indican el centro y la dispersión de la distribución de los datos. Cuando se habla de una función de densidad, se definen la media y la desviación estándar de esta manera:

MEDIA O VALOR ESPERADO

Es una medida de localización central de la variable aleatoria. Se calcula de la forma siguiente:

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

VARIANZA Y DESVIACIÓN ESTÁNDAR

Son medidas de dispersión o variabilidad de la variable aleatoria. Se calculan de esta manera:

$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Cabe recordar que la varianza también puede calcularse así:

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2$$

De este modo determinamos la *desviación estándar* calculando únicamente la raíz cuadrada de la varianza.

Ejemplo 2

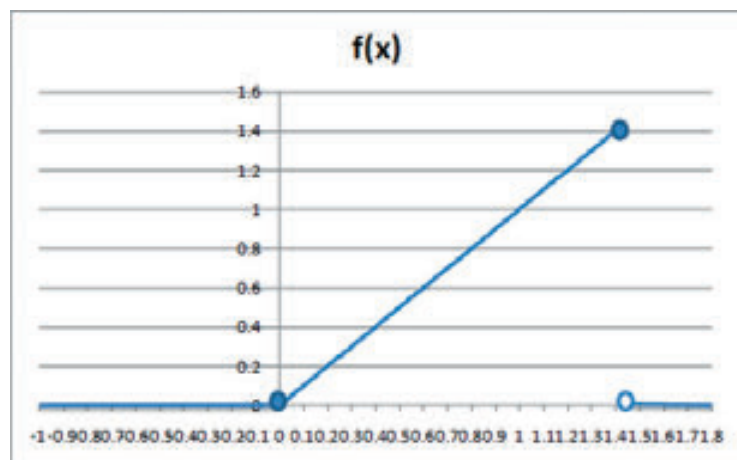
Sea la función de densidad:

$$f(x) = \begin{cases} x & \text{si } 0 \leq x \leq \sqrt{2} \\ 0 & \text{= en otro caso} \end{cases}$$

- Traza la gráfica correspondiente.
- Encuentra $P\{0.5 < x < 1\}$
- Determina la media.
- Calcula la varianza y la desviación estándar.

Solución

- La gráfica de la función de densidad es:



b.

$$P(0.5 < x < 1) = \int_{0.5}^1 x dx = \frac{x^2}{2} \Big|_{0.5}^1 = \frac{1^2}{2} - \frac{0.5^2}{2} = 0.375$$

c. Media:

$$\begin{aligned} \mu = E(X) &= \int_{-\infty}^{\infty} xf(x) dx \\ &= \int_{-\infty}^0 xx dx + \int_0^{\sqrt{2}} xx dx + \int_{\sqrt{2}}^{\infty} xx dx \\ &= 0 + \int_0^{\sqrt{2}} x^2 dx + 0 \\ &= \frac{x^3}{3} \Big|_0^{\sqrt{2}} \\ &= \frac{\sqrt{2}^3}{3} - 0 = 0.9428 \end{aligned}$$

d. Varianza: utilizando $\sigma^2 = \text{Var}\{X\} = E\{X^2\} - [E\{X\}]^2$

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= \int_{-\infty}^0 x^2 x dx + \int_0^{\sqrt{2}} x^2 x dx + \int_{\sqrt{2}}^{\infty} x^2 x dx \\ &= 0 + \int_0^{\sqrt{2}} x^3 dx + 0 \\ &= \frac{x^4}{4} \Big|_0^{\sqrt{2}} \\ &= \frac{\sqrt{2}^4}{4} - 0 = 1 \end{aligned}$$

Así,

$$\sigma^2 = \text{Var}\{X\} = E\{X^2\} - [E\{X\}]^2 = 1 - (0.9428)^2 = 0.1111$$

Por tanto, la desviación estándar es:

$$\sigma = \sqrt{0.1111} = 0.3333$$

Distribución normal

Entre las funciones de densidad, la distribución normal es la más importante, debido principalmente a que hay muchas variables asociadas a fenómenos naturales que siguen un comportamiento que, al graficarse, tienen forma de campana. Estas variables pueden referirse a:

- Caracteres morfológicos de personas, animales o plantas, como tallas, pesos, diámetros, perímetros, etcétera.
- Caracteres fisiológicos, por ejemplo, efecto de una misma dosis de un fármaco o de una misma cantidad de abono.
- Caracteres sociológicos, por citar algunos casos, consumo de cierto producto por un mismo grupo de individuos o las puntuaciones obtenidas en un examen.
- Caracteres psicológicos, como el cociente intelectual o el grado de adaptación a un medio, entre otros.
- Errores cometidos al medir ciertas magnitudes.
- Otras distribuciones como la binomial o la de Poisson pueden aproximarse a una distribución normal.

Entre las características de la distribución normal podemos citar las siguientes (figura 1):

- La distribución normal se conoce también como *campana de Gauss* en honor de Karl F. Gauss y debido a que tiene forma de campana.
- Por ser una distribución de probabilidad, el área bajo la curva es 1.
- Es una distribución asintótica, es decir, cuando x tiende a $-\infty$ o a $+\infty$, la función tiende a cero.
- Es una distribución simétrica respecto a la media (μ). Además, la media, la moda y la mediana tienen el mismo valor.
- El punto máximo se encuentra en μ .
- La ubicación de la distribución normal se determina por μ .
- La dispersión de la distribución normal se determina por σ .
- Los puntos de inflexión, donde cambia de curvatura la función, se encuentran en $\mu \pm \sigma$.

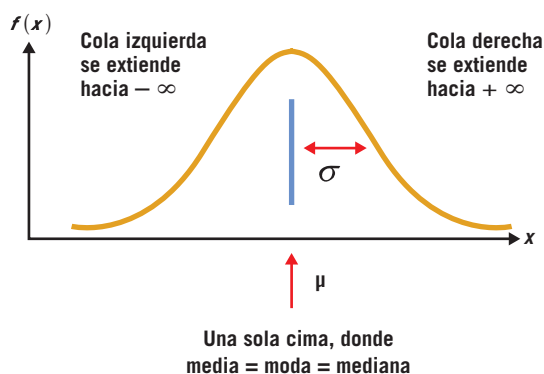


Figura 1 Gráfica de campana de una distribución normal.

Función de densidad

La función de densidad para una variable aleatoria X que se distribuye normalmente es

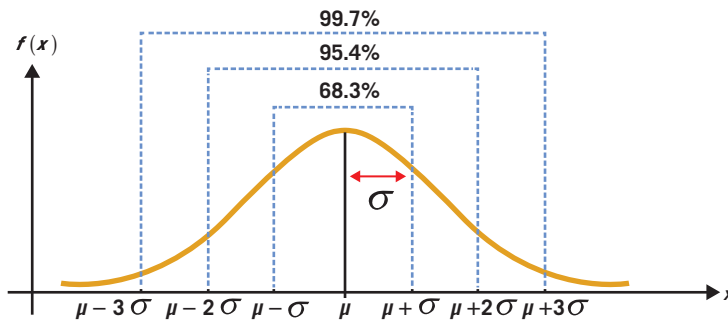
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x-\mu)^2/2\sigma^2]}$$

donde $f(x)$ = función de densidad de la variable aleatoria x , donde $x \in (-\infty, +\infty)$
 σ = desviación estándar poblacional
 μ = media poblacional

Dicho en otras palabras, necesitamos conocer la media y la desviación estándar o la varianza para tener definida la distribución normal.

Entonces, al ser la distribución normal una función de densidad, el cálculo de probabilidades se traduce en determinar el área bajo la curva entre dos valores; esto es, se integra la función.

En la gráfica mostrada en la figura 2 se presentan las probabilidades de que la variable X se encuentre en $\pm$ una, dos o tres desviaciones estándar respecto a la media; es decir, 68.3% de las observaciones se encuentran en el intervalo $(\mu - \sigma, \mu + \sigma)$; 95.4% están en el intervalo $(\mu - 2\sigma, \mu + 2\sigma)$ y 99.7% se hallan en el intervalo $(\mu - 3\sigma, \mu + 3\sigma)$.



Cuando un valor de X está por debajo de $\mu - 3\sigma$ o por arriba de $\mu + 3\sigma$ se llama *outlier* o dato aberrante.

Figura 2 Probabilidades de que la variable X se encuentre en $\pm$ una, dos o tres desviaciones estándar respecto a la media.

Si varían los parámetros μ y σ se obtiene distribuciones normales *diferentes*; por tanto, se tendrá un número infinito de distribuciones, como se muestra en la figura 3.

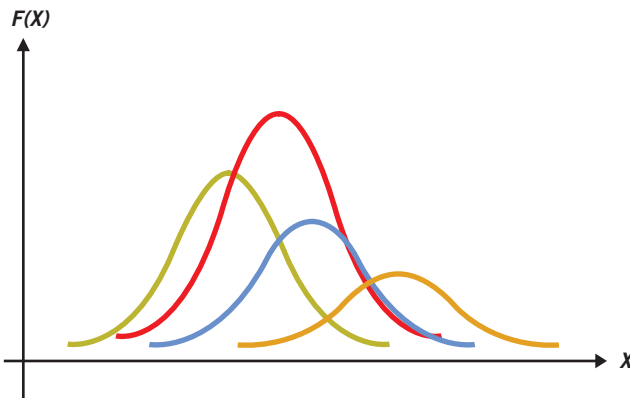


Figura 3 Un número infinito de distribuciones normales debido a la variación de la media y de la desviación estándar poblacionales (μ y σ , respectivamente).

Pero como no es sencillo integrar la función de densidad normal, para calcular probabilidades debemos utilizar la *distribución normal estándar*.

Distribución normal estándar

Esta distribución tiene media igual a cero y desviación estándar igual a uno. Es importante señalar que a la variable aleatoria distribuida como una normal estándar se le llama Z para diferenciarla de la X .

Función de densidad

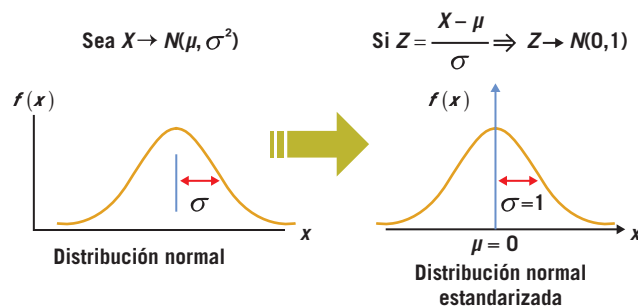
La función de densidad para una variable aleatoria Z que se distribuye como una normal estándar es

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-(z^2/2)}$$

donde $f(z)$ = función de densidad de la variable aleatoria Z , tal que $z \in (-\infty, +\infty)$.

Área bajo la curva

Recordemos que el área bajo la curva es, en Estadística, una probabilidad, en tanto que en cálculo es una integral. Entonces, para trabajar con la normal estándar hay que realizar la transformación de variable siguiente:



Este proceso de transformar una variable X con distribución normal en una variable Z se llama *estandarización*.

En el anexo B del libro encontrarás la tabla de probabilidad de una normal estándar. En esa tabla se muestra la probabilidad acumulada a la izquierda de un valor, es decir, se presenta la distribución acumulada de la normal estándar. De esta forma, sólo transformamos la variable y ya no es necesario realizar los cálculos del área bajo curva.

Ejemplo 3

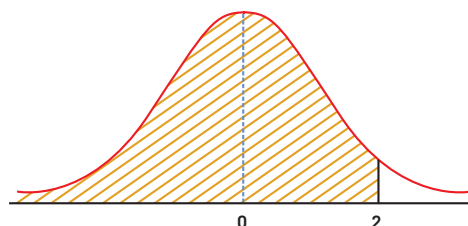
Determina las probabilidades siguientes:

- $P\{Z \leq 2\}$
- $P\{Z > -1\}$
- $P\{Z \leq 1.2\}$
- $P\{-0.9 \leq Z \leq 2.2\}$

Solución

Como se trata de una variable aleatoria Z , entonces empleamos directamente la tabla del anexo B:

- La probabilidad que debemos determinar se presenta en la gráfica, y el área que nos interesa es la de la izquierda de 2; por tanto, buscamos directamente el valor en la tabla:



z	0.00	0.01
1.9	0.971283	0.971933
2.0	0.977250	0.977784
2.1	0.982136	0.982571
2.2	0.986097	0.986447
2.3	0.989276	0.989556
2.4	0.991802	0.992024

De esta manera,

$$P(Z \leq 2) = 0.977250$$

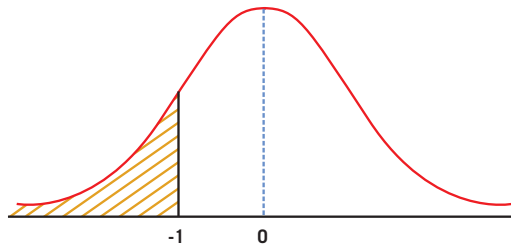
es decir, 97.72% de las observaciones son menores o iguales a 2.

Si quieres calcularlo con Excel, sólo debes escribir `=DISTR.NORM.ESTAND()` y entre paréntesis indicar el valor de Z que buscas, como se muestra enseguida:

`=DISTR.NORM.ESTAND(2)`

Cuando apliques la función `=DISTR.NORM.ESTAND()` de Excel, recuerda que se trata la probabilidad acumulada, es decir, $P(Z < z)$.

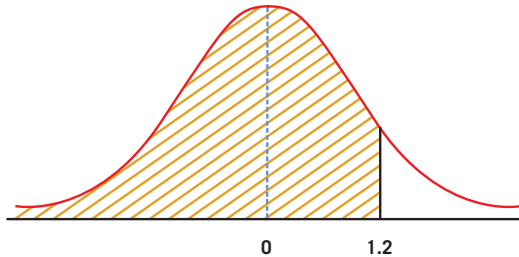
- b. Para calcular $P(Z > -1)$, primero debemos hacer $P(Z > -1) = 1 - P(Z \leq -1)$, ya que la tabla presenta las probabilidades a la izquierda del valor que se nos indica. Entonces,



z	0.00	0.01
-1.2	0.11507	0.113139
-1.1	0.135666	0.133500
-1.0	0.158655	0.156248
-0.9	0.184060	0.181411
-0.8	0.211855	0.208970
-0.7	0.241964	0.238852

$$P(Z > -1) = 1 - P(Z \leq -1) = 1 - 0.158655 = 0.841345$$

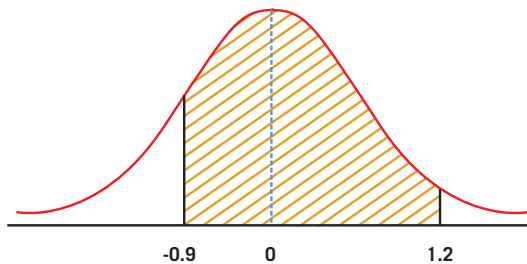
- c. $P(Z \leq 1.2) = 0.884930$



z	0.00	0.01
1.0	0.841345	0.843752
1.1	0.864334	0.866500
1.2	0.884930	0.886861
1.3	0.903200	0.904902
1.4	0.919243	0.920730

- d. Para calcular la probabilidad entre dos valores hacemos

$$P(-0.9 \leq Z \leq 2.2) = P(Z \leq 2.2) - P(Z \leq -0.9) = 0.986097 - 0.184060 = 0.802036$$



z	0.00	0.01
1.9	0.971283	0.971933
2.0	0.977250	0.977784
2.1	0.982136	0.982571
2.2	0.986097	0.986447
2.3	0.989276	0.989556
2.4	0.991802	0.992924

Esto se hace así porque a la probabilidad acumulada hasta 2.2 le quitamos la probabilidad acumulada hasta -0.9 y queda el área que nos interesa.

z	0.00	0.01
-1.1	0.135666	0.133500
-1.0	0.158655	0.156248
-0.9	0.184060	0.181411
-0.8	0.211855	0.208970

Ejemplo 4

Sea X una variable aleatoria distribuida normalmente, con $\mu = 3$ y $\sigma = 2$. Calcula:

- $P(X \leq 2.5)$
- $P(X > -3)$
- $P(X \leq 3.3)$
- $P(-2 \leq X \leq 2)$

Solución

Recuerda que estandarizar es transformar X en Z de esta forma:

$$Z = \frac{X - \mu}{\sigma}$$

Como la media es distinta de 0 y la desviación estándar diferente de 1, entonces hay que estandarizar la X para poder utilizar la tabla de la distribución normal estándar:

a.

$$P(X \leq 2.5) = P\left(Z \leq \frac{2.5 - 3}{2}\right) = P(Z \leq -0.25) = 0.401294$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
-0.3	0.382089	0.378280	0.374484	0.370700	0.366928	0.363169	0.359424
-0.2	0.420740	0.416834	0.412936	0.409046	0.405165	0.401294	0.397432
-0.1	0.460172	0.456205	0.452242	0.448283	0.444330	0.440382	0.436441
-0.0	0.500000	0.496011	0.492022	0.488034	0.484047	0.480061	0.476078

b.

$$P(X > -3) = 1 - P(X \leq -3) = 1 - P\left(Z \leq \frac{-3 - 3}{2}\right) = 1 - P(Z \leq -3) = 1 - 0.001350 = 0.99865$$

z	0.00	0.01
-3.1	0.000968	0.000935
-3.0	0.001350	0.001306
-2.9	0.001866	0.001807
-2.8	0.002555	0.002477

c.

$$P(X \leq 3.3) = P\left(Z \leq \frac{3.3 - 3}{2}\right) = P(Z \leq 0.15) = 0.559618$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
0.0	0.500000	0.503989	0.507978	0.511966	0.515953	0.519939	0.523922
0.1	0.539828	0.543795	0.547758	0.551717	0.555670	0.559618	0.563559
0.2	0.57926	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568
0.3	0.617911	0.621720	0.625516	0.628300	0.633072	0.636831	0.640574

d.

$$P(-2 \leq X \leq 2) = P(X \leq 2) - P(X \leq -2)$$

$$= P\left(Z \leq \frac{2 - 3}{2}\right) - P\left(Z \leq \frac{-2 - 3}{2}\right)$$

$$= P(Z \leq -0.5) - P(Z \leq -2.5)$$

$$= 0.308538 - 0.006210$$

$$= 0.302328$$

z	0.00	0.01
-0.6	0.274253	0.270931
-0.5	0.308538	0.305026
-0.4	0.344578	0.340903
-0.3	0.382089	0.378280

z	0.00	0.01
-2.6	0.004661	0.004527
-2.5	0.006210	0.006037
-2.4	0.008198	0.007976
-2.3	0.010724	0.010444

Una persona que tiene ingresos variables gana a la semana, en promedio, \$1250 con una desviación estándar de \$350. Si se sabe que lo que gana esta persona se distribuye normalmente, determina:

Ejemplo 5

- ¿Cuál es la probabilidad de que en una semana gane más de \$1500?
- ¿Cuál es la probabilidad de que gane entre \$1000 y \$1750 en una semana?
- ¿A partir de cuánto se le podría considerar el 10% de los mejores ingresos semanales?

Solución

Sabemos que $\mu = 1250$ y $\sigma = 350$, entonces:

a.

$$P(X > 1500) = 1 - P(X \leq 1500) = 1 - P\left(Z \leq \frac{1500 - 1250}{350}\right) = 1 - P(Z \leq 0.71) = 0.761148$$

Cuando realices la estandarización, redondea a dos decimales para poder utilizar la tabla de la normal.

z	0.00	0.01
0.6	0.725747	0.729069
0.7	0.758036	0.761148
0.8	0.788145	0.791030
0.9	0.815940	0.818589

La persona tiene 76.11% de probabilidad de ganar más de \$1500 en una semana.

b.

$$P(1000 \leq X \leq 1750) = P(X \leq 1750) - P(X \leq 1000)$$

$$= P\left(Z \leq \frac{1750 - 1250}{350}\right) - P\left(Z \leq \frac{1000 - 1250}{350}\right)$$

$$= P(Z \leq 1.43) - P(Z \leq -0.71)$$

$$= 0.923641 - 0.235762$$

$$= 0.687879$$

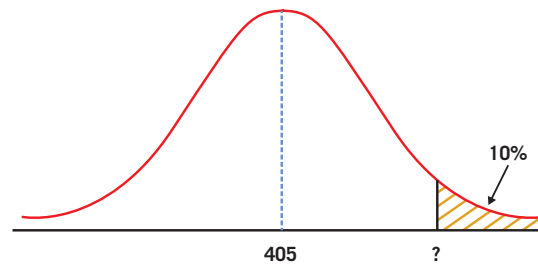
z	0.00	0.01	0.02	0.03	0.04
1.2	0.884930	0.886861	0.888768	0.890651	0.892512
1.3	0.903200	0.904902	0.906582	0.908241	0.909877
1.4	0.919243	0.920730	0.922196	0.923641	0.925066
1.5	0.933193	0.934478	0.935745	0.936992	0.938220
1.6	0.945201	0.946301	0.947384	0.948449	0.949497

z	0.00	0.01	0.02
-0.8	0.211855	0.208970	0.206108
-0.7	0.241964	0.238852	0.235762
-0.6	0.274253	0.270931	0.267629

La probabilidad de que esta persona de ganar entre \$1000 y \$1750 es de 68.79%.

- En este inciso debemos determinar el salario semanal a partir del cual se acumula el 10%, es decir:

$$P(X \geq x) = 0.10, \text{ o bien, } P(X < x) = 0.90$$



Como es lo único que conocemos para calcular probabilidades, debemos estandarizar:

$$P\left(Z \leq \frac{x - 1250}{350}\right) = 0.90$$

donde se acumula el 90% es en $Z = 1.28$ (buscamos en las probabilidades lo más cercano a 0.90 y vemos el valor de Z):

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	0.841345	0.843752	0.846136	0.848495	0.85083	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.879000	0.881000	0.882977
1.2	0.884930	0.886861	0.888768	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903200	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736

Así,

$$\frac{x - 1250}{350} = 1.88 \Rightarrow x = 1.88(350) + 1250 = 1908$$

es decir, a partir de \$1908 se encuentra el 10% superior de sus ingresos semanales.

Aproximación de la normal a la binomial

Muchas veces podemos utilizar la distribución normal como una aproximación a distribuciones discretas, por ejemplo, a la distribución binomial. Para que sea una muy buena aproximación se lleva a cabo una corrección por continuidad. Ello debido a que la distribución normal es continua y si ponemos rectángulos para calcular el área bajo la curva, quedan pequeños espacios que no se tienen en cuenta o que sobran, como se advierte en la figura siguiente:



Cuando el histograma de la distribución binomial no esté demasiado sesgado, la variable aleatoria X se puede distribuir mediante una normal con $\mu = np$ y $\sigma^2 = npq$, utilizando un factor de corrección de 0.5 aplicable al estandarizar la variable, es decir:

$$\begin{aligned} \blacktriangleright P(X \leq x) &= P\left(Z \leq \frac{x + 0.5 - \mu}{\sigma}\right) \\ \blacktriangleright P(X \geq x) &= P\left(Z \geq \frac{x - 0.5 - \mu}{\sigma}\right) \end{aligned}$$

Esta aproximación es adecuada siempre que $np \geq 5$ y $nq \geq 5$

Se tiene una variable aleatoria X que sigue una distribución binomial, con $n = 50$ y $p = 0.2$. Determina $P(5 \leq X \leq 7)$, mediante:

Ejemplo 6

- La distribución binomial.
- La aproximación de la normal.

Solución

- Calculando esta probabilidad mediante la distribución binomial, tenemos:

$$P(5 \leq X \leq 7) = P(X = 5) + P(X = 6) + P(X = 7) = 0.1719$$

- Como $np \geq 5$ y $nq \geq 5$, podemos usar la aproximación de la normal, determinando $\mu = np = 10$ y $\sigma^2 = npq = 8$. Así,

$$P(5 \leq X \leq 7) = P(X \leq 7) - P(X \leq 5)$$

$$= P\left(Z \leq \frac{7 + 0.5 - 10}{\sqrt{8}}\right) - P\left(Z \leq \frac{5 - 0.5 - 10}{\sqrt{8}}\right)$$

$$= P(Z \leq -0.88) - P(Z \leq -1.94)$$

$$= 0.189430 - 0.026190$$

$$= 0.16324$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-2.0	0.022750	0.022216	0.021692	0.021178	0.020675	0.020182	0.019699	0.019226	0.018763	0.018309
-1.9	0.028717	0.028067	0.027429	0.026803	0.026190	0.025588	0.024998	0.024419	0.023852	0.023295
-1.8	0.035930	0.035148	0.034380	0.033625	0.032884	0.032157	0.031443	0.030742	0.030054	0.029379

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0.9	0.184060	0.181411	0.178786	0.176186	0.173609	0.171056	0.168528	0.166023	0.163543	0.161087
-0.8	0.211855	0.208970	0.206108	0.203269	0.200454	0.197663	0.194895	0.192150	0.189430	0.186733
-0.7	0.241964	0.238852	0.235762	0.232695	0.22965	0.226627	0.223627	0.220650	0.217695	0.214764

Como puedes observar, es una buena aproximación.

Ejercicios

I. TEORÍA Y ALGO MÁS

1. Subraya la opción correcta:

- a) Una distribución normal es (simétrica, asimétrica) respecto a la media.
- b) La distribución normal estándar tiene media de (0, 1) y desviación estándar de (0, 1).
- c) La $P(Z \leq 0)$ es (1, 0.5).
- d) Los puntos de inflexión de una distribución normal se encuentran a $(+/-2, +/-1)$ desviaciones estándar.
- e) El proceso de cambiar la variable aleatoria X por $Z = \frac{X - \mu}{\sigma}$ se llama (transformación, estandarización).
- f) En el punto máximo de una distribución normal se cumple que (media = moda = mediana, media $\neq$ moda $\neq$ mediana).
- g) Si Z es una variable aleatoria con distribución normal estándar, entonces $P(-2 < Z < 2)$ es aproximadamente igual a (0.99, 0.95, 0.68).
- h) Si se transforma una distribución normal en una distribución normal estándar, entonces lo que no cambia es (la media, la forma, la varianza).
- i) Si la probabilidad es menor a 0.5, entonces el valor de Z es (negativo, positivo).
- j) En una distribución normal, si la variable aleatoria X tiene un valor positivo muy grande, entonces tiene una (mayor, menor) probabilidad de ocurrir.

II. PIENSO, ... ¿Y LUEGO?

2. De acuerdo con el artículo “Otra vez la educación”, de Alejandro Villagómez (Villagómez, 2010), un tema recurrente en la agenda nacional es la educación, debido a los malos resultados obtenidos en pruebas tipo ENLACE (Evaluación Nacional del Logro Académico en Centros Escolares) o PISA (Programa Internacional para la Evaluación del Estudiante). Según un estudio realizado por la OCDE (Organización para la Cooperación y el Desarrollo Económico), existe un vínculo entre las habilidades cognitivas reflejadas en la prueba de PISA y el crecimiento económico, es decir, “se estimaba la ganancia que se obtendría por mayor crecimiento, en términos del PIB actual, esto es, la ganancia en valor presente”. De acuerdo con los resultados obtenidos en la prueba de PISA, México ocupa los últimos lugares junto con países como Turquía.

En palabras de Alejandro Villagómez:

Considerando el escenario más moderado (y probablemente realista) que fue suponer reformas en 2010 que permitan aumentar el puntaje promedio en 25 puntos PISA hacia 2030, se estimaba que la ganancia (en valor presente) para México por un mayor crecimiento sería equivalente a 268% del PIB actual. Este valor podría superar el 2000% en el escenario más demandante.

También hace unos días hice referencia en mi blog a otro estudio que se enfocaba en el área de matemáticas utilizando resultados PISA y analizaba la situación en las colas de la distribución, es decir, considerando a nuestros mejores estudiantes en comparación con los de otros países.

Sin entrar en detalles, sólo menciono que el estudio utilizaba como referencia un puntaje de 650 que coloca a un estudiante en el 10% superior de la OCDE.

Por cierto, la media mexicana fue de 405, pero el punto es que con esta referencia de 650 tendríamos que sólo tres de cada mil estudiantes mexicanos se ubicarían en este círculo exclusivo que es el 10% superior entre los países de la OCDE.

- Se sabe que las calificaciones o puntajes de evaluaciones siguen una distribución normal. De acuerdo con los datos proporcionados en el fragmento anterior, ¿cuál es la desviación estándar mexicana?
- ¿Cuáles son los puntajes en los que se encuentra el 90% central de los estudiantes mexicanos?
- Junto con tus compañeros (y según tu experiencia) expongan cuáles son las fortalezas y debilidades del sistema educativo mexicano. ¿Cómo te gustaría que fueran las clases?
- Por medio de una “lluvia de ideas”, propongan medidas para mejorar la educación en nuestro país.

III. PURO DEPORTE

- Sólo para practicar.* Como vimos en este bloque, el área bajo la curva de una función se traduce en la probabilidad entre dos valores. Calcula las integrales definidas siguientes:

- $\int_1^{1.5} (2x + 1) dx$

- $\int_0^{0.5} \frac{x^2}{3} dx$

- $\int_1^{1.2} e^x dx$

- $\int_{\frac{\pi}{6}}^{\frac{\pi}{4}} \sin x dx$

- Sólo para practicar.* Sea Z una variable aleatoria que tiene una distribución normal estándar; encuentra:

- $P(Z \leq 1.54)$
- $P(Z > 1.54)$
- $P(-0.1 \leq Z \leq 0.9)$
- $P(-3.66 \leq Z \leq -2.15)$

- Sólo para practicar.* Determina el valor z :

- $P(Z \leq z) = 0.6293$
- $P(-z \leq Z \leq +z) = 0.90$
- $P(-z \leq Z \leq +z) = 0.95$
- $P(-z \leq Z \leq +z) = 0.99$

- Sólo para practicar.* Sea X una variable aleatoria que se distribuye de manera normal con $\mu = 4$ y $\sigma = 1.5$. Calcula:

- $P(X \leq 5)$
- $P(X \geq 4.5)$
- $P(0 \leq X \leq 2)$

7. *Demografía.* De acuerdo con la Organización de Naciones Unidas (ONU), en Latinoamérica se tiene una esperanza de vida al nacer de 73.4 años y una desviación estándar de 4.3714 años. Si la esperanza de vida de una persona sigue una distribución normal, calcula:
- La probabilidad de que una persona sobrepase los 80 años de vida.
 - La probabilidad de que una persona al nacer tenga una esperanza de vida entre 70 y 75 años.
 - ¿A partir de qué valor se tiene el 10% más alto de la esperanza de vida al nacer?
 - Junto con tus compañeros, determina cuáles son los factores que influyen en el aumento de esperanza de vida al nacer, así como las consecuencias que representa para el gobierno el tener una población “vieja”.
8. *Finanzas.* Entre mayo y junio de 2011, la cotización del dólar a la venta tuvo una media de \$16.9205 y una desviación estándar de \$0.2695. Si la cotización del dólar sigue una distribución normal, determina:
- La probabilidad de que la venta de un dólar sobrepase los \$18.
 - La probabilidad de que la cotización del dólar a la venta sea menor a los \$15.
 - Con tu equipo, investiga qué factores intervienen en la cotización del dólar. Discute con ellos qué consecuencias trae una devaluación del peso frente al dólar.
9. *Ecología.* Uno de los graves problemas que se enfrentan actualmente es la falta de agua y la contaminación de los acuíferos. De acuerdo con cifras de la SEMARNAT (Secretaría del Medio Ambiente y Recursos Naturales) a través de CONAGUA (*Estadísticas del Agua en México, 2010*, México, 2010), de 2002 a 2008 en el país las descargas residuales tuvieron una media de 7.795 km³/año con una desviación estándar de 0.2311 km³/año (1 km³ equivale a mil millones de metros cúbicos).
- Suponiendo que las descargas residuales se distribuyen normalmente, encuentra la probabilidad de que, en un año, las descargas residuales sean:
- Menores a 7.5 km³
 - Estén entre 7.5 y 7.9 km³
 - ¿A partir de qué valor se tiene el 10% más bajo de descargas residuales?
 - De acuerdo con estudios, 30% de la contaminación de agua es causada por arrojar basura a ríos y lagos. Además, cada persona consume casi 150 litros de agua al día, los cuales, una vez que se contaminan, se pierden (*redacción*, 2011). Investiga con tus compañeros qué porcentaje de agua es tratada en México. Realicen un debate sobre las acciones que debemos emprender como sociedad para cuidar el agua. ¿Qué acciones realizas tú en tu comunidad?
10. *Educación.* De acuerdo con la Organización de Naciones Unidas (ONU), los años de educación promedio son de 8.1 y una varianza de 3.1 en países latinoamericanos. Suponiendo que los años de educación promedio se distribuyen como una normal, calcula la probabilidad de que un país:
- Tenga más de 10 años de educación promedio.
 - Tenga entre 7.5 y 10.5 años de educación promedio.
 - ¿A partir de qué valor se tiene el 5% más bajo de educación?
 - “Tan sólo por la educación puede el hombre llegar a ser hombre. El hombre no es más que lo que la educación hace de él.” (Kant). Discute con tus compañeros sobre esta frase y cómo afecta la educación en el crecimiento económico de un país.

11. *Economía.* La industria petrolera es una de las principales fuentes de ingresos de países como Brasil, México, Venezuela y Arabia Saudita, entre otros. A continuación se presenta la producción de petróleo crudo de las 15 principales empresas en el mundo:

Producción de petróleo crudo	
Principales empresas	Barriles diarios (miles)
Exxon Mobil Estados Unidos	6271
Sinopec China	4559
Royal Dutch Shell Reino Unido/Holanda	3639
PDVSA Venezuela	3035
CNPC China	2913
Conoco Phillips Estados Unidos	2902
Valero Energy Estados Unidos	2780
BP Reino Unido	2666
Total Francia	2594
Saudi Aramco Arabia Saudita	2374
Petrobras Brasil	2161
Chevron Estados Unidos	2158
NIOC Irán	1566
Petróleos Mexicanos México	1540
Nippon Japón	1317

Fuente: PEMEX, *Anuario estadístico 2011*.

- a) Determina la media y la desviación estándar poblacional.

Con la información del inciso (a), y suponiendo que la producción de petróleo crudo se distribuye normalmente, calcula la probabilidad de que en un día, la producción de petróleo crudo:

- b) Sea mayor a 5500 barriles.
 c) Esté entre 2500 y 4000 barriles.
 d) ¿Cuántos barriles de petróleo crudo deben producirse en un día para que éste sea considerado “excelente”? Debe entenderse por “excelente” que esté en el 5% superior de la distribución.
 e) El petróleo ocupa “el primer lugar en cuanto a contaminantes marítimos. Su proceso de extracción, refinación y transporte provoca que grandes cantidades terminen en el mar. Además, tres años tardan en recuperarse los ecosistemas marinos expuestos a grandes cantidades de crudo; los contaminados por refinado, en especial los estuarios, requieren 10 años o más”. (Alba, “Numeralia”: *Petróleo*, 2010). ¿Qué se podría hacer para evitar la contaminación marítima? ¿Qué energías alternativas conoces? ¿Crees que los biocombustibles sean la solución? Discútelos con tus compañeros y expongan sus inquietudes y conclusiones.

12. *Meteorología.* Supongamos que la precipitación pluvial se distribuye normalmente. Observa la tabla siguiente:

Precipitación pluvial por región hidrológica (1971–2000)	
Zona	Precipitación (ml)
Península de Baja California	168.3
Noroeste	448.1
Pacífico norte	747.7
Balsas	963.0
Pacífico sur	1184.6
Río Bravo	435.9
Cuencas centrales norte	427.6
Lerma Santiago Pacífico	817.9
Golfo norte	910.9
Golfo centro	1552.8
Frontera sur	1845.6
Península de Yucatán	1219.2
Aguas del Valle de México	610.2

Fuente: SEMARNAT, CONAGUA. *Estadísticas del agua en México*, 2008.

Con esta información:

- Calcula la media y la desviación estándar poblacionales.
 - Determina la probabilidad de que la precipitación pluvial sea mayor a 800 ml.
 - ¿Cuál es la probabilidad de que la precipitación pluvial esté entre 500 y 1000 ml?
 - Encuentra la cantidad de lluvia a partir de la cual se encuentra el 5% superior de la distribución (es decir, cantidad peligrosa de lluvia).
 - Actualmente se ha visto que los fenómenos meteorológicos son más “agresivos”, de acuerdo con los registros históricos, lo cual es ocasionado por la deforestación y el cambio climático, entre otras causas. Por ejemplo, las zonas sur y sureste de México han sufrido inundaciones severas en los últimos años durante la época de lluvias. Las inundaciones ocurridas entre 1997 y 2006 afectaron a cerca de 1.6 millones de personas y los daños materiales sumaron 3 millones de dólares. ¿Cuáles son los principales factores que causan las inundaciones? ¿Qué acciones propones para evitarlas? Realiza una presentación a tus compañeros de clase.
13. *Salud.* La falta de sueño reparador puede afectar nuestra capacidad para llevar a cabo las actividades diarias, ya que estamos demasiado cansados o tenemos problemas de concentración. Todos los tipos de insomnio pueden llevar a somnolencia diurna, mala concentración e incapacidad para sentirnos renovados y descansados en la mañana.
- Pregunta a tus compañeros de clase cuántas horas duermen al día.
 - Calcula la media y la desviación estándar poblacionales, considerando a todos tus compañeros de clase.
 - ¿Qué porcentaje de tus compañeros duermen menos de seis horas?
 - ¿Debajo de qué valor se encuentra el 10% de los que menos horas duermen? ¿A partir de qué valor se encuentran los “dormilones” (10% superior)?
 - Investiga en equipo cuáles son las causas y las consecuencias que trae la falta de sueño. Comenta qué acciones se deben tomar para evitar este problema de salud.

14. *Psicología.* El cociente intelectual (IQ) es la puntuación obtenida en tests que permiten medir la inteligencia de una persona. Si los puntajes del cociente intelectual se distribuyen normalmente con media igual a 100, determina:
- La desviación estándar, si se sabe que el 2% se consideran personas superdotadas al tener un puntaje mayor a 130.
 - ¿Cuál es la probabilidad de que una persona tenga un puntaje por debajo de 70 (retardo mental)?
 - ¿Qué proporción de la población se le considera “genio”? Una persona que obtiene un puntaje de 165 o mayor es generalmente considerada genio.
 - ¿Crees que una prueba realmente puede medir cuán inteligente es una persona? ¿Qué es para ti ser inteligente? ¿Crees que se nace inteligente o se cultiva la inteligencia a través del tiempo?
15. *Entretenimiento.* Johnny Depp es uno de los mejores actores de Hollywood y en 2010 fue considerado el hombre más sexy. Ha trabajado en más de cuarenta películas. Si la duración de sus filmes sigue una distribución normal con una media de 112.5111 minutos y una varianza de 383.8054, calcula:
- La probabilidad de que una película dure menos de 90 minutos.
 - La probabilidad de que una película dure entre 95 y 125 minutos.
 - La probabilidad de que dure más de 150 minutos.
 - El cine es cultura.* ¿Cómo crees que influye en tu vida ver películas de otros países? ¿Consideras que se deba apoyar más el cine mexicano? Explica tu respuesta.
16. *Derechos de los animales.* Cualquier tipo de mascota requiere cuidados y tiempo, además de que implica gastos de alimentación y salud, entre otros. Por ello, se debe ser consciente al tener una mascota; por ejemplo, un perro. De acuerdo con las estadísticas de refugios para perros callejeros, ocho de cada 10 perros que se compran terminan en la calle. Esto representa no sólo un problema de salud sino de falta de responsabilidad por parte del dueño. Si en el mes siguiente se venden 50 perros, determina, utilizando la aproximación a la normal:
- ¿Cuál es la probabilidad de que menos de 30 perros terminen en la calle?
 - ¿Cuál es la probabilidad de que entre 30 y 45 perros terminen en la calle?
 - Actualmente hay muchas organizaciones que luchan por los derechos de los animales (<http://www.facebook.com/group.php?gid=43882148077>). Comenta con tus compañeros las obligaciones y responsabilidades que tenemos como dueños, en este caso, de perros. ¿Qué se debe pensar antes de tener un perro?
17. *Medios de comunicación.* En México hay dos grandes empresas de televisión abierta: Televisa y TV Azteca. Se sabe que TV Azteca tiene aproximadamente 40% de la audiencia. Si se realiza una encuesta a 100 personas, usando la aproximación de la normal, determina:
- La probabilidad de que más de 55 personas vean TV Azteca.
 - La probabilidad de que menos de 30 personas sintonicen TV Azteca.
 - Realiza una mesa redonda con tus compañeros sobre la influencia que tienen los medios de comunicación en la sociedad.

Evaluación sumativa

Un escalón hacia arriba

Contesta los reactivos siguientes, que te permitirán establecer cuánto has comprendido de lo expuesto en este bloque. ¡Adelante!

1. Determina el valor de Z , tal que:
 - a) $P(-z \leq Z \leq z) = 0.80$
 - b) $P(Z \geq z) = 0.15$
2. Supón que el tiempo que duran las llamadas por celular que realiza una persona se distribuyen normalmente. El tiempo promedio es de 5.5 minutos y una desviación estándar de 2.5 minutos. Determina:
 - a) La probabilidad de que una llamada dure entre 4 y 6 minutos.
 - b) De las siguientes 20 llamadas que haga la persona, ¿cuántas durarán menos de 5 minutos?
 - c) ¿A partir de cuánto se puede considerar una llamada “tardada”? (*Nota:* se entiende por “tardada” una llamada que dura más de 15 minutos.)
3. Una cadena de supermercados (WM) tiene 48% de participación en el mercado. Si en los próximos meses se piensan abrir 150 supermercados en el país, ¿cuál es la probabilidad de que entre 50 y 60 de ellos sean de WM? (*Observación:* para responder a esta pregunta usa la aproximación de la normal.)

Proyecto 8

RESERVAS DE PETRÓLEO CRUDO

El petróleo es un líquido oleoso de origen natural compuesto por diferentes sustancias orgánicas. Se emplea como combustible y materia prima para la industria química, así como para la fabricación de medicinas, fertilizantes, productos alimenticios, objetos de plástico, materiales de construcción, pinturas o textiles y para generar electricidad.

Algunas operaciones de la industria petrolera han sido causantes de la contaminación del agua debido a los desechos; incluso esa industria ocupa el primer lugar en cuanto a contaminantes marítimos, pues se la relaciona con los productos derivados de la refinación y con derrames de petróleo; por ejemplo, 700 millones de litros fueron derramados en el Golfo de México en el accidente del pozo Ixtoc-1, en 1979.

La recuperación de los ecosistemas marinos expuestos a grandes cantidades de crudo es de tres años, aproximadamente. Mientras tanto, los contaminados por refinado, en especial los estuarios, requieren de 10 años o más.

La combustión de combustibles fósiles produce gases de efecto invernadero y otros contaminantes del aire. Actualmente, en México 45% de la electricidad se genera con petróleo, y 95% del transporte depende de él. Los contaminantes incluyen óxidos de nitrógeno, dióxido de azufre, compuestos orgánicos volátiles y metales pesados.

Forma un equipo y consulta la base de datos [petróleo.xls](#), donde se presenta el número de barriles que los países productores de petróleo tienen como reservas y realiza un análisis estadístico, suponiendo que las reservas se distribuyen normalmente. ¿A qué conclusiones llegas?

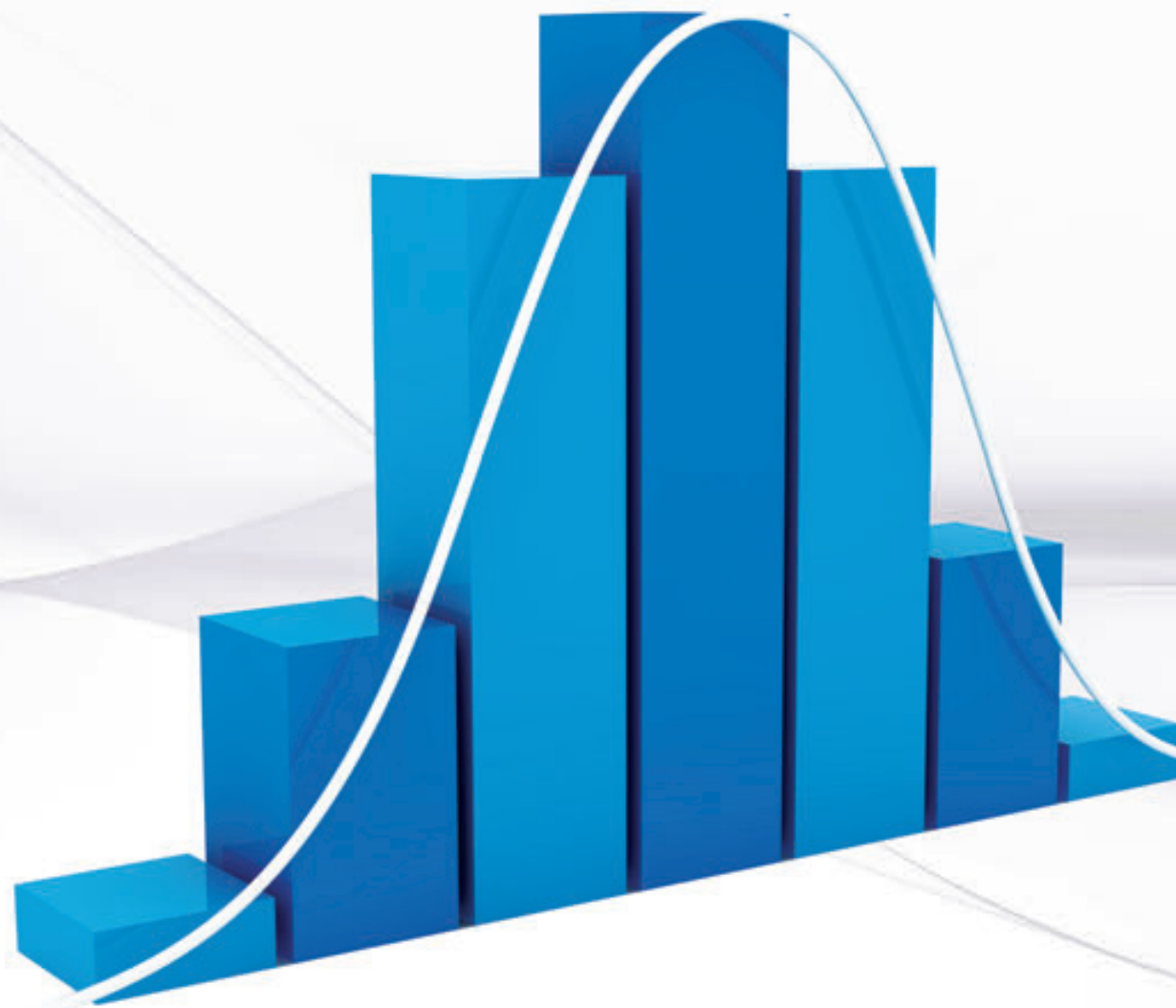
Para realizar el análisis estadístico, se sugiere:

- Trazar las gráficas correspondientes (histogramas, por ejemplo)
- Calcular las medidas de tendencia central y de dispersión (poblacionales).
- Determinar los valores donde se concentre 90, 95 o 99% de las reservas.
- Determinar probabilidades que tú y tus compañeros de equipo especifiquen.

Se sabe que las reservas petroleras en el país alcanzarán para 10 años. Supón que se debe tomar la decisión de explorar en aguas profundas, lo cual representa una inversión muy alta, además de que se necesitaría de inversión extranjera, lo cual generaría una molestia social, porque “el petróleo es de los mexicanos”. La otra opción es fomentar el desarrollo de biocombustibles, con las consecuencias ecológicas que esto generaría. ¿Qué decisión tomarías?

Preguntas para dirigir la discusión:

- ¿Qué beneficios económicos, ecológicos y sociales se obtendrían en cada situación?
- ¿Qué pérdidas económicas, ecológicas y sociales habrían en cada escenario?
- ¿Qué principios éticos estarían en juego en cada caso?
- ¿Cuál crees que es la responsabilidad social que se tiene como ciudadano?
- ¿Qué recomendaciones podrías hacer al gobierno federal en cada escenario?



Competencias

- Enfrenta las dificultades que se le presentan y es consciente de sus valores, fortalezas y debilidades
- Construye e interpreta modelos matemáticos, estructurando ideas y argumentos de manera clara, coherente y sintética en el análisis de dos variables, mediante la aplicación de procedimientos aritméticos, algebraicos, geométricos y variacionales para la comprensión y análisis de situaciones reales e hipotéticas y expresando ideas y conceptos mediante representaciones lingüísticas, matemáticas o gráficas para analizar los resultados de un problema
- Identifica las ideas clave en un texto sobre la representación de datos de dos variables e infiere conclusiones a partir de ellas
- Propone maneras de resolver un problema en equipo, definiendo un curso de acción con pasos específicos
- Formula y resuelve problemas matemáticos, aplicando los diagramas de dispersión y los coeficientes de correlación
- Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos para comprender el comportamiento de los datos de dos variables

9. Análisis de datos de dos variables

Contenido

- Representación de datos
- Correlación lineal

El científico italiano Santorio Santorio, también conocido como Sanctorius (1561-1636), llevó a cabo estudios para determinar la correlación que existe entre la ingesta de alimentos y el peso corporal. Sanctorius disponía de una plataforma sostenida por cuerdas que le permitía pesarse mientras permanecía sentado en ella.

Así, realizaba mediciones cuidadosas de su peso a intervalos regulares a lo largo de todo un día, y consiguió relacionar sus cambios de peso con el peso de alimentos sólidos y líquidos que había ingerido. Asimismo pesaba sus excreciones sólidas y líquidas, y siguiendo métodos numéricos determinó la pérdida de peso debido a la transpiración. (Cohen, 2005)

Ubícate

Antes de abordar el análisis de datos de dos variables, contesta los ejercicios siguientes, que te permitirán establecer si tienes los conocimientos necesarios para comprender cabalmente los temas que se exponen en el presente bloque.

1. Las calificaciones obtenidas por 100 alumnos en las clases de Matemáticas y Química fueron las siguientes:
 - De 60 hombres, 35 aprobaron matemáticas
 - 57 personas aprobaron matemáticas

Con base en esta información elabora una tabla de contingencia.

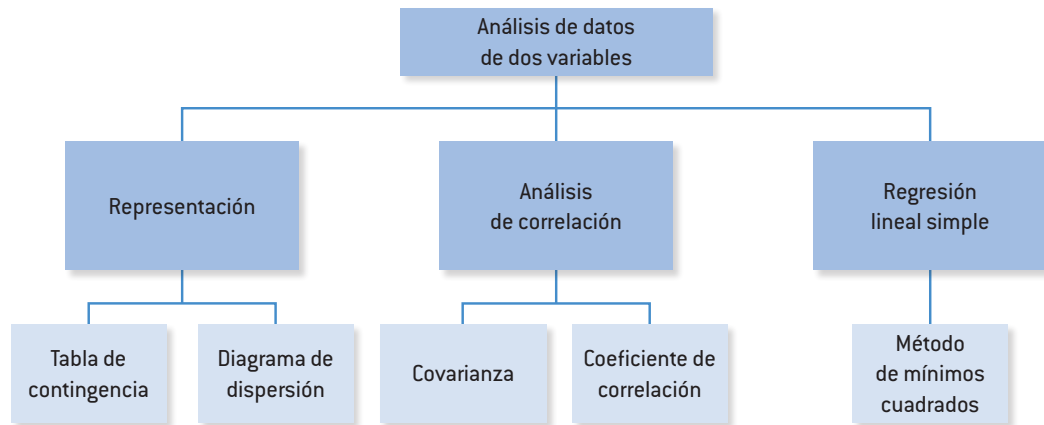
2. Dados los puntos $A(-1, 1)$ y $S(1, 5)$:
 - a) Determina la ecuación de la recta, indicando el valor de la pendiente y la ordenada al origen.
 - b) Traza la gráfica correspondiente.
3. A partir de la muestra siguiente determina la media y la desviación estándar:

3	5	1	-3	-2	-3
---	---	---	----	----	----

Si no tuviste problemas o dudas con los ejercicios anteriores, ahora sí comienza con el bloque 9.

Análisis de datos de dos variables

En este bloque estudiaremos la asociación entre dos variables, así como su representación gráfica. Además, haremos una introducción a un método que se utiliza para estimar (explicar) una variable: regresión lineal simple. En el esquema siguiente se muestran los temas que abordaremos.



Desempeños

Al terminar de estudiar este bloque y realizar todas las actividades propuestas habrás aprendido lo siguiente:

- Resolver problemas que impliquen el comportamiento de datos de dos variables en situaciones de tu propio interés en el ámbito escolar o personal.
- Representar datos de dos variables mediante la elaboración de tablas de contingencias y diagramas de dispersión.
- Analizar la representación tabular y gráfica del comportamiento de dos variables.
- A partir del análisis de la tendencia, utilizando el diagrama de dispersión y el cálculo del coeficiente de correlación, resolver problemas de tu entorno.

Representación

En esta sección nos centraremos en revisar la asociación que hay entre dos variables, a las que llamaremos X y Y . A veces se pueden representar los datos mediante una tabla de contingencia, o bien, simplemente como una lista de valores para cada variable.

Tabla de contingencia

La de contingencia es una tabla que se utiliza para clasificar el número de observaciones respecto a dos características o variables de interés. Las variables pueden ser cualitativas o cuantitativas (si tienes dudas acerca de estos conceptos, puedes repasar el bloque 1).

Alicia es una persona que gusta de la lectura y la estadística. Un día, se pregunta si podría asociar el número de páginas de un libro con su precio; es decir, considera que el precio de un libro depende del número de páginas.

Ejemplo 1

Para comprobarlo, decide recopilar la información correspondiente a ocho libros y construir una tabla de contingencia, lo cual queda de esta manera:

Título	Núm. de págs	Precio
El sueño del celta	464	\$269
El asedio	736	\$289
El perfume	232	\$188
Cartas a una joven matemática	238	\$219
La mano del fuego	368	\$219
El psicoanalista	574	\$165
Johnny Depp	192	\$149
El lector	208	\$113

Fuente: www.gandhi.com.mx

En este caso, “Título” es una variable cualitativa que sirve de identificador, en tanto que “Núm. de págs.” y “Precio” son variables cuantitativas.

Muchas veces no es necesario incluir en la tabla de contingencia variables cualitativas, sólo se presentan los valores de las variables cuantitativas; de esta forma, la tabla anterior quedaría como sigue:

Núm. de págs	Precio
464	\$269
736	\$289
232	\$188
238	\$219
368	\$219
574	\$165
192	\$149
208	\$113

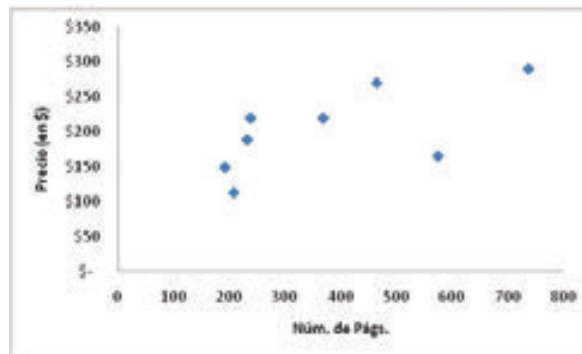
Diagrama de dispersión

Un diagrama de dispersión es una gráfica de puntos representados en el plano cartesiano. Cada punto indica un par de valores (x, y) . Este diagrama permite observar cómo se relacionan dos variables; generalmente, lo que se busca al usar un diagrama de este

tipo es determinar si los puntos siguen una línea recta y si ésta tiene pendiente positiva o negativa.

Ejemplo 2

Ya que Alicia ha recopilado la información sobre el precio de los libros y el número de páginas, realiza un diagrama de dispersión para determinar cómo es la relación que hay entre ambas variables. En el diagrama de dispersión, se asigna una variable al eje x y la otra al eje y . En el diagrama siguiente, nota que la variable “Núm de págs” se asignó al eje x y la variable “Precio (en \$)” al eje y :



Como se observa en la gráfica, podría haber una relación lineal y ésta sería positiva, ya que la posible recta sería creciente (pendiente positiva).

Análisis de correlación

Muchas veces, los diagramas de dispersión no muestran claramente si existe una relación lineal entre dos variables, por lo que es necesario medir el grado de asociación de ellas, es decir, calcular un valor numérico que indique el tipo de relación que hay entre ellas.

Los siguientes son los tipos de asociación lineal que puede haber entre dos variables:

- *Directa*, si la línea recta es creciente (pendiente positiva).
- *Inversa*, si la línea recta es decreciente (pendiente negativa).
- *Inexistente*, que es, obviamente, cuando no hay relación entre las variables (pendiente igual a cero).

Covarianza

La covarianza es una medida descriptiva que permite determinar el tipo de asociación lineal entre dos variables. La covarianza poblacional se obtiene mediante la fórmula:

$$\sigma_{xy} = \frac{\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{N}$$

donde x_i = valor i -ésimo de la variable x
 μ_x = media poblacional de la variable x
 y_i = valor i -ésimo de la variable y
 μ_y = media poblacional de la variable y
 N = tamaño de la población.

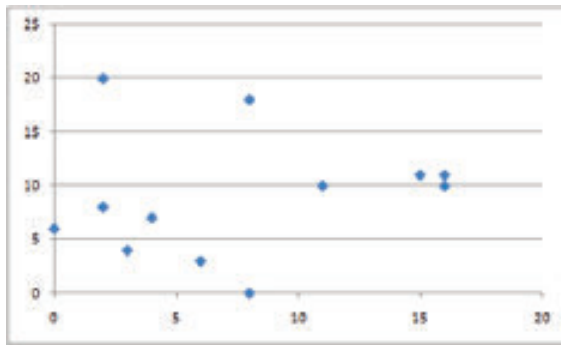
Por su parte, la covarianza muestral se calcula de la manera siguiente:

$$s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

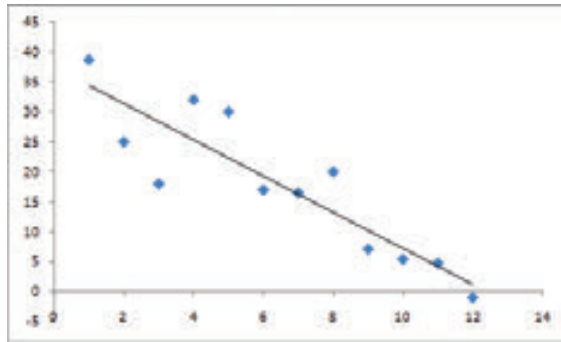
donde x_i = valor i -ésimo de la variable x
 $\bar{x}$ = media muestral de la variable x
 y_i = valor i -ésimo de la variable y
 $\bar{y}$ = media muestral de la variable y
 n = tamaño de muestra

La interpretación de este valor es muy sencilla:

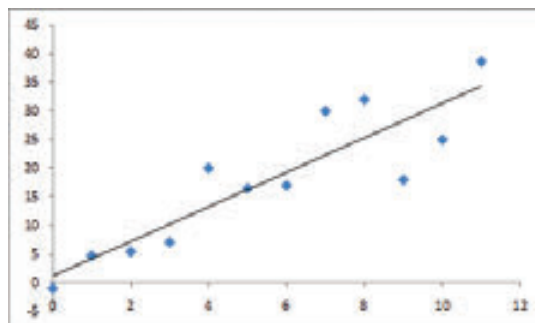
- Si s_{xy} es cero, no existe relación entre las variables.



- Si s_{xy} es negativo, hay una relación inversa entre las variables. Esto indica que cuando X aumenta de valor, Y disminuye, y viceversa.



- Si s_{xy} es positivo, hay una relación directa entre las variables. Esto indica que cuando el valor de X aumenta, el de Y también crece, y viceversa.



Ejemplo 3

Una vez que Alicia dibujó el diagrama de dispersión para relacionar el número de páginas y el precio de los libros, calcula la covarianza entre estas variables para determinar numéricamente, cómo es la relación lineal, si ésta existe, entre ellas. Como cree que el número de páginas determina el precio de un libro, entonces:

Núm. de págs.	Precio			
X	Y	$(X - \bar{X})$	$(Y - \bar{Y})$	$(X - \bar{X})(Y - \bar{Y})$
464	\$269	87.500	67.625	5917.188
736	\$289	359.500	87.625	31501.188
232	\$188	-144.500	-13.375	1932.688
238	\$219	-138.500	17.625	-2441.063
368	\$219	-8.500	17.625	-149.813
574	\$165	197.500	-36.375	-7184.063
192	149	-184.500	-52.375	9663.188
208	113	-168.500	-88.375	14891.188
$\Sigma(X - \bar{X})(Y - \bar{Y})$				54130.500

$$\bar{X} = 376.500$$

$$\bar{Y} = 201.375$$

Por tanto,

$$s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1} = \frac{54130.500}{8 - 1} = 7732.9286$$

Ahora, utilizando Excel:

# Pág	Precio
X	Y
464	\$269
736	289
232	188
238	219
368	219
574	165
192	149
208	113

=COVAR(B19:B26,C19:C26)
COVAR(matriz1, matriz2)

Se escribe =COVAR(datos X, datos Y), con lo que se obtiene el resultado de la covarianza poblacional: 6766.3125.

Interpretación. Como el valor de la covarianza es positivo, indica que la relación lineal entre el número de páginas y el precio del libro es directa, es decir, cuando aumenta el número de páginas, el precio del libro aumenta también.

Coeficiente de correlación

Aunque la covarianza indica el tipo de relación lineal que hay entre dos variables, no podemos saber la fortaleza de esa relación. Para eso debemos calcular otro valor, llamado *coeficiente de correlación*. El coeficiente de correlación se utiliza para medir la magnitud de la relación lineal entre dos variables, es decir, indica cuán fuerte o débil es una relación lineal.

Se denota con la letra r y también se le conoce como r de Pearson, en honor a Karl Pearson. Se calcula de esta forma:

$$r = \frac{s_{xy}}{s_x s_y}$$

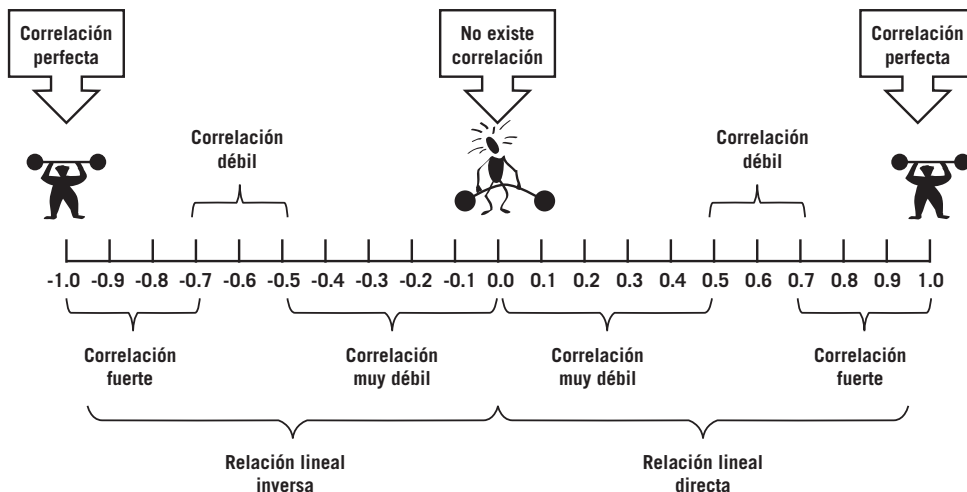
donde s_{xy} = la covarianza muestral entre las variables X y Y
 s_x = desviación estándar muestral de x
 s_y = desviación estándar muestral de y

Si lo que se desea (aunque no es muy usual) es determinar el coeficiente de correlación poblacional, la fórmula es la siguiente:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

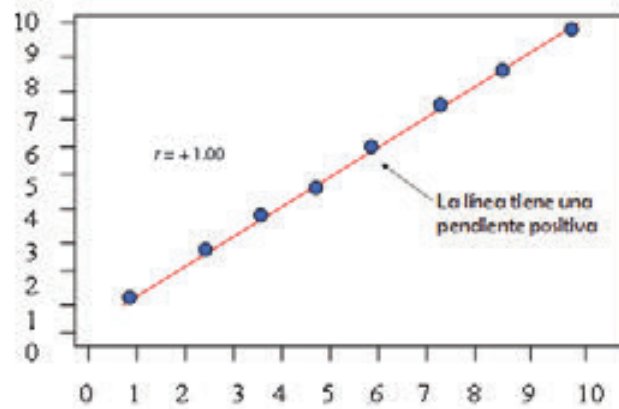
donde σ_{xy} = la covarianza poblacional entre las variables X y Y
 σ_x = desviación estándar poblacional de x
 σ_y = desviación estándar poblacional de y

El rango de valores está entre -1 y $+1$; los valores intermedios pueden interpretarse, de forma intuitiva, de esta manera:

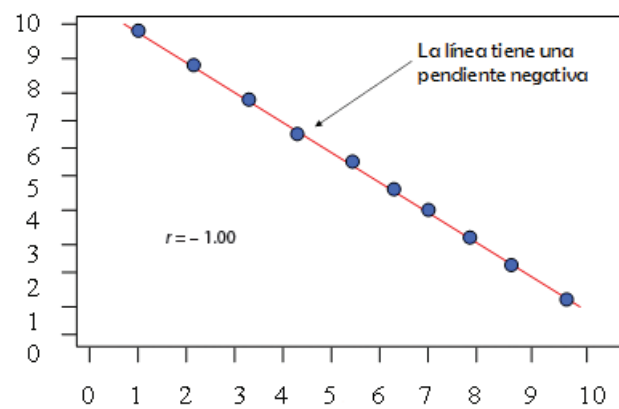


Gráficamente se verían así:

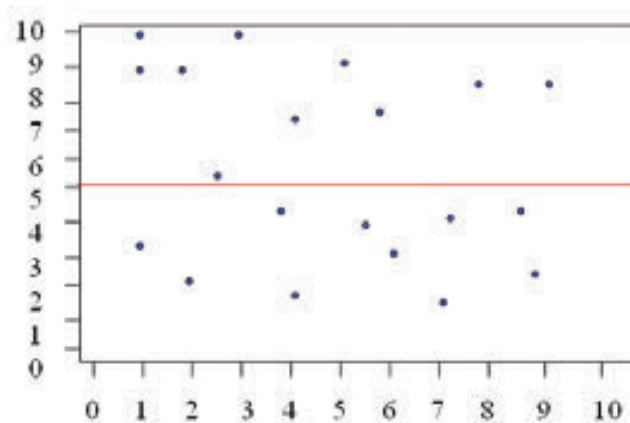
- Correlación positiva perfecta



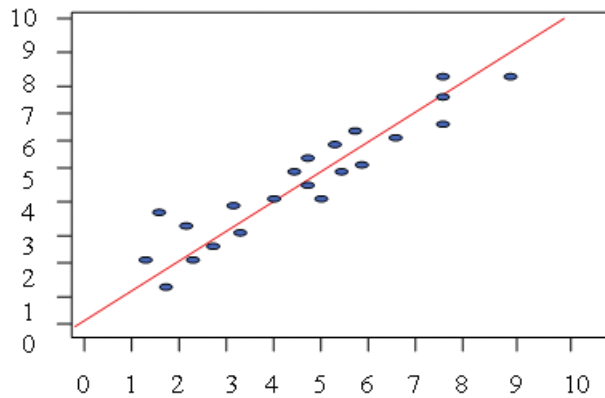
- Correlación negativa perfecta



- Correlación cero



► Correlación positiva fuerte

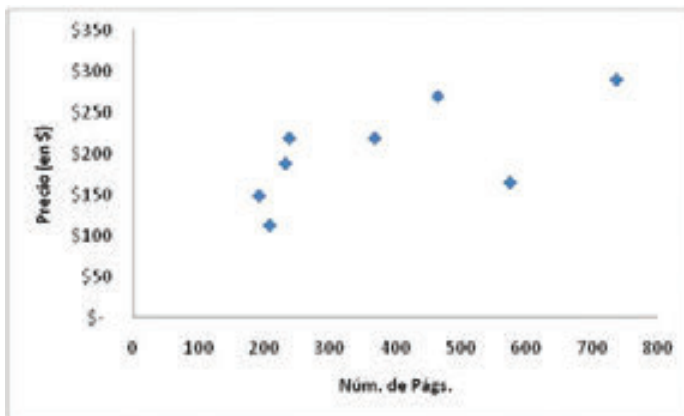


Ahora, Alicia calcula el coeficiente de correlación entre número de páginas y el precio de los libros para establecer cuán fuerte (o débil) es la relación lineal entre las variables:

Ejemplo 4

Datos:

- $s_{xy} = 7732.9286$
- $s_x = 199.4528$
- $s_y = 59.6752$



Entonces,

$$r = \frac{7732.9286}{(199.4528)(59.6752)} = 0.6497$$

Utilizando una hoja de cálculo como Excel, para calcular el coeficiente de correlación podemos emplear la función `=COEF.DE.CORREL(datos x, datos y)`, o bien, la función `=PEARSON(datos x, datos y)`, como se muestra a continuación:

Cuando calculas el coeficiente de correlación, no importa cuál variable hayas establecido como X y cuál como Y , ya que sólo estás midiendo el grado de asociación que hay entre ellas.

	A	B	C	D
16		# Pág	Precio	
17		X	Y	
18		464	\$269	
19		736	289	
20		232	188	
21		238	219	
22		368	219	
23		574	165	
24		192	149	
25		208	113	
26				
27		=COEF.DE.CORREL(B18:B25,C18:C25)		

Interpretación. El coeficiente de correlación obtenido es 0.6497, lo que indica que la asociación lineal entre el número de páginas y el precio del libro es directa (cuanto mayor el número de páginas, mayor el precio, y viceversa) y débil.

Coeficiente de determinación

Cuando lo que interesa es analizar una relación de causalidad entre dos variables, primero debemos definir cuál de ellas es la variable Y , variable dependiente, y cuál es la variable X , variable independiente.

La variable dependiente Y es la que se busca explicar; en términos estadísticos, es la que se busca estimar o pronosticar. A su vez, la variable independiente X es la que brinda información para explicar Y y recibe el nombre de *variable de predicción*.

Para saber si una variable X es “buena” para explicar la variable Y se calcula el coeficiente de determinación, que representaremos con r^2 y que tiene las características siguientes:

- Es el cuadrado del coeficiente de correlación.
- Su rango de valores está entre 0 a 1.
- No da ninguna información sobre la dirección de la relación entre las variables.

Cuanto más cerca esté de 1, la variable independiente X será una buena variable para explicar Y . Es decir, es un factor determinante para Y . En contraparte, conforme r^2 se acerca a 0, indica que X no es un factor significativo para explicar Y .

El coeficiente de determinación puede expresarse como porcentaje.

Ejemplo 5

Siguiendo con el ejemplo, ahora Alicia calcula el coeficiente de determinación:

$$r^2 = 0.4221$$

Interpretación. El coeficiente de determinación indica que la variable “Núm. de páginas” explica el 42.21% de la variabilidad del “Precio del libro”. En otras palabras, para poder explicar completamente el precio de un libro debemos considerar otros factores, no sólo el número de páginas.

Regresión lineal simple

El término *regresión* fue introducido por Francis Galton (1822-1911). En un famoso artículo, Galton planteó que, a pesar de que había una tendencia en la que los padres de estatura alta tenían hijos altos y padres de estatura baja tenían hijos bajos, la estatura promedio de los niños nacidos de padres de una estatura dada tendía a moverse o *regresar* hacia la estatura promedio de la población total. En palabras de Galton, se trataba de una regresión a la mediocridad.

Fue tal el éxito de la contribución de las investigaciones de Galton que originaron la *ley de regresión universal*, que fue confirmada por Karl Pearson (1857-1936) y que constituye el pilar del análisis de regresión.

En general, el análisis de regresión se centra en la exploración, explicación y estudio de dependencia de una variable mediante una o más variables explicativas, de ahí el nombre del *método de predicción* basado en este modelo.

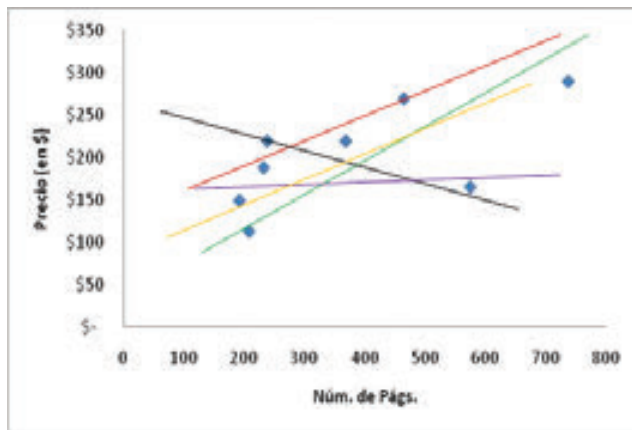
En este tema haremos una introducción al método de regresión lineal simple. Recibe este nombre, porque:

- *Regresión*: utilizaremos información pasada
- *Lineal*: bajo el supuesto de que entre dos variables (X y Y) existe una relación lineal
- *Simple*: usaremos sólo una variable independiente para tratar de explicar la variable dependiente.

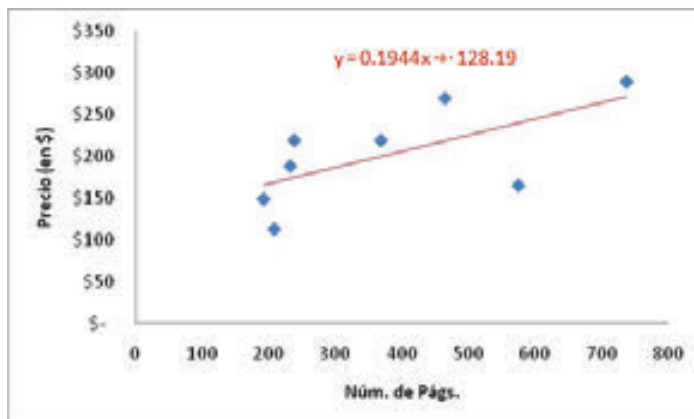
En otras palabras, ajustaremos una recta a los datos. “Ajustar” se refiere a construir la única recta que pase lo más cerca de todos los puntos ubicados en el diagrama de dispersión.

Alicia decide ahora determinar la ecuación de la recta y para ello comienza trazando rectas en un diagrama de dispersión, como se muestra enseguida:

Ejemplo 6



Trazar rectas es una forma muy lenta de encontrar la que pase lo más cerca de todos los puntos. ¿Cómo entonces podría determinarse esa recta? Precisamente, mediante el método de mínimos cuadrados:



Método de mínimos cuadrados

El análisis de regresión consiste en definir la variable independiente X que ayude a explicar (estimar) la variable dependiente Y , siempre que exista una relación lineal entre ellas, además de que ambas variables deben ser cuantitativas.

El método de mínimos cuadrados se usa para determinar la ecuación de la recta de regresión, es decir, por medio de él se encuentra la única recta que pasa lo más cerca que se puede de todos los puntos (observaciones) ubicados en un diagrama. La ecuación del método de mínimos cuadrados es

$$Y' = b_1 X + b_0$$

donde Y' = valor estimado (aproximado) de Y

b_0 = ordenada al origen; es el valor de Y' cuando X es igual a cero.

b_1 = pendiente de la recta; es el cambio en Y' cuando X aumenta en una unidad.

Para obtener la pendiente y la ordenada al origen de dicha recta, llamada *recta de regresión* o *de mínimos cuadrados*, hay que minimizar el término:

Recuerda: "minimizar" se refiere a derivar la expresión

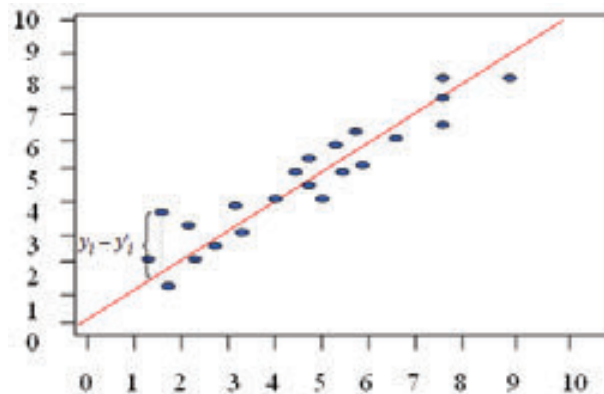
$$\sum_{i=1}^n (Y_i - Y'_i)^2$$

para obtener los puntos críticos y, después, determinar cuál es mínimo, mediante el criterio de la segunda derivada.

$$\sum_{i=1}^n (Y_i - Y'_i)^2$$

donde Y_i = el valor i de Y (valor real)

Y'_i = el valor i estimado de Y (valor sobre la recta de regresión), es decir, es la distancia que hay entre cada punto y la recta de regresión. Gráficamente,



Recuerda: con el método de mínimos cuadrados se calcula lo que comúnmente se conoce como la *recta del mejor ajuste*. El método subjetivo no es confiable.

Minimizando esas distancias se obtienen b_0 , la ordenada al origen, y b_1 , la pendiente de la recta. Así, las ecuaciones para determinar b_0 y b_1 son

$$b_1 = \frac{rs_y}{s_x} \text{ y } b_0 = \bar{Y} - b_1 \bar{X}$$

donde r = coeficiente de correlación

s_y = desviación estándar muestral de Y

s_x = desviación estándar muestral de X

$\bar{Y}$ = media muestral de Y

$\bar{X}$ = media muestral de X

A b_0 y b_1 se les denomina *coeficientes de regresión*.

Observa que para calcular la pendiente se necesita del coeficiente de correlación r . Por ello, la pendiente y el coeficiente de correlación tienen el mismo signo.

De esta forma, utilizando la ecuación de regresión se estiman los valores de Y , es decir, sólo debemos sustituir la variable independiente con algún valor y realizar los cálculos correspondientes.

Ahora sí, como Alicia ya aprendió el método de mínimos cuadrados, puede determinar la ecuación de la recta para el precio del libro.

Datos:

- $\bar{X} = 376.5$
- $\bar{Y} = 201.375$
- $s_x = 199.4528$
- $s_y = 59.6752$
- $r = 0.6407$

Se sustituyen los valores en la fórmula y queda:

$$b_1 = \frac{rs_y}{s_x} = \frac{0.6497(59.6752)}{199.4528} = 0.1944$$

y

$$b_0 = \bar{Y} - b_1\bar{X} = 201.375 - (0.1944)(376.5) = 128.19$$

La ecuación de regresión es:

$$Y' = 0.1944X + 128.19$$

Si quieres calcular la ecuación de regresión con Excel, utiliza las funciones =INTERSECCION.EJE(*datos y, datos x*), de este modo:

Excel screenshot showing the formula bar with `=INTERSECCION.EJE(C18:C25,B18:B25)` and the resulting value `b1= 0.1944` in cell F20. The data table is as follows:

# Pág	Precio
X	Y
464	\$269
736	289
232	188
238	219
368	219
574	165
192	149
208	113

y =PENDIENTE(*datos y, datos x*)

Excel screenshot showing the formula bar with `=PENDIENTE(C18:C25,B18:B25)` and the resulting value `b1= 0.1944` in cell F19. The data table is the same as in the previous screenshot.

# Pág	Precio
X	Y
464	\$269
736	289
232	188
238	219
368	219
574	165
192	149
208	113

Ejemplo 7

Sólo ten presente que este modelo no es muy bueno. Su correlación es débil; por tanto, la estimación del precio puede ser muy distinta de la del precio real.

Interpretación. Cuando el libro tiene cero páginas, el precio estimado es de \$128.19. Por cada página que aumente el libro, el precio estimado aumenta en \$0.1944.

Si se tiene un libro con 666 páginas, entonces el precio estimado será de

$$Y' = 0.1944(666) + 128.19 = 257.66$$

es decir, de \$257.66.

Es muy importante señalar que para validar este “modelo” deben realizarse pruebas estadísticas, cuya explicación rebasa el ámbito de esta obra (se contempla en cursos superiores).

En muchas ocasiones se utiliza el método de mínimos cuadrados para determinar la tendencia de una serie de tiempo. Una serie de tiempo es una lista de datos ordenados por fecha; por ejemplo, la cotización del dólar o la tasa de interés a lo largo de varios días.

La tendencia es el crecimiento o decrecimiento del nivel medio de la serie a largo plazo; además de que es uno de los componentes más importantes para pronosticar una serie de tiempo. Pero todo esto se estudia en cursos posteriores de Estadística.

Ejemplo 8

La tasa de interés interbancaria de equilibrio (TIIE) fue creada por el Banco de México en 1996 a fin de establecer una tasa que lograra representar de manera más fiel las condiciones del mercado. Se establece según las cotizaciones de las tasas de interés a diferentes plazos de al menos seis entidades bancarias, por lo que constituye una buena referencia de análisis de los créditos bancarios, ya que ofrece un panorama muy amplio de las cotizaciones de las tasas de interés que establecen los bancos del país.

Con base en la información mensual de 2010 y 2011, determina la tendencia para la TIIE.

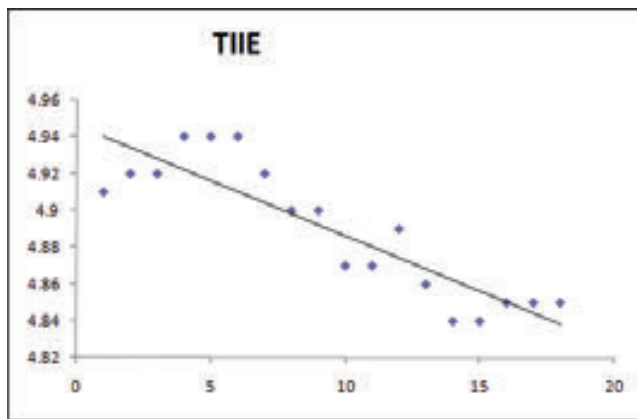
Solución

Lo primero que debemos hacer es asignar un número consecutivo a cada mes. Ese número se denota con *t* y representa el periodo al que corresponde cada mes:

Periodo		t	TIIE 28 días
2010	Enero	1	4.91
	Febrero	2	4.92
	Marzo	3	4.92
	Abril	4	4.94
	Mayo	5	4.94
	Junio	6	4.94
2010	Julio	7	4.92
	Agosto	8	4.90
	Septiembre	9	4.90
	Octubre	10	4.87
	Noviembre	11	4.87
	Diciembre	12	4.89
2011	Enero	13	4.89
	Febrero	14	4.84
	Marzo	15	4.84
	Abril	16	4.85
	Mayo	17	4.85
	Junio	18	4.85

Fuente: Banco de México. *Estadísticas. Sistema financiero. Tasas y precios de referencia. Costos de captación.*

Trazamos la gráfica de dispersión correspondiente, tomando como variable independiente t y como variable dependiente la TIIE:



Calculamos e interpretamos los coeficientes de correlación y de determinación:

- $r = -0.8839$ La relación lineal entre la TIIE y el periodo es fuerte e inversa, es decir, conforme pasa el tiempo la TIIE disminuye
- $r^2 = 0.7813$ El periodo de tiempo explica el 78.13% de la variabilidad de la TIIE

Determinamos la pendiente y la ordenada al origen mediante el método de mínimos cuadrados:

$$b_1 = 0.0060 \text{ y } b_0 = 4.9464$$

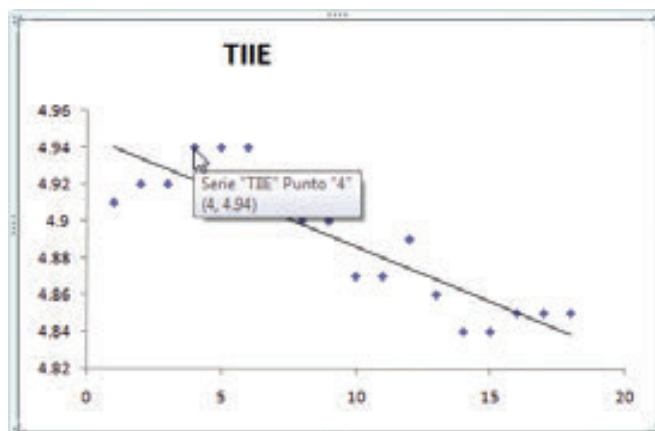
Por tanto, la ecuación de regresión es:

$$Y' = -0.0060t + 4.9464$$

es decir, inicialmente la TIIE es de 4.9464 puntos porcentuales. Por cada periodo (cada mes) que pase, la TIIE estimada disminuirá 0.006 puntos porcentuales.

Hay una forma de determinar la ecuación de regresión. Sigue estos pasos:

1. Cuando se traza la gráfica de dispersión en Excel, se coloca el cursor sobre alguno de los puntos y se hace clic con el botón izquierdo del ratón. Con esto se marcan todos los puntos de la gráfica:



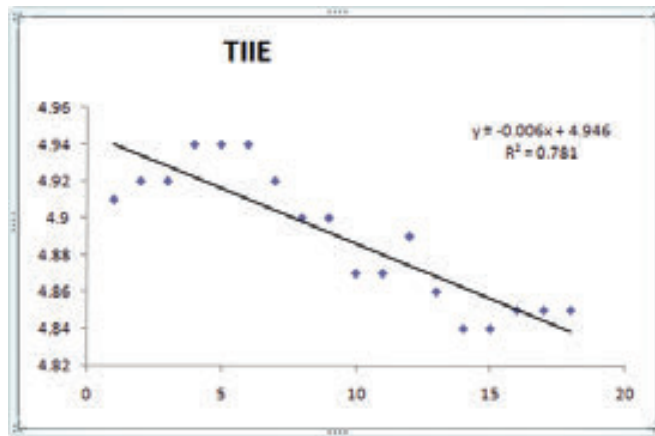
- Se hace clic con el botón derecho del ratón para abrir un menú contextual, del que se elige la opción “Agregar línea de tendencia”:



- Se abre una ventana, en la que se marca que el tipo de tendencia sea lineal; en la parte de abajo se pide presentar la ecuación en la gráfica y el valor R cuadrado:

The screenshot shows the 'Formato de línea de tendencia' (Trendline Format) dialog box. On the left, there are tabs for 'Opciones de línea de tendencia', 'Color de línea', 'Estilo de línea', and 'Sombra'. The 'Opciones de línea de tendencia' tab is active. Under 'Tipo de tendencia o regresión', the 'Lineal' option is selected. Below this, the 'Nombre de la línea de tendencia' section shows 'Automático' selected, resulting in the name 'Lineal (TIIE)'. The 'Extrapolar' section has 'Adelante: 0.0' and 'Hacia atrás: 0.0' periods. At the bottom, the 'Señalar intersección = 0.0' checkbox is unchecked, while the 'Presentar ecuación en el gráfico' and 'Presentar el valor R cuadrado en el gráfico' checkboxes are checked. A 'Cerrar' button is at the bottom right.

4. De esta manera obtenemos la ecuación de regresión y el coeficiente de determinación, y como se ve son los mismos resultados:



Ejercicios

I. TEORÍA Y ALGO MÁS

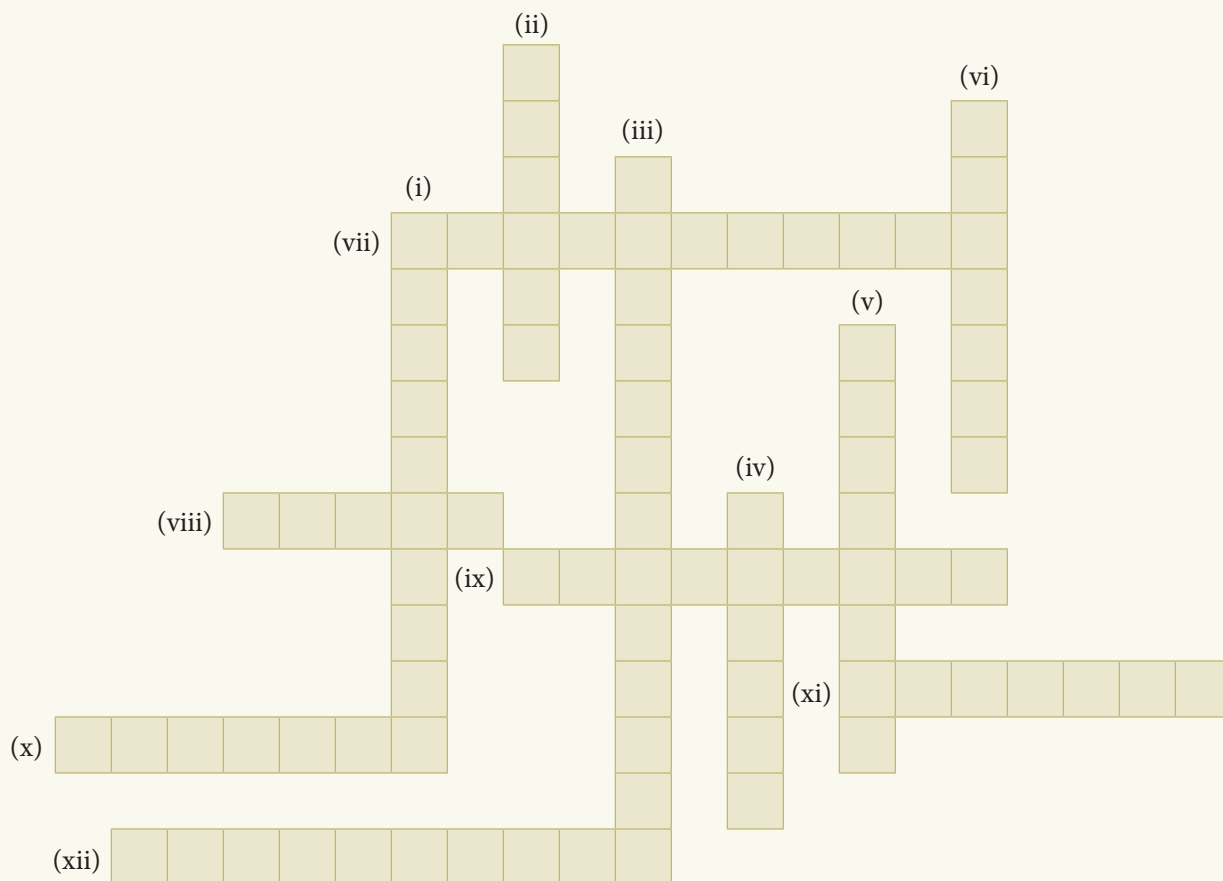
1. Completa el crucigrama:

Verticales

- i. Medida descriptiva que permite determinar la asociación lineal entre dos variables.
- ii. Sea $r = -0.85$, lo que indica que la relación lineal entre dos variables es inversa y _____.
- iii. Coeficiente que indica el porcentaje que se explica de la variable Y.
- iv. Sea $r = 1$ o $r = -1$, lo que indica que la relación _____ entre las variables X y Y es perfecta.
- v. En la ecuación de regresión, b_0 es la _____ al origen.
- vi. El método para determinar la “mejor” recta se llama de _____ cuadrados.

Horizontales

- vii. Coeficiente que indica la magnitud de la relación lineal entre dos variables.
- viii. Sea $r = -0.4$, lo que indica que la relación lineal entre dos variables es inversa y _____.
- ix. En la ecuación de regresión, b_1 es la _____.
- x. Sea $r < 0$, lo que indica que la relación entre dos variables es _____.
- xi. Sea $r > 0$, lo que indica que la relación entre dos variables es _____.
- xii. Diagrama que permite observar gráficamente la relación entre dos variables.



II. PIENSO, ... ¿Y LUEGO?

2. En el artículo “Lady Gaga: la elegida del pop”, escrito por Fausto Alzati Fernández (Fernández, 2011) se establece que, hoy en día, la atención es un factor que

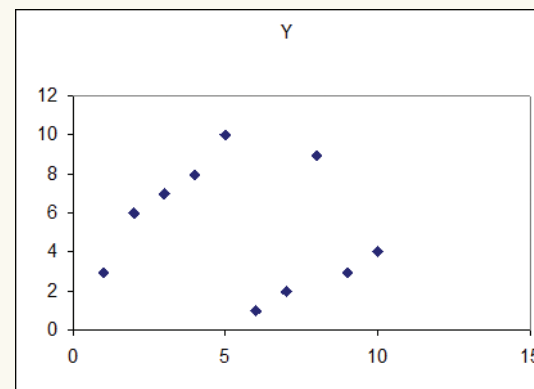
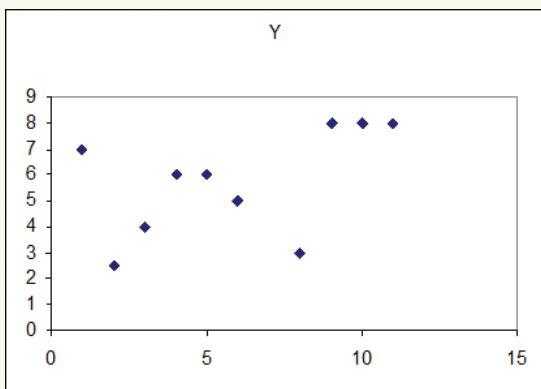
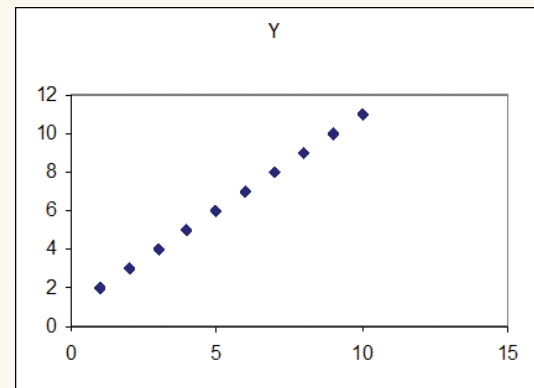
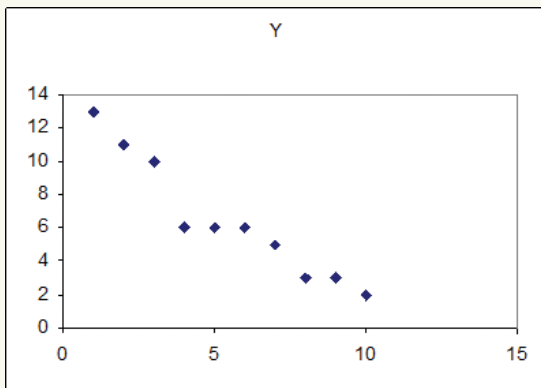
pretende medirse más que en cualquier otro punto de la historia, debido al modo en que las tecnologías dan cuenta de las tendencias y movimientos de la atención global y a la cantidad de horas que un número creciente de personas nos encontramos enchufadas a algún tipo de medio. Pasa que el vertiginoso ascenso de Gaga a la cumbre de la celebridad, en sólo tres años desde su álbum debut, titulado con y sin ironía *The Fame* (2008), tiene una correlación significativa con la tecnología. Gaga es la primera gran diva del pop en la era de las redes sociales, pero más aún, gran parte de su pronta y exuberante fama se debe a que ella apela a la generación de nativos de la red.

- ¿Consideras que el éxito de un artista tiene una relación “fuerte y directa” con la tecnología? ¿Hay otros factores que influyen considerablemente en el éxito de un artista? Explica tu respuesta.
- En nuestro país, un grupo que surgió de las redes sociales es Austin TV. ¿Qué papel desempeñan actualmente las redes sociales en el logro de la fama de un artista? ¿Cómo podrías usarlas tú para alcanzar alguno de tus objetivos? Forma un equipo, comenten al respecto, obtengan conclusiones y preséntenlas al grupo.
- Pese al éxito de intérpretes como Lady Gaga, en la actualidad hay una tendencia a lo *retro*. ¿Qué factores existen para que artistas de la década de 1980 estén hoy “de moda”?

III. PURO DEPORTE

3. *Sólo para practicar.* A continuación se muestran cuatro coeficientes de correlación. Analiza cada uno de ellos y decide a qué diagrama de dispersión corresponden:

- a) $r = -0.17$ b) $r = 0.505$ c) $r = 1$ d) $r = -0.96$



4. *Sólo para practicar.* Las observaciones siguientes en una muestra se seleccionaron al azar:

X	7	8	9	11	6
Y	2	1	2	1	3

- a) Traza el diagrama de dispersión correspondiente.
- b) Calcula los coeficientes de correlación y de determinación. Interpreta el resultado.
- c) Determina la ecuación de regresión. Interpreta la pendiente y la ordenada al origen.

5. *Sólo para practicar.* Las observaciones siguientes en una muestra se seleccionaron al azar:

X	6	10	11	12	15
Y	4	6	10	16	20

- a) Traza el diagrama de dispersión correspondiente.
- b) Calcula los coeficientes de correlación y de determinación. Interpreta.
- c) Determina la ecuación de regresión. Interpreta la pendiente y la ordenada al origen.

6. *Negocios.* Una empresa desea desarrollar un modelo para pronosticar sus ventas con base en los gastos que realiza en publicidad. La siguiente es la información con la que cuenta:

Ventas (millones de pesos)	Gastos de publicidad (millones de pesos)
2	4
3	5
4	4
5	6
7	7
6	6
5	7
8	9
9	11

- a) Traza el diagrama de dispersión correspondiente.
- b) Calcula los coeficientes de correlación y de determinación. Interpreta el resultado.
- c) Determina la ecuación de regresión. Interpreta los coeficientes de regresión.

7. *Crecimiento económico.* Los políticos han coincidido últimamente en que la educación es el camino para lograr el desarrollo del país. Con esta finalidad, se presentan a continuación los datos sobre años promedio de educación en 17 países de América Latina, así como el producto interno bruto per cápita (en dólares) para cada uno de esos países:

País	Promedio educativo (años), 2010	PIB per cápita (2010)
Argentina	9.3	14 931
Bolivia	9.2	4 502

Brasil	7.2	10 847
Chile	9.7	14 780
Colombia	7.4	8 959
Costa Rica	8.3	11 143
Ecuador	7.6	8 170
El Salvador	7.7	6 660
Guatemala	4.1	4 761
Honduras	6.5	3 845
México	8.7	14 192
Nicaragua	5.7	2 632
Panamá	9.4	13 210
Paraguay	7.8	4 629
Perú	9.6	9 016
Uruguay	8.4	14 022
Venezuela	6.2	11 820

Fuente: <http://hdr.undp.org>

- Traza el diagrama de dispersión correspondiente.
 - Calcula los coeficientes de correlación y de determinación. Interpreta.
 - Determina la ecuación de regresión. Interpreta los coeficientes de regresión.
 - Junto con tus compañeros, discute si la educación es el único medio para lograr el desarrollo del país, o si considerarían otros aspectos para lograrlo.
 - Investiguen cuántos años de educación se tienen en promedio países como Alemania, Estados Unidos y Corea del Sur. ¿Cómo podría nuestro país alcanzarlos en este rubro?
8. *Seguridad.* Se desea realizar un estudio para ver de qué manera se puede reducir el número de delitos en el país. Se ha considerado que el número de policías al servicio de la sociedad depende del número de delitos cometidos en cada entidad federativa. Para empezar, se cuenta con la información siguiente:

Entidad federativa	Víctimas de algún delito por entidad federativa de residencia de la víctima	Estado por fuerza por cada mil habitantes
Aguascalientes	105 889	2.9
Baja California Norte	282 001	3.15
Baja California Sur	46 907	5.56
Campeche	31 568	3.04
Coahuila	154 452	2.05
Colima	33 788	4.47

Chiapas	77 963	2.97
Chihuahua	278 589	3.01
Distrito Federal	1 358 393	10.25
Durango	81 256	3.03
Guanajuato	329 407	2.56
Guerrero	93 879	4.27
Hidalgo	59 517	3.01
Jalisco	457 127	3.34
México	1 323 259	4.48
Michoacán	246 062	2.72
Morelos	92 144	3.60
Nayarit	39 380	3.36
Nuevo León	265 635	2.40
Oaxaca	130 188	3.24
Puebla	368 371	2.83
Querétaro	88 968	2.13
Quintana Roo	106 691	4.23
San Luis Potosí	137 182	3.41
Sinaloa	126 053	3.05
Sonora	218 599	3.16
Tabasco	100 104	4.64
Tamaulipas	127 483	2.92
Tlaxcala	49 588	3.53
Veracruz	261 305	2.59
Yucatán	94 889	3.71
Zacatecas	52 092	2.53

Fuente: INEGI. Resultados preliminares (censo poblacional 2010) e INEGI. Encuesta nacional sobre inseguridad.

- Realiza un análisis de correlación. Expón tus conclusiones. ¿Encuentras lógico el resultado obtenido?
- Lleva a cabo un análisis de regresión. ¿A qué conclusiones llegas?
- Supón que una entidad federativa registra 161 072 delitos, ¿cuántos policías debe tener?
- A partir de los resultados anteriores, organiza una mesa redonda con el tema a discutir “¿Cómo disminuir el número de delitos?” Propón otros factores que consideres importantes para lograr ese objetivo.

9. *Democracia.* Se realizó un estudio para estimar cómo se percibe la democracia, para lo que se creó el índice de confianza en los partidos políticos. La información que se recabó es la siguiente:

País	Confianza en los partidos políticos	Percepción de la democracia (promedio simple. Escala de 1 a 10, donde 1 = nada y 10 totalmente democrático)
Argentina	14.6	5.8
Bolivia	20.2	5.5
Brasil	20.4	6.0
Chile	21.8	6.0
Colombia	25.3	6.1
Costa Rica	23.9	6.5
Ecuador	14.4	6.0
El Salvador	38.3	5.4
Guatemala	21.6	5.2
Honduras	16.2	5.5
México	18.7	5.0
Nicaragua	14.3	5.4
Panamá	31.2	5.9
Paraguay	31.2	5.0
Perú	11.2	5.0
Uruguay	46.1	7.1
Venezuela	33.8	6.6

Fuente: Comisión Económica para América Latina y el Caribe (CEPAL). Cálculos realizados sobre la base de las encuestas de opinión de la Corporación Erobarómetro/Latinbarómetro

- Traza el diagrama de dispersión correspondiente.
- Calcula los coeficientes de correlación y de determinación. Interpretalos.
- Determina la ecuación de regresión. Interpreta los coeficientes de regresión.
- ¿Cuál sería el índice de percepción de la democracia si un país presenta un índice de confianza en los partidos políticos de 30 puntos?
- Investiga cuáles son las características de un gobierno democrático para, posteriormente, discutir con tus compañeros sobre qué otras variables habría que considerar para estimar la percepción de la democracia. Además, ¿qué acciones se deben tomar en México para aumentar la percepción de la democracia?

10. *Educación.* Se desea analizar el número de años promedio de escolaridad en función del gasto educativo. Los datos se presentan a continuación:

Entidad federativa	Gasto total en educación (millones de pesos)	Grado promedio de escolaridad de la pobla- ción de 15 años y más
Aguascalientes	3 039.0	9.2
Baja California Norte	6 958.0	9.3
Baja California Sur	2 060.4	9.4
Campeche	2 908.5	8.5
Coahuila	6 735.8	9.5
Colima	2 044.7	9.0
Chiapas	11 351.8	6.7
Chihuahua	6 823.7	8.8
Distrito Federal	21 850.2	10.5
Durango	4 857.1	8.6
Guanajuato	8 661.6	7.7
Guerrero	11 016.5	7.3
Hidalgo	7 464.8	8.1
Jalisco	11 376.3	8.8
México	21 003.9	9.1
Michoacán	10 330.4	7.4
Morelos	4 046.6	8.9
Nayarit	3 340.2	8.6
Nuevo León	7 332.0	9.8
Oaxaca	11 757.6	6.9
Puebla	9 863.0	8.0
Querétaro	3 791.5	8.9
Quintana Roo	3 166.3	9.1
San Luis Potosí	7 091.9	8.3
Sinaloa	5 818.1	9.1
Sonora	6 010.3	9.4
Tabasco	5 082.6	8.6
Tamaulipas	8 152.4	9.1
Tlaxcala	2 893.4	8.8
Veracruz	16 352.9	7.7
Yucatán	4 146.7	8.2
Zacatecas	4 657.3	7.9

- a) Traza el diagrama de dispersión correspondiente.
- b) Calcula los coeficientes de correlación y de determinación. Interpretalos.
- c) Determina la ecuación de regresión. Interpreta los coeficientes de regresión. ¿Te parecen lógicos los resultados?
- d) Supón que en el próximo presupuesto de egresos de la federación se decide otorgar a un estado la cantidad de 12 000 millones de pesos para mejorar la educación. ¿Cuál sería el grado de escolaridad estimado?
- e) De acuerdo con los resultados anteriores, realiza un debate sobre cómo optimar el uso de los recursos para la educación y que esto se vea reflejado en mejores resultados en evaluaciones como la de PISA y ENLACE.

11. Divisas. A continuación se presenta la cotización promedio mensual del dólar a la compra:

Mes	Compra	Mes	Compra
enero 2009	13.9	abril 2010	11.80
febrero 2009	14.6	mayo 2010	12.50
marzo 2009	13.8	junio 2010	12.35
abril 2009	13.5	julio 2010	12.20
mayo 2009	12.8	agosto 2010	12.80
junio 2009	12.8	septiembre 2010	12.20
julio 2009	12.8	octubre 2010	12.00
agosto 2009	12.9	noviembre 2010	12.00
septiembre 2009	13.1	diciembre 2010	11.98
octubre 2009	12.8	enero 2011	11.70
noviembre 2009	12.5	febrero 2011	11.70
diciembre 2009	12.5	marzo 2011	11.58
enero 2010	12.6	abril 2011	11.20
febrero 2010	12.4	mayo 2011	11.15
marzo 2010	11.9	julio 2011	11.30

- a) Asigna a cada mes el periodo t correspondiente y grafica los datos. ¿Cómo es la tendencia de éstos?
- b) Determina la ecuación de la tendencia por medio del método de mínimos cuadrados.
- c) Pronostica la cotización del dólar a la compra para los meses de octubre, noviembre y diciembre de 2011.
- d) Forma un equipo y escribe con tus compañeros un ensayo sobre las ventajas y desventajas de una devaluación del peso frente al dólar.

12. *Demografía.* De acuerdo con los censos realizados por el INEGI, el crecimiento de la población en México se presenta a continuación:

Año	Población total (miles de habitantes)
1950	25 791
1960	34 923
1970	48 225
1980	68 776
1990	81 250
2000	97 483
2010	112 337

Fuente: <http://www.inegi.org.mx>

- Traza la gráfica correspondiente. ¿Cómo es la tendencia de los datos?
 - Calcula la ecuación de regresión para explicar la tendencia del crecimiento de la población.
 - ¿Cuál es el número estimado de habitantes para el 2015?
 - De acuerdo con los resultados obtenidos en los incisos anteriores, ¿por qué crees que es importante estimar el crecimiento poblacional? Además, consulta el artículo en <http://www.reforma.com/nacional/articulo/615/1229162/> y expón tus conclusiones ante el grupo.
13. *Salud.* Diversos estudios científicos han mostrado que el dormir poco es un factor para el sobrepeso. Para poder comprobarlo realiza el ejercicio siguiente:
- Elige una muestra de 15 a 20 personas.
 - Registra el número de horas que duermen en promedio, así como su peso.
 - Con esta información, analiza los coeficientes de correlación y de determinación; traza el diagrama de dispersión correspondiente. Interpreta los coeficientes.
 - Determina la ecuación de regresión e interpreta los coeficientes.
 - Con la ecuación de regresión, calcula los pesos estimados utilizando las horas de sueño que tienes registradas y compara. ¿Se parece tu estimación al valor real?
 - ¿Qué conclusiones obtienes? ¿Qué otros factores influyen en el peso de una persona? ¿Qué acciones deben tomarse para evitar el sobrepeso? ¿Qué consecuencias tiene el que un porcentaje considerable de personas presenten sobrepeso/obesidad?
14. *Desempeño escolar.* Para comprobar si el tiempo que dedica una persona a estudiar se ve reflejado en buenos resultados en sus calificaciones realiza lo siguiente:
- Elige una materia; por ejemplo, Matemáticas.
 - Elige una muestra de 15 a 20 personas.
 - Registra el tiempo promedio de estudio para un examen, así como la calificación obtenida.
 - Con esta información, analiza los coeficientes de correlación y de determinación; traza el diagrama de dispersión correspondiente. Interpreta los coeficientes.
 - Determina la ecuación de regresión e interpreta los coeficientes.
 - Con la ecuación de regresión, calcula las calificaciones estimadas utilizando la información que tienes sobre tiempo de estudio y compara tus resultados. ¿Se parece la estimación al valor real?
 - ¿A qué conclusiones llegas? ¿Qué otros factores influyen para obtener buenas calificaciones?

15. *Economía*. La teoría keynesiana¹ establece que “cuando el ingreso aumenta, el consumo también lo hace, aunque en una proporción inferior al incremento de este ingreso”. Para probar esta teoría:
- a) Elige una muestra de 15 a 20 personas.
 - b) Registra el consumo que realiza una persona, así como el ingreso que tiene en una semana.
 - c) Con esta información, analiza los coeficientes de correlación y de determinación; traza el diagrama de dispersión correspondiente. Interpreta los coeficientes.
 - d) Determina la ecuación de regresión e interpreta los coeficientes.
 - e) De acuerdo a los resultados que obtuviste, ¿estás de acuerdo con lo que establece la teoría keynesiana?

¹ Teoría económica basada en las ideas de John Maynard Keynes.

Evaluación sumativa

Un escalón hacia arriba

Contesta los reactivos siguientes, lo que te permitirá establecer si has comprendido cabalmente lo expuesto en este bloque o debes repasar algún tema. ¡Adelante!

1. Contesta verdadero o falso a las siguientes cuestiones. En caso de ser falso, da la respuesta correcta.
 - a) Por el método de mínimos cuadrados se encuentran las rectas de ajuste que cumplen con la condición de que la distancia de todos los puntos a ella sea mínima.
 - b) El valor r^2 para un caso vale 0.81; entonces, la correlación entre las variables puede ser -0.9 o +0.9.
 - c) La recta de regresión $Y' = 50 - 2X$ significa que si X se incrementa en 1, Y' disminuye en 48 unidades.
2. ¿Cuál de las situaciones siguientes es verdadera? Justifica tu respuesta.
 - a) $Y' = 499 + 0.21X$ $r = 0.75$
 - b) $Y' = 100 + 0.9X$ $r = -0.70$
 - c) $Y' = -20 - 1.1X$ $r = 0.40$
 - d) $Y' = -7 - 4.2X$ $r = 0.90$
3. Una empresa quiere estimar las ventas de su producto con base en los minutos de publicidad que contrata en radio. La información es la siguiente:

Publicidad (minutos)	Ventas (miles de \$)
4	3.0
9	3.2
7	2.9
5	2.3
8	3.5

- a) ¿Cuál es la variable independiente y cuál la dependiente?
- b) Calcula:
 - i. La media y la desviación estándar muestral de cada variable.
 - ii. La covarianza. Interpreta.
 - iii. El coeficiente de correlación. Interpreta.
 - iv. El coeficiente de determinación. Interpreta.
- c) Determina la ecuación de regresión e interpreta los coeficientes de regresión.
- d) Si la empresa decide contratar seis minutos de publicidad en radio, ¿a cuánto ascienden las ventas estimadas?

Proyecto 9

ACCIDENTES DE TRÁNSITO

Un accidente de tránsito, accidente automovilístico o siniestro de tráfico es un percance en el que se ve implicado al menos un automóvil u otro tipo de vehículo de transporte por carretera. Los tipos de colisiones que se han presentado son: con vehículo automotor, con peatón (atropellamiento), con animal, con objeto fijo, volcadura, caída de pasajero, salida del camino, incendio, con ferrocarril, con motocicleta/ciclista y otros.

Esto se ha convertido hoy en día en uno de los principales problemas de las zonas urbanas por las pérdidas tanto humanas como materiales que generan. Por ello, resulta de interés conocer el número de accidentes de tránsito que sucederán, ya que eso repercute directamente en el planteamiento de medidas preventivas contra esta situación, así como en el costo de seguros de automóvil.

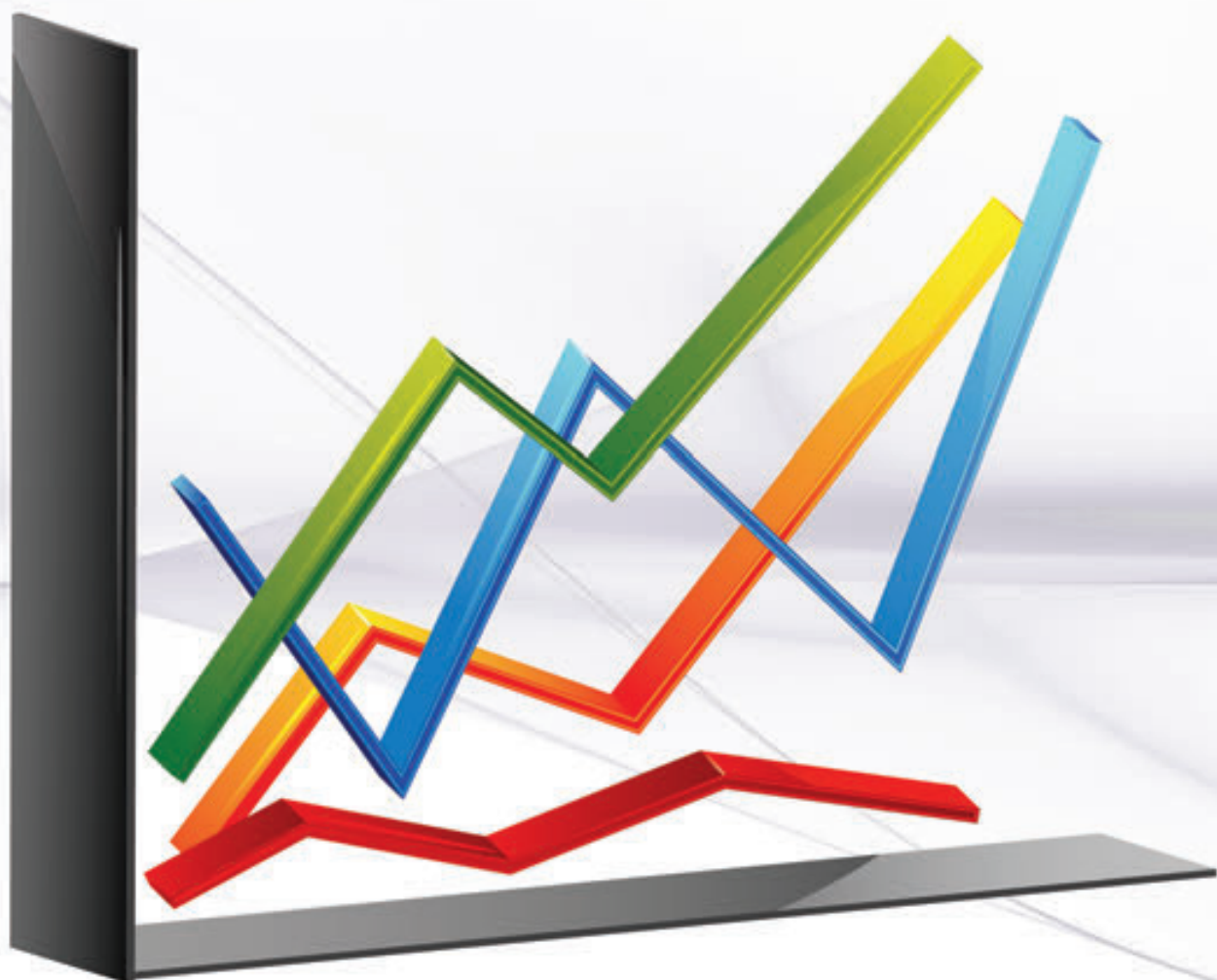
Consulta la base de datos [accidentes.xls](#) y supón que te ha contratado el gobierno federal para realizar un estudio sobre el número de accidentes de tránsito, en el que se te pide que desarrolles un modelo para pronosticar cuántos accidentes de tránsito se esperan en los próximos tres años. Esto servirá como base para presentar una iniciativa sobre reglamentos viales en el nivel federal, que se lanzaría con pretexto de las próximas elecciones.

¿Cuál sería el pronóstico de accidentes para los próximos tres años? Justifica tu modelo.

Realizas este proyecto y te das cuenta que la tendencia de accidentes de tránsito es creciente. Si entregas el resultado tal como lo obtuviste, ello implicaría que se establecieran reglas más estrictas al conducir, lo que conllevaría un alto costo político para el gobierno. Si alteras los resultados, el gobierno obtendría un porcentaje mayor de aceptación en las encuestas, pero el costo social y económico por accidentes de tránsito sería mayor.

Preguntas para dirigir la discusión:

- ¿Qué harías? ¿Modificarías las cifras? ¿Las presentarías como el INEGI?
- ¿Qué beneficios económicos y sociales se obtendrían en cada situación?
- ¿Qué pérdidas económicas y sociales habrían en cada escenario?
- ¿Qué principios éticos estarían en juego en cada caso?
- ¿Cuál es tu responsabilidad como profesionista?
- ¿Cuál crees que es la responsabilidad social que se tiene como ciudadano?
- ¿Crees que una cifra de accidentes es sólo un número?



10. La sumatoria

Por cierto...

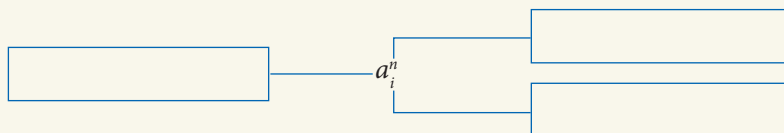
El profesor Büttner les había mandado sumar todas las cifras de uno a cien. Eso costaría horas y ni con la mejor voluntad lo lograrían sin cometer tarde o temprano algún fallo en la suma que los haría acreedores al castigo... A los tres minutos se encontraba Gauss (Brunswick, 1777-Gotinga, 1855) con su pizarrita, que contenía una sola línea escrita, ante el pupitre del maestro.

Veamos, dijo Büttner agarrando la palmeta. Su mirada cayó sobre el resultado, y se quedó petrificado. Preguntó qué significaba eso. Cinco mil cincuenta. ¿Qué? A Gauss le falló la voz, carraspeó, sudaba. Sólo ansiaba volver a su sitio y sumar como los demás, que permanecían sentados con la cabeza gacha como si no escuchasen (Kehlmann: 2007).

Ubícate

Antes de comenzar el estudio de la sumatoria, resuelve los ejercicios siguientes para cerciorarte de que tienes los conocimientos que te facilitarán aprender los temas que se expondrán en el bloque.

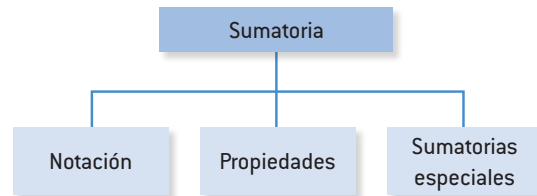
1. ¿Qué es una suma? ¿Cuál es el nombre de las partes que la componen?
2. ¿Qué es una sucesión?
3. ¿Cuál de los siguientes es un número entero?
 - a) $\sqrt{2}$
 - b) $\frac{1}{4}$
 - c) -3
 - d) 0.125
4. Escribe los nombres de cada parte de la expresión siguiente:



5. Desarrolla las expresiones algebraicas que siguen:
 - a) $(x - 3)^2$
 - b) $(x + 1)(x - 1)$
 - c) $(x + 2)^3$
6. Factoriza:
 - a) $ax^2 + xy$
 - b) $2xy^2 - 4y^3$

Propiedades de la sumatoria

En este bloque abordaremos todo lo referente a la sumatoria y cómo fue que Friederich Gauss llegó tan rápido al resultado. En el esquema siguiente se muestran los temas que se estudiarán.



Notación

La sumatoria es una operación matemática que permite representar la adición de un número (finito) de elementos, llamados *sumandos*. Se acostumbra utilizar la letra griega sigma mayúscula (Σ) para simbolizar la sumatoria.

Considera la suma siguiente:

$$1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10$$

Exprésala de forma compacta, o sea, en notación sigma o de sumatoria.

Solución

$$1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 = \sum_{i=1}^{10} i$$

Considera la suma siguiente:

$$c_1 + c_2 + c_3 + c_4 + c_5 + c_6$$

Exprésala de forma compacta [notación sigma].

Solución

$$c_1 + c_2 + c_3 + c_4 + c_5 + c_6 = \sum_{i=1}^6 c_i$$

Ejemplo 1

Se utiliza Σ para señalar que vamos a sumar todos los números de la forma indicada cuando el índice i empieza en 1 y termina en 10. Estos límites, 1 y 10, se indican arriba y abajo de Σ , respectivamente.

Ejemplo 2

Observa que en esta suma, el subíndice (el número pequeño a la derecha de c) es el que toma valores enteros positivos.

Propiedades de la sumatoria

La sumatoria es un operador de forma lineal. Por ello, tenemos el teorema siguiente sobre la linealidad de Σ .

TEOREMA 3.1 Sean $\{a_i\}$ y $\{b_i\}$ los símbolos de dos sucesiones y sea c una constante. Entonces,

- i. $\sum_{i=1}^n c = nc$
- ii. $\sum_{i=1}^n ca_i = c \sum_{i=1}^n a_i$
- iii. $\sum_{i=1}^n (a_i + b_i) = \sum_{i=1}^n a_i + \sum_{i=1}^n b_i$
- iv. $\sum_{i=1}^n (a_i - b_i) = \sum_{i=1}^n a_i - \sum_{i=1}^n b_i$

DEMOSTRACIÓN

Para (i):

$$\sum_{i=1}^n c = \underbrace{c + c + c + \cdots + c}_{c \text{ se repite } n \text{ veces}} = nc$$

Para denotar los índices se utilizan letras de la mitad del alfabeto, como i y k .

Para (ii):

$$\sum_{i=1}^n c a_i = c a_1 + c a_2 + c a_3 + \cdots + c a_{n-1} + c a_n = c(a_1 + a_2 + a_3 + \cdots + a_{n-1} + a_n) = c \sum_{i=1}^n a_i$$

■

Ejemplo 3

Se sabe que $\sum_{i=1}^{20} a_i = 50$ y $\sum_{i=1}^{20} b_i = 10$. Calcula el valor de:

a. $\sum_{i=1}^{20} (3a_i - 2b_i + 6)$

b. $\sum_{i=1}^{20} (a_i + 3b_i)$

Solución

Aplicamos las propiedades vistas en el teorema 3.1:

a.

$$\sum_{i=1}^{20} (3a_i - 2b_i + 6) = \sum_{i=1}^{20} 3a_i - \sum_{i=1}^{20} 2b_i + \sum_{i=1}^{20} 6 \quad \text{Usa los incisos (iii) y (iv)}$$

$$= 3 \sum_{i=1}^{20} a_i - 2 \sum_{i=1}^{20} b_i + \sum_{i=1}^{20} 6 \quad \text{Usa los incisos (i) y (ii)}$$

$$= 3(50) - 2(10) + 20(6)$$

$$= 150 - 20 + 120$$

$$= 250$$

b. Igual que el inciso anterior:

$$\sum_{i=1}^{20} (a_i + 3b_i) = \sum_{i=1}^{20} a_i + \sum_{i=1}^{20} 3b_i$$

$$= \sum_{i=1}^{20} a_i + 3 \sum_{i=1}^{20} b_i$$

$$= 50 + 3(10) = 80$$

Sumatorias especiales

En esta sección se presentan fórmulas para la suma de los primeros n enteros positivos.

i. $\sum_{i=1}^n i = 1 + 2 + 3 + 4 + \cdots + (n-1) + n = \frac{n(n+1)}{2}$

$$\text{ii. } \sum_{i=1}^n i^2 = 1^2 + 2^2 + 3^2 + 4^2 + \cdots + (n-1)^2 + n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\text{iii. } \sum_{i=1}^n i^3 = 1^3 + 2^3 + 3^3 + 4^3 + \cdots + (n-1)^3 + n^3 = \left[\frac{n(n+1)(2n+1)}{6} \right]^2$$

$$\text{iv. } \sum_{i=1}^n i^4 = 1^4 + 2^4 + 3^4 + 4^4 + \cdots + (n-1)^4 + n^4 = \frac{n(n+1)(6n^3 + 9n^2 + n - 1)}{30}$$

$$\text{v. } \sum_{i=0}^n cr^i = c + cr + cr^2 + cr^3 + \cdots + cr^{n-1} + cr^n = \frac{c(1 - r^{n+1})}{1 - r}, \text{ donde } c \text{ es una constante y } r \neq 1$$

Observa que en la suma especial (v), el índice inferior empieza en 0.

Indica los sumandos y resuelve:

Ejemplo 4

$$\text{a. } \sum_{i=1}^6 i$$

$$\text{b. } \sum_{i=1}^6 i^2$$

$$\text{c. } \sum_{i=1}^6 i^3$$

$$\text{d. } \sum_{i=1}^6 i^4$$

$$\text{e. } \sum_{i=0}^6 \left(\frac{1}{3}\right)$$

Solución

Aplicamos las fórmulas de las sumas especiales:

a. Utilizamos la suma especial (i), con $n = 6$:

$$\sum_{i=1}^6 i = 1 + 2 + 3 + 4 + 5 + 6 = \frac{6(6+1)}{2} = \frac{42}{2} = 21$$

b. Aplicamos la suma especial (ii), con $n = 6$:

$$\sum_{i=1}^6 i^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 = \frac{6(6+1)(2 \times 6 + 1)}{6} = \frac{546}{6} = 91$$

c. Con la suma especial (iii) y $n = 6$, obtenemos:

$$\sum_{i=1}^6 i^3 = 1^3 + 2^3 + 3^3 + 4^3 + 5^3 + 6^3 = \left[\frac{6(6+1)}{2} \right]^2 = [21]^2 = 441$$

d. Usamos la suma especial (iv) y $n = 6$:

$$\sum_{i=1}^6 i^4 = 1^4 + 2^4 + 3^4 + 4^4 + 5^4 + 6^4 = \frac{6(6+1)(6 \times 6^3 + 9 \times 6^2 + 6 - 1)}{30} = \frac{68250}{30} = 2275$$

e. Por último, puesto que $c = 1$, $r = \frac{1}{3}$ y $n = 6$, calculamos:

$$\sum_{i=0}^6 \left(\frac{1}{3}\right)^i = \left(\frac{1}{3}\right)^0 + \left(\frac{1}{3}\right)^1 + \left(\frac{1}{3}\right)^2 + \left(\frac{1}{3}\right)^3 + \left(\frac{1}{3}\right)^4 + \left(\frac{1}{3}\right)^5 + \left(\frac{1}{3}\right)^6 = \frac{1\left(1 - \left(\frac{1}{3}\right)^{6+1}\right)}{1 - \frac{1}{3}} = \frac{1093}{729} = 1.499314$$

Ejemplo 5

Resuelve

$$\sum_{i=1}^5 (i-2)^2$$

Solución

Primero desarrollamos el binomio al cuadrado:

$$\sum_{i=1}^5 (i-2)^2 = \sum_{i=1}^5 (i^2 - 4i + 4)$$

Después, aplicamos las propiedades vistas en el teorema 3.1:

$$\sum_{i=1}^5 (i^2 - 4i + 4) = \sum_{i=1}^5 i^2 - 4 \sum_{i=1}^5 i + \sum_{i=1}^5 4 = \sum_{i=1}^5 i^2 - 4 \sum_{i=1}^5 i + 5(4)$$

Por último, aplicamos las fórmulas de sumas especiales:

$$\begin{aligned} \sum_{i=1}^5 i^2 - 4 \sum_{i=1}^5 i + 5(4) &= \frac{5(5+1)(2 \times 5 + 1)}{6} - 4 \left[\frac{5(5+1)}{2} \right] + 20 \\ &= \frac{5 \times 6 \times 11}{6} - 4 \left[\frac{30}{2} \right] + 20 = 15 \end{aligned}$$

Ejercicios

I. TEORÍA Y ALGO MÁS

1. Demuestra las propiedades siguientes:

$$\text{a) } \sum_{i=1}^n (a_i + b_i) = \sum_{i=1}^n a_i + \sum_{i=1}^n b_i$$

$$\text{b) } \sum_{i=1}^n (a_i - b_i) = \sum_{i=1}^n a_i - \sum_{i=1}^n b_i$$

2. Encuentra una fórmula para:

$$\text{a) } \sum_{k=1}^n (k-1)(k+6)$$

$$\text{b) } \sum_{k=1}^n (k^3 - k^2)$$

II. PIENSO,... ¿Y LUEGO?

3. *Cuadrados mágicos.* Según una antigua leyenda china, una vez se desbordó el río Lo y los habitantes de la zona imploraron al dios del río que los salvara. Entonces surgió del agua una tortuga con una extraña inscripción en el caparazón (cuadro de abajo). Si se suma cualquier fila de tres números (horizontal, vertical o diagonal), siempre se obtiene el mismo resultado (en este caso, 15). (Anna Claybourne: 2009) ¿Podrías encontrar y dibujar otro cuadrado mágico?

4	9	2
3	5	7
8	1	6

4. Cuando el profesor Büttner dejó a sus alumnos sumar los números de 1 a 100, Gauss lo hizo muy rápido, mientras sus compañeros se pusieron a sumar uno a uno. Lo que Gauss hizo fue escribir horizontalmente los números, en un renglón:

1 2 3 4 ... 97 98 99 100

Posteriormente, en un renglón abajo del primero escribió los mismos números, pero ahora de forma descendente, y luego sumó de este modo

1 2 3 4 ... 97 98 99 100

100 99 98 97... 4 3 2 1

101 101 101 101 ... 101 101 101 101

Las sumas daban el mismo número, por tanto, lo que pedía el profesor Büttner era multiplicar cien veces el número ciento uno, obtenido de dos series: $\frac{100 \cdot 101}{2} = 5050$.

Hazlo con cualquier suma de números consecutivos y verás que siempre se aplica este razonamiento.

5. Si tienes ocho ochos, ¿cómo podrías sumarlos para que el resultado sea 1000? (Escandón: 1969)

III. PURO DEPORTE

6. Escribe en notación sigma las sumas siguientes:

a) $1 + 2 + 3 + 4 + 5 + \dots + 65 + 66$

b) $3 + 6 + 9 + 12 + \dots + 78 + 81$

c) $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots + \frac{1}{64} + \frac{1}{128}$

d) $a_1 + a_2 + a_3 + \dots + a_{38} + a_{39}$

e) $-1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5}$

7. Sean $\sum_{k=1}^{15} a_i = 30$ y $\sum_{k=1}^{15} b_i = 25$. Encuentra:

a) $\sum_{k=1}^{15} (4a_i - 3b_i)$

b) $\sum_{k=1}^{15} \left(2a_i + \frac{b_i}{5} \right)$

c) $\sum_{k=1}^{15} (b_i + 3)$

8. Calcula:

a) $\sum_{i=1}^{12} (2i + 2)^2$

b) $\sum_{k=1}^9 (k + 2)(k - 2)$

c) $\sum_{i=1}^{11} 3i(i^2 + 4)$

9. *Gimnasia.* En una pirámide humana, se colocan 10 personas en la base; después, en el nivel siguiente hay nueve personas. Sobre ellas, se colocan ocho y así sucesivamente hasta que sólo hay una persona en la “cúspide”. ¿Cuántas personas hay en total?
10. *Ajedrez.* Hay una leyenda sobre la creación del ajedrez. Se dice que hace muchos siglos, en un país del Oriente vivía un rey que había perdido a su hijo en una batalla. Uno de sus ministros llamó a todos los científicos y filósofos del reino para que buscaran una posible solución a la tristeza que ese hecho había causado al rey. Uno de ellos inventó un juego de estrategias, el ajedrez. El rey quedó tan feliz con el invento que decidió recompensar al inventor con lo que le solicitara. El joven que había creado el ajedrez pidió lo siguiente: un grano de trigo por la primera casilla o escaque del tablero, dos granos por la segunda, cuatro por la tercera, ocho por la cuarta, dieciséis por la quinta y así sucesivamente hasta completar los 64 escaques del tablero de ajedrez. El rey pidió a los matemáticos del reino que calcularan el número de granos de trigo que debían entregar al joven. Después de un rato, los matemáticos dijeron al rey que no alcanzaba todo el trigo del mundo para dar la recompensa por el juego de ajedrez.
- Expresa la suma de las casillas de ajedrez en forma desarrollada y en notación sigma.
 - Con ayuda de una calculadora o de una computadora, obtén la suma total de granos de trigo que el rey debía entregar.
 - Utiliza la suma especial correspondiente y compara este resultado con el obtenido en el inciso anterior.
 - Un grano de trigo pesa entre 0.8 y 1 gramo, según su variedad. Convierte la cantidad obtenida en toneladas. ¿Crees que el rey pudo haber juntado esa cantidad?

Evaluación

Un escalón hacia arriba

Contesta los reactivos siguientes que te permitirán determinar lo que has aprendido. ¡Adelante!

1. Indica los sumandos de la siguiente expresión:

$$\sum_{k=1}^5 \frac{7}{k+1}$$

2. Escribe en notación sigma la expresión:

$$2 + 4 + 6 + 8 + \dots + 90$$

3. Dadas $\sum_{i=1}^{100} a_i = 10$ y $\sum_{i=1}^{100} b_i = 30$, determina:

a) $\sum_{i=1}^{100} \left(3a_i + \frac{2}{3}b_i \right)$

b) $\sum_{i=1}^{100} (2a_i + 5)$

4. Encuentra el valor de las sumas siguientes y simplifica el resultado:

a) $\sum_{k=1}^8 (3k - 1)^2$

a) $\sum_{i=1}^7 (i + 2)(i - 3)$

5. Determina una fórmula para $\sum_{k=1}^n k^2 + k$ y simplifica el resultado.

A. Breve cronología del desarrollo de la estadística

3000 a.C. Los babilonios recopilan datos en tablas sobre la producción agrícola y los productos intercambiados mediante trueque.



3050 a.C. Los egipcios analizan los datos de la población y la renta del país, llevando cuenta de los movimientos poblacionales y realizando continuamente censos.

En la Biblia, en el libro de “Números” aparece el censo que realizó Moisés después de la salida de Egipto.



2200 a.C. En China había censos ordenados por el emperador Tao.

500 a.C. Se realizaron censos en Roma para conocer la población existente en ese momento; ello dio lugar a la creación de la figura de censor.



1662 Antonio von Leeuwenhoek calcula mediante la aritmética el número máximo de personas que habitan la Tierra, tomando como base la densidad de población en Holanda y utilizando la información de la “capitación” (impuesto por cabeza).

1662 John Graunt publica *Observaciones naturales y políticas mencionadas en un índice adjunto y basados en las listas de mortalidad*, donde muestra las cifras brutas de nacimientos y defunciones ocurridas en Londres durante el periodo 1604-1661, lo que puede considerarse el primer trabajo estadístico serio sobre la población.





1690 William Petty publica *Aritmética política*, texto sobre un método nuevo de gobierno basado en los números y en el que busca entender las relaciones entre población y economía.



El astrónomo Edmund Halley (1656-1742) presenta las tablas numéricas para el cálculo de rentas vitalicias para las compañías de seguros.



1731 Benjamin Franklin presenta en la *Pennsylvania Gazette* compilaciones de datos relacionados con el comercio, la navegación y la demografía.

1790 En Estados Unidos se lleva a cabo el Primer Censo Federal.



1800 En Francia se crea la Oficina de Estadística General.

1812 Pierre-Simon Laplace publica su *Teoría analítica de las probabilidades*, donde formula, entre otros, el método de mínimos cuadrados, fundamental en el área de pronósticos.



1820 Florence Nightingale aplica sus conocimientos de estadística a la epidemiología y a la estadística sanitaria, pues emplea gráficas para un mejor entendimiento. Fue la primera mujer admitida en la *Royal Statistical Society* británica y miembro honorario de la *American Statistical Association*.

c. 1825 Pierre Charles Alexandre Louis introdujo el método numérico, en el que establece que es necesario contar, es decir, llevar la cuantificación a la medicina, con lo que se convierte en *padre de la estadística médica* y precursor de la estadística clínica.



1833 Joseph Fourier realiza aportaciones a la teoría matemática de la probabilidad.

1835 André Michel Guerry publica su *Ensayo sobre la estadística moral de Francia*. Junto con Quetelet, es fundador de la estadística moral, ya que utiliza la estadística en la criminalidad. Esto permitió, posteriormente, el desarrollo de la criminología, la sociología y las ciencias sociales.



Quetelet aplicó el método estadístico al estudio del comportamiento humano. Publicó *Sobre el hombre y el desarrollo de las facultades humanas*, *Ensayo sobre física social* (1835) y *La antropometría, o medida de las diferentes facultades del hombre* (1871).

1853 Se realiza el Primer Congreso Internacional de Estadística.

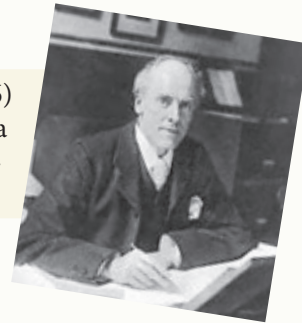


1864 William Farr desarrolla las tablas de vida inglesas y escribe ensayos anuales sobre estadísticas “vitales”.

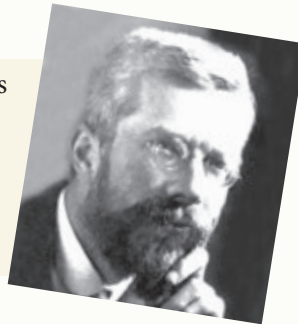
En el siglo XIX, la estadística entra en una nueva fase de desarrollo con la generalización del método para estudiar fenómenos de las ciencias naturales y sociales.



Galton (1822-1911) y Pearson (1857-1936) pueden considerarse los padres de la estadística moderna, pues se les debe el paso de la estadística deductiva a la inductiva.



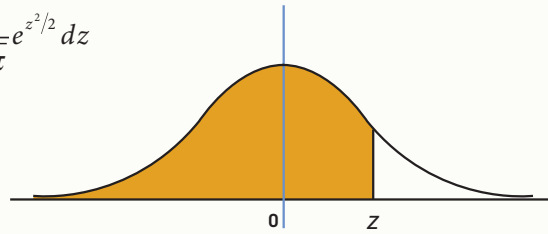
Los fundamentos de la estadística actual y muchos de los métodos de inferencia se deben a R. A. Fisher, cuyos trabajos culminan con la publicación de la obra *Métodos estadísticos para investigaciones*, donde aparece el método de la estadística como se conoce hoy.



Mediados del siglo XX Se origina lo que podemos denominar *estadística moderna*. Entre los factores determinantes para su surgimiento se halla la aparición y popularización de las computadoras.

B. Tabla de distribución acumulada normal estándar $N(0, 1)$

$$P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.6	0.000159	0.000153	0.000147	0.000142	0.000136	0.000131	0.000126	0.000121	0.000117	0.000112
-3.5	0.000233	0.000224	0.000216	0.000208	0.000200	0.000193	0.000185	0.000178	0.000172	0.000165
-3.4	0.000337	0.000325	0.000313	0.000302	0.000291	0.000280	0.000270	0.000260	0.000251	0.000242
-3.3	0.000483	0.000466	0.000450	0.000434	0.000419	0.000404	0.000390	0.000376	0.000362	0.000349
-3.2	0.000687	0.000664	0.000641	0.000619	0.000598	0.000577	0.000557	0.000538	0.000519	0.000501
-3.1	0.000968	0.000935	0.000904	0.000874	0.000845	0.000816	0.000789	0.000762	0.000736	0.000711
-3.0	0.001350	0.001306	0.001264	0.001223	0.001183	0.001144	0.001107	0.001070	0.001035	0.001001
-2.9	0.001866	0.001807	0.001750	0.001695	0.001641	0.001589	0.001538	0.001489	0.001441	0.001395
-2.8	0.002555	0.002477	0.002401	0.002327	0.002256	0.002186	0.002118	0.002052	0.001988	0.001926
-2.7	0.003467	0.003364	0.003264	0.003167	0.003072	0.002980	0.002890	0.002803	0.002718	0.002635
-2.6	0.004661	0.004527	0.004396	0.004269	0.004145	0.004025	0.003907	0.003793	0.003681	0.003573
-2.5	0.006210	0.006037	0.005868	0.005703	0.005543	0.005386	0.005234	0.005085	0.004940	0.004799
-2.4	0.008198	0.007976	0.007760	0.007549	0.007344	0.007143	0.006947	0.006756	0.006569	0.006387
-2.3	0.010724	0.010444	0.010170	0.009903	0.009642	0.009387	0.009137	0.008894	0.008656	0.008424
-2.2	0.013903	0.013553	0.013209	0.012874	0.012545	0.012224	0.011911	0.011604	0.011304	0.011011
-2.1	0.017864	0.017429	0.017003	0.016586	0.016177	0.015778	0.015386	0.015003	0.014629	0.014262
-2.0	0.022750	0.022216	0.021692	0.021178	0.020675	0.020182	0.019699	0.019226	0.018763	0.018309
-1.9	0.028717	0.028067	0.027429	0.026803	0.026190	0.025588	0.024998	0.024419	0.023852	0.023295
-1.8	0.035930	0.035148	0.034380	0.033625	0.032884	0.032157	0.031443	0.030742	0.030054	0.029379
-1.7	0.044565	0.043633	0.042716	0.041815	0.040930	0.040059	0.039204	0.038364	0.037538	0.036727
-1.6	0.054799	0.053699	0.052616	0.051551	0.050503	0.049471	0.048457	0.047460	0.046479	0.045514
-1.5	0.066807	0.065522	0.064255	0.063008	0.061780	0.060571	0.059380	0.058208	0.057053	0.055917
-1.4	0.080757	0.079270	0.077804	0.076359	0.074934	0.073529	0.072145	0.070781	0.069437	0.068112
-1.3	0.096800	0.095098	0.093418	0.091759	0.090123	0.088508	0.086915	0.085343	0.083793	0.082264
-1.2	0.115070	0.113139	0.111232	0.109349	0.107488	0.105650	0.103835	0.102042	0.100273	0.098525
-1.1	0.135666	0.133500	0.131357	0.129238	0.127143	0.125072	0.123024	0.121000	0.119000	0.117023
-1.0	0.158655	0.156248	0.153864	0.151505	0.149170	0.146859	0.144572	0.142310	0.140071	0.137857
-0.9	0.184060	0.181411	0.178786	0.176186	0.173609	0.171056	0.168528	0.166023	0.163543	0.161087
-0.8	0.211855	0.208970	0.206108	0.203269	0.200454	0.197663	0.194895	0.192150	0.189430	0.186733
-0.7	0.241964	0.238852	0.235762	0.232695	0.229650	0.226627	0.223627	0.220650	0.217695	0.214764
-0.6	0.274253	0.270931	0.267629	0.264347	0.261086	0.257846	0.254627	0.251429	0.248252	0.245097
-0.5	0.308538	0.305026	0.301532	0.298056	0.294599	0.291160	0.287740	0.284339	0.280957	0.277595
-0.4	0.344578	0.340903	0.337243	0.333598	0.329969	0.326355	0.322758	0.319178	0.315614	0.312067
-0.3	0.382089	0.378280	0.374484	0.370700	0.366928	0.363169	0.359424	0.355691	0.351973	0.348268
-0.2	0.420740	0.416834	0.412936	0.409046	0.405165	0.401294	0.397432	0.393580	0.389739	0.385908
-0.1	0.460172	0.456205	0.452242	0.448283	0.444330	0.440382	0.436441	0.432505	0.428576	0.424655
-0.0	0.500000	0.496011	0.492022	0.488034	0.484047	0.480061	0.476078	0.472097	0.468119	0.464144
0.0	0.500000	0.503989	0.507978	0.511966	0.515953	0.519939	0.523922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555670	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621720	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732

0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802337	0.805105	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823814	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.879000	0.881000	0.882977
1.2	0.884930	0.886861	0.888768	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903200	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935745	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959070	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965620	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999534	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999651
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999822	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888

C. Instalación del complemento

Análisis de datos de Excel

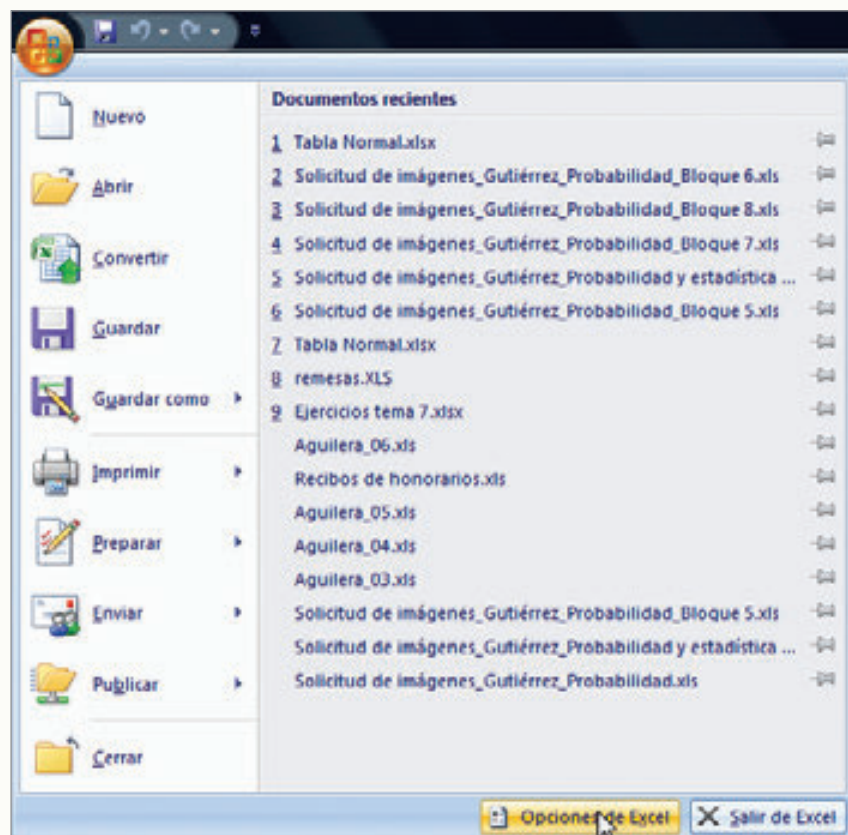
Excel cuenta con algunas aplicaciones, denominadas *complementos*, que le añaden funciones específicas. Estos complementos *no se instalan* de manera automática cuando instalas Excel, de modo que debes agregarlos después.

En este apéndice te explicaremos cómo instalar el complemento Análisis de datos, que incluye gran parte de las herramientas con las que puedes hacer buena parte de los ejercicios de este libro.

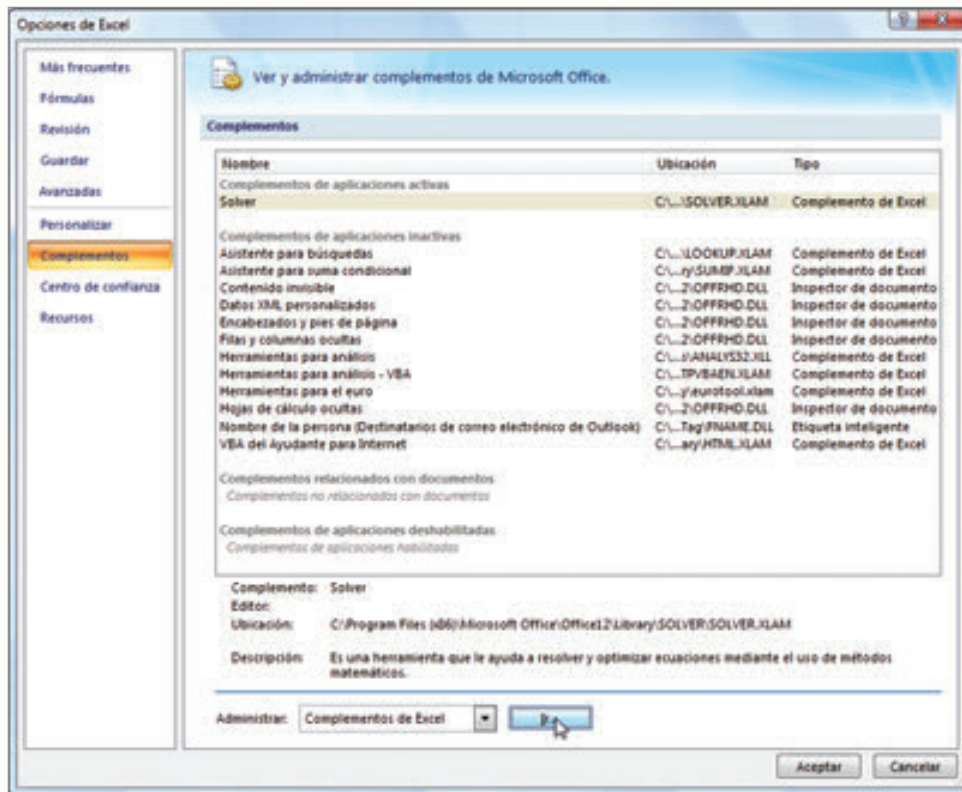
Es importante precisar que este procedimiento se aplica a las versiones recientes de Microsoft Office (2007 y 2010). Si utilizas una versión anterior, el procedimiento varía un poco, por lo que te sugerimos consultar la ayuda del programa.

Al instalar el complemento Análisis de datos como se indica a continuación, es probable que el programa te pida el disco de Office para terminar la instalación del complemento, de modo que trata de tenerlo a mano cuando sigas estos pasos:

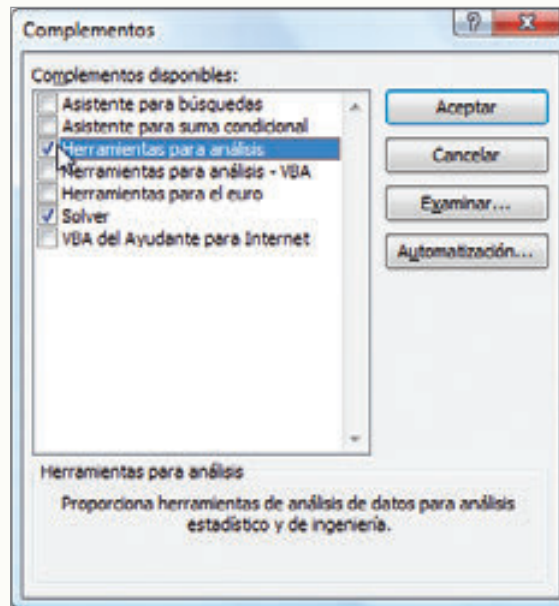
1. Pon en marcha Excel.
2. Haz clic en el botón de Office y después en el botón Opciones de Excel.



3. En el cuadro de diálogo Opciones de Excel, haz clic en la opción Complementos del panel que aparece a la izquierda y en seguida en el botón Ir.



4. En el cuadro de diálogo que aparece, selecciona la opción Herramientas para análisis; cerciérate de que aparezca una marca de verificación en el cuadro correspondiente para que la función quede habilitada. Concluye su instalación haciendo clic en el botón Aceptar.



Eso es todo.

Bibliografía

- Aguirre, V. Y. (2006). *Fundamentos de probabilidad y estadística*. México: Jit Press.
- Alba, R. D. (2010). "Numeralia. Petróleo." *Equilibrio*, 6-7.
- _____. (2011). "México ¿aprende?" *Equilibrio*, 36-38.
- Alsina, C. (2008). *El club de la hipotenusa*. México: Ariel-Paidós Mexicana.
- Anna Claybourne, A. L. (2009). *La historia de la ciencia*. Londres: Usborne Publishing.
- Cohen, I. B. (2005). *El triunfo de los números*. Madrid: Alianza Editorial.
- Escandón, R. (1969). *Curiosidades matemáticas*. México: Novaro.
- Fernández, F. A. (8 de mayo de 2011). "Lady Gaga: la elegida del pop" en "El Ángel", suplemento cultural de *Reforma*, pág. 1.
- Hoffman, J. (9 de abril de 2011). "Cuando el sexting es viral", *Reforma (The New York Times)*, págs. 1-2.
- Kehlmann, D. (2007). *La medición del mundo*. México: Diana.
- Livio, M. (2009). *¿Es Dios un matemático?* Barcelona: Ariel.
- Mitofsky, C. (2010). *Encuesta nacional sobre la percepción de seguridad ciudadana en México*. México.
- Organización Mundial de la Salud [s.f.]. oms. Recuperado el 15 de 09 de 2010, <http://www.who.int>.
- Redacción, I. (2011). "Acciones individuales: con la conciencia limpia." *Equilibrio*, 39.
- Sarmiento, S. (10 de junio de 2010). "Futbolcifras", en la columna Jaque Mate, *Reforma*, pág. 12.
- Stewart, I. (2001). *El laberinto mágico*. Barcelona: Drakontos Crítica.
- Tapia, Oliver y Juan José Cortés (noviembre de 2010). "Suiza, unión en la diversidad". Suplemento *Reforma*, págs. 8-9.
- Torres, K. (2010). "Ecotweet. 140 caracteres atacan el cambio climático." *Equilibrio*, 33.
- Villagómez, A. (8 de septiembre de 2010). "Otra vez la educación", en la columna Tintero Económico, *El Universal*.

Sitios de interés en internet

En la OLC hallarás también esta lista de ligas

La importancia de la información estadística:

<http://www.youtube.com/watch?v=b2joia0VtA>

Historia de la estadística:

<http://www.youtube.com/watch?v=oXP47A-T0zs&feature=related>

Laboratorio virtual de probabilidad y estadística:

http://www.planetamatematico.com/index.php?option=com_content&task=view&id=118&Itemid=158

Estadística animada:

<http://www.math.ucsd.edu/~anistat/mindex.html>

Trazo de diagramas de barras:

<http://www.sopadenumeros.com/content/view/493/65/>

Trazo de histogramas:

<http://www.sopadenumeros.com/content/view/530/65/>

<http://www.youtube.com/watch?v=FWHoxj6Tyq4>

<http://www.youtube.com/watch?v=OMrLRVcqsOM&feature=related>

Calculadora de frecuencias y porcentajes:

http://www.elosiodelosantos.com/calculadoras/calculadora_conteo_y_porcentaje.htm

Cálculo de la media, la moda y la mediana:

<http://es.easycalculation.com/statistics/mean-median-mode.php>

Calculadora para estadística descriptiva:

<http://www.elosidelosantos.com/descriptiva.html>

Simulador de media y mediana con histograma:

http://virtual.uptc.edu.co/ova/estadistica/applets/RVLS_simulations/stat_sim/descriptive/index.html

Cálculo de la media armónica:

<http://es.easycalculation.com/statistics/harmonic-mean.php>

Definiciones de los principales conceptos estadísticos:

<http://www.aaamaticas.com/sta.htm>

Juegos, usando los principales estadísticos:

<http://www.aaamaticas.com/sta418x5.htm#section3>

Conceptos básicos de conjuntos:

<http://www.revistaciencias.com/publicaciones/EpZEIAkFkADxLLOHfF.php>

Definiciones de conjuntos:

<http://sipan.inictel.gob.pe/internet/av/introduccion.htm>

Diagramas de Venn

<http://sipan.inictel.gob.pe/internet/av/venn.htm>

<http://www.youtube.com/watch?v=S0ojkaWZhWo&feature=related>

Calculadora de combinaciones y permutaciones:

<http://www.elosidelosantos.com/combinaciones.html>

Simulador de combinaciones y permutaciones:

<http://www.mameworld.info/robbie/calculador.htm>

Construcción de diagrama de árbol y ejercicios:

<http://laprofematematica.com/blog/ejercicios-de-diagrama-de-arbol/>

Actividades de probabilidad condicional:

<http://www.eduteka.org/MI/master/interactivate/lessons/pm4.html>

Probabilidad condicional usando diagramas de Venn:

http://www.selectividad.tv/S_M_3_1_6_S_probabilidad_condicional.html

Teorema de Bayes:

<http://www.youtube.com/watch?v=8WgFRPOuKsU&feature=related>

Ejercicios resueltos de distribuciones discretas:

http://www.vitutor.com/pro/3/a_a.html

Distribución binomial

http://recursostic.educacion.es/descartes/web/materiales_didacticos/Distribucion_binomial/binomial.htm

Calculadora de distribución binomial:

<http://www.elosidelosantos.com/binomial.html>

Trazo de una distribución binomial, dando n y p :

<http://www.aulademate.com/contentid-300.html>

Simulador de la distribución binomial:

<http://virtual.uptc.edu.co/ova/estadistica/applets/simulacionbino/Simulacion.htm>

Distribución de Poisson:

<http://www.youtube.com/watch?v=nhxMLs5fwHU&feature=related>

Distribución de Poisson usando Excel:

<http://www.youtube.com/watch?v=-miu0V0fdj0&feature=related>

Cálculo de probabilidades con una distribución normal:

<http://www.sopadenumeros.com/content/view/495/65/>

La distribución normal

<http://www.abciencia.com.ar/matematicas/distribucion-normal>

Simulador de la distribución normal:

http://www.ruf.rice.edu/~lane/stat_sim/normal_approx/index.html

Calculadora de probabilidades de una distribución normal estándar:

<http://www.elosiodelosantos.com/distribucionnormalestandar.html>

Simulador del coeficiente de correlación:

http://virtual.uptc.edu.co/ova/estadistica/applets/RVLS_simulations/stat_sim/comp_r/index.html

Calculadora del coeficiente de correlación y de la recta de regresión:

<http://www.elosiodelosantos.com/regresionlineal.html>

