Estadística Descriptiva y Probabilidad

Leandro González Támara

Universidad de Bogotá
JORGE TADEO LOZANO
Facultad de Ciencias Naturales e Ingeniería
Departamento de Ciencias Básicas

Estadística Descriptiva y Probabilidad

Universidad de Bogotá
JORGE TADEO LOZANO
Facultad de Ciencias Naturales e Ingeniería
Departamento de Ciencias Básicas

González Támara, Leandro
Estadística descriptiva y probabilidad / Leandro González Támara. – Bogotá: Universidad de Bogotá Jorge Tadeo Lozano. Facultad de Ciencias Naturales e Ingeniería, 2013.
236 p.; 28 cm.

ISBN: 978-958-725-114-2

1. ESTADÍSTICA. 2. PROBABILIDADES. II. tit.

CDD519.53"G643"

Fundación Universidad de Bogotá Jorge Tadeo Lozano
Carrera 4 Nº 22-61 - pbx: 242 7030 - www.utadeo.edu.co

Estadística

ISBN: 978-958-725-114-2

Primera edición: 2013

Rectora: Cecilia María Vélez White
Vicerrector Académico: Diógenes Campos Romero
Decano de la Facultad de Ciencias Naturales e Ingeniería: Daniel Bogoya Maldonado
Director Departamento de Ciencias Básicas: Favio Cala Vitery
Director Editorial (E): Jaime Melo Castiblanco

Coordinación editorial: Henry Colmenares Melgarejo
Revisión de textos: Henry Colmenares Melgarejo
Diseño de portada: Francisco Jiménez
Diseño y diagramación: Francisco Jiménez
Impresión: D'vinni S.A.

Estadística Descriptiva y Probabilidad

Leandro González Támara

Universidad de Bogotá
Jorge Tadeo Lozano
Facultad de Ciencias Naturales e Ingeniería
Departamento de Ciencias Básicas

CONTENIDO

PRESENTACIÓN

La estadística es una disciplina que ha tomado parte importante en muchas áreas del conocimiento en ámbitos que van desde la determinación de los mecanismos de obtención y organización de los datos, hasta el ajuste a normas comunes en procesos de investigación. La estadística se ha convertido en una necesidad de ciudadanos y de profesionales para tomar decisiones con base en el análisis de información, con el objetivo de generar conocimiento.

Este es el libro guía de un curso universitario de Estadística Descriptiva y Probabilidad introductorio. Está escrito en un lenguaje sencillo y claro para favorecer el encuentro inicial con los conceptos básicos y también propone ideas de tipo didáctico a diferentes tipos de profesionales que se dedican a la enseñanza de esta disciplina. Es un libro diseñado para ser seguido en una clase presencial que motiva la discusión de diferentes temáticas y que puede ser considerado como texto de transición, porque permite a los estudiantes acercarse a los conceptos por medio de aplicaciones, mucho antes de concentrarse en aspectos teóricos. Por esta última razón se aconseja acompañarlo con otras referencias.

El libro contiene seis capítulos. El primero, sobre estadística descriptiva, introduce el lenguaje de la estadística y muestra diferentes formas gráficas de resumir datos. El segundo capítulo gira en torno a los estadísticos de centro y dispersión, con énfasis en su interpretación. El tercer capítulo hace una introducción al concepto de regresión y muestra al estudiante cómo puede ser usada esta técnica. El capítulo cuatro aborda el concepto de probabilidad como fundamento para el estudio de las variables aleatorias. Las distribuciones de probabilidad binomial, de Poisson e hipergeométrica son tratadas en el capítulo cinco. Por último, en el capítulo seis, se estudia la distribución de probabilidad normal.

Cada capítulo consta de un ejemplo o situación inicial que motiva el estudio de algunos de los contenidos. En lugar de un desarrollo teórico detallado, se presentan explicaciones breves para que los detalles sean tema de discusión en las clases. Hay abundantes ejemplos con estrategias para resolver problemas de análisis de datos y de probabilidad, los cuales, en su mayoría, abordan situaciones cercanas a los estudiantes con el propósito de despertar su interés. Cada concepto tratado se cierra con una sección de ejercicios para afianzar habilidades y generar discusión en las clases. Se espera que los estudiantes puedan realizar la mayoría de los cálculos aritméticos y las gráficas con un programa de computador o una calculadora con funciones estadísticas; esto, con el objeto de dar más tiempo a los significados y a la interpretación, y así realzar la importancia de esta tarea.

Al final del libro se incluye una sección con seis pruebas de conocimiento que son útiles para evaluar lo aprendido. Los datos presentados en el libro son en su mayoría reales y fueron obtenidos en páginas web de diferentes entidades reconocidas, entre ellas se anotan:

- Index Mundi. Disponible en http://www.indexmundi.com
- Instituto Nacional de Estadísticas INE (España). Disponible en www.ine.es
- Departamento Nacional de Estadísticas DANE (Colombia). Disponible en www.dane.gov.co
- Oficina del Censo de Estados Unidos. Disponible en http://www.census.gov/
- Link con las agencias de estadísticas internacionales. Disponible en http://www.census.gov/aboutus/stat_int.html
- Estadísticas internacionales del U.S. Census Bureau. Disponible en http://www.census.gov/population/international/data/idb/country.php
- Bases de datos internacionales del U.S. Census Bureau. Disponible en http://www.census.gov/population/international/data/idb/informationGateway.php
- Banco de la República (Colombia). Disponible en http://www.banrep.gov.co/
- Organización Mundial de la Salud. Disponible en http://www.who.int/research/en/
- Estadística de búsquedas de Google. Disponible en http://www.google.com/insights/search/?hl=es#
- Tendencias de Google. Disponible en http://www.google.es/trends

Varios ejercicios hacen referencia a bases de datos que se encuentran en Avata. Su tamaño obliga a un procesamiento con un programa de computador. Se advierte que se sigue la convención utilizada en Centroamérica, Suramérica y Europa de escribir la coma como separador decimal y el punto como separador de miles. No obstante, el lector encontrará algunas gráficas en las cuales esto no ocurre porque el software con el que fueron generadas proviene de un país donde el punto es el separador decimal.

CAPÍTULO 1
ESTADÍSTICA DESCRIPTIVA

¿Qué es la estadística?

No hace falta ser un experto para estar en contacto con la estadística. Con frecuencia, los medios de comunicación presentan información de naturaleza estadística, como encuestas de intención de voto y de opinión, información de *rating*, evolución de las tasas de cambio de diferentes divisas y otros indicadores económicos, datos del clima, etc. Los resultados que muestran los medios solo son una parte del trabajo estadístico y detrás de ellos existe una amplia disciplina, una forma de razonar y un método de generación de conocimiento. Los problemas que resuelve la estadística surgen de preguntas en diversas áreas que se fundamentan en el conocimiento inductivo. En un nivel más pragmático, la estadística tiene que ver con la forma más conveniente de obtener, resumir y analizar información. Algunos autores la definen como el arte de aprender de los datos. En la siguiente tabla se muestran algunas preguntas, de diversas áreas, en la cual la estadística resulta muy útil.

Economía	¿Cuál es el ingreso promedio de un profesional egresado en Cartagena?
	¿Cuántas personas viajan en avión en Colombia?
	¿Cómo han variado los precios de los alimentos el último mes?
	¿Cuántas personas desempleadas hay en el país?
	¿Cuál será la tasa de cambio del dólar el próximo mes?
Publicidad	¿Cuántas personas ven un determinado programa de televisión?
	¿Los adolescentes del sur del país leen prensa escrita?
	¿Qué percepción tienen los consumidores de cierta marca?
Mercadeo	¿Cuántas tiendas hay en Medellín?
	¿A qué precio se debe vender un producto?
	¿Cuánto vende una marca de bebida gaseosa en Ibagué?
Política	¿Cuál es la favorabilidad de un gobernante?
	¿Cuántas personas pueden participar en las siguientes elecciones?
Sector público	¿Cuántos niños menores de cinco años hay en actualidad en Boyacá?
	¿Cuántos ancianos hay en Santa Marta?
	¿Cuánto le cuesta al Estado una elección popular?
	¿Qué características tienen los evasores del impuesto predial de Bogotá?

La estadística se ha convertido en el lenguaje aceptado por la comunidad científica para la puesta a prueba, validación o rechazo de hipótesis de investigación; una descripción estadística es una forma de comunicación con un lenguaje especial. Los términos estadísticos tienen significados precisos y se asimilan paso a paso. La estadística se puede clasificar por su intención como **descriptiva** o **inferencial**. La primera tiene que ver con la mención de los hechos observados o la descripción de características de un conjunto de datos. La segunda se refiere a la obtención de propiedades generales, basadas en una muestra de un conjunto de datos. En este libro, se estudian algunas técnicas para hacer descripciones estadísticas y se realiza una introducción a la probabilidad, que es la base de la inferencia estadística. Para desarrollar las ideas anteriores, considere la siguiente situación:

Imagine que se desea estudiar cuáles son los hábitos y los gastos de los estudiantes de una universidad en telefonía celular. En algún momento, esto requerirá realizar un sondeo dentro de la población de estudiantes; no obstante, esto corresponde a una etapa posterior. La tarea inicial se debe concentrar en la determinación de un objetivo y la definición clara del concepto central del estudio. En este caso, definir con exactitud qué se entiende por hábitos y gastos. Las características del problema particular determinarán si basta con una descripción de la información recolectada o si con esta se pueden realizar conclusiones más generales o atribuir causas a los comportamientos observados. Este proceso está más allá del alcance de este libro, pero los lectores interesados en el tema pueden remitirse a Mendenhall (2006) y Lohr (2000) para un estudio detallado.

Población, muestra y variables

En estadística se llama **población** a la colección total de elementos con algunas características comunes y sobre la que se desea obtener alguna información o realizar algún análisis. El tamaño de la población es el número de elementos de esta colección y, generalmente, se denota con la letra N. Con frecuencia, el tamaño de la población es demasiado grande como para intentar examinar a todos sus elementos, por esta razón, es conveniente seleccionar solo una parte de los elementos de la población, es decir, una **muestra**, que refleje el comportamiento general de la población con respecto a lo que se desea estudiar.

A las medidas que se refieren a los elementos de toda la población se les llama **parámetros**; mientras que a las que son obtenidas de los elementos de la muestra se les llama **estadísticos.**

Sobre los elementos de la población o de la muestra, se miden características que recogen información del concepto en estudio; como estas características son diferentes en cada elemento se denominan **variables**. Por ejemplo, si se está interesado en averiguar sobre la actividad laboral en un grupo de individuos en Bucaramanga, se pueden medir variables como nivel de estudio o salario en los individuos empleados, o edad y tiempo de inactividad entre los desempleados.

Aquella variable que mida un atributo o característica que se pueda clasificar en categorías se le denomina **cualitativa**. Si examina una característica que se pueda medir o contar se le llama **cuantitativa**. Existen diversas clasificaciones de las variables según su naturaleza y escala. Una de estas es la división de las variables cualitativas en ordinales y nominales y las cuantitativas en discretas y continuas:

Cualitativa	Ordinal
	Nominal
Cuantitativa	Discreta
	Continua

Variable cualitativa ordinal

Una variable cualitativa ordinal identifica características que se pueden clasificar en grupos que admiten orden. Por ejemplo: resultado de una entrevista (rechazado, normal y sobresaliente) o premio en una prueba olímpica (oro, plata, bronce).

Variable cualitativa nominal

Una variable cualitativa nominal identifica características que se pueden clasificar en grupos que difieren o son iguales en una propiedad. Por ejemplo, municipio de nacimiento o el color de los ojos. Observe que en las variables cualitativas nominales las características no admiten un orden natural.

Variable cuantitativa discreta

Una variable cuantitativa discreta es aquella que toma un número finito o contable[1] de valores. Por ejemplo: número de respuestas correctas en un examen de cinco preguntas {0, 1, 2, 3, 4, 5} o el número de hijos de una persona {0, 1, 2, ...}.

Variable cuantitativa continua

Una variable cuantitativa continua es aquella que puede tomar infinitos valores en cualquier intervalo razonable de la variable. La estatura de un hombre adulto es una variable aleatoria continua. Note que existe una infinidad de posibilidades para la estatura en un intervalo arbitrario, por ejemplo, entre los 160 y 161 centímetros. Obviamente, las personas suelen responder con un número entero cuando se les pregunta por su estatura (en centímetros), pero esta se puede dar con la precisión deseada utilizando el instrumento de medida adecuado.

1 En términos matemáticos, un conjunto se dice contable si es posible construir una función biyectiva entre dicho conjunto y los números naturales.

Tabla 1.1 Tipos de variables

		Los valores que toma la variable:
Cualitativa	**Ordinal**	Se pueden clasificar en grupos ordenados.
	Nominal	Se pueden clasificar en grupos, pero no hay un orden entre ellos.
Cuantitativa	**Discreta**	Son finitos o contables.
	Continua	Son infinitos en cualquier intervalo razonable.

Adicionalmente, algunos autores consideran una clasificación adicional según la escala de medición. Si la elección del punto correspondiente al cero de la variable no es arbitraria y este significa la ausencia de la característica que se mide, se dice que es **escala de razón**. Por ejemplo, el peso o la estatura. Si el cero de la escala es arbitrario, se dice que es de **escala de intervalo**. Por ejemplo, la temperatura en grados centígrados o el resultado de conocimientos en una prueba de matemáticas.

Ejemplo 1.1

Enunciado

¿Cree que el color rosado es femenino? Existe una percepción generalizada acerca de que efectivamente lo es. Es posible que en el pasado esto no fuera así o que ciertas culturas actuales no vean el rosado como un color femenino. Para estudiar esta idea, se llevó a cabo un experimento con 65 integrantes de la comunidad emberá, cuarenta hombres y veinticinco mujeres. A cada uno se le pidió señalar su predilección entre tres prendas idénticas salvo por el color (azul, verde o rosado).Se obtuvieron los siguientes resultados:

	Color escogido		
	Azul	**Verde**	**Rosado**
Hombres	22	12	6
Mujeres	4	9	12

En general, ¿las mujeres emberá se sienten más atraídas por el color rosado que los hombres?

Solución

Observe que la pregunta hace referencia a la población emberá, cuyo tamaño puede llegar a los 70.000 individuos. En el estudio, se consideró una muestra de 65 elementos, es decir, 65 miembros de esta comunidad. La pregunta hace referencia a la población, es decir, se intenta realizar una inferencia estadística con base en la muestra.

No es posible responder a la pregunta a partir de esta información. Se desconoce la razón por la cual se seleccionó una muestra de 65 individuos o si este número es suficiente para tener una visión de las preferencias de toda la comunidad. No se menciona si la muestra fue elegida al azar o si por el contrario fue autoelegida —los que decidieron participar por voluntad propia—. Tampoco se revela la manera como se permitió a los individuos la elección del color, si fue de manera individual o grupal. Este aspecto es importante porque si se hiciera de manera grupal, las últimas elecciones se pueden ver influenciadas por las primeras. Todo esto sugiere que para realizar una inferencia estadística de esta naturaleza se requiere un diseño o plan mucho más cuidadoso. Por lo pronto solo es posible hacer descripciones estadísticas como las siguientes: la mayoría de los hombres emberá de la muestra escogieron el color azul (55%), mientras que la mayoría de las mujeres optaron por el color rosado (48%). Tan solo el 15% de los hombres de la muestra fueron atraídos por el color rosado, porcentaje similar al de las mujeres que escogieron el color azul (16%).▪

Gráficos para resumir datos cualitativos

En esta parte del libro se ilustra la utilidad de varias gráficas estadísticas a la hora de resumir datos. Se mencionan los gráficos de barras, de sectores y de puntos. Se acompañan los ejemplos con el análisis de la situación considerada. Como regla general, cuando se presenta algún tipo de información estadística, las gráficas deben ir acompañadas de un comentario o explicación.

Ejemplo 1.2

Enunciado

Construya una gráfica para representar la distribución del sexo en un salón con diez hombres y seis mujeres.

Solución

Primero, se resume la información en una tabla de frecuencias:

Sexo	Frecuencia absoluta	Frecuencia relativa
Hombres	10	$\frac{10}{16} \times 100 = 62,5\%$
Mujeres	6	$\frac{10}{16} \times 100 = 37,5\%$

Cuando se desea comparar algún valor en diferentes categorías se utiliza un **diagrama de barras** (figura1.1) y cuando lo importante es resaltar la contribución de cada categoría se utiliza un gráfico de sectores (figura 1.2). Las dos gráficas son igualmente informativas, pero con los énfasis mencionados. Observe que las barras aparecen separadas en el diagrama, porque no existen ni orden ni distancia numérica entre los sexos.▪

Figura 1.1 Diagrama de barras. Distribución del sexo

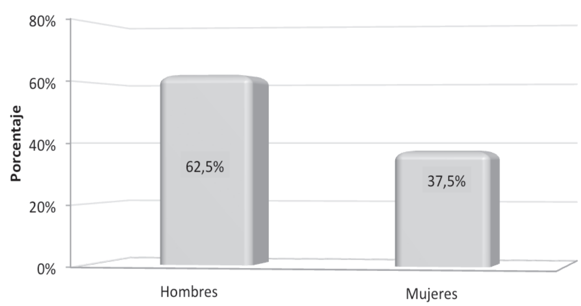
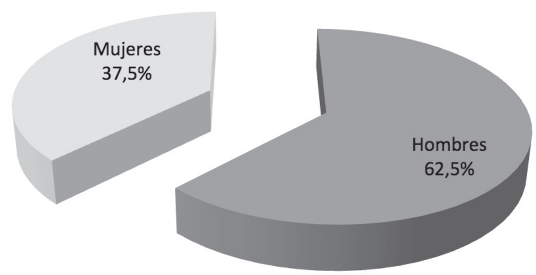


Figura 1.2 Gráfico de sectores. Distribución del sexo



Ejemplo 1.3

Enunciado

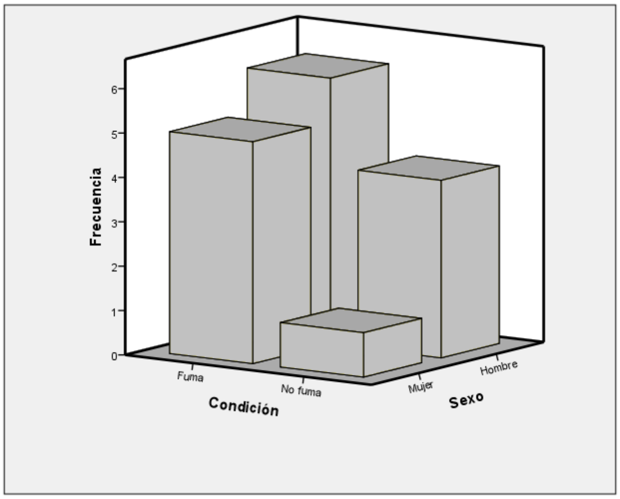
Construya una gráfica para representar las variables sexo y condición de fumador consignadas en la tabla 1.2.

Tabla 1.2 Frecuencias de sexo y condición de fumador

	Masculino	Femenino	Total
Fuma	6	5	11
No fuma	4	1	5
Total	10	6	16

Solución

Figura 1.3 Gráfico de rascacielos.
Sexo *vs.* condición de fumador

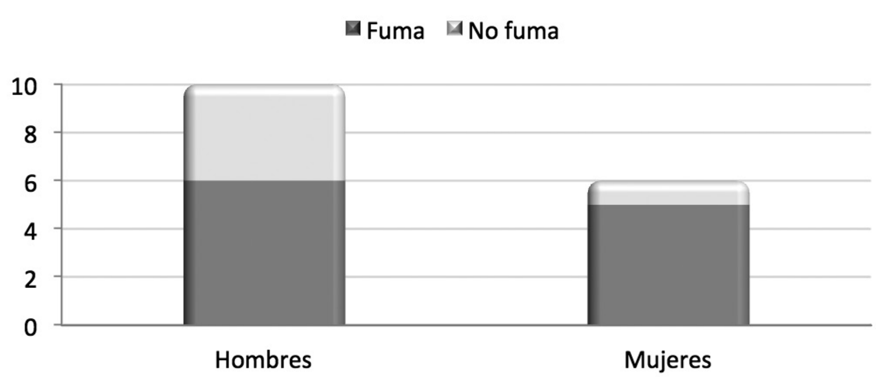


Un gráfico de *rascacielos*, como el de la figura 1.3, es adecuado para estos datos. Las alturas de los prismas rectangulares corresponden a las frecuencias de cada uno de los cuatro grupos formados por las dos variables:

Seis hombres fumadores.
Cuatro hombres no fumadores.
Cinco mujeres fumadoras.
Una mujer no fumadora.▪

Otra gráfica para el mismo conjunto de datos es la de la figura 1.4. A esta se le llama diagrama de **barras bivariado** o de **columnas apiladas** y representa los mismos cuatro grupos de la figura 1.3.

Figura 1.4 Gráfico de barras bivariado. Sexo *vs.* condición de fumador



En los dos últimos gráficos, sobresale más el grupo de hombres fumadores. Esto podría sugerir, erróneamente, que hay mayor número de fumadores entre los hombres. En cifras absolutas, esto es cierto, pero en términos relativos existe un mayor porcentaje de fumadores en las mujeres. Para evitar malentendidos con las cifras absolutas, se construye un gráfico de **barras de frecuencias relativas o mosaico** que represente cada sexo en columnas de la misma altura. Para ello, se parte de la tabla 1.2. Por su parte, en la tabla 1.3 se muestran las frecuencias relativas a la tabla, la fila (condición de fumador) y la columna (sexo).

Tabla 1.3 Frecuencias relativas de sexo y condición de fumador

	Hombres	Mujeres
Fuma	6	5
	37,5%	31,3%
	54,5%	45,5%
	60,0%	83,3%
No fuma	4	1
	25,0%	6,3%
	80,0%	20,0%
	40,0%	16,7%

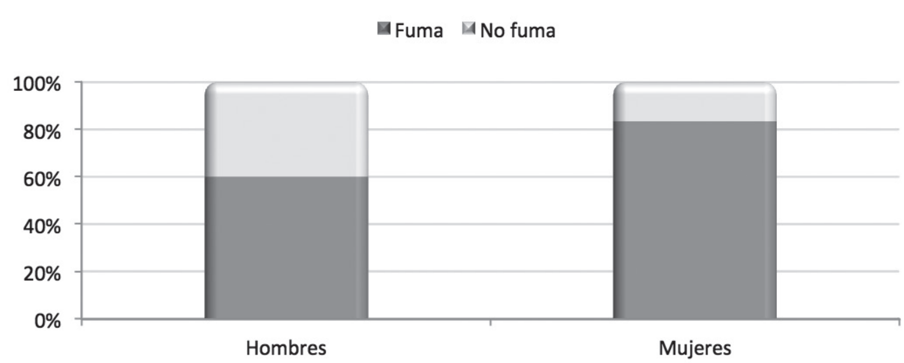
Celda
Frecuencia absoluta
Porcentaje de la tabla
Porcentaje de la fila
Porcentaje de la columna

A manera de ejemplo, se muestran los significados de los valores en la última celda:

1	Número de mujeres no fumadoras.
6,3%	Porcentaje de mujeres no fumadoras.
20,0%	Porcentaje de mujeres dentro de los no fumadores.
16,7%	Porcentaje de no fumadoras dentro de las mujeres.

Con base en la información de las celdas correspondientes a los porcentajes de la columna de la tabla 1.3 se construye la figura 1.5.

Figura 1.5 Mosaico. Sexo *vs.* condición de fumador



En este último gráfico, se observa que existe mayor porcentaje de fumadores en las mujeres que en los hombres, en la muestra de dieciséis personas de la tabla 1.2.▪

Ejemplo 1.4

Enunciado

La mortalidad perinatal (MP) está constituida por todas las muertes ocurridas entre la semana dieciocho de gestación y el séptimo día de nacimiento. El desarrollo social, acceso, cobertura y calidad en los servicios de salud en una población pueden ser evaluados por medio de este tipo de información. En la tabla 1.4, se presentan los casos de MP en los departamentos del país durante el año 2009.

Construir una gráfica con las estadísticas de MP en Colombia contenidas en la tabla 1.4.

Solución

No es posible realizar una comparación de MP entre los departamentos colombianos solo con el número de casos registrados durante el año 2009. Esta cifra absoluta no es comparable, ya que resulta lógico que se presenten más casos en aquellos departamentos donde hay mayor población y esto puede suceder de manera independiente de las condiciones de los servicios de salud. Para realizar una comparación objetiva, la estadística MP se debe acompañar del tamaño de la población en cada departamento. De esta manera, se calculan los casos de MP por cada cien mil habitantes, cifra que resulta comparable y da un indicio de lo que ocurre en cada departamento.

Como en otras cifras de salud pública, no se muestran las estadísticas por cada cien habitantes debido a la baja incidencia del fenómeno. Esto se debe a que, tratándose de casos, es deseable expresar los resultados en valores enteros. Se agregó a la tabla la población total y el cálculo del número de casos por cada 100.000 habitantes.

Tabla 1.4 Mortalidad perinatal en Colombia 2009

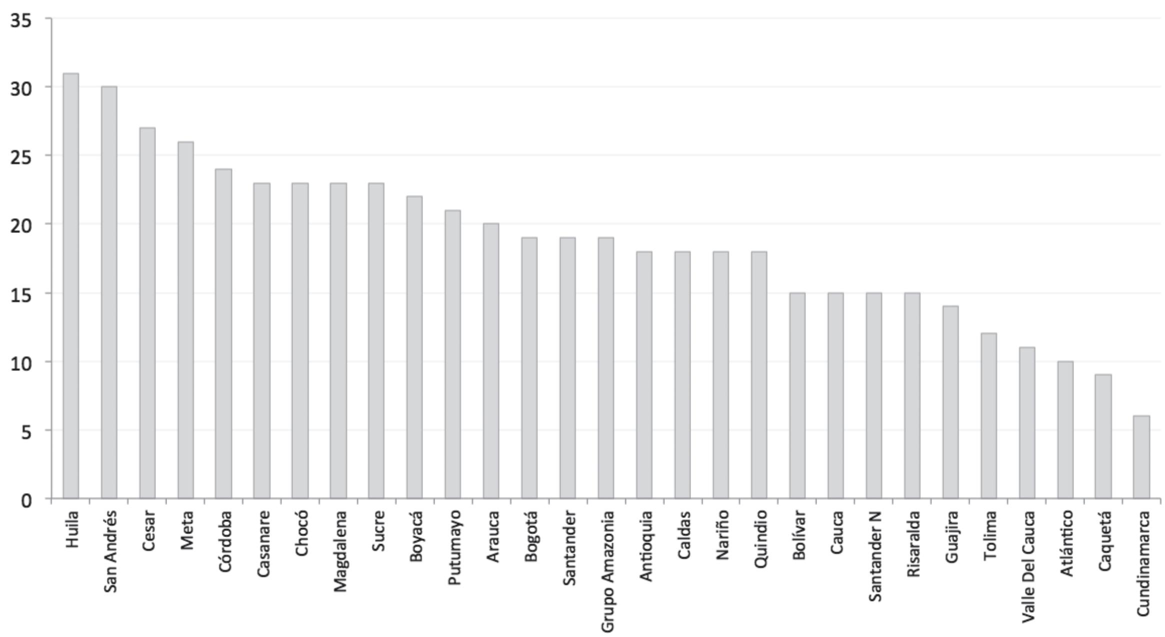
Depto.	Antioquia	Arauca	Atlántico	Bogotá	Bolívar	Boyacá	Caldas	Caquetá	Casanare	Cauca
Casos de MP	1.059	49	231	1.353	292	283	173	41	73	199
Población total	5.988.458	244.507	2284840	7.259.597	1.958.224	1.265.517	976.438	442.033	319.502	1.308.090
MP x 100.000	18	20	10	19	15	22	18	9	23	15

Depto.	Cesar	Chocó	Córdoba	C/marca	Huila	Guajira	Magdalena	Meta	Nariño	Santander N
Casos de MP	255	109	372	146	328	112	277	219	298	189
Población total	953.827	471.601	1.558.793	2.437.151	1.068.820	791.027	1.190.585	853.115	1.619.464	1.286.728
MP x 100.000	27	23	24	6	31	14	23	26	18	15

Depto.	Putumayo	Quindío	Risaralda	San Andrés	Santander	Sucre	Tolima	Valle	Amazonia
Casos de MP	69	101	139	22	382	183	168	490	61
Población total	322.681	546.566	919.653	72.735	2.000.045	802.733	1.383.323	4.337.909	313.796
MP x 100.000	21	18	15	30	19	23	12	11	19

De esta manera, se representa en la figura 1.6 cada uno de los departamentos junto con el número de casos de MP por cada 100.000 habitantes. Para que la lectura de la gráfica sea más sencilla, se ordenaron los departamentos según la mortalidad perinatal. Note que los dos departamentos con mayor MP son Huila y San Andrés, mientras que los dos con menor MP son Caquetá y Cundinamarca. Bogotá es una región en la que se presenta una MP intermedia.▪

Figura 1.6 Mortalidad perinatal en Colombia 2009. Casos por cada 100.000 habitantes



Ejercicio 1.1 Descriptiva

1. Plantee una pregunta de investigación en la que resulte útil un estudio estadístico.

2. Discuta el proceso que se debería adelantar para estudiar cuáles son los hábitos y gastos de los estudiantes de la Universidad en telefonía celular.

3. Suponga que desea averiguar cuáles son las empresas en el país con mejores salarios. Discuta cuál es la población y cómo se debería seleccionar una muestra.

4. Determine el elemento sobre el que se deben medir las siguientes variables:
   a. Candidato por el que se votará en una elección.
   b. Número de goles de un jugador.
   c. Calificación de un estudiante en un parcial.
   d. Resultado de una prueba de embarazo.
   e. Precipitación. (La cantidad de lluvia se mide en mm y equivale al espesor de la lámina de agua que se formaría por su causa sobre una superficie de 1 m^2 plana e impermeable).

5. Diga si las siguientes variables son cuantitativas o cualitativas.
   a. Saldo en las cuentas de ahorro de un banco.
   b. Goles marcados en los partidos de fútbol de un campeonato.
   c. Estado en el que inician sesión los usuarios de *messenger.*
   d. Nacimientos mensuales en el hospital San Ignacio.
   e. Tiempo de duración de una llamada.
   f. Porcentaje de estudiantes fumadores en las universidades de Bogotá.

6. Explique la diferencia entre variables cuantitativas y cualitativas. Dé un ejemplo de cada una.

7. Explique la diferencia entre variables cualitativas nominales y ordinales. Dé un ejemplo de cada una.

8. Diga si las siguientes variables son discretas o continuas.
   a. Número de dulces en un paquete.
   b. Saldo en una cuenta corriente.
   c. Calificación de un estudiante en el primer corte del semestre.
   d. Número de titulares de una tarjeta de crédito en mora.

9. Explique la diferencia entre variables discretas y continuas. Dé un ejemplo de cada una.

10. Los siguientes datos corresponden a titulares de créditos de un banco:

Identificación	Nombre	Edad	Profesión	Ingresos	Vivienda	Vehículo
802342…	María C…	23	Abogado	6’540.000	Si	CRE543
791234…	Juan An..	54	Traductor	5’000.400	Si	SAP232
102345…	Ricardo…	35	Ingeniero	4’850.000	No	-
323212…	Mario…	29	-	2’350.000	No	HDR342

a. ¿Cuáles son los elementos de la población?
b. ¿Cuáles son las variables?
c. ¿Qué tipo de variable es cada una?

11. Dos mil amas de casa del sur de la ciudad escogieron el supermercado de su preferencia, con los siguientes resultados:

Supermercado	Amas de casa
Carrefour	100
Cafam	543
Éxito	754
Carulla	84
Olímpica	179
Colsubsidio	340
Total	2000

a. ¿Cuáles son los elementos de la población?
b. ¿Estos datos corresponden a una población o a una muestra?
c. ¿Cuál es la variable que se mide? ¿Es cualitativa o cuantitativa?
d. Construya un gráfico de barras para describir los datos.
e. Construya un gráfico de sectores para describir los datos.
f. ¿Cuál es la proporción de amas de casa que tienen por preferencia el Éxito?
g. ¿Cuál es la proporción de amas de casa que no tienen por preferencia a Cafam o a Colsubsidio?

12. Analice los datos de mortalidad perinatal contenidos en la figura 1.6.

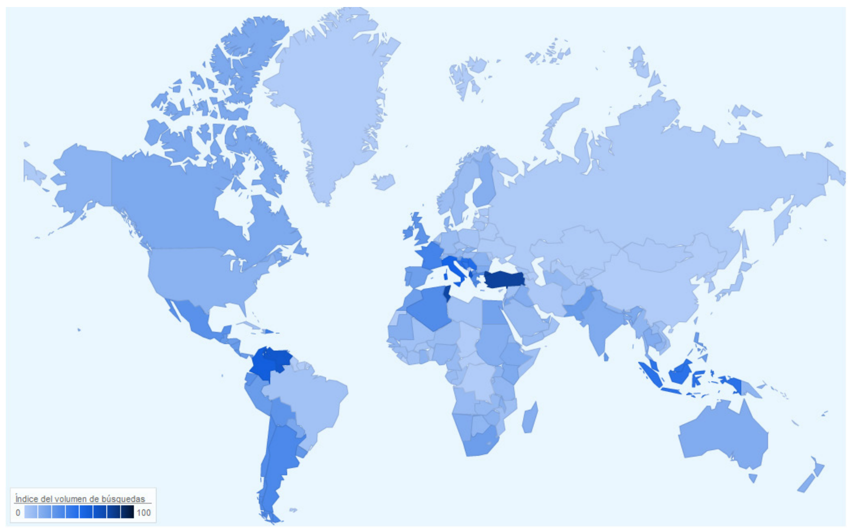
13. En un salón de clases los estudiantes registraron el lugar de fabricación de sus teléfonos celulares, con los siguientes resultados:

Shanghái	Estados Unidos	Shanghái	México
Shanghái	Shanghái	Hong Kong	Taiwán
Shanghái	México	Taiwán	Taiwán
Taiwán	Shanghái	Taiwán	Shanghái
Hong Kong	Estados Unidos	Hong Kong	Estados Unidos
Shanghái	Shanghái	Taiwán	Taiwán
Shanghái	Shanghái	Taiwán	Taiwán
Hong Kong	Shanghái	México	Taiwán

a. ¿Cuáles son los elementos de la población?
b. ¿Cuál es la variable que se mide?
c. Construya un gráfico para describir los datos.
d. ¿Qué proporción de teléfonos fueron construidos fuera de China?
e. ¿Qué proporción de teléfonos fueron construidos en América?

14. La clasificación de países con mayor número de búsquedas en la red del término "facebook" está encabezado por Turquía. Si su volumen de búsquedas se toma como base (100), Colombia tendría el puesto séptimo con 58. La clasificación completo se muestra en la siguiente tabla:

1	Turquía	100
2	Túnez	79
3	Italia	75
4	Venezuela	74
5	Albania	70
6	Croacia	60
7	Colombia	58
8	Francia	54
9	Bosnia	52
10	Indonesia	47



a. Construya un gráfico de sectores y otro de barras para describir la información.
b. ¿Cuál de los dos gráficos es más adecuado para estos datos?

15. Lo que un grupo de mujeres colombianas respondió a la pregunta ¿qué condición prefiere en un hombre?, de la encuesta titulada *Lo que las mujeres quieren*, se resume en la siguiente tabla:

Que sea sociable	22%
Que sea talentoso	60%
Que tenga dinero	11%
Que tenga pinta	5%

a. ¿Están todas las condiciones enumeradas en la tabla de resultados? Al agregar otra condición, ¿cambiarían los porcentajes de las demás?
b. Construya un gráfico de barras para describir los datos.

c. Si usted le realiza esta misma pregunta a una nueva mujer dándole las mismas opciones, ¿cuál cree que es la respuesta más probable?

16. Imagine que se desea averiguar cuál es el destino que eligieron los estudiantes de una universidad en sus últimas vacaciones. Redacte la pregunta que se debería incluir en una encuesta dirigida a estudiantes de diferentes facultades, jornadas y semestres de esa universidad, de manera que los encuestados puedan elegir una y solo una de las opciones incluidas en ella. ¿Qué características debe tener la pregunta? Pídale a su profesor un ejemplo de categorías de respuestas exhaustivas y excluyentes.

17. En estudios de opinión, es frecuente la utilización de la escala Likert para determinar la orientación de las respuestas de un grupo de personas con respecto a un tema. Por ejemplo, en un estudio acerca de la posición de estudiantes de Administración y Relaciones Internacionales frente a la posibilidad de incluir en su carrera la certificación en un tercer idioma, se eligió la siguiente escala de respuesta:

1	Bastante en desacuerdo
2	Algo en desacuerdo
3	Neutral
4	Algo de acuerdo
5	Bastante de acuerdo
6	No aplica

Un resumen de las respuestas dadas por 480 estudiantes se presenta en la siguiente tabla:

	Administración	Relaciones	Total
Bastante en desacuerdo	93	9	102
	19,6%	1,9%	21,5%
	91,2%	8,8%	
	31,5%	5,0%	
Algo en desacuerdo	136	16	152
	28,6%	3,4%	32,0%
	89,5%	10,5%	
	46,1%	8,9%	
Neutral	30	26	56
	6,3%	5,5%	11,8%
	53,6%	46,4%	
	10,2%	14,4%	
Algo de acuerdo	22	57	79
	4,6%	12,0%	16,6%
	27,8%	72,2%	
	7,5%	31,7%	
Bastante de acuerdo	8	67	75
	1,7%	14,1%	15,8%
	10,7%	89,3%	
	2,7%	37,2%	
No aplica	6	5	11
	1,3%	1,1%	2,3%
	54,5%	45,5%	
	2,0%	2,8%	
Total	295	180	475
	62,1%	37,9%	100,0%

Frecuencia
% tabla
% fila
% columna

a. Solo hay 475 respuestas de estudiantes en el recuento que aparece en la tabla. Existen varias razones para que esto ocurra, ¿puede mencionar algunas de ellas?
b. ¿Qué circunstancias pueden estar cobijadas dentro de la categoría: "No aplica"?
c. ¿Cuántos estudiantes de Administración tuvieron una posición neutral?
d. ¿Qué porcentaje de respuestas corresponde a estudiantes de Relaciones?
e. ¿Qué porcentaje de respuestas corresponde a estudiantes de relaciones y que estén algo de acuerdo?
f. Entre los estudiantes de Administración, ¿cuál es el porcentaje de los que estuvieron bastante en desacuerdo?
g. Entre quienes estuvieron bastante de acuerdo, ¿cuál es el porcentaje de estudiantes de Relaciones?
h. Si se elije un estudiante al azar, ¿cuál es su opinión más probable?
i. Si se elije un estudiante al azar y se sabe que estudia relaciones, ¿cuál es su opinión más probable?
j. Desde un punto de vista intuitivo, ¿parece probable que los estudiantes de Administración y de Relaciones difieran sustancialmente en su opinión frente al tema consultado?

18. En un banco, a final de mes, se registró el sexo y el estado de cumplimiento de un grupo de usuarios de tarjetas de crédito.

Sexo	Estado
F	Incumplido
F	Cumplido
F	Incumplido
M	Incumplido
F	Cumplido
F	Cumplido
F	Incumplido
M	Cumplido

Sexo	Estado
F	Cumplido
F	Incumplido
M	Cumplido
F	Incumplido
F	Incumplido
M	Cumplido
F	Incumplido
M	Cumplido

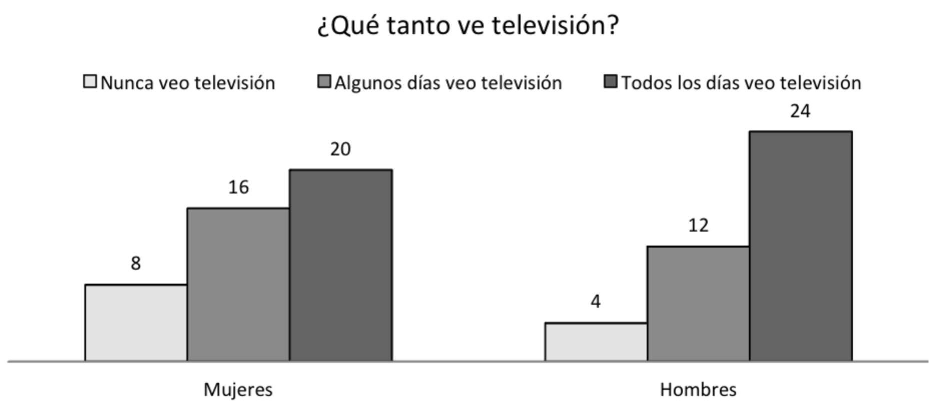
Sexo	Estado
F	Incumplido
M	Cumplido
F	Incumplido
F	Cumplido
F	Incumplido
M	Cumplido
F	Incumplido
F	Cumplido

a. Construya una tabla de sexo contra estado.
b. Comente cualquier diferencia aparente que observe.

19. La paradoja de Simpson. Se realizó un ensayo clínico para determinar la eficacia de dos drogas con los resultados de la tabla. ¿Cuál de las dos drogas es mejor?

	Mujeres		**Hombres**	
	Droga 1	**Droga 2**	**Droga 1**	**Droga 2**
Surtió efecto	200	10	19	1000
No surtió efecto	1800	190	1	1000

20. En el año 2002, se adelantó un estudio para examinar, entre otros aspectos, la relación del sexo en el gusto por la televisión en los estudiantes de colegios públicos en el municipio de La Mesa. El gráfico muestra los resultados de las respuestas de una muestra de estudiantes acerca del sexo y si ven televisión.



a. ¿Cuál es el número de estudiantes en la muestra?
b. ¿Qué porcentaje de estudiantes nunca ve televisión?
c. ¿Qué porcentaje de mujeres nunca ve televisión?
d. De quienes todos los días ven televisión, ¿qué porcentaje son hombres?

21. Los diez términos más buscados en Google el último año en Colombia son:

1	Facebook en español	100
2	Hotmail correo	16
3	facebook.com	10
4	Taringa	8
5	Google traductor	7
6	Youtube.com	4
7	Caracol	3
8	Face	3
9	Facebook	3
10	Juegos de Mario	3

Estos términos pueden corresponder a un grupo menor de categorías. Por ejemplo, taringa puede pertenecer a la categoría de búsqueda de información y software.

a. Construya un grupo reducido de categorías que reúna en grupos los términos de la tabla.

b. Explique por qué es de utilidad una agrupación de este tipo.

c. Dé un ejemplo en el que sea útil crear categorías para resumir información.

22. En el ejercicio 21 se mencionaron los diez términos más buscados en Google el último año en Colombia.

a. Construya una gráfica adecuada para resumir esta información.

b. Escriba una conclusión del gráfico construido.

23. Los siguientes datos corresponden a las edades y el sexo de veinte asistentes a un curso de inglés avanzado:

Edad y sexo de una muestra de los asistentes a un curso avanzado de inglés (M: masculino, F: femenino)									
15	15	18	25	22	42	40	32	24	28
F	M	M	F	M	F	M	F	M	F
18	18	28	20	16	26	45	30	25	20
M	F	F	M	M	M	F	M	F	F

a. Construya un diagrama de barras para el sexo.

b. Suponga que se desea clasificar a cada persona como "joven" o "mayor" y se determina que una persona es "joven" si tiene menos de 21 años y "mayor" si tiene 21 años o más. Construya una tabla que resuma esta clasificación junto con la variable sexo.

24. Con la información contenida en Avata *Resultados del Mundial,* ¿qué gráfica podría construir con el país y el número de partidos jugados?

25. De acuerdo con las cifras entregadas por la Secretaría de Movilidad de Bogotá, las infracciones o comparendos más comunes cometidos por conductores y motociclistas en la capital, de enero a marzo de 2011 son:

Infracción	Comparendos
Incumplimiento de normas para motocicletas	13.361
Estacionar un vehículo en sitios prohibidos	9.998
Usar el celular al momento de conducir	5.829
No realizar la revisión técnico-mecánica o emisión de gases	5.535
Conducir un vehículo sin portar la licencia de conducción	4.682
No utilizar el cinturón de seguridad	3.593
Transitar por sitios restringidos o en horas prohibidas	3.342
Bloquear una calzada o intersección con un vehículo	2.470
Conducir un vehículo de transporte público de pasajeros incumpliendo las normas	1.360
Dejar o recoger pasajeros en sitios prohibidos	180

a. ¿Cuáles son los elementos de la población?
b. ¿Estos datos corresponden a una población o a una muestra?
c. ¿Cuál es la variable que se mide? ¿Es cualitativa o cuantitativa?
d. Construya un gráfico adecuado para describir los datos.

Gráficos para resumir datos cuantitativos

Una técnica útil para la observación inicial de los datos y la comprensión de lo que representan y cómo se distribuyen es el **diagrama de tallo y hojas**. En este diagrama, los datos se reproducen fielmente y consta de varias columnas, la primera (o tallo) representa en orden la mayor porción de los datos y las siguientes separadas por una línea vertical representan la menor porción de cada dato en particular. Para ver como se construye y su utilidad se incluye el siguiente ejemplo.

Ejemplo 1.5

Enunciado

Construir un diagrama de tallo y hojas con los datos de la tabla 1.5, que corresponden al reporte sobre fertilidad de la Organización Mundial de la Salud (2010), en 109 países. La columna **Fer.** (fertilidad), es el número promedio de hijos que una cohorte hipotética de mujeres tendría al final de su periodo reproductivo.

Tabla 1.5 Fertilidad mundial 2010 (hijos promedio por mujer)

Ubicación	Fer.	Ubicación	Fer.	Ubicación	Fer.	Ubicación	Fer.	Ubicación	Fer.
Bosnia	1,2	Rusia	1,4	Corea del Norte	1,9	Sudáfrica	2,5	Samoa	4,0
Corea del Sur	1,2	España	1,4	Líbano	1,9	Venezuela	2,5	Guatemala	4,1
Andorra	1,3	Barbados	1,5	Noruega	1,9	Ecuador	2,6	Irak	4,1
Alemania	1,3	Cuba	1,5	Suecia	1,9	Panamá	2,6	Sudán	4,2
Japón	1,3	Suiza	1,5	Costa Rica	2,0	Perú	2,6	Congo	4,4
Malta	1,3	Canadá	1,6	Irlanda	2,0	República Dominicana	2,7	Camerún	4,6
Polonia	1,3	Serbia	1,6	Maldivas	2,0	India	2,7	Kenia	4,9
Rumania	1,3	Trinidad	1,6	Mongolia	2,0	Nicaragua	2,7	Senegal	5,0
Singapur	1,3	Armenia	1,7	Nueva Zelanda	2,0	Israel	2,8	Liberia	5,1
Eslovaquia	1,3	Estonia	1,7	Dominica	2,1	Camboya	2,9	Sierra Leona	5,2
Ucrania	1,3	Países Bajos	1,7	Turquía	2,1	Egipto	2,9	Yemen	5,2
Austria	1,4	Australia	1,8	Uruguay	2,1	Nepal	2,9	Etiopía	5,3
Bulgaria	1,4	Bélgica	1,8	Estados Unidos	2,1	Paraguay	3,1	Nigeria	5,3
Croacia	1,4	China	1,8	Argentina	2,2	Filipinas	3,1	Angola	5,8
Rep. Checa	1,4	Dinamarca	1,8	Indonesia	2,2	Arabia Saudita	3,1	Zambia	5,8
Grecia	1,4	Finlandia	1,8	Kuwait	2,2	Honduras	3,3	Congo	6,0
Hungría	1,4	Irán	1,8	México	2,2	Siria	3,3	Chad	6,2
Italia	1,4	Tailandia	1,8	El Salvador	2,3	Namibia	3,4	Uganda	6,3
Letonia	1,4	Gran Bretaña	1,8	Sri Lanka	2,3	Zimbabue	3,4	Somalia	6,4
Lituania	1,4	Brasil	1,9	Colombia	2,4	Bolivia	3,5	Afganistán	6,6
Macedonia	1,4	Chile	1,9	Jamaica	2,4	Haití	3,5	Nigeria	7,1
Portugal	1,4	Francia	1,9	Qatar	2,4	Pakistán	4,0		

Solución

El tallo del diagrama está constituido por los números acompañados de las comas decimales. Esta es la parte común de los números de la correspondiente fila, por ejemplo, los veinticuatro números en la primera fila van desde 1,2 hasta 1,4 y los dieciocho de la tercera fila del 2,0 al 2,4. En la primera columna, está el número de datos hasta la correspondiente fila contando de izquierda a derecha o de derecha a izquierda salvo en la fila que contiene al dato en la posición central. En la siguiente sección se discutirá en más detalle este último hecho.

24	1,	2 2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4
48	1,	5 5 5 6 6 6 7 7 7 8 8 8 8 8 8 8 8 9 9 9 9 9 9 9
(18)	2,	0 0 0 0 0 1 1 1 1 2 2 2 2 3 3 4 4 4
43	2,	5 5 6 6 6 7 7 7 8 9 9 9
31	3,	1 1 1 3 3 4 4
24	3,	5 5
22	4,	0 0 1 1 2 4
16	4,	6 9
14	5,	0 1 2 2 3 3
8	5,	8 8
6	6,	0 2 3 4
2	6,	6
1	7,	1

Se puede ver que existe gran concentración de países donde el número promedio de hijos por mujer es menor a 3,5 (primeras cinco filas). Este diagrama de tallo y hojas es útil para determinar la posición exacta de cada una de las observaciones; por ejemplo, el máximo de las tasas de fertilidad es 7,1 y el mínimo es 1,2 que se repite dos veces.

Si no resulta sencillo agrupar los datos en categoría o intervalos resultan muy útiles los dos siguientes tipos de gráficos: **diagrama de dispersión** (una variable) y **gráfico de puntos**.

Ejemplo 1.6

Enunciado

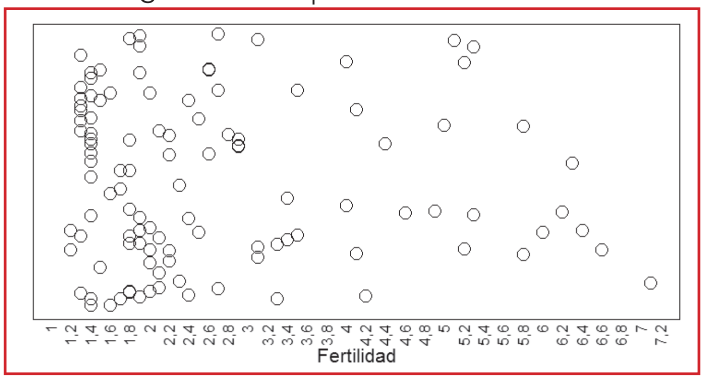
Construir un diagrama de dispersión de los datos de fertilidad.

Solución

Para construir el diagrama, primero se ubica una escala en el eje *X*, que corresponde al rango de los datos de fertilidad. Cada uno de los puntos de la figura 1.7

representa un país y su ubicación refleja el número de hijos promedio por mujer. La altura de los puntos se decide al azar para tener una mejor visualización de los datos. Se puede ver una gran concentración de países donde las mujeres tienen en promedio entre 1,2 y 3,4 hijos. También se destaca Nigeria con un número promedio de hijos por mujer muy alto.▪

Figura 1.7 Diagrama de dispersión de la fertilidad mundial



Ejemplo 1.7

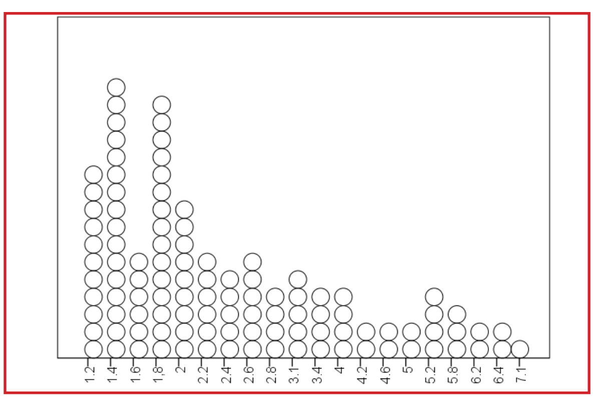
Enunciado

Construir una gráfica de puntos con los datos de fertilidad del ejemplo 1.5.

Solución

En la figura 1.8 cada dato se representa con un punto ubicado sobre el eje X. Valores coincidentes se ubican uno sobre otro. En este caso, por comodidad, la primera pila de puntos corresponde a valores 1,2 o 1,3, la segunda a valores entre 1,4 y 1,5 y así sucesivamente.

Figura 1.8 Gráfica de puntos de la fertilidad mundial



Se vuelve a notar la concentración de países con un número promedio de hijos por mujer en los valores más bajos. Sin embargo, se resalta que hay un número considerable de ellos donde las mujeres tienen en promedio entre 1,4 y 1,5 hijos.▪

Ejemplo 1.8

Enunciado

En la tabla 1.6, se muestra un resumen de las base de datos de cédulas de la Registraduría Nacional del Estado Civil recopilados por Meisel y Vega (2004).

Tabla 1.6 Estatura en Colombia 1910 a 1984

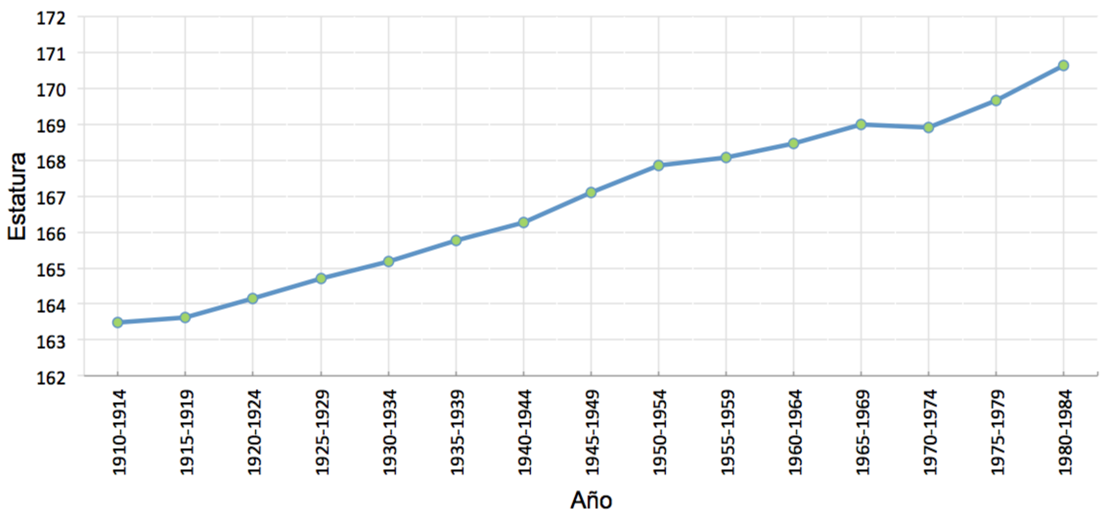
Año de nacimiento	Estatura promedio hombres	Estatura promedio mujeres	Número de hombres	Número de mujeres	Total
1910-1914	163,48	150,78	1.751	2.197	3.948
1915-1919	163,61	151,49	4.582	4.993	9.575
1920-1924	164,16	152,38	9.086	9.779	18.865
1925-1929	164,7	153,06	15659	16.299	31.958
1930-1934	165,17	153,48	22.219	24.619	46.838
1935-1939	165,76	154,21	34.637	33.820	68.457
1940-1944	166,26	154,69	40.186	45.623	85.809
1945-1949	167,1	155,59	53.164	60.723	113.887
1950-1954	167,84	156,4	73.835	80.863	154.698
1955-1959	168,07	156,81	101.613	111.278	212.891
1960-1964	168,47	157,17	142.054	157.950	300.004
1965-1969	169	157,34	177.464	221.024	398.488
1970-1974	168,91	157,21	298.908	426.660	725.568
1975-1979	169,66	157,81	1’639.499	1’646.878	3’286.377
1980-1984	170,64	158,65	1’528.875	1’468.110	2’996.985
Total	169,664	157,778	4’143.532	4’310.816	8’454.348

Construir y analizar una gráfica para resumir la información de las estaturas de los hombres.

Solución

En este ejemplo resulta importante resaltar la tendencia de la estatura promedio en el tiempo. En general, cuando este es el objetivo se elige un **gráfico de líneas**, como el que se muestra enseguida.

Figura 1.9 Evolución de la estatura promedio de los hombres en Colombia



En la figura 1.9, se puede ver que existe una tendencia de aumento de la estatura a lo largo del tiempo. Algunos autores utilizan la variable estatura adulta como medida indirecta de la calidad de vida, porque ella refleja la situación nutricional durante los años en los que la persona está en crecimiento. Si esto es así, es posible concluir que se evidencia un aumento en la calidad de vida de este grupo de hombres en el periodo considerado.▪

Histogramas

Cuando una variable es cuantitativa de tipo continuo, la representación gráfica del conteo de frecuencias resulta inadecuada, debido a la gran cantidad o infinidad de valores que puede tomar la variable. En este caso, es útil dividir en intervalos los valores de la variable, **intervalos de clase**, para luego graficar las frecuencias observadas en cada uno de ellos. A la anterior representación, se le denomina **histograma**. Es aconsejable que el número de clases no sea muy pequeño, a fin de no perder información y que no sea demasiado grande como para no poder mostrar un patrón distinguible en la distribución de los datos. El número adecuado de clases para la construcción de un histograma es una elección subjetiva y resulta conveniente probar con diferentes números de clases para determinar cuál de los histogramas descubre algún patrón en los datos. No obstante, una propuesta para determinar el número de clases, sugerida por Herbert Sturges, es incluida en los siguientes pasos para la construcción de un histograma.

Histograma: representación gráfica de una distribución de frecuencias por medio de rectángulos, cuyas anchuras representan intervalos de la clasificación y cuyas alturas representan las correspondientes frecuencias.

La forma como se reparten los porcentajes a través del histograma se suele llamar informalmente la **distribución de la variable o distribución de frecuencias.**

Pasos para la construcción de un histograma:

1. Determinar el número de clases de la distribución de frecuencias. La sugerencia de Sturges es:

Tabla 1.7 Número de clases en un histograma

Número de datos	**Número de clases**
De 1 a 15	Muy pocos datos
De 16 a 31	5
De 32 a 63	6
De 64 a 127	7
$\vdots$	$\vdots$
n	$1 + 3,322 \log_{10} n$

2. Calcular el rango de los datos, que corresponde a la diferencia entre el dato mayor y el menor:

$$\text{Rango} = \text{Máximo} - \text{Mínimo}$$

3. Calcular la amplitud de las clases y redondear por exceso a la misma unidad de los datos.

$$\text{Amplitud} = \frac{\text{Rango}}{\text{número de clases}}$$

La razón de redondear por exceso es que con ello se garantiza cubrir por completo el rango de los datos. Si la amplitud llega a tener la misma precisión de los datos, esta se debe aumentar en una unidad completa.

4. Determinar el límite inferior de la primera clase. Para garantizar que ningún dato de los observados caiga en un límite de clase, estos contendrán un decimal más. Es así como la primera clase comenzará media unidad por debajo del valor mínimo.

Precisión de los datos	**Unidad**	**$\frac{1}{2}$ Unidad**
Enteros	1	0,5
Décimas	0,1	0,05
Centésimas	0,01	0,005
Milésimas	0,001	0,0005
Diezmilésimas	0,0001	0,00005
⋮	⋮	⋮

5. Determinar los límites de cada uno de los intervalos de clases y realizar el conteo de frecuencias de los datos en cada una de ellas. Al promedio de los límites de cada clase se la denomina **punto medio de la clase** o **marca de clase**.

Para ilustrar estos pasos se incluye el siguiente ejemplo.

Ejemplo 1.9

Enunciado

Construir el histograma de los datos de la tabla 1.5 sobre las tasas de fertilidad de 109 países. Por conveniencia se reproduce nuevamente la tabla de datos.

Ubicación	**Fer.**	**Ubicación**	**Fer.**	**Ubicación**	**Fer.**	**Ubicación**	**Fer.**	**Ubicación**	**Fer.**
Bosnia	1,2	Rusia	1,4	Corea del Norte	1,9	Sudáfrica	2,5	Samoa	4,0
Corea del Sur	1,2	España	1,4	Líbano	1,9	Venezuela	2,5	Guatemala	4,1
Andorra	1,3	Barbados	1,5	Noruega	1,9	Ecuador	2,6	Irak	4,1
Alemania	1,3	Cuba	1,5	Suecia	1,9	Panamá	2,6	Sudán	4,2
Japón	1,3	Suiza	1,5	Costa Rica	2,0	Perú	2,6	Congo	4,4
Malta	1,3	Canadá	1,6	Irlanda	2,0	República Dominicana	2,7	Camerún	4,6
Polonia	1,3	Serbia	1,6	Maldivas	2,0	India	2,7	Kenia	4,9
Rumania	1,3	Trinidad	1,6	Mongolia	2,0	Nicaragua	2,7	Senegal	5,0
Singapur	1,3	Armenia	1,7	Nueva Zelanda	2,0	Israel	2,8	Liberia	5,1
Eslovaquia	1,3	Estonia	1,7	Dominica	2,1	Camboya	2,9	Sierra Leona	5,2

Ubicación	Fer.
Ucrania	1,3
Austria	1,4
Bulgaria	1,4
Croacia	1,4
Rep. Checa	1,4
Grecia	1,4
Hungria	1,4
Italia	1,4
Letonia	1,4
Lituania	1,4
Macedonia	1,4
Portugal	1,4

Ubicación	Fer.
Países Bajos	1,7
Australia	1,8
Bélgica	1,8
China	1,8
Dinamarca	1,8
Finlandia	1,8
Irán	1,8
Tailandia	1,8
Reino Unido	1,8
Brasil	1,9
Chile	1,9
Francia	1,9

Ubicación	Fer.
Tuquía	2,1
Uruguay	2,1
Estados Unidos	2,1
Argentina	2,2
Indonesia	2,2
Kuwait	2,2
México	2,2
El Salvador	2,3
Sri Lanka	2,3
Colombia	2,4
Jamaica	2,4
Qatar	2,4

Ubicación	Fer.
Egipto	2,9
Nepal	2,9
Paraguay	3,1
Filipinas	3,1
Arabia Saudita	3,1
Honduras	3,3
Siria	3,3
Namibia	3,4
Zimbabue	3,4
Bolivia	3,5
Haití	3,5
Paquistán	4,0

Ubicación	Fer.
Yemen	5,2
Etiopía	5,3
Nigeria	5,3
Angola	5,8
Zambia	5,8
Congo	6,0
Chad	6,2
Uganda	6,3
Somalia	6,4
Afganistán	6,6
Nigeria	7,1

Solución

Paso 1: según la tabla 1.7, el número de clases para este histograma es siete.

Paso 2:

$$\textit{Rango} = \textit{Máximo} - \textit{Mínimo} = 7{,}1 - 1{,}2 = 5{,}9$$

Paso 3:

$$\text{Amplitud} = \frac{\text{Rango}}{\text{número de clases}} = \frac{5,9}{7} = 0,842857143 \approx 0,9$$

El ancho real utilizado se halla redondeando la amplitud hasta la misma cantidad de decimales que los datos.

Paso 4: el límite inferior de la primera clase debe estar media unidad por debajo del valor mínimo. Como los datos tienen un decimal, la unidad es 0,1 y media unidad 0,05; esto significa que el límite inferior de la primera clase es 1,2 – 0,05= 1,15. Los límites de las demás clases se obtienen sumando sucesivamente la amplitud.

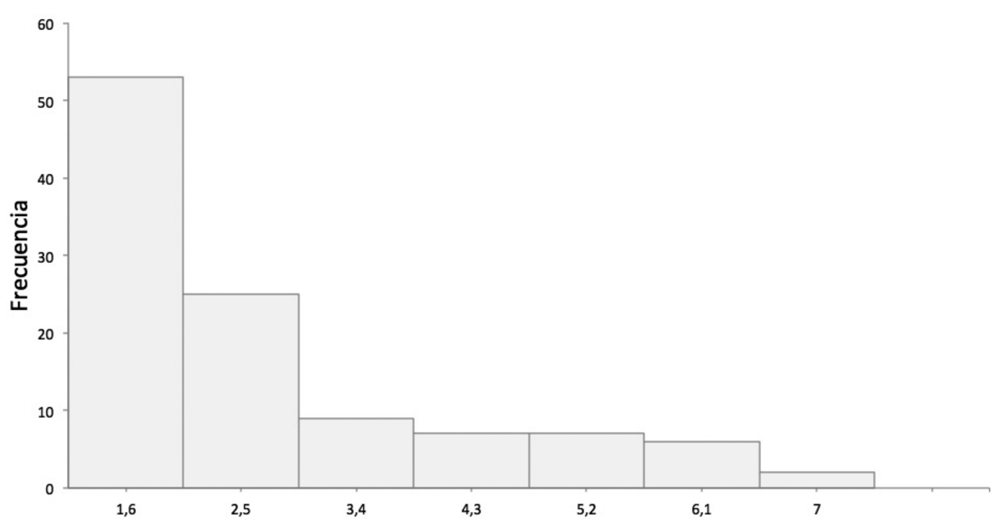
Paso 5: por último, se examina la frecuencia de observaciones en cada una de las clases conformadas y se construye el histograma de la figura 1.10 con las frecuencias absolutas. También se puede construir con las frecuencias relativas.

Al seguir los pasos anteriores, se obtienen los datos de la tabla 1.8, los cuales se resumen en la figura 1.10.

Tabla 1.8 Distribución de frecuencias de la fertilidad

Clase	Límite inferior	Límite superior	Frecuencia	Marca de clase
1	1,15	2,05	53	1,6
2	2,05	2,95	25	2,5
3	2,95	3,85	9	3,4
4	3,85	4,75	7	4,3
5	4,75	5,65	7	5,2
6	5,65	6,55	6	6,1
7	6,55	7,45	2	7,0

Figura 1.10 Distribución de la fertilidad. Hijos promedio por mujer



El análisis de este histograma revela que un alto porcentaje de los países está concentrado en las tres primeras clases.▪

Algunas formas típicas de distribuciones de frecuencias se muestran en la figura 1.11. En su orden, distribución de frecuencias sesgada a la derecha, simétrica y sesgada a la izquierda.▪

Figura 1.11. Forma de las distribuciones

Sesgada a la derecha   Distribución simétrica   Sesgada a la izquierda

En el anterior ejemplo, se observa que la distribución de frecuencias de la variable fertilidad es sesgada a la derecha. Esto significa que la mayoría de países tienen una fertilidad que se concentra en los valores más bajos y que solo unos pocos tienen una fertilidad muy alta.

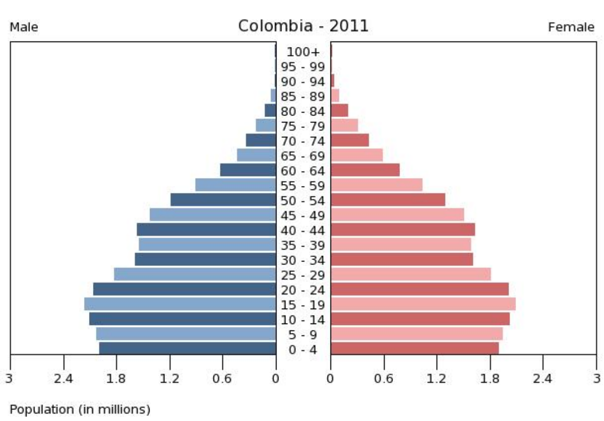
Pirámides de población

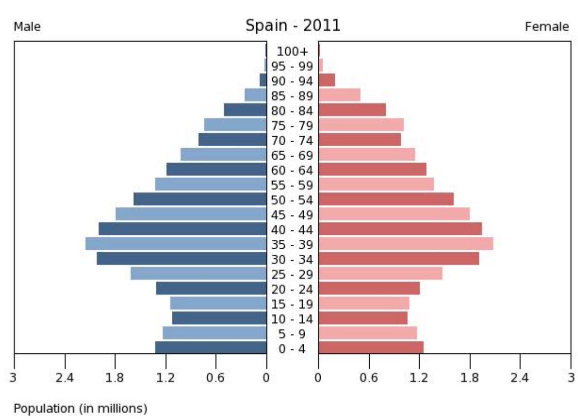
Un tipo especial de histogramas que se utiliza en demografía son las pirámides de población. En estas gráficas generalmente se compara la distribución de edades de una población por sexo. Las barras de los histogramas de cada sexo aparecen horizontales y yuxtapuestas. Cada barra representa la frecuencia absoluta o relativa del grupo de personas en un rango de edad. El nombre de pirámide es consecuencia del hecho de que la mayoría de individuos de una población se concentra en los grupos de edad inferiores y solo una pequeña porción de los mismos en los grupos de mayor edad.

Ejemplo 1.10

Enunciado

Interpretar y comparar las pirámides de población de Colombia y España en 2011.





Solución

La pirámide de población de Colombia 2011, tiene una base ancha y una cima angosta. Esto significa que hay una gran cantidad de población joven. Sobresale el grupo de colombianos entre 15 y 19 años con la mayor frecuencia y que solo una reducida porción alcanza edades superiores a los 85 años.

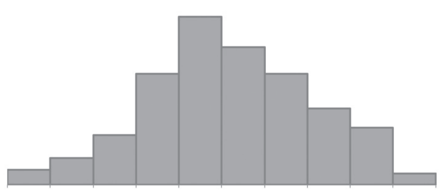
En España se observa una base muy angosta, lo que significa una presencia menor de jóvenes y un abultamiento notable en grupos de edades cercanas al rango de los 35 a los 39 años. También se puede destacar que la población puede alcanzar con facilidad edades superiores a los 89 años.

Según datos de los censos de población, el número de habitantes de Colombia y España es semejante: aproximadamente 45 y 46 millones respectivamente. No obstante, saltan a la vista dos diferencias importantes en la distribución de las edades. La primera, el envejecimiento de la población española y la segunda, su mayor esperanza de vida. Se puede consultar que la esperanza de vida en Colombia en 2011 era de 75 años mientras que en España era de 81 años.▪

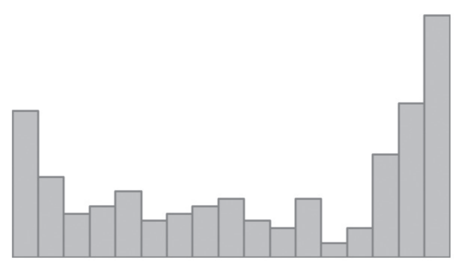
Ejercicio 1.2 Descriptiva 2

1. Empareje cada histograma con una de las variables I, II o III. Mencione las unidades y el rango adecuado.

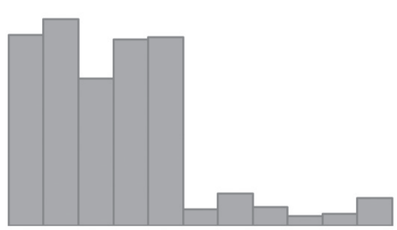
A



C



B



I. Edad de una persona al morir en una población durante un año.
II. Estatura de un grupo de niños de una escuela primaria.
III. Salario de los trabajadores de una empresa.

4. Dibuje histogramas que se ajusten a las distribuciones de cada una de las siguientes variables:

a. Salarios de trabajadores en una empresa en la cual la mayoría gana poco.

b. Tiempos de deportistas que compiten en 100 metros planos.

c. Edad al primer parto de las madres chocoanas.

3. Las siguientes son las calificaciones en el examen final de microeconomía de cien estudiantes, pertenecientes a dos sedes de una universidad, cada una con cincuenta estudiantes.

SEDE A	3,7	3,1	4,9	3,5	3,3	4,2	3,9	3,6	3,7	3,8
	4,2	4,5	3,0	3,8	3,8	4,0	3,4	4,4	3,3	4,6
	3,8	4,3	4,3	4,5	4,2	4,1	4,1	3,9	4,0	3,4
	3,9	4,1	3,8	4,4	3,9	4,3	3,8	3,2	3,2	3,9
	4,3	3,7	3,4	3,5	4,1	3,1	3,3	4,2	4,3	4,6

SEDE B	3,7	3,0	4,1	2,7	2,9	2,9	3,7	3,1	3,3	2,4
	2,9	3,0	3,2	3,1	2,4	3,2	2,7	2,6	4,0	3,7
	3,1	2,1	2,5	3,3	3,1	3,0	3,1	3,4	3,3	1,6
	2,2	3,1	2,6	2,5	3,2	3,3	3,3	2,6	3,1	3,7
	3,7	3,5	3,8	3,2	2,9	3,0	3,4	1,9	2,2	2,0

a. Construya un diagrama de tallo y hojas para las calificaciones en cada una de las sedes.

b. Describa la forma de la distribución de las calificaciones en cada una de las sedes.

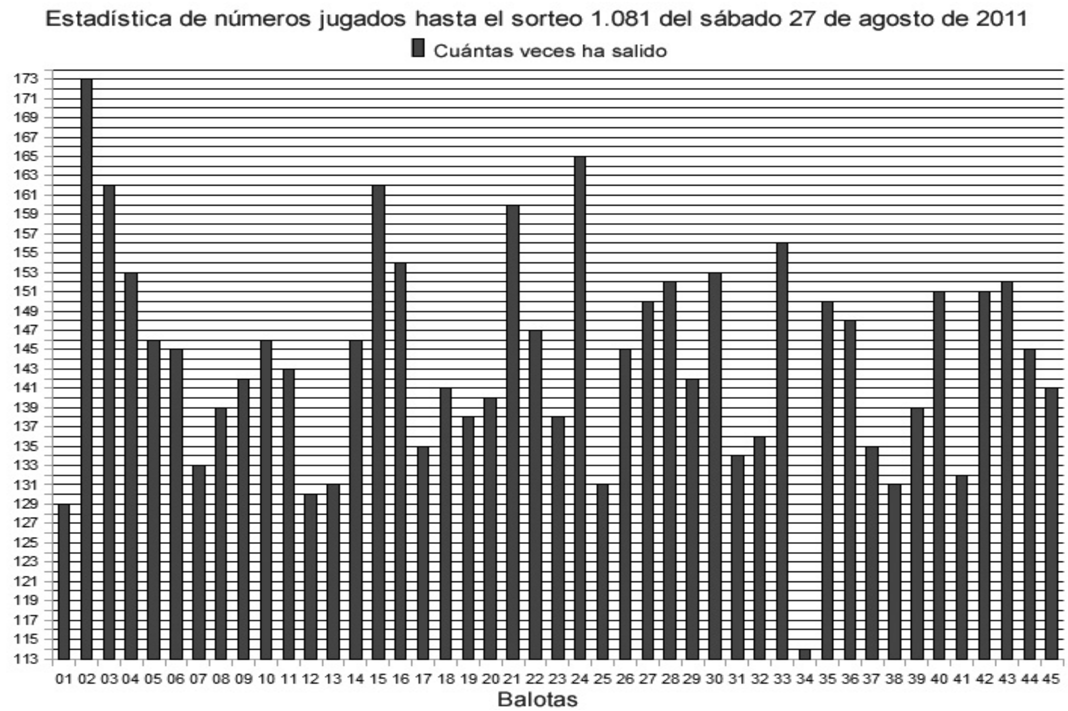
c. En la sede A ¿qué proporción de los estudiantes obtuvieron calificaciones inferiores a 3,0?

d. En la sede B, ¿qué proporción de los estudiantes obtuvieron calificaciones superiores a 3,0?

e. ¿En cuál sede se obtuvieron mejores calificaciones?

f. Utilice Excel para construir un histograma de las calificaciones en cada una de las sedes.

4. La siguiente gráfica muestra el número de veces que ha salido cada número en el baloto. ¿La gráfica es simétrica o sesgada?



5. Un vendedor ambulante de minutos a celular registró la duración de las últimas cuarenta llamadas de sus clientes. Los tiempos en segundos se muestran a continuación:

18	155	171	31	142	310	229	24
70	4	195	155	279	87	8	33
55	23	145	256	15	38	62	24
79	40	5	24	85	124	100	571
12	132	60	39	14	123	222	38

a. Construya un histograma de frecuencias relativas para describir los datos.
b. La distribución de la duración de las cuarenta llamadas, ¿es simétrica, sesgada a la derecha o a la izquierda?
c. ¿Qué porcentaje de las llamadas de los clientes es de más de un minuto?
d. ¿Qué porcentaje de las llamadas de los clientes es de un minuto o menos?

6. La siguiente tabla muestra los valores de mercado del dólar y del euro el día uno de cada mes entre enero de 2009 y agosto de 2010.

Fecha	Dólar (en pesos)	Euro (en pesos)
01/01/2009	2.246	3.122
01/02/2009	2.420	3.102
01/03/2009	2.555	3.245
01/04/2009	2.544	3.366
01/05/2009	2.288	3.033
01/06/2009	2.140	3.030
01/07/2009	2.145	3.032
01/08/2009	2.040	2.893
01/09/2009	2.057	2.943
01/10/2009	1.925	2.803
01/11/2009	1.993	2.941
01/12/2009	1.998	3.016
01/01/2010	2.044	2.933
01/02/2010	1.982	2.754
01/03/2010	1.932	2.604
01/04/2010	1.921	2.601
01/05/2010	1.950	2.593
01/06/2010	1.971	2.422
01/07/2010	1.913	2.383
01/08/2010	1.842	2.400

a. Construya una gráfica para describir las dos variable en el tiempo.

b. Describa el comportamiento de las tasas de cambio a partir de la gráfica construida.

7. ¿Cuál es la edad promedio en la que se gradúan los bachilleres de Santander? Para responder esta pregunta, se analizaron las edades de cincuenta estudiantes al momento de graduarse en un colegio de ese departamento.

17	18	18	19	19
16	16	16	20	15
18	17	14	16	15
18	17	17	16	19
16	19	18	16	15
17	16	16	16	15
17	15	17	19	19
16	15	16	16	15
17	17	16	16	16
19	17	15	16	17

a. Construya un gráfico de tallo y hojas para resumir los datos

b. ¿Qué proporción de los estudiantes de la muestra se graduaron antes de los 18 años?
c. ¿Se puede responder a la pregunta inicial con esta muestra?
d. Al tomar la muestra, ¿qué consideraciones se deberían tener en cuenta?

8. Cincuenta consumidores de una marca de crema para manos calificaron un nuevo aroma del producto en una escala de 00 a 100.

100	93	100	90	92	92	100	94	96	87
92	93	95	94	87	95	90	89	100	100
94	84	88	96	94	93	94	97	96	79
85	94	89	88	95	96	96	89	94	100
100	98	100	95	92	93	97	82	85	83

a. Use una gráfica adecuada para describir los datos.
b. Describa la forma de la distribución de las calificaciones.

9. El observatorio Laboral para la Educación desde el año 2005 hace en Colombia un seguimiento de los profesionales graduados en las instituciones de educación superior del país. En particular, desarrolla una encuesta acerca de la situación laboral de los profesionales en diferentes departamentos del país e intenta estimar los salarios promedio por carreras. A mediados de 2010, el promedio salarial, según el nivel de estudios adquirido era el siguiente:

Carrera	Con pregrado	Con posgrado
Ingeniería Mecánica	1’715.448	3’684.858
Ingeniería Eléctrica	1’617.200	3’658.732
Economía	1’512.443	3’266.283
Medicina	2’087.414	3’020.385
Ingeniería de Sistemas	1’327.901	3’020.385
Ingeniería de Minas	2’729.324	2’960.705
Administración	1’370.824	2’917.220
Física	1’363.108	2’754.601
Derecho	1’508.722	2’748.672
Ingeniería Electrónica	1’550.308	2’712.256
Artes Representativas	895.286	2’673.090
Ingeniería Industrial	1’621.338	2’666.941
Comunicación Social	1’391.117	2’597.326
Antropología	1’339.619	2’397.698

Carrera	Con pregrado	Con posgrado
Biología	1'149.402	2'213.130
Arquitectura	1'330.175	2'167.563
Ingeniería Agroindustrial	1'205.460	2'061.817
Sicología	1'232.454	1'869.674
Artes Plásticas y Visuales	1'020.648	1'759.221
Educación	935.055	1'629.697
Diseño	1'097.046	1'560.887
Música	1'358.499	1'427.333
Deportes	1'055.316	1'244.101

a. ¿En cuáles carreras resulta más rentable estudiar un posgrado?

b. ¿Qué tipo de gráfica utilizaría para describir los datos? Constrúyala.

c. Resuma la información de la gráfica en relación con las carreras, los salarios y el nivel de estudios.

10. Se sabe que se ha arrojado material tóxico a una ciénaga. Los investigadores midieron la cantidad de material tóxico (en partes por millón) en peces capturados en tres lugares diferentes, con los siguientes resultados:

Lugar 1	31	26	20	20	29	28	21	26	25
Lugar 2	10	15	10	12	11	15	13	15	
Lugar 3	22	26	24	26	27	26	24		

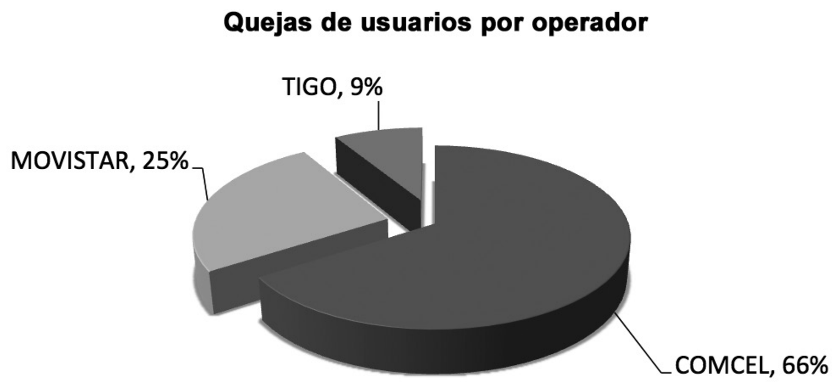
a. Use una gráfica apropiada para describir los datos.

b. ¿Qué sugieren los datos?

c. ¿Entre cuáles lugares existe mayor diferencia en la medición del material tóxico?

11. Quejas de los usuarios de las empresas de telefonía móvil en Colombia. En el año 2009 el número global de quejas según operador fue el siguiente:

Empresa	Quejas
Comcel	15.008
Movistar	5.825
Tigo	1.997
Usuarios: Comcel (28'495.932), Movistar (9'537.552),Tigo (4'429.875) Fuente de los datos: *Portafolio*, 27 de octubre de 2010	



a. ¿Considera adecuado el anterior gráfico de sectores para describir el número de quejas por empresa?
b. Construya un gráfico de barras para los mismos datos.
c. ¿De qué manera influye el número de usuarios por operador en la construcción de la gráfica?

12. Los nombres y edades de elección y muerte de los últimos veinte papas se muestran en la siguiente tabla:

Nombre	Inicio	Fin	Edad al inicio	Edad de fallecimiento	Años como Papa
Benedicto XVI	19 de abril de 2005	Presente	78		7
Juan Pablo II	16 de octubre de 1978	2 de abril de 2005	58	84	26
Juan Pablo I	26 de agosto de 1978	28 de septiembre de 1978	65	65	0 (33 días)
Pablo VI	21 de junio de 1963	6 de agosto de 1978	65	80	15
Juan XXIII	28 de octubre de 1958	3 de junio de 1963	76	81	4
Pío XII	2 de marzo de 1939	9 de octubre de 1958	63	82	19
Pío XI	6 de febrero de 1922	10 de febrero de 1939	64	81	17
Benedicto XV	3 de septiembre de 1914	22 de enero de 1922	59	67	7
Pío X	4 de agosto de 1903	20 de agosto de 1914	68	79	11
León XIII	20 de febrero de 1878	20 de julio de 1903	67	93	25
Pío IX	16 de junio de 1846	7 de febrero de 1878	54	85	31
Gregorio XVI	2 de febrero de 1831	1 de junio de 1846	65	80	15
Pío VIII	31 de marzo de 1829	1 de diciembre de 1830	67	69	1
León XII	28 de septiembre de 1823	10 de febrero de 1829	63	68	5
Pío VII	14 de marzo de 1800	20 de agosto de 1823	59	83	23
Pío VI	15 de febrero de 1775	29 de agosto de 1799	57	81	24
Clemente XIV	19 de mayo de 1769	22 de septiembre de 1774	63	68	5
Clemente XIII	6 de julio de 1758	2 de febrero de 1769	65	75	10
Benedicto XIV	17 de agosto de 1740	3 de mayo de 1758	65	83	17
Inocencio VII	17 de octubre de 1404	6 de noviembre de 1406	67	69	2

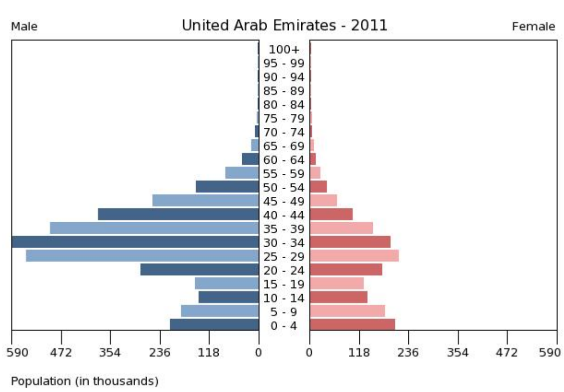
a. Construya diagramas de tallos y hojas para las variables contenidas en la anterior tabla.
b. ¿Existen valores inusuales en alguna de ellas?

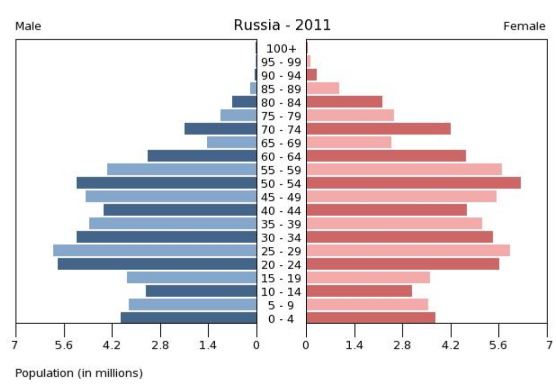
13. En un servicio de atención al cliente de una empresa de telefonía celular, las llamadas son clasificadas según el tipo de servicio que requiere el usuario. Setenta y dos llamadas clasificadas como “promociones” fueron monitoreadas para evaluar el desempeño de los agentes de servicio y, entre otras cosas, se registró la duración (en segundos) de estas, como se muestra aquí:

50	52	112	113	114	102	103	103	103	104
55	56	57	34	124	106	106	107	107	84
63	33	33	72	145	68	69	69	79	89
65	65	68	76	91	73	74	74	83	93
71	72	90	81	98	77	78	79	86	99
76	76	97	85	91	81	81	83	92	98
80	80	90	90	99	85	86	86	91	97
85	85								

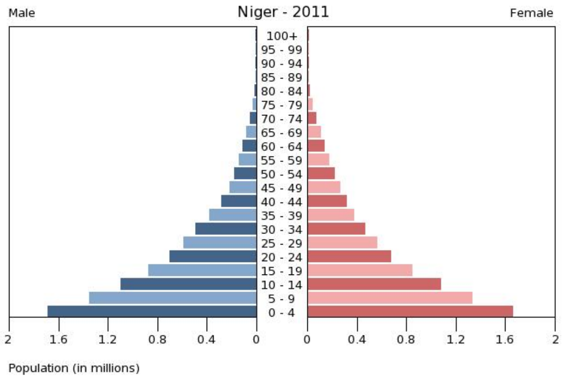
a. Construya un diagrama de tallo y hojas para los datos.
b. Determine el valor máximo y mínimo de los tiempos de duración de las llamadas.
c. Determine la proporción de llamadas de más de un minuto.
d. Divida los datos en llamadas cortas, normales y largas, de manera que cada clase tenga el mismo número de llamadas. ¿Cuáles son los límites de cada clase?
e. ¿Encuentra alguno de los datos fuera de lo común?

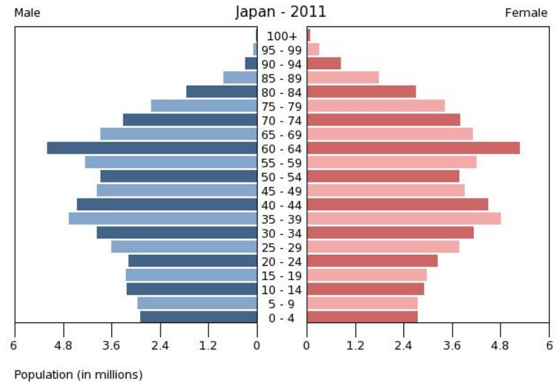
14. Compare las pirámides de población de Emiratos Árabes y Rusia 2011.



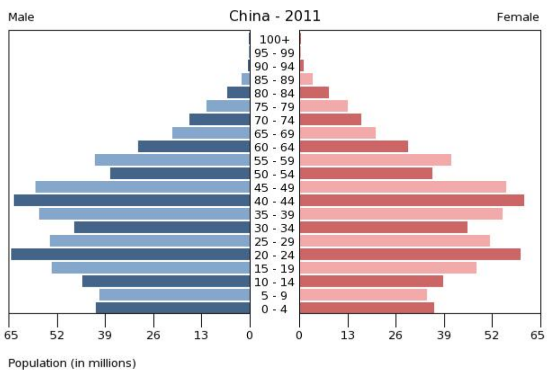


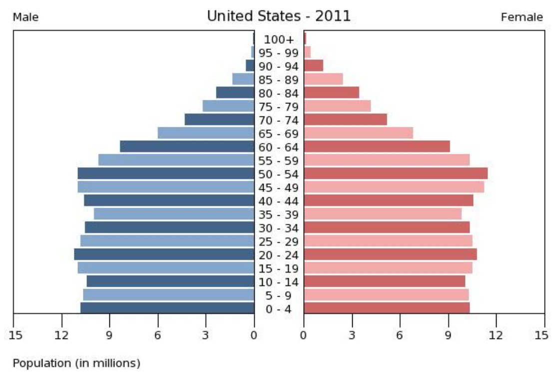
15. Compare las pirámides de población de Nigeria y Japón 2011.





16. Compare las pirámides de población de China y Estados Unidos 2011.





17. Averigüe las esperanzas de vida de los países del mundo, construya un histograma e interprételo.

18. Lea y discuta la lectura "Buenas prácticas estadísticas" que se encuentra en el apéndice del libro.

CAPÍTULO 2
ESTADÍSTICOS DE CENTRO Y VARIABILIDAD

Estadísticos de centro

La media, la mediana y la moda son estadísticos que tratan de ubicar la posición central de un conjunto de valores. Cada uno tiene una utilidad particular que depende de las circunstancias. A continuación, se presenta una breve descripción de cada uno de ellos acompañada de varios ejemplos. Se hace énfasis en la interpretación y en la utilidad de los estadísticos de centro y dispersión más que en la manera de calcularlos.

La media o promedio $\overline{x}$

Sea $x_1, x_2, x_3, \ldots, x_n$, una muestra de tamaño n de una variable X . La **media** *o* **promedio aritmético** $\overline{x}$ de la muestra es un estadístico que intenta representar el punto central de los datos y se calcula como sigue:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

La media es muy utilizada, porque resume con sencillez los datos y tiene fácil interpretación. Este estadístico es utilizado inclusive en variables de tipo cualitativo que sean binarias (codificadas con cero y uno). En este caso representa el porcentaje de observaciones en la muestra que tienen la característica codificada como 1.

Sobre un conjunto de *n* datos $x_1, x_2, x_3, \ldots, x_n$, la media o promedio aritmético es:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Ejemplo 2.1

Enunciado

El precio del galón de gasolina en Bogotá varía según la fecha, la estación de servicio y la localización de esta. El día 7 de junio de 2011 se registró el precio del galón de gasolina corriente en cinco estaciones de servicio (tabla 2.1).

Tabla 2.1 Precio del galón de gasolina

Estación	Precio por galón
Texaco	7.995
Brio	8.310
Esso	8.180
Terpel	7.959
Mobil	8.099
Total	**40.543**

Calcular el promedio del precio de venta del galón de gasolina en estas cinco estaciones.

Solución

Si x_i representa el precio del galón de gasolina en la *i-ésima* estación, el total de la tabla se puede escribir como:

$$\sum_{i=1}^{5} x_i = 40.543$$

Este sería el precio de cinco galones de gasolina comprados en cada una de las cinco estaciones. Para calcular el promedio solo basta dividir por el número de estaciones:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{40.543}{5} = 8.108,6$$

Esto significa que el promedio del precio de venta de un galón de gasolina en estas cinco estaciones de servicio es de $8.108,6.▪

Ejemplo 2.2

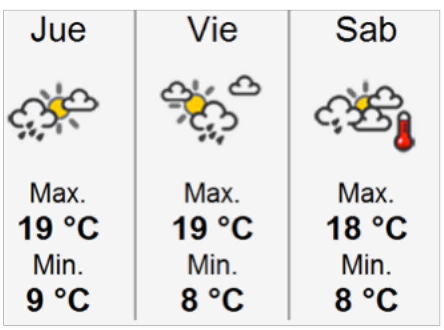
La media o promedio se puede presentar con ciertos matices en diferentes contextos, por ejemplo:

1. El producto interno bruto (PIB) es la suma de todos los bienes y servicios finales producidos por un país en un periodo de un año. Si esta suma es dividida por el número de habitantes del país, se obtiene una estadística que representa el PIB por persona. A esta estadística se le denomina PIB per cápita**.** Per cápita significa "por cabeza", lo que alude al ingreso promedio, es decir, el PIB per cápita representa los bienes o servicios producidos en promedio por una persona de la población.

2. La densidad de una población se define como el número de individuos, dividido entre la superficie que ocupan. Si se habla de la densidad de un país, donde la superficie se mide en kilómetros cuadrados, la densidad se refiere al número de habitantes promedio por kilometro cuadrado. Por ejemplo, Colombia tiene un territorio continental de 1'141.748 km^2 y una población estimada de 46 millones de habitantes (2011), lo que conduce a una densidad aproximada de cuarenta habitantes por kilometro cuadrado.

3. La nota final de un estudiante depende de las calificaciones en una evaluación escrita C_1, un trabajo C_2 y una exposición C_3. La nota final no necesariamente es el promedio aritmético $\frac{C_1 + C_2 + C_3}{3}$ porque se le puede dar mayor peso a una de las calificaciones. Por ejemplo, si se desea que la evaluación escrita tenga un peso del 50%, el trabajo 30% y la exposición 20%, se calcula una **media ponderada** $0,5C_1 + 0,3C_2 + 0,2C_3 = \frac{50C_1 + 30C_2 + 20C_3}{100}$. La media ponderada es una medida versátil y una condición para su cálculo es que la suma de los pesos sea igual a uno. Observe que si las ponderaciones son idénticas, se obtiene la media aritmética.▪

Estadísticas de orden



En ocasiones, de todos los valores que puede tomar una variable solo son de interés unos de ellos. Por ejemplo, en el pronóstico de la temperatura, se suelen dar solamente los valores **máximo** y **mínimo** o en las acciones negociadas en la bolsa de valores se destacan las cotizaciones más y menos valoradas.

Dada una muestra tamaño n, $x_1, x_2, x_3, \ldots, x_n$, de una variable X, las **estadísticas de orden** son los mismos valores de la muestra pero ordenados de menor a mayor. El valor mínimo de la muestra se escribe $x_{(1)}$ y el máximo $x_{(n)}$. Las estadísticas de orden satisfacen que:

$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \ldots \leq x_{(n)}$$

A $x_{(1)}, x_{(2)}, x_{(3)}, \ldots, x_{(n)}$, se le llama muestra ordenada. Si dos o más elementos tienen el mismo valor, su orden de ubicación no es importante.

Ejemplo 2.3

Enunciado

Explicar cómo se evalúa la configuración del hardware y software de un computador con sistema operativo Windows.

Solución

La evaluación de la configuración del hardware y software de un computador se expresa en forma de un número denominado puntuación base. Esta puntuación puede ser consultada en la opción "sistema" dentro del panel de control. Una puntuación alta significa que el equipo funciona mejor y más rápido que otro equipo con una puntuación básica inferior. Cada componente de hardware del computador recibe una subpuntuación individual. Los componentes evaluados son: procesador, memoria, gráficos, gráficos de juego y disco duro. La puntuación base del equipo viene determinada por la subpuntuación **mínima**. Se puede usar la puntuación base para instalar un programa que coincida con la puntuación base del equipo. Por ejemplo, si el equipo tiene una puntuación básica de 4,1, como en la figura 2.1, se puede instalar en él cualquier software diseñado que requiera un equipo con una puntuación básica de 4,1 o inferior.▪

Figura 2.1. Evaluación de la configuración de hardware y software de un computador

Componente	Qué se evalúa	Puntuación	Puntuación total
Procesador:	Cálculos por segundo	6.9	4.1 Determinado por la puntuación más baja
Memoria (RAM):	Operaciones de memoria por segundo	7.2	
Gráficos:	Rendimiento del escritorio de Windows Aero	4.1	
Gráficos de juego:	Rendimiento de gráficos en 3D para negocios y juegos	5.5	
Disco duro principal:	Velocidad de transferencia de datos en el disco	5.9	

Ejemplo 2.4

Enunciado

La siguiente tabla muestra a los diez países con más dominios en internet (enero de 2011).

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
14'900.0000	7'920.000	12'300.000	80'100.000	17'400.000	31'900.000	8'170.000	15'000.000	12'200.000	23'400.000
Rusia	Brasil	Polonia	Alemania	Holanda	Reino Unido	Francia	China	Italia	Japón

Encuentre las estadísticas de orden.

Solución

Las estadísticas de orden muestran el número de dominios en internet de menor a mayor. Observe que de los diez países, el que menos dominios tiene es Brasil y el que más tiene es Alemania.▪

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
7'920.0000	8'170.000	12'200.000	12'300.000	14'900.000	15'000.000	17'400.000	23'400.000	31'900.000	23'400.000
Brasil	Francia	Italia	Polonia	Rusia	China	Holanda	Japón	R. Unido	Alemania

La mediana *m*

Sea $x_1, x_2, x_3, \ldots, x_n$, una muestra de tamaño n de una variable X. La **mediana** m de la muestra es un valor que se ubica en la posición central de las estadísticas de orden. Esta medida es especialmente útil para describir el centro de la distribución de un conjunto de datos cuando hay presencia de observaciones muy grandes o muy pequeñas que modifican considerablemente el promedio.

Para hallar la mediana m de una muestra se pueden seguir los siguientes pasos:

1. Obtener las estadísticas de orden $x_{(1)}, x_{(2)}, x_{(3)}, \ldots, x_{(n)}$.

Si n es impar, la mediana es $m = x_{\left(\frac{n+1}{2}\right)}$, si n es par la mediana es $m = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$

Sobre un conjunto de n datos ordenados $x_{(1)}, x_{(2)}, x_{(3)}, \ldots, x_{(n)}$ **la mediana es:**

$x_{\left(\frac{n+1}{2}\right)}$ **si *n* es impar**

$\frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$ **si *n* es par**

Ejemplo 2.5

Enunciado

Una empresa reporta que los salarios de sus diez empleados son los siguientes:

800.000	800.000	7'000.000	650.000	700.000	700.000	800.000	800.000	650.000	720.000

Calcular el salario promedio y el salario mediano de dicha empresa.

Solución

Si x_i representa el salario del *i-ésimo* empleado, la nómina de la empresa es:

$$\sum_{i=1}^{10} x_i = 13'620.000$$

Entonces el salario promedio es:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{13'620.000}{10} = \$1'362.000$$

Observe que un salario es comparativamente más grande que los demás ($7’000.000). El promedio ($1.362.000) no representa fielmente el centro de este conjunto de datos, porque casi todos ellos son menores que él. El salario mayor podría ser el del dueño de la empresa o su gerente y al mezclarlo con los otros distorsiona el centro de los datos. En este caso, es aconsejable calcular la mediana para determinar un punto central más razonable.

Para calcular la mediana se siguen los siguientes pasos:

1. Las estadísticas de orden de los datos son:

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$	$x_{(10)}$
650.000	650.000	700.000	700.000	720.000	800.000	800.000	800.000	800.000	7’000.000

2. Como $n = 10$ es par y $\frac{n}{2} = \frac{10}{2} = 5$ y $\frac{n}{2} + 1 = \frac{10}{2} + 1 = 6$. La mediana es el promedio de los datos en las posiciones 5 y 6, es decir,

$$\frac{x_{(5)} + x_{(6)}}{2} = \frac{720.000 + 800.000}{2} = \$760.000$$

Observe que este dato refleja mejor el centro de los salarios de los trabajadores de la empresa y se mantiene constante aun ante la presencia de valores extremos (inusualmente grandes o pequeños) Por esta razón, se dice que la mediana es un estadístico resistente a la presencia de datos extremos mientras que el promedio es muy sensible a ellos.▪

La moda

La moda de un conjunto de datos es aquel que ocurre con más frecuencia. Si no se dispone de la totalidad de los datos, sino solamente de una tabla de frecuencias como la que se construye cuando se elabora un histograma, la moda se asigna al valor central de la clase de mayor frecuencia. Es posible que, en un conjunto de datos, exista más de una moda. Si hay dos modas se dice que el conjunto de datos es bimodal, si hay tres o más se dice multimodal.

Sobre un conjunto de x_1, x_2, x_3, ..., x_n, la moda es el valor que ocurre con mayor frecuencia.

Ejemplo 2.6

Enunciado

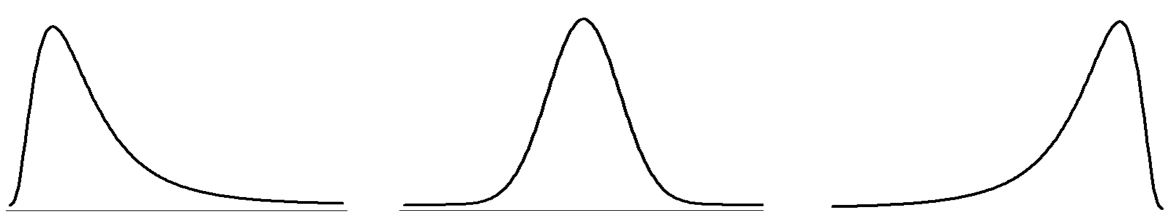
Calcular la moda de los salarios del ejemplo 2.5.

Solución

La suma de \$800.000 es un salario que se repite cuatro veces. Ningún otro se repite tantas veces como él. Por tanto, la moda es \$800.000. En este conjunto de datos, este estadístico refleja mejor el centro de los datos que el promedio aritmético.

Las posiciones relativas de la media y la mediana contienen información acerca de la forma de la distribución de la variable que están resumiendo. Por ejemplo, si la distribución es sesgada a la derecha es porque existe una pequeña porción de valores mucho más grandes que la mayoría de datos. En este caso, la media es sensible a estos valores y es más grande que la mediana. Si la distribución es sesgada a la izquierda, es porque existe una pequeña proporción de valores mucho más pequeños que la mayoría de los datos. En este caso, la media es más pequeña que la mediana. Si las dos mediciones son aproximadamente iguales, la distribución puede ser simétrica. Esto se resume en la figura 2.2.▪

Figura 2.2 Posiciones relativas de la media y la mediana



Sesgada a la derecha $m < \bar{x}$ | Distribución simétrica $m \approx \bar{x}$ | Sesgada a la izquierda $\bar{x} < m$

Ejercicio 2.1 Estadísticos de centro

1. El costo por minuto de una llamada desde un teléfono celular varía dependiendo del operador, del tipo de plan y del destino (fijo nacional, celular nacional, fijo internacional, celular internacional). Se seleccionó a un grupo de personas con planes pospago y se les preguntó por el valor aproximado de una llamada a otro operador desde su plan con los siguientes resultados:

Movistar	**Comcel**	**Tigo**
143	178	131
124	114	96
154	182	111
175	54	115
173	193	141
184	213	115
106	126	115
145	168	105
175	181	115
200	156	100

a. Calcule la media, la mediana y la moda de cada uno de los tres grupos.
b. ¿Qué medida describe mejor el centro de los datos en cada grupo?
c. Calcule la media, la mediana y la moda de todos los datos.

2. El dueño de una panadería está interesado en observar las diferencias en el peso de los panes que debería ser de 200 g. Durante 5 días, toma muestras de pan elaborado por dos panaderos y dos máquinas. Los resultados se muestran en la siguiente tabla:

	Máquina 1		**Máquina 2**	
Día	**Panadero 1**	**Panadero 2**	**Panadero 1**	**Panadero 2**
1	196,3	215,2	197,8	216,7
2	198,5	213,3	199,8	214,8
3	200,6	211,2	203,7	213,6
4	204,3	209,9	205,4	211,4
5	206,7	208,4	206,7	210,6

Estudie las diferencias de los pesos de los panes dadas por las variables analizadas por el dueño de la panadería. Construya gráficas adecuadas para resumir sus conclusiones.

3. ¿Cuántos televisores hay en su casa? Esta fue la pregunta que se le hizo a un grupo de universitarios de estratos 3 y 4 con los siguientes resultados:

2	4	3	2	4	2	3	3	2	5
2	4	2	6	1	3	8	2	4	1
3	3	6	2	2	3	3	1	1	5
2	5	4	3	3	3	3	2	5	4

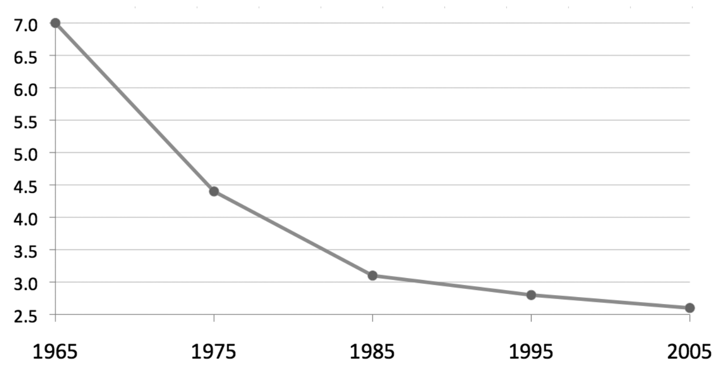
a. ¿La distribución del número de televisores es simétrica o sesgada?
b. Construya un histograma de frecuencias del conjunto de datos.
c. Localice en el anterior gráfico la media y la mediana de las mediciones.
d. ¿La localización de la media y la mediana concuerdan con su respuesta en el inciso a?

4. Se cuenta con los registros médicos de pacientes con cáncer de pulmón. Los tiempos de sobrevivencia en meses desde el momento de detección de la enfermedad de 35 pacientes fallecidos por esta causa fueron:

10	1	3	7	9	17	9
1	10	9	5	15	34	5
27	17	5	4	3	15	13
3	1	10	6	13	15	3
1	19	1	3	0	2	76

a. Calcule el tiempo de supervivencia promedio y mediano.
b. ¿Qué forma tiene la distribución de estos datos?
c. Entre el promedio y la mediana, ¿cuál describe mejor el centro de los datos?

5. Según Profamilia, las mujeres colombianas desean cada vez tener menos hijos. Las colombianas consideran que el número ideal de hijos debe ser 2,3 en promedio, cifra que aún no se ajusta a la tasa de fecundidad actual del país que es 2,6 hijos por mujer. Las mujeres actualmente casadas o unidas dijeron preferir 2,5, mientras que las que no tienen ningún hijo afirmaron que lo ideal sería 2.



¿Qué significa 2,3 hijos en promedio? ¿Cómo se obtiene este dato?

6. El precio del galón de gasolina en Bogotá varía según la fecha, la estación de servicio y la localización de esta. El día 30 de julio del 2010 se registró el precio del galón de gasolina corriente en catorce estaciones de servicio en la ciudad.

Estación	Precio por galón
Texaco calle 62 con 24	7.890
Mobil Calle 170	7.865
Mobil Tunjuelito	7.799
Esso Fontibón	7.800
Terpel Boyacá con 77	7.718
Terpel Ciudad de Cali	7.749
Texaco Teusaquillo	7.999
Texaco Calle 13 con 32	7.607
Brio Calle 70 con 66	7.740
Mobil Los Lagartos	7.930
Mobil Salitre	7.599
Esso Calle 20 con 3	7.850
Texaco Calle 80 con 26	7.930
Terpel Las Villas	7.538

Fuente: http://www.ahorrecomparando.com/precio_de%20gasolina_bogota.php

a. Encuentre el precio promedio del galón de gasolina en las catorce estaciones.
b. Encuentre el precio mediano del galón de gasolina en las catorce estaciones.
c. Con base en los resultados anteriores, ¿cree que la distribución de los precios de la gasolina es sesgada?

7. Los salarios mensuales en el año 2009 de Lionel Messi, Andrés Iniesta y Zlatan Ibrahimović, jugadores de la liga española, fueron: €1’272.000, €416.000 y €1’017.000 respectivamente. ¿Usted cree que la distribución de los salarios de los futbolistas de la liga española es simétrica o sesgada? ¿Qué medida utilizaría para describir su centro?

8. Calcule y compare las densidades de población de los países de América.

9. Calcule y compare los ingresos per cápita de los países de América.

10. Si sus calificaciones son $C_1 = 5{,}0$ en una evaluación escrita, $C_2 = 4{,}0$ en un trabajo, $C_3 = 2{,}9$ en una exposición y los pesos de las calificaciones sobre la nota final son 50%, 10% y 40%, respectivamente. ¿Qué le conviene más, la media ponderada o la media aritmética?

11. En el ejercicio 1.2.3 calificaciones en el examen final de microeconomía.
    a. Calcule el promedio y la mediana de las calificaciones en cada una de las sedes.
    b. Compare sus resultados con la respuesta dada en el literal e. del ejercicio 1.2.3.

12. El precio de los vehículos nuevos en el año 2010 con cilindraje 1600, 1800 y 2000 de venta en Colombia se muestra a continuación:

Marca	Línea	Cilindraje	Precio (miles)
Audi	A3	1600	56.000
Audi	A3	2000	78.000
Audi	A5	2000	111.000
Audi	TT TFSI 200 HP	2000	103.000
BMW	116 I	1600	59.000
BMW	120 I	2000	76.000
BMW	320 I	2000	81.000
Chevrolet	Optra	1600	36.000
Chevrolet	Optra	1800	38.000
Fiat	Siena	1800	33.000
Ford	Ecosport	2000	52.000
Hyundai	Getz	1600	26.000
Kia	Rio	1600	32.500
Mazda	5	2000	55.000
Mercedes Benz	C 200 K	2000	90.000

Marca	Línea	Cilindraje	Precio (miles)
Mercedes Benz	E 200 K	1800	105.000
Nissan	Almera	1600	36.000
Nissan	Sentra	2000	43.000
Nissan	Tiida	1800	40.000
Peugeot	308	1600	62.000
Peugeot	407	2000	54.000
Renault	Clio	1600	29.000
Renault	Scenic	1600	45.000
Seat	Leon	1600	52.000
Ssangyong	Actyon	2000	56.000
Ssangyong	Kyron	2000	63.000
Subaru	Impreza	2000	60.000
Volkswagen	Bora Turbo	2000	71.000
Volkswagen	Jetta	1800	47.000
Volkswagen	Jetta	2000	39.000

a. ¿Cuál es el precio promedio y mediano de estos vehículos según su cilindraje?

b. ¿Es importante agrupar los vehículos según alguna característica para dar una estadística con un mayor nivel de detalle? Si es así hágalo.

c. Clasifique según su criterio cada uno de los vehículos en una nueva categoría llamada "gama" y dé para ella las estadísticas del precio.

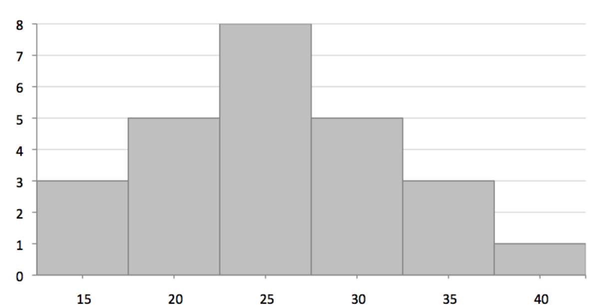
13. Con respecto a la duración de las llamadas al servicio de atención al cliente del ejercicio 1.2.13 calcule:

a. El promedio, la mediana y la moda.

b. Compare la mediana y el promedio. ¿Qué puede concluir?

c. Construya una gráfica de puntos para los datos. ¿Esto confirma su apreciación anterior?

14. El siguiente histograma muestra la distribución de edades de un conjunto de personas:



a. ¿Con cuántas personas se construyó el histograma? ¿Cuál es la edad promedio?

b. Aproximadamente ¿qué porcentaje de individuos tiene una edad entre 18 y 37 años?

15. En un ascensor hay diez personas, seis mujeres y cuatro hombres. El peso promedio de los hombres es 70 kg y el de las mujeres 60 kg. Determine el peso promedio de las personas en el ascensor.

16. Utilice los datos de la tabla 1.6. Estatura en Colombia 1910 a 1984 para:

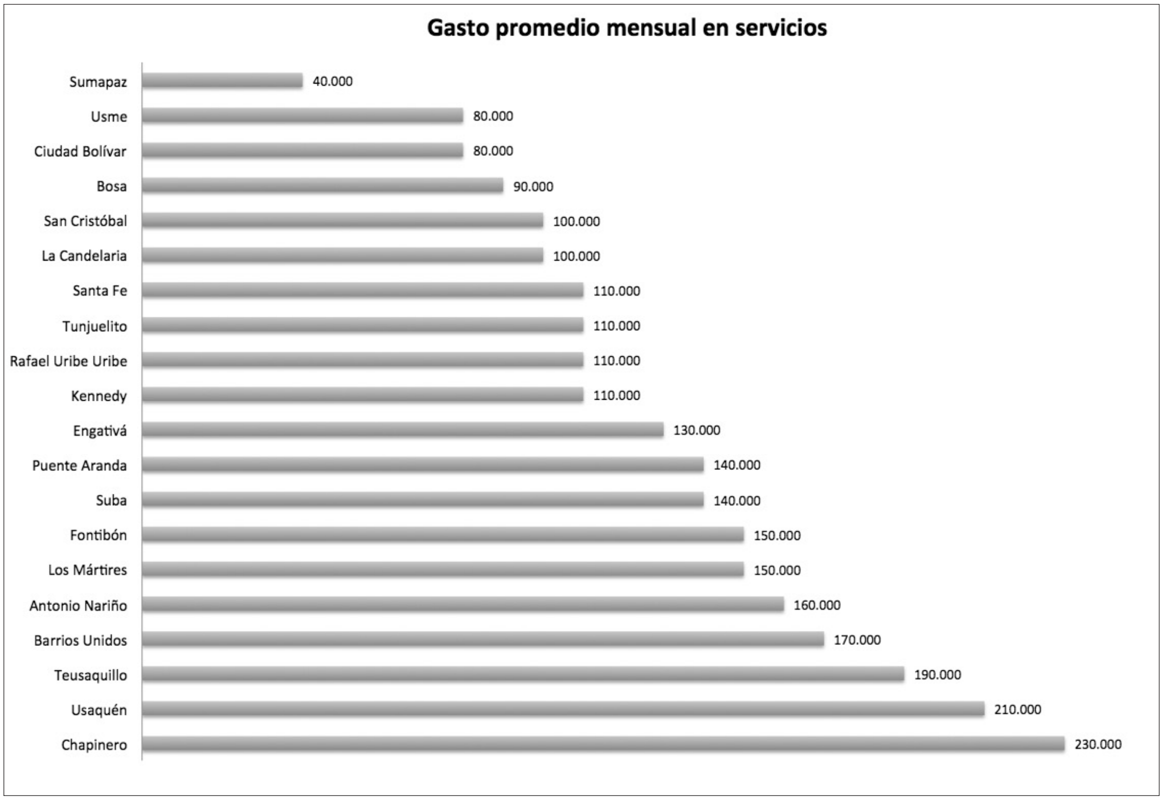
a. Explicar cómo se calcula el promedio total de la estatura de hombres y mujeres.

b. Encuentre el promedio de la estatura global de los 8’454.348 individuos de la tabla.

Tabla 1.6 Estatura en Colombia 1910 a 1984

Año de nacimiento	Estatura promedio hombres	Estatura promedio mujeres	Número de hombres	Número de mujeres	Total
1910-1914	163,48	150,78	1.751	2.197	3.948
1915-1919	163,61	151,49	4.582	4.993	9.575
1920-1924	164,16	152,38	9.086	9.779	18.865
1925-1929	164,7	153,06	15.659	16.299	31.958
1930-1934	165,17	153,48	22.219	24.619	46.838
1935-1939	165,76	154,21	34.637	33.820	68.457
1940-1944	166,26	154,69	40.186	45.623	85.809
1945-1949	167,1	155,59	53.164	60.723	113.887
1950-1954	167,84	156,4	73.835	80.863	154.698
1955-1959	168,07	156,81	101.613	111.278	212.891
1960-1964	168,47	157,17	142.054	157.950	300.004
1965-1969	169	157,34	177.464	221.024	398.488
1970-1974	168,91	157,21	298.908	426.660	725.568
1975-1979	169,66	157,81	1’639.499	1’646.878	3’286.377
1980-1984	170,64	158,65	1’528.875	1’468.110	2’996.985
Total	169,664	157,778	4’143.532	4’310.816	8’454.348

17. La encuesta de calidad de vida de Bogotá 2007, desarrollada por el Dane publicó los siguientes resultados de los gastos mensuales relacionados con servicios públicos, por persona, en Bogotá, en veinte localidades.



Con respecto a los anteriores datos, calcule:

a. Media, mediana y moda.
b. Determine si esta variable es simétrica, sesgada a la derecha o a la izquierda. Justifique.

18. Calcule el promedio de la edad de elección de los papas del ejercicio 1.2.12.

19. Utilice en una hoja de cálculo las funciones promedio, mediana y moda para encontrar las medidas de tendencia central de las variables cuantitativas de la base de datos *Trabajadores*, que se encuentra en la plataforma virtual de este curso.

20. Calcule e interprete las medidas de tendencia central de los partidos jugados, goles a favor y goles en contra de la base de datos *Resultados del mundial*.

21. En un estudio médico se probó un nuevo medicamento y tratamiento para inhibir el crecimiento de tumores. De una población de 500 pacientes se tomó una muestra aleatoria de 120 pacientes, cincuenta hombres y setenta mujeres con patologías similares. A ellos se les aplicó un tratamiento y se les suministró medicamento durante un año.

Posteriormente, se midió la masa del tumor y se registró el crecimiento. Los datos se encuentran en la base de datos *Crecimiento*.

a. Construya un histograma para los crecimientos de los tumores de los 120 pacientes. ¿Tiene alguna forma particular este histograma? ¿Cuál cree que puede ser la razón para esta forma?
b. Calcule la media y la mediana del crecimiento del tumor en hombres y mujeres. ¿Encuentra diferencias que sean de resaltar?
c. Construya histogramas para los crecimientos de los tumores en cada uno de los géneros.
d. Determine si las siguientes afirmaciones son verdaderas o falsas:
   i. En promedio el tumor de estos 120 pacientes presentó un crecimiento de 8,5 g.
   ii. El grupo de 120 pacientes es muy homogéneo frente al crecimiento del tumor.
   iii. Existe diferencia entre pacientes hombres y mujeres en relación con el crecimiento del tumor.
   iv. En el grupo de las mujeres el crecimiento del tumor es mucho mayor que en el de los hombres.

Estadísticos de variabilidad

Imagine que se conforman tres grupos G_1, G_2 y G_3 de personas en una caminata. En cada grupo la edad promedio es 20 años. ¿Se puede afirmar que los grupos tienen edades similares?

No. El promedio tan solo ofrece información acerca del centro, pero no de las diferencias entre las edades. Por ejemplo, si las edades de cada grupo son:

G_1	G_2	G_3
20	21	38
20	19	20
20	20	25
20	22	15
20	18	2
$\overline{x}_1 = 20$	$\overline{x}_2 = 20$	$\overline{x}_3 = 20$

Se observa que aunque los promedios son exactamente iguales a 20 años, cada grupo tiene una configuración de edades diferente. En unos las edades difieren mucho —son muy variables— en otros son más parecidas —son poco variables—.

- En el primer grupo, todos tienen 20 años, se puede decir que no existe variabilidad en las edades de estas personas.
- En el tercer grupo las edades son muy diferentes.
- El segundo grupo tiene una configuración intermedia, es decir, sus edades varían entre sí, pero no tanto como las del grupo 3.

Para tener una idea de cómo se comporta un conjunto de datos, es decir, para caracterizarlo es importante contar con un estadístico adicional a los de centro. Estos son los estadísticos de variabilidad. A continuación, se mencionan tres de ellos: el rango, la varianza y la desviación estándar.

El rango

El rango de un conjunto de datos es la diferencia entre el valor máximo y el mínimo. Este estadístico da la primera y más rápida información acerca de lo variable que es un conjunto de datos.

$$Rango = Máximo - Mínimo$$

En el ejemplo inicial, el primer grupo tiene rango cero, mientras que el segundo grupo tiene un rango de 4 años, es decir, que la diferencia máxima de edades entre dos personas de este grupo es de cuatro años. En el tercer grupo, el rango es 36 años; como se ve, mucho mayor que en los dos grupos anteriores.

La varianza

La varianza de una población de tamaño *N* es el promedio de los cuadrados de las desviaciones con respecto a la media. Al ser un parámetro de la población se usa una letra griega para su designación:

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

En una muestra, el verdadero promedio de los cuadrados de las desviaciones con respecto al promedio es $\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}$. Sin embargo y por tener mejores propiedades estadísticas, se utiliza la cuasi varianza muestral $\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$. Con el propósito de hacer sencillo el lenguaje, se define simplemente como la varianza muestral:

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

La desviación estándar muestral

Para expresar la variabilidad en las mismas unidades de los datos, se calcula la desviación estándar muestral:

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

Ejemplo 2.7

Enunciado

Calcular la desviación estándar muestral de las edades del grupo 2 en el ejemplo inicial.

Solución

En la siguiente tabla, la primera columna tiene las edades de las n=5 personas del grupo. En la segunda columna, se calculan las desviaciones de la edad de cada persona con respecto a la media. Este valor da una idea de lo variables que son los datos, sin embargo, al calcular su promedio se obtiene cero. Esto se debe a que las desviaciones positivas y negativas se cancelan entre sí. Para evitar este inconveniente, en la tercera columna se calculan los cuadrados de las desviaciones con respecto a la media.

Edad x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
21	1	1
19	-1	1
20	0	0
22	2	4
18	-2	4
	$\sum(x_i - \overline{x})^2 =$	**10**

Por último, se suman estos cuadrados y se divide entre n – 1 para obtener la varianza muestral s^2:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1} = \frac{10}{4} = 2,5$$

Observe que si la edad está medida en años, la varianza muestral tiene como unidad años al cuadrado. La desviación estándar es la raíz cuadrada de este valor $s = \sqrt{2,5} \approx 1,58$ y tiene las mismas unidades que los datos, es decir, años. Esto significa que el promedio de los cuadrados de las desviaciones con respecto a la media es 1,58 años. El lector puede probar que la desviación estándar muestral del primer y tercer grupo son 0 y 13,2, lo que refleja los niveles de variabilidad en cada uno de ellos.▪

Ejemplo 2.8

Enunciado

Los pesos en kilogramos de dos grupos de personas, fumadores y no fumadores, entre los 20 y 25 años se registran a continuación:

	Peso kg									
No fumadores	71	75	78	74	71	70	79	81	75	70
Fumadores	80	77	72	83	86	87	67	56	81	51

Comparar las medias y las desviaciones estándar de los pesos de los dos grupos.

Solución

En promedio, los pesos de los dos grupos no son muy diferentes: 74 y 74,4 kg, respectivamente. Sin embargo, sus desviaciones sí lo son. En los no fumadores, la desviación estándar no es superior a 4 kg, mientras que en el grupo de los fumadores sobrepasa los 12 kg. Esto significa que los pesos de los fumadores considerados son tres veces más variables que los del grupo de no fumadores. Los datos se resumen a continuación:

	Promedio	**Desviación estándar**
No fumadores	74,4	3,95
Fumadores	74,0	12,45

▪

Rango

Rango = Máximo – Mínimo

Varianza muestral

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

Desviación estándar muestral

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

El coeficiente de variación (CV)

El coeficiente de variación (CV) es una estadística que mide la dispersión relativa y sirve para comparar la variabilidad de diferentes grupos. Se define como la desviación sobre el valor absoluto de la media —siempre que esta sea diferente de cero—: $CV = \frac{s}{\bar{x}} \times 100$. Como la desviación y la media tienen las mismas unidades que los datos, estas se cancelan y por tanto, el coeficiente está libre de unidades. Esta medida es especialmente útil para valorar la dispersión de dos conjuntos de datos cuando estos difieren mucho en el promedio.

Coeficiente de variación

$$CV = \frac{s}{\bar{x}} \times 100$$

Ejemplo 2.9

Enunciado

Compare la dispersión del salario de los dos grupos de la tabla 2.2.

Solución

La tabla contiene los salarios de dos grupos, uno con salarios bajos y otro con salarios muy altos. Los dos grupos difieren en el salario medio, sin embargo, las desviaciones estándar son idénticas. ¿Tienen los salarios variaciones iguales en los dos grupos? Si se observa con más detalle, se puede ver, por ejemplo, que la diferencia de los dos primeros salarios es la misma en los dos grupos ($100.000). Sin embargo porcentualmente, las diferencias son muy grandes 20% y 0,2%, respectivamente —en relación con el salario del primer trabajador de cada grupo—; esto sugiere que los salarios difieren más entre sí, en el primer grupo que en el segundo. El hecho de que la desviación estándar sea igual es porque los salarios del segundo grupo difieren en una constante de los del primero. Esto modifica la media, pero no la varianza. Para reflejar las diferencias porcentuales existentes, se puede calcular el coeficiente de variación en cada grupo: $CV_1 = \frac{452.769}{1.000.000} \times 100 = 45,3\%$ y $CV_2 = \frac{452.769}{51.000.000} \times 100 = 0,9\%$. Por tanto, existe una mayor variabilidad relativa en los salarios del primer grupo que en los del segundo.▪

Tabla 2.2 Coeficiente de variación

Grupo 1 Salarios bajos	Grupo 2 Salarios muy altos
500.000	50’500.000
600.000	50’600.000
1.500.000	51’500.000
1.400.000	51’400.000
1.000.000	51’000.000

Medias	
$\overline{x}_1$ = 1'000.000	$\overline{x}_2$ = 51'000.000
Desviaciones	
s_1 = 452.769	s_2 = 452.769
Coeficientes de variación	
CV_1 = 45,3%	CV_2 = 0,9%

Ejercicio 2.2 Medidas de variabilidad

1. A continuación, se muestra el precio de las acciones de Ecopetrol entre el 21 de julio y el 20 de agosto del año 2010.

Fecha	Precio
21 de julio del 2010	3.100
22 de julio del 2010	3.135
23 de julio del 2010	3.085
26 de julio del 2010	3.165
27 de julio del 2010	3.130
28 de julio del 2010	3.135
29 de julio del 2010	3.095
30 de julio del 2010	3.105

Fecha	Precio
02 de agosto del 2010	3.145
03 de agosto del 2010	3.245
04 de agosto del 2010	3.310
05 de agosto del 2010	3.450
06 de agosto del 2010	3.365
09 de agosto del 2010	3.350
10 de agosto del 2010	3.305
11 de agosto del 2010	3.250

Fecha	Precio
12 de agosto del 2010	3.280
13 de agosto del 2010	3.320
17 de agosto del 2010	3.340
18 de agosto del 2010	3.310
19 de agosto del 2010	3.310
20 de agosto del 2010	3.300

Con respecto a los precios de las acciones de Ecopetrol, calcule e interprete:

a. El rango.
b. La varianza muestral.
c. La desviación estándar muestral.
d. En el ámbito financiero, la desviación estándar de los precios de las acciones en un horizonte temporal suele llamarse volatilidad. ¿Qué interpretación le da a este nombre?

2. Se obtuvieron los datos de talla y peso de los veintisiete inscritos al gimnasio de la Universidad:

ID	Sexo	Estatura (cm)	Peso (kg)
1	Mujer	159	49
2	Hombre	164	62
3	Mujer	172	65
4	Mujer	167	52
5	Mujer	164	51
6	Mujer	161	67
7	Mujer	168	48
8	Hombre	181	74
9	Hombre	183	74
10	Mujer	158	50
11	Mujer	156	65
12	Hombre	173	64
13	Mujer	158	43
14	Hombre	178	74

ID	Sexo	Estatura (cm)	Peso (kg)
15	Hombre	181	76
16	Hombre	182	91
17	Hombre	176	73
18	Mujer	162	68
19	Mujer	156	52
20	Mujer	152	45
21	Hombre	181	80
22	Hombre	173	69
23	Mujer	155	53
24	Hombre	189	87
25	Mujer	170	70
26	Hombre	170	67
27	Mujer	168	56

a. Calcule el rango de los veintisiete datos de las estaturas.

b. ¿Existe alguna razón para que el anterior rango sea menor o mayor que el rango de las estaturas del grupo de las mujeres? Verifíquelo.

c. Calcule e interprete el rango de los pesos de las mujeres.

d. Calcule la varianza y la desviación muestral de los pesos de las mujeres.

e. Compare el rango con la desviación estándar de los pesos. ¿Aproximadamente, cuántas desviaciones estándar tiene el rango?

f. Compare e interprete el coeficiente de variación de los pesos de los hombres y las mujeres.

3. Una familia de estrato 3 en Cali pagó por concepto de agua y alcantarillado, los siguientes valores de enero a diciembre de 2009.

Enero	Febrero	Marzo	Abril	Mayo	Junio
102.132	98.935	98.388	91.710	87.898	81.056

Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
97.665	88.361	85.169	82.393	78.167	74.715

a. Calcule el rango de los valores pagados por concepto de agua y alcantarillado.

b. Calcule el promedio del valor pagado.

c. Calcule e interprete la desviación estándar muestral de los valores pagados.

4. En el ejercicio 1.2.3 se mostraron las calificaciones en el examen final de microeconomía de cien estudiantes, pertenecientes a dos sedes de una universidad, cada una con cincuenta estudiantes.

SEDE A	3,7	3,1	4,9	3,5	3,3	4,2	3,9	3,6	3,7	3,8
	4,2	4,5	3,0	3,8	3,8	4,0	3,4	4,4	3,3	4,6
	3,8	4,3	4,3	4,5	4,2	4,1	4,1	3,9	4,0	3,4
	3,9	4,1	3,8	4,4	3,9	4,3	3,8	3,2	3,2	3,9
	4,3	3,7	3,4	3,5	4,1	3,1	3,3	4,2	4,3	4,6

SEDE B	3,7	3,0	4,1	2,7	2,9	2,9	3,7	3,1	3,3	2,4
	2,9	3,0	3,2	3,1	2,4	3,2	2,7	2,6	4,0	3,7
	3,1	2,1	2,5	3,3	3,1	3,0	3,1	3,4	3,3	1,6
	2,2	3,1	2,6	2,5	3,2	3,3	3,3	2,6	3,1	3,7
	3,7	3,5	3,8	3,2	2,9	3,0	3,4	1,9	2,2	2,0

a. Calcule la media y la desviación estándar muestral de las calificaciones en cada una de las sedes.

b. Obtenga una conclusión al comparar las anteriores medias y desviaciones estándar.

5. Dos competidores se preparan para la prueba atlética de los 100 metros planos.[1] En las últimas quince carreras cada uno ha obtenido los siguientes tiempos (en segundos):

Atleta 1	12,01	10,95	10,31
	11,09	11,49	13,51
	10,79	12,22	12,03
	10,62	13,18	10,27
	12,67	10,54	10,82

Atleta 2	11,59	11,01	11,24
	10,91	11,35	11,18
	11,10	11,20	11,26
	11,00	11,29	11,55
	11,14	11,25	10,90

a. Calcule la media y la desviación estándar muestral de los tiempos de cada uno de los atletas.

b. Calcule, compare e interprete el coeficiente de variación de los tiempos de cada uno de los atletas. En promedio, ¿cuál de los dos atletas es mejor?

1 El 16 de agosto de 2009 UsainBolt batió el récord mundial de atletismo de los 100 metros planos con un tiempo de 9,578 segundos.

c. ¿Cree usted que se debe considerar otro estadístico para analizar los tiempos de los atletas? ¿Cuál? ¿Cuál de los dos atletas es mejor?

6. En el ejercicio 2.1.6 se mostraron los precios del galón de gasolina corriente en catorce estaciones de servicio en Bogotá

Estación	**Precio por galón**
Texaco calle 62 con 24	7.890
Mobil Calle 170	7.865
Mobil Tunjuelito	7.799
Esso Fontibón	7.800
Terpel Boyacá con 77	7.718
TerpelCiudad de Cali	7.749
Texaco Teusaquillo	7.999
Texaco Calle 13 con 32	7.607
Brio Calle 70 con 66	7.740
Mobil Los Lagartos	7.930
Mobil Salitre	7.599
Esso Calle 20 con 3	7.850
Texaco Calle 80 con 26	7.930
Terpel Las Villas	7.538

Para el precio del galón de gasolina en las catorce estaciones, calcule e interprete:

a. La media.
b. La varianza muestral.
c. La desviación muestral.
d. El rango.

7. En el ejercicio 2.1.1 se mostró el valor aproximado de una llamada a otro operador desde un celular en plan pospago de las empresas Movistar, Tigo y Comcel:

Movistar	**Comcel**	**Tigo**
143	178	131
124	114	96
154	182	111
175	54	115
173	193	141
184	213	115
106	126	115
145	168	105
175	181	115
200	156	100

a. Calcule la media y la desviación estándar muestral de los precios de las llamadas.

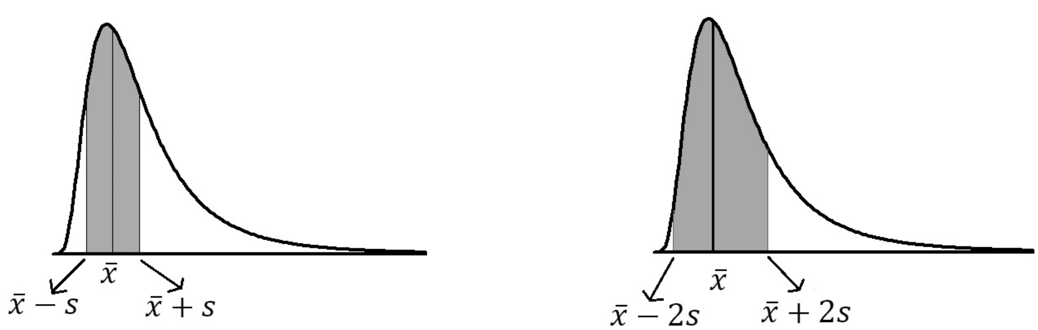
b. Calcule los coeficientes de variación en cada una de las empresas e interprételos.

8. El ingreso promedio de los operarios de una fábrica es de \$900.000, con una desviación estándar de 9000. En la misma fábrica, el personal administrativo tiene un salario promedio de \$3'000.000 y una desviación estándar de \$30.000. ¿En cuál grupo de trabajadores hay mayor dispersión relativa?
9. Con respecto al ejercicio 2.1.2, compare las desviaciones estándar en relación con las máquinas y los panaderos. ¿Qué se puede concluir de estos resultados?
10. Utilice las funciones PROMEDIO, VAR y DESVEST de Excel para encontrar la media, la varianza, la desviación estándar muestral y el coeficiente de variación de todas las variables cuantitativas de la base de datos *Trabajadores*. Determine en cuál de ellas existe mayor dispersión relativa.

La desigualdad de Chevyshev

En el capítulo 1, se hizo alusión a la distribución de una variable. En los ejemplos, se resaltó la manera como se distribuyen las frecuencias en el rango de un histograma. Conocer los porcentajes de datos que se encuentran en determinados intervalos del rango de la variable aumenta el conocimiento de la distribución. La desigualdad de Chevyshev es un resultado que permite dar porcentajes mínimos en intervalos en torno a la media. Por ejemplo, en la figura 2.3 se sombrea a la izquierda, el porcentaje de datos que se encuentran a una desviación estándar de la media. A la derecha se sombrea el porcentaje de datos que se encuentra a dos desviaciones estándar de la media.

Figura 2.3



Porcentaje de datos que se encuentra a una desviación estándar de la media o menos.

Porcentaje de datos que se encuentra a dos desviaciones estándar de la media o menos.

En general, se puede construir un intervalo que abarque k desviaciones estándar en torno a la media; esto es, el intervalo entre $\overline{x}-ks$ y $\overline{x}+ks$. Se está interesado en el porcentaje de datos que se encuentra en este intervalo. La desigualdad de Chevyshev establece los porcentajes mínimos en dichos intervalos.

Desigualdad de Chevyshev

Dado un conjunto de mediciones y un número $k \geq 1$, la fracción incluida en el intervalo $\overline{x}-ks$ a $\overline{x}+ks$ **es por lo menos** de $100\left(1-\frac{1}{k^2}\right)\%$.

Una justificación de esta desigualdad es la siguiente:

Considere todos los x_i, que se encuentran a más de k desviaciones estándar de la media, es decir, aquellos donde $|x_i-\overline{x}| > ks$. Es posible dividir cualquier conjunto de datos en dos grupos:

Grupo 1 (G1): los que están a más de k desviaciones estándar de la media.
Grupo 2 (G2): los que están a k o menos desviaciones estándar de la media.

De esta manera, la varianza se puede descomponer así:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 = \frac{1}{n-1}\sum_{x_i \in G_1}(x_i-\overline{x})^2 + \frac{1}{n-1}\sum_{x_i \in G1}(x_i-\overline{x})^2$$

Ahora, como las dos últimas cantidades son positivas

$$s^2 \geq \frac{1}{n-1}\sum_{x_i \in G_1}^{n}(x_i-\overline{x})^2$$

Donde cada x_i está a k o más desviaciones de la media, por tanto, $(x_i-x)^2 > (ks)^2$. Así que:

$$s^2 \geq \frac{1}{n-1}\sum_{x_i \in G_1}^{n}(x_i-\overline{x})^2 > k^2 s^2 f(|x_i-\overline{x}| > ks)$$

Donde $f(|x_i-\overline{x}| > ks)$ es la frecuencia de datos que se encuentra a más de k desviaciones estándar de la media.

En conclusión:

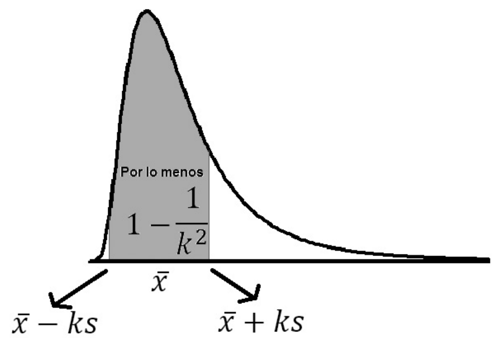
$$k^2 s^2 f(|x_i-\overline{x}| > ks) < s^2$$

$$f(|x_i-\overline{x}| > ks) < \frac{1}{k^2}$$

O lo que es lo mismo

$$f(|x_i-\overline{x}| \leq ks) \geq 1-\frac{1}{k^2}$$

Figura 2.4 Desigualdad de Chevyshev



En cualquier conjunto de n datos con media $\overline{x}$ y desviación estándar s, el porcentaje de datos contenido en el intervalo de $\overline{x}-ks$ a $\overline{x}+ks\,(k \geq 1)$ **es por lo menos** $100\left(1-\frac{1}{k^2}\right)\%$

Es importante anotar que la desigualdad es aplicable a cualquier conjunto de datos sin importar que su distribución sea simétrica o no. Esta desigualdad establece frecuencias mínimas para cada intervalo, como se muestra en la tabla 2.3.

Tabla 2.3 Porcentaje mínimo de datos entre $x-ks$ y $\overline{x}+ks$

k	$100\left(1-\frac{1}{k^2}\right)\%$
1	$100\left(1-\frac{1}{1^2}\right)\%=0$ (no es informativo)
2	$100\left(1-\frac{1}{2^2}\right)\%=75\%$
3	$100\left(1-\frac{1}{3^2}\right)\%\approx 88,9\%$
4	$100\left(1-\frac{1}{4^2}\right)\%\approx 93,8\%$
5	$100\left(1-\frac{1}{5^2}\right)\%\approx 96\%$

En la práctica, los anteriores valores se rebasan con facilidad. Si adicionalmente se supone que la distribución tiene forma de campana se logran unos porcentajes mayores en cada intervalo. A esto último, se le llama la regla para las distribuciones acampanadas o regla empírica:

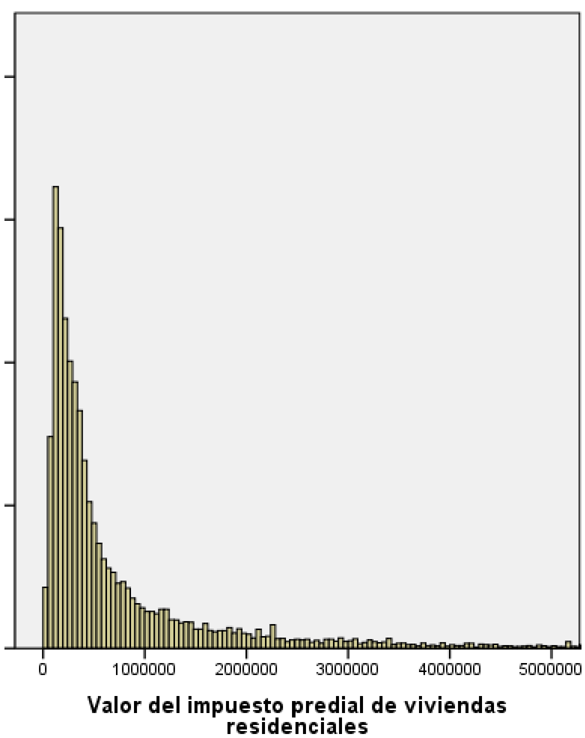
Regla para las distribuciones acampanadas o regla empírica

Si la distribución de frecuencias de un conjunto de datos tiene forma parecida a la de una campana, entonces los porcentajes mínimos dados por la regla de Chevyshev de los intervalos alrededor de la media se aumentan como se muestra enseguida:

Descripción	Intervalo	Porcentaje
A una desviación estándar o menos de la media se encontrarán aproximadamente 68,26% de los datos.	Entre $\overline{x}-s$ y $\overline{x}+s$	68,26% ≈ 68%*
A dos desviaciones estándar o menos de la media se encontrarán aproximadamente 95,44% de los datos.	Entre $\overline{x}-2s$ y $\overline{x}+2s$	95,44% ≈ 95%
A tres desviaciones estándar o menos de la media se encontrarán aproximadamente 99,74% de los datos.	Entre $\overline{x}-3s$ y $\overline{x}+3s$	99,74% ≈ 99%

*Cuando estudie la distribución de probabilidad normal, se dará una justificación para estos porcentajes.

Ejemplo 2.10



Enunciado

En Bogotá, los predios residenciales pagan, en promedio, por concepto de impuesto predial \$343.000 con una desviación estándar de \$95.000. Construya e interprete un intervalo que abarque tres desviaciones estándar en torno a la media e interprételo.

Solución

Como $\overline{x}=343.000$ y $s=95.000$, los límites de un intervalo que abarca tres desviaciones estándar en torno a la media son:

$$\overline{x}-3s=343.000-3\times 95.000=58.000$$

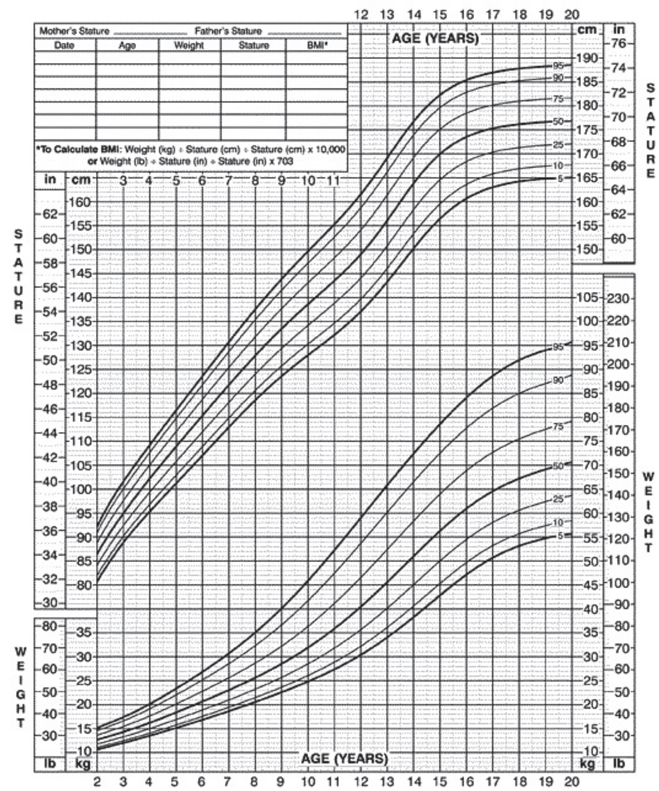
$$\overline{x}+3s=343.000+3\times 95.000=628.000$$

Esto significa que en Bogotá, por lo menos, el 88,9% de los predios residenciales pagan entre \$58.000 y \$628.000. Observe que este intervalo permite tener una idea del comportamiento de la distribución del valor del impuesto predial en este segmento.▪

Ejemplo 2.11

Enunciado

El peso de los niños varones de 5 años, en Villavicencio, es 20 kg, en promedio, con desviación estándar de 3 kg. Asuma que el peso de los niños de 5 años tiene una distribución parecida a una campana. Construya un intervalo que abarque dos desviaciones estándar en torno a la media e interprételo.



Solución

Como $\overline{x} = 20$ y $s = 3$, los límites de un intervalo que abarca dos desviaciones estándar en torno a la media son:

$$\overline{x} - 2s = 20 - 2 \times 3 = 14$$
$$\overline{x} + 2s = 20 + 2 \times 3 = 26$$

Esto significa que, en Villavicencio, aproximadamente el 95% de los niños varones de 5 años pesan entre 14 y 25 kg. Esta información permite tener una idea global del peso de los niños y determinar, por ejemplo, si en particular un niño tiene un peso dentro de los límites normales.▪

Ejercicio 2.3 Desigualdad de Chevyshev

1. La siguiente tabla corresponde a los consumos con tarjeta de crédito de una muestra de veinte individuos: diez hombres y diez mujeres.

Grupo	Consumo con tarjeta según género (valores en miles de pesos)									
Mujeres	400	200	600	150	280	120	350	160	400	400
Hombres	90	50	40	120	300	40	55	75	50	60

a. ¿Cuál de los dos grupos es más variable con respecto al consumo con tarjeta de crédito? Explique.

b. Suponga que la distribución de los consumos de las mujeres con tarjeta de crédito tiene aproximadamente forma de campana. Encuentre un intervalo que contenga el 95% de las mediciones de los consumos con tarjeta de crédito correspondiente a las mujeres.

2. Las tarifas por minuto de veinticinco parqueaderos públicos cubiertos en la localidad de Chapinero se muestran en la tabla.

#	Precio	#	Precio	#	Precio	#	Precio	#	Precio
1	65	6	55	11	60	16	70	21	67
2	70	7	50	12	60	17	65	22	55
3	60	8	50	13	50	18	55	23	55
4	55	9	60	14	60	19	60	24	45
5	63	10	60	15	70	20	60	25	80

a. Estime la desviación estándar muestral de los datos basada en el rango.

b. Calcule la desviación estándar y compárela con el resultado del inciso a.

c. ¿Qué porcentaje de tarifas se encuentran a una o menos desviaciones estándar de la media? ¿Qué porcentaje a dos o menos? ¿Qué porcentaje a tres o menos?

d. Compare los resultados anteriores con los de la regla de las distribuciones acampanadas.

3. En cierta empresa, el salario promedio de los trabajadores profesionales es de \$1'409.309 con una desviación estándar de \$296.131. Asuma que la distribución de estos salarios tiene forma de campana y determine el porcentaje de salarios que en dicha empresa se encontrarían en los siguientes intervalos:

Entre \$1'113.178 y \$1'113.178.	Más de \$1'705.440.
Entre \$817.047 y \$2'001.571.	Menos de \$817.047.
Entre \$520.916 y \$2'297.702.	Menos de \$520.916 o más de \$1'705.440.

4. Como estrategia de mercado, una empresa de medicina prepagada enviará por correo un obsequio, el día de su cumpleaños, a sus afiliados. Se eligió una muestra de sesenta afiliados y se anotó el mes del cumpleaños de cada uno.

agosto	agosto	febrero	julio	julio	febrero	mayo	julio	octubre	noviembre
agosto	noviembre	enero	octubre	octubre	marzo	mayo	agosto	noviembre	septiembre
diciembre	noviembre	junio	noviembre	febrero	febrero	septiembre	julio	noviembre	septiembre
septiembre	diciembre	junio	mayo	noviembre	diciembre	diciembre	julio	agosto	diciembre
abril	septiembre	febrero	mayo	marzo	marzo	marzo	mayo	abril	abril
diciembre	enero	febrero	abril	enero	junio	diciembre	marzo	enero	febrero

a. ¿Habría alguna razón para pensar que en algún mes la empresa debe dar más obsequios?

b. ¿Qué forma puede tener la distribución del mes del cumpleaños?

5. Los tiempos de sobrevivencia en meses de 35 fallecidos por cáncer de pulmón mostrados en el ejercicio 2.1.4 son:

10	1	3	7	9	17	9
1	10	9	5	15	34	5
27	17	5	4	3	15	13
3	1	10	6	13	15	3
1	19	1	3	0	2	76

a. Visualice mentalmente la distribución de los tiempos de supervivencia. ¿Su distribución será sesgada?

b. Determine los límites de un intervalo que recoja el 75% de los tiempos de supervivencia.

c. Utilice la regla de las distribuciones acampanadas para determinar por qué la distribución de los tiempos de supervivencia no puede tener forma de campana.

6. En el ejercicio 2.2.3, se mencionó el valor pagado por concepto de agua y alcantarillado de una familia de estrato 3 en Cali.

Enero	Febrero	Marzo	Abril	Mayo	Junio
102.132	98.935	98.388	91.710	87.898	81.056

Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
97.665	88.361	85.169	82.393	78.167	74.715

a. Estime la desviación estándar muestral de los datos basada en el rango.

b. Compare el resultado anterior con el verdadero valor de la desviación estándar.

7. Las ventas brutas de la Lotería de Bogotá durante los últimos 24 meses entre 2008 y 2010 fueron:

843’675.000	990’855.000
812’425.000	875’805.000
870’345.000	972’760.000
935’675.000	999’970.000
870’250.000	997’105.000
805’855.000	923’818.000
858’560.000	824’680.500
871’760.000	832’849.500
860’165.000	795’126.500
802’355.000	847’710.500
865’020.000	808’409.000
894’795.000	790’982.500

a. Calcule e interprete el rango.
b. Estime la desviación estándar muestral de los datos basada en el rango.
c. Calcule la desviación estándar muestral.
d. ¿Qué proporción de las ventas brutas se encuentran a dos desviaciones estándar de la media o menos?; ¿a tres desviaciones o menos?
e. ¿Concuerdan los anteriores porcentajes con los que se obtienen con la desigualdad de Chevyshev?

8. Use la información estadística de la base de datos *Indicadores demográficos según departamento*, para calcular e interpretar la varianza y desviación estándar muestrales de las siguientes variables:
a. Edad media de fecundidad.
b. Esperanza de vida de hombres y de mujeres.

9. Use la información estadística de la base de datos *Nacimientos por departamento*, para determinar:
a. La media de nacimientos por periodo en los departamentos.
b. La desviación estándar muestral de los nacimientos por periodo en los departamentos.
c. ¿Qué proporción de los nacimientos por departamento se encuentran a dos desviaciones estándar de la media o menos? ¿A tres desviaciones o menos?

Nota: obtenga la información demográfica de los departamentos para poder determinar las estadísticas por habitante.

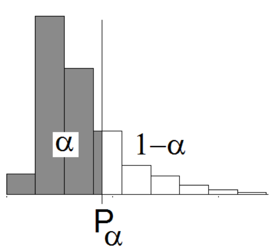
Percentiles muestrales

Considere la siguiente afirmación: "**Promedio de 381 para los estudiantes colombianos en la prueba de matemáticas Pisa 2009 en donde participaron 66 países**". Observe que la afirmación menciona el puntaje obtenido por los estudiantes colombianos, pero no informa acerca de la escala en la que son evaluados. No se sabe si 381 puntos en promedio es una calificación alta o baja. Es más, aun conociendo la escala no se sabría cómo es ese resultado en comparación con los demás países. En situaciones como esta, es importante dar una medida estadística que represente la posición que ocupa un dato con respecto a los demás o la posición relativa. Por ejemplo, podríamos reescribir la afirmación así: "**Promedio de los estudiantes colombianos en la prueba de matemáticas Pisa 2009, entre los diez más bajos de los 66 países participantes**". Note que no es necesario determinar el puntaje promedio obtenido por los colombianos para darse una idea de cómo fue el resultado en comparación con los otros países y que se resalta la posición en la que se ubicó el país.

A continuación se analizan los principales estadísticos que dan cuenta de la posición relativa en un conjunto de datos: los percentiles y cuartiles muestrales. Recuerde que la mediana es un valor que se ubica en la posición central de un conjunto de datos. La mediana divide los datos en dos grupos, cada uno con igual frecuencia. Esta idea se puede extende al dividir el conjunto de datos en cien grupos de igual frecuencia para dar lugar a los percentiles.Por su construcción los percentiles son útiles en la medida en que el conjunto de datos sea grande, simplemente, porque si se cuenta con pocos datos no es razonable dividirlos en cien grupos.

El percentil muestral P_α de un conjunto de datos es un valor tal que α por ciento de los datos son menores o iguales que él y $1-\alpha$ por ciento de los valores son mayores que él.

Figura 2.5. Percentil muestral P_α



Percentil P_α

El percentil muestral P_α es un valor mayor o igual que al menos α por ciento de los datos y menor que al menos $1-\alpha$ por ciento de los datos.

Para calcular un percentil α se sigue un razonamiento similar al utilizado en el cálculo de la mediana, solo que el percentil se ubicará en proporción aritmética entre los valores enteros que su posición indique. La posición del percentil α, es:

$$\textit{Posición del percentil } \alpha = \frac{\alpha}{100}(n+1)$$

Los detalles del cálculo de un percentil se muestran en el siguiente ejemplo.

Ejemplo 2.12

Enunciado

En la tabla 2.4, se organizaron de menor a mayor los salarios del primer empleo de cincuenta egresados de contaduría de diferentes universidades del país durante el año 2010. Encontrar e interpretar el percentil 90.

Tabla 2.4

1	11	21	31	41
1'426.000	1'795.000	2'008.000	2'095.000	2'267.000
2	12	22	32	42
1'491.000	1'801.000	2'009.000	2'110.000	2'347.000
3	13	23	33	43
1'727.000	1'810.000	2.011.000	2'112.000	2'424.000
4	14	24	34	44
1'729.000	1'811.000	2'014.000	2'122.000	2'436.000
5	15	25	35	45
1'735.000	1'862.000	2'020.000	2'125.000	2'518.000
6	16	26	36	46
1'750.000	1'966.000	2'025.000	2130.000	2'523.000
7	17	27	37	47
1'756.000	1'972.000	2'030.000	2180.000	2'531.000
8	18	28	38	48
1'756.000	1'982.000	2'054.000	2200.000	2'573.000
9	19	29	39	49
1'763.000	1'986.000	2'061.000	2'211.000	2'683.000
10	20	30	40	50
1'781.000	1'991.000	2'086.000	2'238.000	2'745.000

Solución

Como se cuenta con $n = 50$ datos, la posición del percentil 90 es $\frac{\alpha}{100}(n+1) = \frac{90}{100}(51) = 45,9$. Esto significa que el percentil se encuentra entre la posición 45 y 46, pero más cercano al salario ubicado en la posición 46. Por tanto, el percentil es:

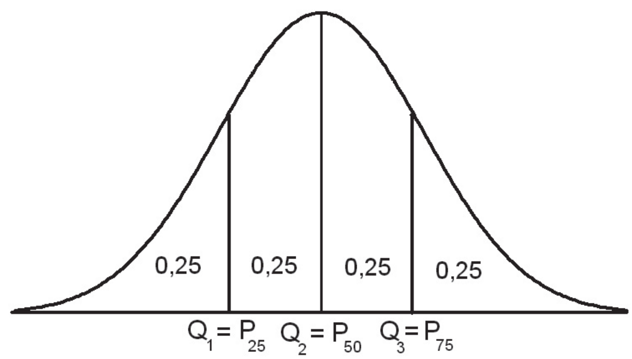
$$P_{90} = \overset{\substack{\text{Salario en la}\\ \text{posición 45}\\ \downarrow}}{2'518.000} + \overset{\substack{\text{Decimal de la}\\ \text{posición}\\ \downarrow}}{0,9} \times (\overset{\substack{\text{Salario en la}\\ \text{posición 46}\\ \downarrow}}{2'523.000} - \overset{\substack{\text{Salario en la}\\ \text{posición 45}\\ \downarrow}}{2'518.000}) = 2'522.500$$

La interpretación es que el 90% de los recién egresados de contaduría de la muestra tienen un salario inferior o igual a \$2'522.500 y solo el 10% de ellos gana más de este valor.▪

Cuartiles

Tres percentiles utilizados con frecuencia son P_{25}, P_{50} y P_{75} porque dividen el conjunto de datos en cuatro partes cada una con un porcentaje aproximado de 25%. Por esta razón, se les denomina cuartiles. El primer cuartil Q_1 es el percentil 25, el segundo cuartil Q_2 es el percentil 50 y el tercer cuartil Q_3 es el percentil 75:

$Q_1 = P_{25}$, $Q_2 = P_{50}$, $Q_3 = P_{75}$



Los cuartiles también son útiles para detectar valores inusualmente grandes o pequeños mediante una regla fundada en la experiencia. Este criterio se muestra enseguida:

Tipo de dato	**Criterio**
Inusualmente grande	Aquellos valores mayores que $Q_3 + 1.5(Q_3 - Q_1)$
Inusualmente pequeño	Aquellos valores menores que $Q_1 - 1.5(Q_3 - Q_1)$

Ejemplo 2.13

Enunciado

Los diez últimos pagos del impuesto de Industria y Comercio de un almacén de zapatos son:

220.000 237.000 215.000 279.000 283.000 249.000 258.000 217.000 205.000 140.000

¿Existe algún valor inusualmente mayor que los demás? ¿El último pago es inusualmente bajo?

Solución

Las estadísticas de orden de los pagos son:

$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$	$X_{(8)}$	$X_{(9)}$	$X_{(10)}$
140.000	205.000	215.000	217.000	220.000	237.000	249.000	258.000	279.000	283.000

Primero se determinan los cuartiles:

Posición	Cuartil
$0,25(10+1)=2,75$	$Q_1 = 205.000 + 0,75(215.000 - 205.000) = 212.500$
$0,5(10+1)=5,5$	$Q_2 = 220.000 + 0,50(237.000 - 220.000) = 228.500$
$0,75(10+1)=8,25$	$Q_3 = 258.000 + 0,25(279.000 - 258.000) = 263.250$

Un valor se considera inusualmente grande si:
Es mayor que $Q_3 + 1.5(Q_3 - Q_1) = 263.250 + 1.5(263.250 - 212.500) = 339.375$

Un valor se considera inusualmente pequeño si:
Es menor que $Q_1 - 1.5(Q_3 - Q_1) = 212.500 - 1.5(263.250 - 212.500) = 136.375$
Por lo tanto en los valores pagados por concepto de Industria y Comercio no se presentan valores inusuales.▪

Gráfico de caja

Como se dijo, los cuartiles dividen los datos en cuatro grupos cada uno con 25% de los mismos. Una representación práctica de los cuartiles es el **gráfico de caja**, el cual revela el rango de variación de los datos, el sesgo o la simetría de la distribución y la presencia de datos inusualmente grandes o pequeños. El gráfico se construye a partir de los valores máximo y mínimo de los datos —siempre y cuando no sean inusuales— y los cuartiles: Q_1, Q_2 y Q_3.

Ejemplo 2.14

Enunciado

Construir el gráfico de caja de los salarios del ejemplo 2.12.

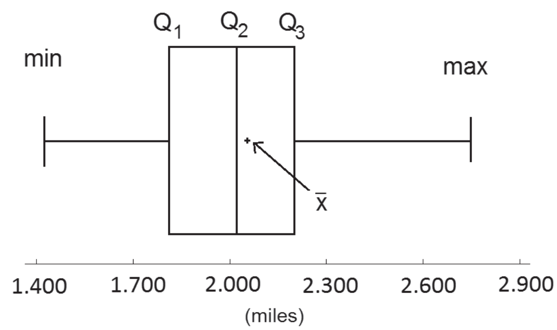
Solución

Los cuartiles de los datos son:

Posición	Cuartil
$0,25(50+1)=12,75$	$Q_1 = 1'801.000 + 0,75(1'.810.000 - 1'801.000) = 1'807.750$
$0,5(50+1)=25,5$	$Q_2 = 2'020.000 + 0,50(2'025.000 - 2'020.000) = 2'022.500$
$0,75(50+1)=38,25$	$Q_3 = 2'200.000 + 0,25(2'211.000 - 2'200.000) = 2'202.750$

El gráfico de caja se construye sobre un eje horizontal que muestra el rango de variación de los datos. Se traza una línea horizontal desde el valor mínimo al máximo (siempre y cuando no sean inusuales) y sobre él una caja cuyos extremos corresponde a Q_1 y Q_3. Por último, se ubica el segundo cuartil y una marca para la media. Los valores inusuales se marcan con asterisco. El máximo y mínimo del gráfico son el valor más grande y el más pequeño, descontando los inusuales. La figura 2.6 presenta el gráfico de caja de los salarios del ejemplo 2.12.

Figura 2.6 Gráfico de caja de los salarios del ejemplo 2.12

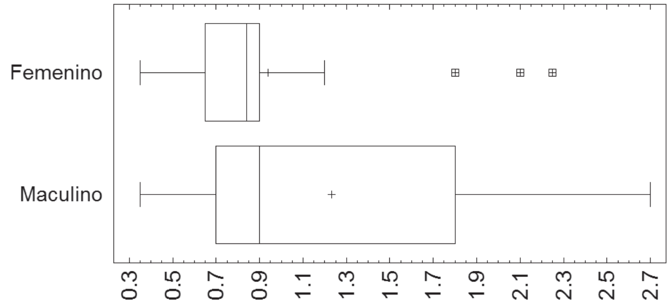


Observe que la ubicación de la caja en el gráfico revela la simetría de la distribución de los pagos. Esto se ve reforzado, porque la media y la mediana son muy parecidas.▪

Ejercicio 2.4 Percentiles

1. Explique el significado de las siguientes expresiones:
   a. La estatura de una persona se encuentra en el percentil 75.
   b. El peso de una adolescente se ubica en el percentil 10.
   c. La nota mínima de aprobación será el percentil 30 de las calificaciones obtenidas por todos los estudiantes en un examen.

2. En el siguiente gráfico, se comparan los salarios (en millones) de hombres y mujeres en una empresa.



a. ¿Qué medida describe mejor el centro de los salarios de los hombres? Justifique.

b. Estime e interprete los cuartiles Q_1, Q_2 y Q_3 del salario de las mujeres.

c. Escriba un breve resumen del gráfico.

3. La siguiente tabla contiene las mediciones del peso (en gramos) de 20 paquetes de café cuya etiqueta anuncia 500 gramos.

467	483	493	503
468	485	494	516
472	488	495	517
474	489	496	519
480	491	502	525

a. Calcule y explique los límites entre los cuales una medición se considera inusual en esta muestra.

b. Calcule e interprete los percentiles 5 y 95.

4. El Ministerio de Transporte colombiano publica todos los años las tablas de avalúos de vehículos. Estos sirven como base para la declaración del impuesto sobre vehículos. Aquí se presenta la lista de avalúos de tres marcas que en el año 2010 eran modelo 2009 con cilindraje menor o igual a 1600 cc.

Marca	Linea	Cilindraje	Avalúo
Chevrolet	Aveo	1400	23.700
Chevrolet	Aveo	1600	24.700
Chevrolet	Aveo Family	1500	21.100
Chevrolet	Aveo Emotion	1400	31.800
Chevrolet	Aveo Emotion	1600	32.200
Chevrolet	Chevy c2	1600	26.200
Chevrolet	Gran Vitara	1600	37.300
Chevrolet	Optra	1600	32.400
Chevrolet	Spark	1000	16.800
Chevrolet	Vitara	1600	33.700
Mazda	Mazda 2	1500	34.900
Mazda	Mazda 3	1600	42.700
Renault	Clio	1600	24.400
Renault	Logan	1600	24.000

Marca	Linea	Cilindraje	Avalúo
Renault	Megane	1600	28.100
Renault	Sandero	1600	24.200
Renault	Scenic	1600	39.100
Renault	Symbol	1400	22.000
Renault	Symbol	1600	22.000
Renault	Twingo U	1200	19.500

a. Elabore una gráfica de caja para los avalúos de los vehículos de la tabla.

b. ¿Hay valores inusuales?

5. En el ejercicio 2.2.2 se relacionaron los datos de talla y peso de veintisiete inscritos al gimnasio de la Universidad:

Sexo	Estatura (cm)	Peso (kg)
Mujer	159	49
Hombre	164	62
Mujer	172	65
Mujer	167	52
Mujer	164	51
Mujer	161	67
Mujer	168	48
Hombre	181	74
Hombre	183	74
Mujer	158	50
Mujer	156	65
Hombre	173	64
Mujer	158	43
Hombre	178	74

Sexo	Estatura (cm)	Peso (kg)
Hombre	181	76
Hombre	182	91
Hombre	176	73
Mujer	162	68
Mujer	156	52
Mujer	152	45
Hombre	181	80
Hombre	173	69
Mujer	155	53
Hombre	189	87
Mujer	170	70
Hombre	170	67
Mujer	168	56

a. Elabore los gráficos de caja del peso de los hombres y de las mujeres.

b. ¿Qué indican los anteriores gráficos sobre la forma de la distribución de los pesos?

c. Calcule los percentiles 30 y 70 de las tallas de los hombres y compárelos con los de las mujeres.

6. En el ejercicio 2.2.5 se mencionaron los tiempos en las últimas quince carreras de dos atletas:

Atleta 1	12,01	10,95	10,31
	11,09	11,49	13,51
	10,79	12,22	12,03
	10,62	13,18	10,27
	12,67	10,54	10,82

Atleta 2	11,59	11,01	11,24
	10,91	11,35	11,18
	11,10	11,20	11,26
	11,00	11,29	11,55
	11,14	11,25	10,90

a. Construya un diagrama de caja para los tiempos de carrera de cada uno de los atletas.
b. Escriba un párrafo corto para comparar los resultados de los dos atletas.

CAPÍTULO 3
CORRELACIÓN Y REGRESIÓN

En este capítulo, se hace un acercamiento a los conceptos de correlación y regresión para mostrar su utilidad y campo de aplicación. Un estudio completo de la regresión requiere conocimientos de inferencia estadística que no son tratados en este libro. El siguiente ejemplo sirve de punto de partida para la discusión del tema.

Ejemplo 3.1

Enunciado

¿Qué variables se pueden tener en cuenta para decidir cuál es el precio de una vivienda?

Solución

El precio de la vivienda se ve afectado por varias variables, entre ellas: ubicación, área, estado, acabados, vías de acceso, antigüedad, etc.

Observe que unas de estas variables son de tipo cualitativo y otras de tipo cuantitativo. Cada una tiene un mayor o menor peso sobre el precio de la vivienda.▪

Figura 3.1. Variables influyentes en el precio de una vivienda

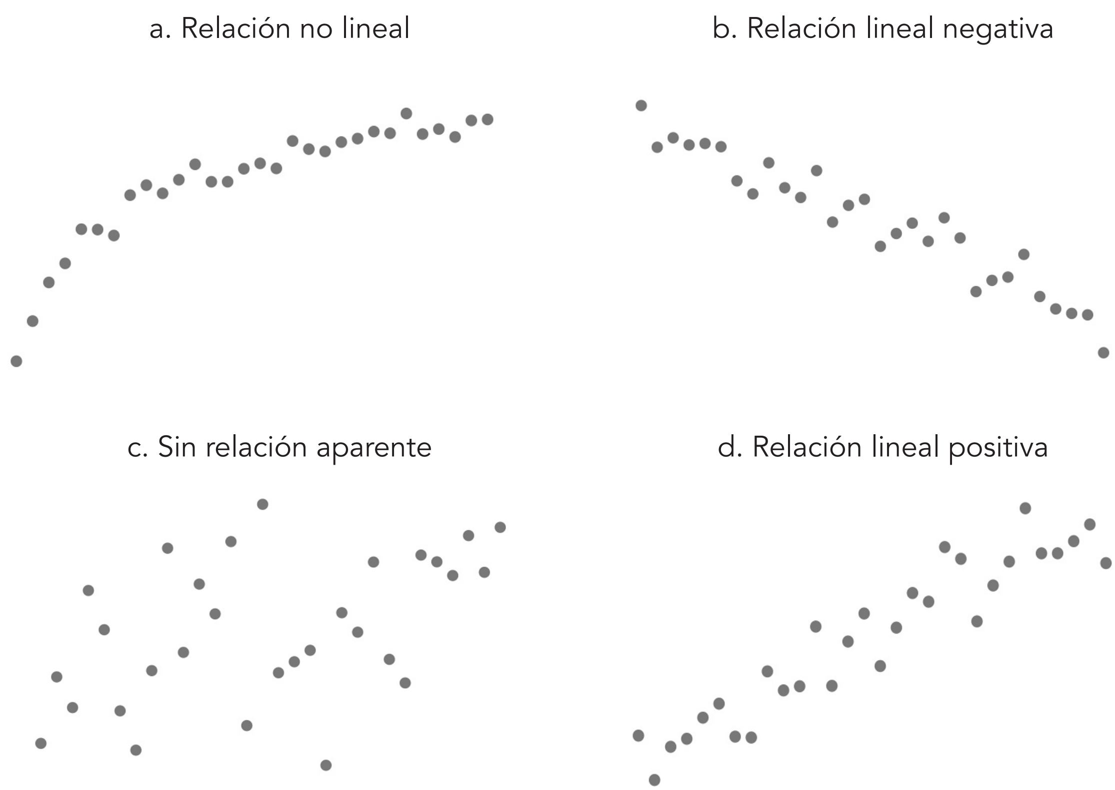
En el anterior ejemplo parece importante explicar el comportamiento de una variable Y (precio) en relación con la dependencia o influencia de otras variables X (ubicación, estado, acabados). A Y se le llama variable respuesta o **dependiente** y a las X **independientes** o estímulos. El análisis de esta situación se puede dividir en dos partes: medir la fuerza de la relación entre las variables (**correlación**) y obtener una ecuación mediante la cual se pueda estimar el valor medio de una variable como función de la otra (**regresión**).

Regresión lineal simple y correlación

Cuando se habla de regresión lineal simple se hace mención a que se está analizando la relación entre dos variables *X* y *Y*. La palabra **regresión** supone que el propósito del estudio es la predicción de una de las variables por medio de una ecuación; **lineal** significa que el patrón que sugieren los datos es el de una línea recta y **simple** que la ecuación contiene solo dos variables.

Antes de intentar construir una ecuación que relacione las dos variables es importante examinar el tipo de relación existente entre las variables por medio de un gráfico de dispersión. Esta es una representación en la cual usualmente en el eje vertical se ubica la variable dependiente y en el horizontal la independiente. Los puntos de la gráfica corresponden a los valores que un elemento de la muestra toma en cada una de las variables. Las formas más típicas de los diagramas de dispersión son las siguientes:

Figura 3.2 Posibles formas de los gráficos de dispersión



En la figura 3.2 parte (c), se muestra un diagrama de dispersión en el que no se observa relación entre las dos variables, porque hay una distribución irregular de los puntos sobre el diagrama. En las partes (b) y (d), los diagramas reflejan una asociación lineal. En la parte (a), el diagrama sugiere una relación entre las variables que no se ajusta a una línea recta.

Las relaciones lineales entre las variables pueden ser de dos tipos:

Si valores pequeños de X tienden a estar asociados con valores pequeños de Y, valores grandes de X tienden a estar asociados con valores grandes de Y y el diagrama de dispersión sugiere una relación lineal. Se dice que hay **asociación lineal positiva** entre X y Y (figura 3.2 d).

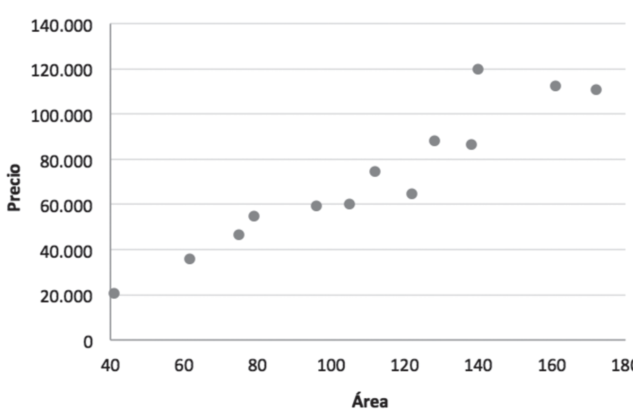
Si valores pequeños de X tienden a estar asociados con valores grandes de Y, valores grandes de X tienden a estar asociados con valores pequeños de Y y el diagrama de dispersión sugiere una relación lineal; se dice que hay **asociación lineal negativa** entre X y Y (figura 3.2 b).

Ejemplo 3.2

Enunciado

La siguiente tabla muestra los precios y áreas de quince viviendas de estrato 3 en Bogotá. Analizar la relación existente entre las variables precio y área.

Área (m^2)	41	62	75	79	96	105	112	122	128	138	140	161	172
Precio (miles de $)	20.800	35.900	46.600	55.000	59.200	60.000	74.600	64.700	88.300	86.700	120.000	112.300	111.000



Solución

El área de una vivienda generalmente es un valor fijo, mientras que el precio puede ser variable. Por esta razón, el precio es la variable dependiente y el área la variable independiente.

El gráfico de dispersión revela el tipo de relación existente entre el precio y el área de estas viviendas.

Se observa que las viviendas con precios altos están relacionadas con áreas grandes y viviendas con precios bajos tienen áreas pequeñas.

Una línea recta parece resumir la relación entre las dos variables, es decir, la forma de la nube de puntos del gráfico de dispersión sugiere que hay una asociación lineal positiva entre el precio y el área de estas viviendas.▪

La asociación lineal entre dos variables se puede medir en la población con la **covarianza** y el **coeficiente de correlación lineal de Pearson**. Como se mencionó, los parámetros poblacionales son desconocidos y se deben estimar con la información disponible en la muestra. A continuación, se presentan las estimaciones muestrales de la covarianza $Cov(X,Y)$ y el coeficiente de correlación lineal de Pearson r.

La covarianza:

$$Cov(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1}$$

Note que si valores pequeños de X tienden a estar asociados con valores pequeños de Y y valores grandes de X tienden a estar asociados con valores grandes de Y, los productos $(x_i - \overline{x})(y_i - \overline{y})$ son en su mayoría positivos. Es decir, si la covarianza es positiva, la asociación lineal entre las variables también lo es.

Igualmente, si valores pequeños de X tienden a estar asociados con valores grandes de Y y valores grandes de X tienden a estar asociados con valores pequeños de Y, los productos $(x_i - \overline{x})(y_i - \overline{y})$ son en su mayoría negativos. Es decir, si la covarianza es negativa, la asociación lineal entre las variables también lo es.

No obstante, con la covarianza es difícil medir la fuerza de asociación lineal entre las variables porque depende de las unidades de medida de las variables. El coeficiente de correlación lineal de Pearson estimado se calcula dividiendo la covarianza entre el producto de las desviaciones estándar de las variables X y Y, lo que deja al estadístico libre de unidades.

$$r = \frac{Cov(X,Y)}{S_x S_y}$$

Donde S_X es la desviación estándar de la variable X y S_Y es la desviación estándar de la variable Y. No se probará, pero el coeficiente de correlación toma valores entre –1 y 1, esto es:

$$-1 \leq r \leq 1$$

Esta es una de las medidas de asociación lineal más utilizadas porque está desligada de las unidades de las variables y porque tiene una interpretación sencilla. Con este coeficiente, el investigador puede decidir acerca de la existencia o no de una asociación lineal entre las variables y sobre su intensidad. El uso que se le da al coeficiente de correlación es algo subjetivo, porque puede depender del objeto de estudio. Sin embargo, Milton (2001) sugiere una escala para su interpretación, la que se presenta en la figura 3.4.

Figura 3.4 Interpretación del coeficiente de correlación

Fuerte –	Moderada –	Débil –	Débil +	Moderada +	Fuerte +
–1 –0,9	–0,5	0	0,5	0,9	1

Covarianza y coeficiente de correlación lineal de Pearson estimados

$$Cov(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1}$$

$$r = \frac{Cov(X,Y)}{S_x S_y}$$

Por ejemplo:

Si la correlación es:	La asociación lineal entre las variables *X* y *Y* es:
-1	Perfecta y negativa
0,5	Moderada y positiva
0,09	Inexistente
1	Perfecta y positiva
-0,95	Fuerte y negativa

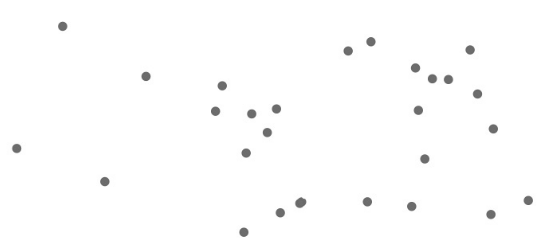
Los gráficos de dispersión también dan una idea de cómo es la correlación:

Figura 3.5 Correlaciones y diagramas de dispersión

a.

$r \to 1$
Asociación lineal positiva fuerte

b.



$r \to 0$
No existe asociación lineal

c. 

$r \to -1$
Asociación lineal
negativa fuerte

No es necesario un ajuste exacto de todos los puntos a una línea recta para asumir la existencia de una asociación lineal; es posible que algunos se alejen de esta tendencia, pero esta debe ser la más frecuente. En la figura 3.5, se muestran dos correlaciones lineales muy fuertes, una positiva (parte a) y otra negativa (parte c). En la parte b, se muestra unos datos no correlacionados.

Ejemplo 3.3

Enunciado

Calcular la covarianza y el coeficiente de correlación muestrales entre las variables precio de la vivienda y área del ejemplo 3.2.

Solución

La siguiente tabla contiene en la columna A los valores de la variable *X* y en la B los de la variable *Y*. Como la expresión de la covarianza contiene en el numerador a $\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})$, primero se calculan las desviaciones de las áreas con respecto al área promedio $x_i-\overline{x}$ (columna C). Luego, las desviaciones de los precios con respecto al precio promedio $y_i-\overline{y}$ (columna D). En la columna E, aparecen los productos de las columnas C y D, es decir, $(x_i-\overline{x})(y_i-\overline{y})$. Puede corroborar algunos valores con una calculadora.

A	B	C	D	E
x_i	y_i	$x_i-\overline{x}$	$y_i-\overline{y}$	$(x_i-\overline{x})(y_i-\overline{y})$
41	20.800	–69,08	–51.130,77	3.531.956,2
62	35.900	–48,08	–36.030,77	1.732.248,5
75	46.600	–35,08	–25.330,77	888.525,4
79	55.000	–31,08	–16.930,77	526.156,2
96	59.200	–14,08	–12.730,77	179.210,1

A	B	C	D	E
105	60.000	–5,08	–11.930,77	60.571,6
112	74.600	1,92	2.669,23	5.133,1
122	64.700	11,92	–7.230,77	–86.213,0
128	88.300	17,92	16.369,23	293.387,0
138	86.700	27,92	14.769,23	412.402,4
140	120.000	29,92	48.069,23	1'438.379,3
161	112.300	50,92	40.369,23	2'055.725,4
172	111.000	61,92	39.069,23	2'419.287,0
$\overline{x} = 110,08$	$\overline{y} = 71.930,8$	$\sum_{i=1}^{13}(x_i-\overline{x})(y_i-\overline{y}) =$		13'456.769
$s_x = 38,808$	$s_y = 30.492,50$			

La última celda de la tabla corresponde al numerador de la covarianza, esto es:

$$Cov(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{n-1} = \frac{13'456.769}{12} = 1'121.397,4$$

Como la covarianza es positiva, se sabe que la asociación lineal también lo es. Sin embargo, con este valor no se puede decidir acerca de la fuerza de dicha asociación. Por esta razón, se calcula el coeficiente de correlación lineal de Pearson que se obtiene estandarizando la covarianza:

$$r = \frac{Cov(X,Y)}{S_X S_Y} = \frac{1'121.397,4}{38,808 \times 30.492,5} = 0,947639178$$

Según la escala de Milton (2001), la asociación lineal entre el precio de la vivienda y el área de la misma es fuerte y positiva. Esto significa que lo que ocurre con más frecuencia es: viviendas con precios altos y áreas grandes o viviendas con precios bajos y áreas pequeñas. Según el coeficiente de correlación, es menos frecuente encontrar alguna vivienda con precio alto y área pequeña o viceversa.▪

Existen formas abreviadas para los cálculos del ejemplo anterior. El autor sugiere no detenerse demasiado en la realización de los cálculos para concentrarse en la interpretación de estos. Una calculadora científica o un programa de computador como Excel son dos herramientas con las que se obtienen estos dos estadísticos fácilmente.

Una vez el investigador tiene certeza de la existencia y de la aceptable fuerza de la asociación lineal entre las variables, puede intentar obtener una ecuación para estimar el valor medio de una de las variables como función de la otra, es decir, puede hacer una regresión.

El modelo de la regresión lineal simple

Si se asume que la asociación es lineal, la ecuación que se utilizará para modelar la relación entre *X* y *Y* será la de una recta $y = \alpha + \beta x$. En la regresión lineal simple se asume que *Y* es una variable aleatoria y que *X* es fija. El concepto de variable aleatoria se estudiará al final del capítulo 4, sin embargo, para lo que aquí concierne suponga que las dos variables no tienen una relación exacta. Para ilustrarlo, remítase al ejemplo 3.1, en el que se planteó la dependencia entre precio de una vivienda y área. No todas las viviendas de 100 m^2 tienen el mismo precio debido a la ubicación, los acabados, el estado, etc., que hacen variar su precio. Entonces, decimos que el área toma un valor fijo, mientras que el precio es aleatorio.

Como el precio de la vivienda depende, en gran medida, del área de esta, se puede suponer que el promedio del precio (Y): $\mu_{Y|X}$ es función del área (*X*). Esto es, se supone que se trabaja con una variable aleatoria condicional $Y|X = x$ que tiene una media $\mu_{Y|X}$. La gráfica de esta función es la que se conoce como la regresión de *Y* sobre *X*. Es importante recalcar que en regresión se modela el valor promedio de una variable aleatoria *Y* en los diferentes valores de la variable *X*. Se le sugiere al lector volver a esta explicación una vez haya estudiado el capítulo 4. Ahora bien, si la relación teórica entre las variables *X* y *Y* es lineal, se puede escribir el modelo como el siguiente:

$$\mu_{Y|X} = \alpha + \beta x$$

Donde α se interpreta como el valor promedio de la variable dependiente cuando la variable independiente toma el valor cero, si es que esta lo admite y β como el aumento o disminución de la variable dependiente por aumento unitario de la variable independiente.

Recuerde que, en regresión lineal simple, se asume que la gráfica del promedio de la variable dependiente en los diferentes valores de la variable independientes se ajusta a una línea recta. Esto es, $\mu_{Y|X} = \alpha + \beta x$ (relación teórica), donde α y β son valores desconocidos.

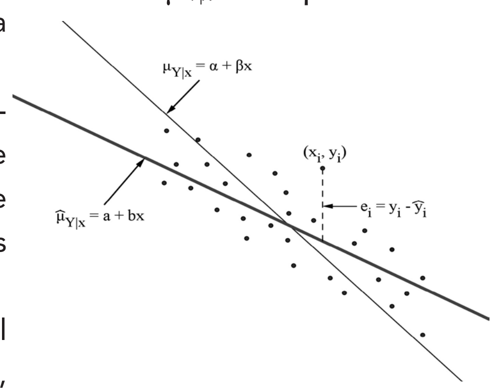
Con base en la muestra, se buscan estimaciones *a* y *b* de los parámetros α y β que mejor reflejen la relación teórica $\mu_{Y|X} = \alpha + \beta x$. Con estas estimaciones se construye la recta de regresión estimada $\widehat{\mu}_{Y|X} = a + bx$.

Esta recta de regresión estimada debe estar lo más cerca posible de los puntos de la muestra. Con base en este criterio los valores de *a* y *b* se escogen de manera que se minimicen los cuadrados de las distancias verticales de los puntos a la recta de regresión.

La figura muestra la diferencia entre un dato puntual y el valor estimado por la recta de regresión, $e_i = y_i - \widehat{y}_i$, donde $\widehat{y}_i = a + bx_i$. A los valores e_i se les llama residuales.

Figura 3.6

$\mu_{Y|X} = \alpha + \beta x$



Las estimaciones de *a* y *b* se obtienen al minimizar la suma de sus cuadrados, es decir, al minimizar $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$. Se utilizan los cuadrados de los residuales para que la suma refleje las diferencias de los puntos a la recta y, así, evitar la cancelación aritmética de residuales negativos y positivos. Se puede demostrar que con base en este criterio, la mejor estimación de α y β es:

$$b = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$a = \overline{y} - b\overline{x}$$

Como el propósito de este capítulo es el acercamiento a los conceptos, se pasará por alto el desarrollo de los cálculos para obtener los valores de *a* y *b*. Por el contrario, se hace énfasis en la interpretación de estos valores.

Casi todas las calculadoras científicas incluyen un modo de regresión con el que se pueden obtener *r, a* y *b*. A continuación, se resume el procedimiento para hacer una regresión en una calculadora:

Modo de regresión en una calculadora científica (Mode REG)

Salvo pequeñas diferencias el procedimiento es el siguiente:

1. Ponga su calculadora en el modo **REG** y escoja la opción **LR** (linear regression).
2. Asegúrese de borrar los datos guardados en memoria con **Mcl** (memory clear).
3. Introduzca los datos por parejas, separados por una coma $\boxed{,}$ y cada vez que ingrese una pareja oprima la tecla $\boxed{M+}$. Por ejemplo: **41, 20800** $\boxed{M+}$.
4. Una vez ingrese todos los datos ubique en la memoria los valores de r, a y b. Generalmente pueden ser llamados de la memoria al oprimir $\boxed{\text{SHIFT}}$ seguido de $\boxed{2}$.

Ejemplo 3.4

Enunciado

Determinar la recta de regresión estimada de los datos de área y precio de las viviendas del ejemplo 3.2 y usarla para estimar el precio promedio de una vivienda de 100 m^2 y otra de 200 m^2.

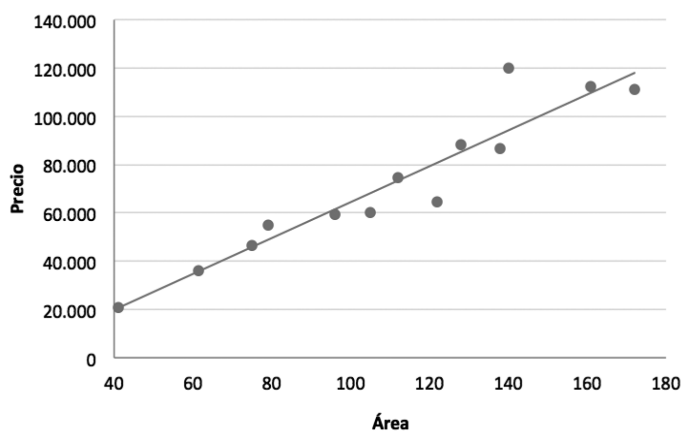
Solución

Como ya se vio, existe una fuerte asociación lineal entre el área y el precio de las viviendas y parece razonable construir una regresión lineal simple. Las estimaciones que ofrece una calculadora de los valores de α y β son respectivamente:

$$a = -10.030,5$$

$$b = 744,582$$

Figura 3.7. Recta de regresión del precio sobre el área



Por lo tanto la recta de regresión estimada es:

$$\mu_{Y|X} = a + bx$$

$$\mu_{Y|X} = -10.030,5 + 744,582x$$

La gráfica de dicha ecuación se muestra sobre el diagrama de dispersión de la figura 3.7. En este caso, no tiene sentido pensar en el precio promedio de una vivienda con cero metros de área, es más, la estimación es negativa.

El valor de $b = 744,582$ significa que cada metro cuadrado adicional en el área de una vivienda representa un aumento de \$744.582 en el precio promedio.

Para estimar el precio promedio de una vivienda de 100 m^2 basta con reemplazar dicho valor en la ecuación de regresión estimada:

$$\mu_{Y|X=100} = -10.030,5 + 744,582 \times 100 = 64.427,7$$

Esto significa que, en promedio, una vivienda de 100 m^2 tiene un precio de \$64'427.700 pesos.

La estimación del precio promedio de la vivienda de 200 m^2 no se puede realizar con la información suministrada por la muestra, porque no se sabe si más allá de los valores estudiados la tendencia entre el área y el avalúo siga siendo lineal.▪

Ejemplo 3.5

Enunciado

Estime la recta de regresión los datos de talla y peso de diez niños que se registran en la siguiente tabla.

Talla (cm)	80	91	95	95	101	106	109	110	115	125
Peso (kg)	14	16	19	17	21	24	24	24	25	30

Solución

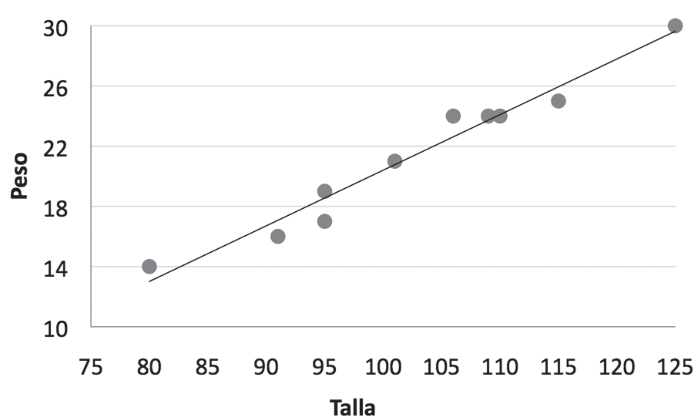
El lector puede comprobar que entre estas dos variables existe una asociación lineal fuerte, porque la correlación es 0,9818 y que una regresión lineal simple es viable. Las estimaciones de mínimos cuadrados que ofrece una calculadora de α y β son respectivamente:

$$a = -16,5682$$

$$b = 0,36970$$

Del valor de *a* no se puede decir mayor cosa, porque no tiene sentido pensar en el peso de un niño de talla cero. Por su parte, *b* indica la variación en el peso promedio por aumento en una unidad de la talla, por ejemplo, una diferencia de un centímetro en la talla está asociada a una diferencia promedio de 0,3697 kg en el peso. Con estos dos valores, se construye la recta de regresión estimada $\mu_{Y|X} = -16,5682 + 0,36970x$. En la figura 3.8 se muestran los datos del problema junto con la recta de regresión.▪

Figura 3.8 Recta de regresión del peso sobre la talla

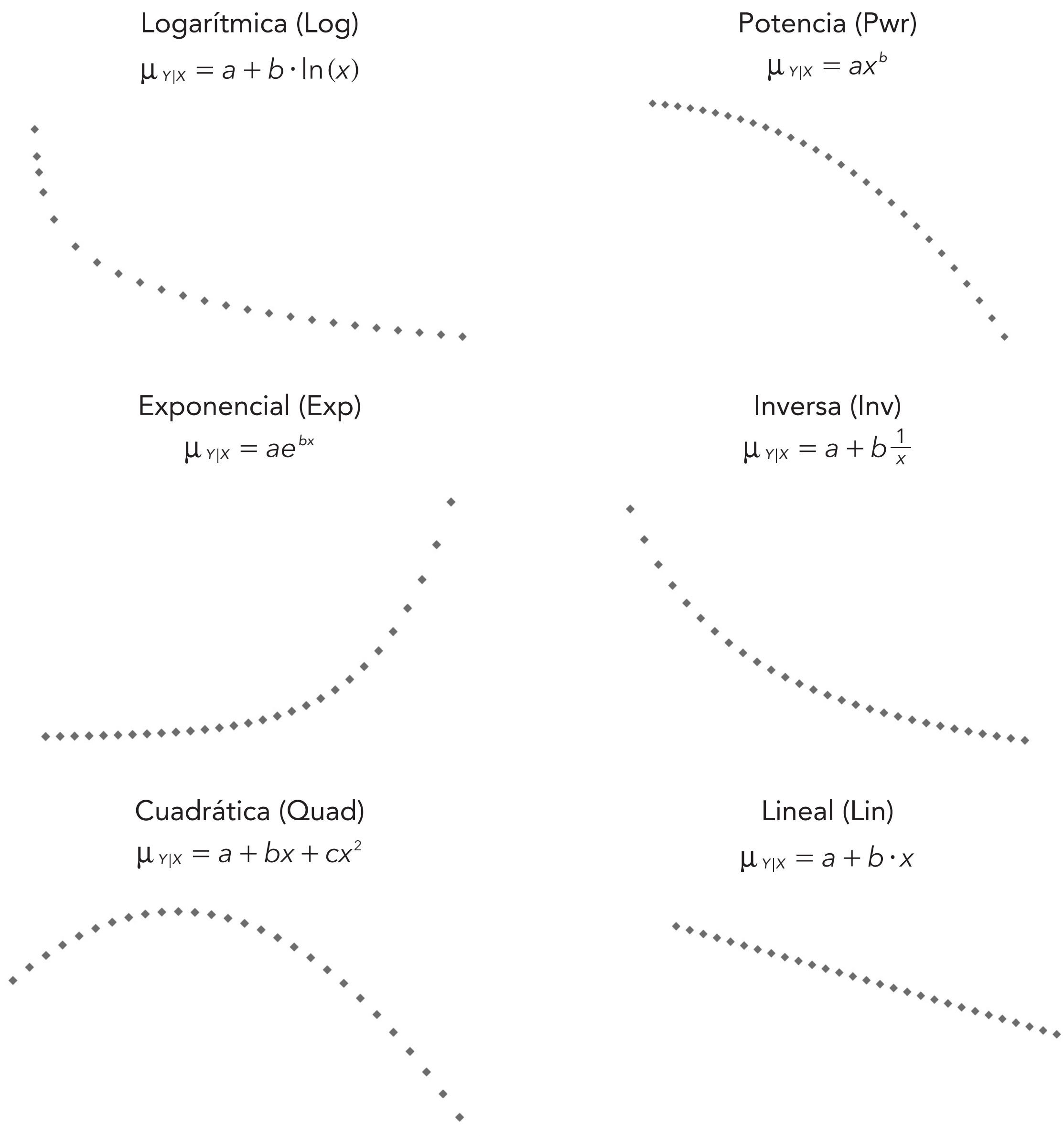


▪

Modelos alternativos a la regresión lineal simple

Si la gráfica de dispersión sugiere una relación entre *X* y *Y* que no queda bien representada por una línea recta, conviene conocer otro tipo de ecuaciones para intentar modelar $\mu_{Y|X}$. Existen muchas posibilidades para escoger un modelo que se ajuste a la relación exhibida por los datos. En la figura 3.9 se muestran varias alternativas, las cuales, generalmente, están disponibles en cualquier calculadora científica y en los programas de computador.

Figura 3.9 Ejemplos de posibles relaciones entre las variables *X* y *Y*



Recuerde que al escribir $y = a + bx$ se plantea una familia de rectas parametrizadas por los valores de a y b. De la misma manera, cada modelo mostrado en la figura anterior corresponde a una familia de curvas de la cual tan solo se ha escogido una como ilustración.

Una calculadora científica estima las regresiones de la figura 3.9. Para activarlas, en el modo REG se escoge alguna de ellas: **Lin** (linear regression), **Log** (Logarítmica), **Exp** (Exponencial), **Pwr** (Potencia), **Inv** (Inversa) y **Quad** (cuadrática).

Además, estas no son las únicas posibilidades para modelar la relación existente entre variables de estudio, ya que existen procedimientos para ajustar un conjunto de datos a el modelo que el investigador considere necesario, varios de estos están expuestos en Montgomery (2001) y Myers (2002).

Determinar qué modelo es el mejor para los datos disponibles, qué tan bueno es el ajuste o cuál es la importancia de cada uno de los parámetros, sobrepasa el alcance de este libro. No obstante, se presenta un ejemplo que centra la atención en la utilidad de esta herramienta.

Ejemplo 3.6

Enunciado

Se realizó un experimento para examinar si el paso del tiempo produce una disminución del número de amigos de una persona. Para ello, se le realizó la siguiente pregunta a un grupo de dieciocho personas de diferentes edades: ¿a cuántas personas considera como verdaderos amigos suyos? Los resultados se muestran en la siguiente tabla:

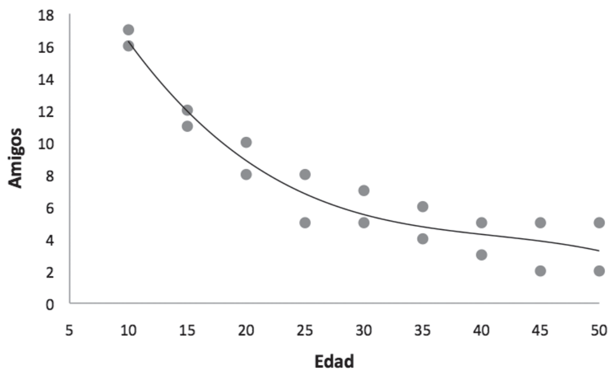
Edad	10	10	15	15	20	20	25	25	30	30	35	35	40	40	45	45	50	50
Amigos	18	17	13	12	11	9	9	6	8	6	7	5	6	4	7	3	6	3

Estime una regresión del número de amigos sobre la edad con base en la muestra seleccionada.

Solución

Un gráfico de dispersión de los datos da información acerca del posible tipo de relación entre el número de amigos y la edad.

Figura 3.10 Regresión del número de amigos sobre la edad



En este grupo de dieciocho personas, se puede observar que el número de amigos va disminuyendo con la edad. Sin embargo, esta disminución no es lineal, sino que disminuye rápidamente entre los 10 y 20 años y muy poco después de los 40 años.

Al comparar el diagrama de dispersión con la figura 3.9, se encuentra que los datos sugieren una relación inversa como:

$$\mu_{Y|X} = a + b\frac{1}{x}$$

Esta no es la única posibilidad para modelar la relación entre estas variables, pero al escogerla se obtienen los siguientes resultados:

$$a = 1,2646$$
$$b = 164,902$$

Por tanto, la ecuación de la regresión queda determinada por:

$$\mu_{Y|X} = 1,2646 + 164,902\frac{1}{x}$$

En la figura 3.10, también se incluye la representación de la ecuación anterior. Observe que si la edad de la persona es muy grande, la segunda parte de la expresión anterior tiende a cero y, por tanto, el número promedio de amigos a edades grandes tiende a 1,2646. ¿Esto significará que las personas tienden a quedarse con un solo amigo? Este no es el único ni el mejor modelo para estudiar la relación entre las dos variables, pero sí puede ser uno útil.▪

Ejercicio 3. Regresión

1. Dé un ejemplo de dos variables que tengan asociación:
   a. Positiva.
   b. Negativa.
   c. Débil o inexistente.

2. La siguiente tabla muestra la tasa de alfabetización de la población adulta de algunos países hispanoamericanos.

País	Total	Varones	Mujeres
Cuba	99,8	99,8	99,8
Uruguay	98,0	97,6	98,3
Argentina	97,6	97,6	97,7
España	97,4	98,6	96,4
Chile	96,5	96,6	96,5
Costa Rica	95,9	95,7	96,2
Paraguay	93,7	94,4	93,1
Colombia	93,6	93,6	93,5
Panamá	93,4	94,0	92,8
Ecuador	92,6	93,6	91,6
México	92,4	94,4	90,6
Brasil	90,5	90,1	90,9
Perú	90,5	95,2	85,8
Bolivia	90,3	95,3	85,4
República Dominicana	89,1	88,8	89,5
El Salvador	85,5	88,4	82,8
Honduras	83,1	82,4	83,7
Nicaragua	80,5	79,7	81,4
Guatemala	73,2	79,0	68,0
Haití	62,1	60,1	64,0

a. ¿Debería estar correlacionada la tasa de alfabetización de hombres con la de mujeres?
b. Trace y analice un diagrama de dispersión. ¿Hay algún país que se aleje un poco de la tendencia?
c. Calcule e interprete el coeficiente de correlación de la tasa de alfabetización entre hombres y mujeres.

3. Los siguientes datos relacionan el número de personas que constituyen doce hogares con el valor pagado en el servicio de energía eléctrica.

Personas	2	4	6	3	4	3
Total pagado	44.715,0	82.393,0	118.388,0	81.710,0	87.898,0	81.056,0

Personas	4	2	4	2	5	7
Total pagado	88.361,0	48.167,0	87.665,0	55.169,0	108.935,0	152.132,0

a. Trace y analice un diagrama de dispersión.
b. Calcule e interprete el coeficiente de correlación entre el número de personas y el total pagado.
c. Encuentre la recta de regresión del total pagado sobre el número de personas.
d. Interprete los coeficientes a y b.
e. Elabore el diagrama de dispersión de los datos y sobre él trace la recta de regresión.
f. Estime el valor medio pagado por un hogar constituido por cinco personas.

4. Un estudiante de biología registró un conjunto de variables antropométricas de sus compañeros, entre ellas, la longitud del brazo derecho y la estatura (ambas en centímetros), como se muestra en la siguiente tabla:

Long. bra. D	74	79	68	72	71	79	70	63	70	70	69	59	74
Estatura	167	182	162	163	158	180	157	153	165	167	166	150	168

Long. bra. D	73	72	66	70	69	75	74	68	76	76	80	82	67
Estatura	170	162	160	164	162	174	173	156	175	170	175	178	160

Long. bra. D	73	71	67	79	71	76	66	68	76	65	73
Estatura	169	168	155	173	163	172	151	163	176	155	167

a. ¿Cree que estas dos variables deberían estar correlacionadas? ¿Qué significa que lo estén?
b. Calcule e interprete el coeficiente de correlación.
c. Encuentre la recta de regresión de la estatura sobre la longitud del brazo. ¿Cuál de las dos variables hará las veces de dependiente?
d. Interprete los coeficientes a y b.

5. Se desea estudiar la relación entre la nota que obtiene un estudiante en un examen final de estadística con el número de horas de estudio. Se indagó una muestra de veinte estudiantes y se observaron las siguientes cifras.

Horas de estudio	1	4	7	6	2	3	3	1	4	7	6	2	3	5	4	7	6	2	3	3
Nota	1,4	3,1	3,9	3,4	1,5	2,0	2,5	2,0	2,7	3,5	4,0	2,8	2,7	3,3	2,5	3,5	3,0	2,0	2,5	2,9

a. Calcule e interprete el coeficiente de correlación, a y b.

b. ¿Cuál es la calificación promedio esperada para quienes estudiaron 5 horas?

6. Resistencia atlética. Se pidió a diez individuos realizar su mejor esfuerzo al correr sobre una pista circular para alcanzar el número máximo de giros. Al final, se registró la edad y el número máximo de giros que cada individuo corrió.

Edad (años)	22	18	41	55	18	21	51	17	29	35
Giros	14	13	10	5	15	13	7	15	13	12

Trace un diagrama de dispersión e interprételo.

a. A partir de la tabla anterior, dé una estimación aproximada del coeficiente de correlación.

b. Calcule e interprete el coeficiente de correlación.

c. Encuentre la recta de regresión.

d. Interprete los coeficientes a y b.

e. ¿Cuántos giros en promedio debería recorrer una persona de 30 años? ¿Cuántos una de 60 años?

f. ¿Cambian los coeficientes a y b al hallar la recta de regresión de la edad sobre el número de giros?

7. La nota del primer parcial de un estudiante en una materia puede ser un indicio de cómo será el desempeño del estudiante durante el resto del curso. Las calificaciones del primer parcial y la definitiva de un grupo de estudiantes se muestran a continuación:

Primer parcial	3,4	2,1	4,5	1,2	1,5	1,5	2,0	3,7	2,7	1,7
Definitiva	3,7	3,3	4,2	0,9	2,5	2,4	2,3	3,3	3,2	1,6

Primer parcial	1,7	4,3	2,3	4,2	2,5	2,8	2,3	2,3	1,7
Definitiva	2,4	3,9	2,4	4,6	2,8	3,5	3,5	2,6	1,7

a. Trace y analice un diagrama de dispersión.

b. Encuentre la recta de regresión de la calificación definitiva sobre la calificación en el primer parcial e interprete los coeficientes a y b.

8. La siguiente tabla contiene el número de casos de tuberculosis (por cada 100.000 habitantes) en la región andina entre 1997 y el 2006. ¿Cree que los casos de tuberculosis han venido disminuyendo?

Año	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Casos	26,3	26,1	24,7	22,8	22,1	20,4	19	17,7	19,3	17,5

9. En la tabla, P denota el peso (en kilogramos) y l la longitud (en centímetros) de la parte posterior de la cabeza a la punta de la nariz de quince caimanes capturados. Como en el hábitat natural es más fácil observar l que P, se ha pensado en un modelo de regresión simple entre la longitud de la cabeza de los caimanes y su peso.

P	119	91	191	76	114	117	79	107	84	89	168	109	102	104	109
l	65	25	202	14	40	55	17	45	18	19	183	42	40	41	35

a. Construya una regresión simple e interprete los coeficientes

b. ¿Sería normal observar en el futuro un caimán con $l = 101{,}6$ y $P = 50$?

10. Se cuenta con los registros médicos de pacientes con cáncer de pulmón. La edad de detección y la edad de muerte de quince pacientes fallecidos por esta enfermedad son:

#	**Edad de la detección**	**Edad de muerte**
1	59	61
2	33	42
3	27	54
4	40	49
5	61	62
6	40	46
7	29	51
8	61	65
9	26	42
10	26	39
11	30	43

#	Edad de la detección	Edad de muerte
12	29	35
13	35	41
14	20	45
15	33	53

a. Trace y analice un diagrama de dispersión.

b. Calcule e interprete el coeficiente de correlación.

11. Las siguientes son las mediciones de seis variables contenidas en las etiquetas informativas de varios quesos vendidos en un supermercado de la ciudad.

Marca	Grasa saturada (g)	Colesterol (mg)	Sodio (mg)	Proteína (g)	Vit A (%)	Calcio (%)
Queso crema Colanta	3	15	80	1	4	2
Queso crema Éxito	6	10	150	2	8	2
Queso crema Alpina	3	15	45	1	4	2
Queso Colanta	5	30	170	5	6	20
Queso Éxito	0	20	210	5	7	18
Queso Alpina	5	30	110	6	0	8

a. Calcule la correlación entre el contenido de calcio y el de sodio.

b. ¿Qué significa esta correlación?

12. Se realizó un estudio para determinar los efectos de no dormir en la capacidad de las personas para resolver problemas sencillos. Diez personas participaron en el estudio. Se dio a cada persona, después de un periodo específico de horas sin dormir (8, 12, 16, 20 o 24 horas), un conjunto de problemas matemáticos sencillos y se registró el número de errores. Se obtuvieron los siguientes resultados.

Errores	8	6	6	10	8	14	14	12	16	12
Horas sin dormir	8	8	12	12	16	16	20	20	24	24

a. Calcule e interprete el coeficiente de correlación.

b. Calcule e interprete *a* y *b*.

c. Determine el número promedio de errores cuando se ha dejado de dormir por 4 horas.

d. Determine el número promedio de errores cuando se ha dejado de dormir por 15 horas.

13. Myers (2002) muestra los datos de un experimento con la mosca de la fruta. El autor analiza el efecto que tiene la concentración de nicotina sobre el porcentaje de muertes. Se sometieron siete grupos de moscas a siete concentraciones diferentes de nicotina, los resultados se muestran en la siguiente tabla.

Concentración de nicotina g/100cc	Moscas expuestas	Moscas muertas	Porcentaje de muertes
0,1	47	8	17,0%
0,2	53	14	26,4%
0,2	55	24	43,6%
0,3	52	32	61,5%
0,5	46	38	82,6%
0,7	54	50	92,6%
1,0	52	50	96,2%

Construya una regresión entre el porcentaje de muertes y la concentración de nicotina y explique los resultados.

14. Cuatro investigadores, por separado, tomaron once muestras y ajustaron un modelo de regresión lineal a las variables x, y.

Investigador 1		Investigador 2		Investigador 3		Investigador 4	
x 1	**y 1**	**x 2**	**y 2**	**x 3**	**y 3**	**x 4**	**y 4**
10	8,04	10	9,14	10	7,46	8,0	6,58
8	6,95	8	8,14	8	6,77	8,0	5,76
13	7,57	13	8,74	13	12,74	8,0	7,71
9	8,81	9	8,77	9	7,11	8,0	8,84
11	8,33	11	9,26	11	7,81	8,0	8,47
14	9,96	14	8,1	14	8,84	8,0	7,04
6	7,24	6	6,13	6	6,08	8,0	5,25
4	4,26	4	3,1	4	5,39	19,0	12,5
12	10,84	12	9,13	12	8,15	8,0	5,56
7	4,82	7	7,26	7	6,42	8,0	7,91
5	5,68	5	4,74	5	5,73	8,0	6,8

a. Determine cuál de los investigadores tiene un conjunto de datos con mayor correlación lineal.

b. Determine cuál de los investigadores tiene un conjunto de datos con menor correlación lineal.

15. La base de datos *Población mundial por país y sexo*, contiene la información del tamaño de la población total de 195 países discriminada por país y sexo en el año 2005. Con base en los datos responda:

  a. Porcentualmente, ¿en cuál país hay más mujeres?
  b. Existe alguna característica especial dentro de los países que porcentualmente tienen más y menos mujeres. ¿Puede dar una explicación a este hecho?
  c. Calcule e interprete el coeficiente de correlación entre el número de mujeres y hombres en los 195 países.

16. La base de datos *Avalúos de predios y áreas construidas,* consiste en una muestra aleatoria de 93 predios ubicados en la ciudad de Bogotá, de los cuales se ha registrado el área del terreno, el área construida, el avalúo del predio, el impuesto liquidado por la administración distrital y el valor pagado por el contribuyente propietario del predio. Con base en esta información:

  a. Use un programa de computador para construir e interpretar los gráficos de dispersión y las correlaciones de las variables.
  b. Encuentre la recta de regresión entre el área del terreno y el avalúo del predio.
  c. Use la recta de regresión para calcular el valor medio del avalúo de un predio cuya área es de 100 m^2.

17. Use la base de datos *Lácteos,* que contiene la información sobre grasas, proteínas, colesterol, vitaminas, etc., de cincuenta productos lácteos. Use un programa de computador para:

  a. Determinar las variables más correlacionadas.
  b. Calcular las variables más correlacionadas dentro del grupo de los quesos.
  c. Calcular las variables más correlacionadas dentro del grupo de las margarinas y las mantequillas.

CAPÍTULO 4
INTRODUCCIÓN A LA PROBABILIDAD

Experimentos aleatorios

Cuando una persona paga la prima para asegurar su vehículo contra robo o accidente, está previendo la posible ocurrencia de dicho suceso. La ocurrencia del robo de un automóvil depende de múltiples causas, por ejemplo:

- De la situación de seguridad en el lugar donde se moviliza.
- De las precauciones que tenga su dueño para cuidarlo.
- De los sistemas de seguridad que posea el vehículo, como por ejemplo, la alarma.

No obstante, un componente de la ocurrencia del robo de un vehículo es aleatoria, es decir, puede que suceda o que no suceda (depende del azar). Otro componente que puede influir en la ocurrencia de un robo es el descuido de su dueño (que lo deje en la calle o que no cierre las puertas, por ejemplo), lo que aumentaría las posibilidades de ser robado. El dueño podría guardarlo en su casa y no sacarlo nunca para disminuir las posibilidades de robo, pero aun así es posible que sea robado.

Si uno quiere saber qué tan posible es que roben un automóvil, puede recurrir a las estadísticas. Por ejemplo, suponga que los reportes de la policía indican que en el último mes han ocurrido doscientos robos de vehículos de una marca particular, en una ciudad donde circulan 20.000 automóviles. La proporción de carros robados de dicha marca, en el último mes es $\frac{200}{20.000} = \frac{1}{100} = 1\%$. Si esta frecuencia es estable a largo plazo, proporciona una medida aproximada e intuitiva de nuestra confianza en la ocurrencia de un evento de esta naturaleza, es decir, si los niveles de seguridad y otros factores permanecen más o menos iguales, se puede esperar que en el siguiente mes la posibilidad de que roben un carro, de la marca considerada, sea del 1%.

Las compañías de seguros usan los datos estadísticos para determinar la probabilidad o posibilidad de ocurrencia de muchos eventos como robo, accidente, incendio, terremoto, etc. Este tipo de cálculo permite determinar cuáles deben ser los pagos de las primas de los usuarios, para que a largo plazo el negocio represente utilidades.

A continuación, se estudiarán situaciones en las que existe un conjunto de posibles resultados y más adelante, se cuantifica la posibilidad o probabilidad de ocurrencia.

La manera como se recoge la información o los datos es mediante un **experimento**. Se entiende por experimento un proceso en que se obtiene exactamente una observación o resultado. Los resultados de un experimento pueden ser cualitativos o cuantitativos. Los siguientes son diferentes tipos de experimentos:

- Preguntar a una persona por su equipo de fútbol favorito.
- Contar el número de niños que ingresan por gripa a un servicio de urgencias en un día.
- Entrevistar a una persona para saber si está o no de acuerdo con una nueva ley.
- Lanzar un dado y observar el número de puntos en la cara superior.
- Contar el número de huecos en una avenida.
- Presentar un examen y obtener una calificación.
- Contar el número de personas que ven un comercial.

Cada resultado obtenido en una sola realización del experimento se denomina **evento simple**. A la unión de varios eventos simples se les denomina **evento compuesto** o simplemente *evento*. Por ejemplo, al lanzar un dado regular se obtienen seis eventos simples: 1, 2, 3, 4 ,5 y 6. Un evento compuesto puede ser aquel en el cual el resultado es número par. Este evento es compuesto y está formado por tres eventos simples: 2, 4 y 6.

Dos eventos son **mutuamente excluyentes** si cuando ocurre uno de ellos, el otro no puede ocurrir. Por ejemplo, suponga que se determina si una persona es menor o mayor de edad y si ha votado o no en las últimas elecciones. Si la persona es menor de edad, se excluye que haya votado y si ha votado, excluye que sea menor de edad, es decir, estos son excluyentes porque si uno ocurre, el otro no puede ocurrir.

Ejemplo 4.1

Enunciado

Se registra el tiempo de un atleta al recorrer un trayecto. Enunciar algunos eventos de este experimento.

Solución

Evento *B:* = registrar un tiempo superior a 60 segundos.
Evento *B:* = registrar un tiempo menor o igual a 60 segundos.
Evento *C:* = registrar entre 20 y 40 segundos.

Si una persona recorre el trayecto y registra un tiempo de 33 segundos, decimos que ocurrió *B* y también *C*, pero que no ocurrió *A*. Advierta que si *A* ocurre no puede ocurrir *B*, es decir, *A* y *B* son mutuamente excluyentes.▪

Ejemplo 4.2

Enunciado

Se le pregunta a un hombre y a una mujer si están de acuerdo con un posible ataque de Estados Unidos a Irán. Enunciar algunos eventos de este experimento.

Solución

E_1: el hombre y la mujer están de acuerdo.
E_2: el hombre está de acuerdo y la mujer no.
E_3: la mujer está de acuerdo y el hombre no.
E_4: hombre y mujer no están de acuerdo.
E_5: el hombre está de acuerdo.
E_6: la mujer está de acuerdo.

Note que:

E_5 y E_6 no son mutuamente excluyentes, porque hombre y mujer pueden estar de acuerdo.
E_1, E_2, E_3 y E_4 forman un conjunto de eventos mutuamente excluyentes porque, por ejemplo, si ocurre E_1 no puede ocurrir E_2.
Cuando ocurre E_1 o E_2 tambien ocurre E_5. Es decir que E_5 se puede descomponer en eventos más simples E_1 y E_2.
Cuando ocurre E_1 o E_3 también ocurren E_6. Es decir que E_6 se puede descomponer en eventos más simples E_1 y E_3.
Los eventos E_1, E_2, E_3 y E_4 se llaman simples porque no se pueden descomponer.▪

Un **evento** es uno de los posibles resultados de un experimento. Cada resultado obtenido en una sola realización del experimento se denomina **evento simple**.
Si *A* es un evento, decimos que *A* ocurre si el resultado del experimento pertenece a *A*.

Espacio muestral

Generalmente, en los experimentos, los resultados no se pueden predecir con absoluta seguridad, aunque sí se conozcan cuáles son todos los posibles resultados. El **espacio muestral** de un experimento es el conjunto de todos sus eventos simples y se denota *S*. En ocasiones, determinar el espacio muestral de un experimento es muy sencillo, por ejemplo, cuando se lanza un dado *S*= {1, 2, 3, 4, 5, 6}, pero, a veces, puede resultar más elaborado, por ejemplo cuando se sacan tres cartas de una baraja.

Espacio muestral
Es el conjunto de todos los eventos simples de un experimento.
El espacio muestral se denota con la letra *S*

Ejemplo 4.3

Enunciado

Encuentre el espacio muestral de los siguientes experimentos:

a. Lanzamiento de una moneda dos veces.
b. Preguntarle a una persona si participará en unas votaciones para la elección de un presidente.

Solución

a. Los resultados posibles son:

Cara en los dos lanzamientos: CC.
Cara en el primer lanzamiento y sello en el segundo: CS.
Sello en el primer lanzamiento y cara en el segundo: SC.
Sello en los dos lanzamientos: SS.
De esta manera el espacio muestral del experimento es: {*CC, CS, SC, SS*}.

b. A la pregunta, la persona podría responder:

S: “Sí voy a participar”.
N: “No voy a participar”.
NS: “No sé o no he decidido participar”.

Es más, la persona podría negarse a responder, en tal caso el espacio muestral para el experimento es S = *{S, N, NS, NR}* donde NR significa no responde.▪

Diagrama de árbol

Cuando un experimento es más complejo, determinar o contar los eventos simples puede ser complicado. Será útil pensar en un método sistemático para obtener todos sus eventos simples. Uno de estos métodos es el **diagrama de árbol**, que resulta útil cuando el experimento se realiza en un conjunto pequeño de etapas o pasos diferentes.

El árbol se construye determinando el número de etapas o pasos en el experimento. Cada etapa del experimento representa una ramificación del árbol. En cada etapa, las ramas del árbol representan las posibilidades en ese punto. Para leer el árbol se siguen las secuencias de sucesos, llamadas trayectorias.

Ejemplo 4.4

Enunciado

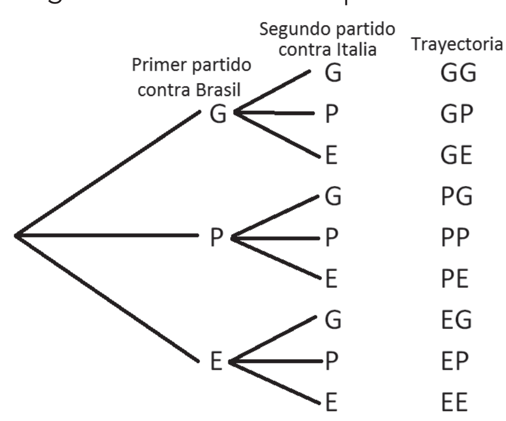
Los dos siguientes partidos la selección alemana de fútbol, los jugará contra Brasil e Italia. Determine todas posibilidades si en cada partido los resultados posibles son ganar, empatar o perder.

Solución

En cada partido en el cual Alemania juega, solo puede ocurrir uno de los siguientes eventos:

G: Alemania gana.
E: Alemania empata.
P: Alemania pierde.

Figura 4.1 Diagrama de árbol de los partidos de fútbol



En la primera etapa, primer partido, hay tres posibilidades, representadas por las tres primeras ramificaciones del árbol. En la segunda etapa, segundo partido, cada una de las tres ramificaciones se divide en tres, dando como resultado nueve trayectorias o resultados posibles.

$$S = \{GG, GP, GE, EG, EE, EP, PG, PE, PP\} \quad ▪$$

Ejemplo 4.5

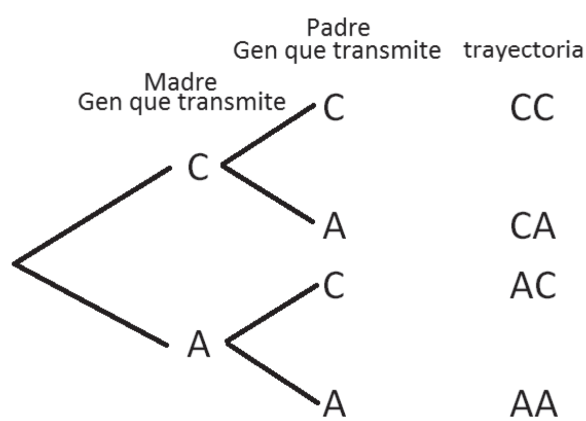
Enunciado

El color de los ojos de una persona se determina por dos genes. El gen de ojos color café C, es dominante sobre el gen de ojos azules A. Esto significa que una persona con dos genes para ojos azules tiene ojos azules (AA); mientras que alguien con dos genes de ojos color café (CC), o uno de ojos color café y otro de ojos azules (CA o AC), tiene ojos color café. Los hijos reciben de sus padres dos genes, uno del padre y otro de la madre, ambos de manera aleatoria. Determine las posibilidades para los genes de un hijo para quien ambos padres tienen un gen de ojos color café y uno de ojos azules.

Solución

Considere la situación como un experimento en dos etapas. En la primera rama del árbol de la figura 4.2 se muestran las posibilidades para el gen que aporta la madre y de cada una de ellas se desprenden las posibilidades para el gen que aporta el padre.

Figura 4.2 Diagrama de árbol genes



Hay cuatro resultados posibles en el espacio muestral *S= {CC, CA, AC, AA}*. Cada uno de sus elementos corresponde al par de genes transmitidos por el padre y la madre al hijo. Si padre y madre tienen un gen de ojos color cafés y uno de ojos azules, solo en uno de los cuatro eventos del espacio muestral, el niño tendrá ojos azules: AA. En los demás, por ser dominante el gen de ojos color café, el hijo tendrá ojos color café: CC, CA y AC. Es decir, es más probable que una pareja como esta tenga un hijo de ojos color café que uno de ojos azules.▪

Ejercicio 4.1 Experimentos aleatorios

1. Liste cada uno de los eventos simples del espacio muestral de los siguientes experimentos.
   a. Un jugador de fútbol cobra dos penaltis.
   b. Preguntar a una persona por el color de sus ojos.
   c. Lanzar dos monedas.
   d. Presentar un examen de matemáticas.

2. Se lanza un dado y se determinan los siguientes eventos.

   A: cae número dos.
   B: cae un número impar.
   C: cae un número par.
   D: cae un número menor que 3.
   E: ocurre A o B.
   F: ocurre B y D.

   a. Liste los eventos simples del experimento.
   b. Liste cada uno de los eventos simples de los eventos A al F.
   c. ¿A y B son mutuamente excluyentes?
   d. ¿C y D son mutuamente excluyentes?
   e. ¿Qué posibilidad de ocurrencia le daría a cada uno de los eventos simples?
   f. ¿Qué posibilidad de ocurrencia le daría a los otros eventos?

3. Suponga que en un nacimiento la probabilidad de ser varón o mujer es la misma. Una familia tiene tres hijos y con respecto al sexo de estos se definen los siguientes eventos:

   A: todos son varones.
   B: todos son mujeres.
   C: por lo menos uno es varón.
   D: por lo menos uno es mujer.
   E: exactamente dos son varones.
   F: exactamente dos son mujeres.

a. Determine todas las posibilidades de sexo de los tres hijos de la familia, teniendo en cuenta el orden de nacimiento de estos.
b. Liste cada uno de los eventos simples de los eventos A al F.
c. ¿A y C son mutuamente excluyentes?
d. ¿ E y F son mutuamente excluyentes?
e. ¿Qué posibilidad de ocurrencia le daría a cada uno de los eventos simples?
f. ¿Qué posibilidad de ocurrencia le daría a los otros eventos?

4. Utilice un diagrama de árbol para encontrar el espacio muestral de los siguientes experimentos:
a. Lanzamiento de una moneda tres veces.
b. Lanzamiento de una moneda y un dado.
c. Lanzamiento de dos dados.

5. Samuel y Daniela jugarán entre ellos ping-pong. El ganador será el primero que complete tres victorias en un máximo de cinco partidos. Sin considerar empates ¿de cuántas maneras diferentes puede finalizar el juego?

6. Un jurado está compuesto por cuatro personas. Cada uno de ellos hará una afirmación sobre el acusado: culpable o inocente. ¿De cuántas maneras pueden responder los jurados? ¿En cuántas de estas posibilidades hay empate en la decisión?

7. A un catador se le sirven tres copas de vino de marcas A, B y C, respectivamente. Él debe jerarquizar la calidad de los tres vinos de mayor a menor. ¿De cuántas maneras puede hacer esto?

Técnicas de conteo

Chance es una palabra que designa la oportunidad o posibilidad de conseguir algo. También, es el nombre que se le ha dado a un tipo de juego de azar, en el que se apuesta una cantidad de dinero a un número de cuatro dígitos (del 0000 al 9999) con la esperanza de recibir una cantidad proporcional al dinero arriesgado.

En el sorteo hay 10.000 resultados posibles, pero para el apostador solo dos resultados son de interés: ganar o perder. Ganar es un evento simple; su único elemento es el número ganador, mientras que perder es un evento compuesto formado por los 9999 números restantes. En estas circunstancias, decimos informalmente que el jugador tiene un chance de uno en diez mil o que la probabilidad de ganar es de una en diez mil.

Contar los eventos simples del juego del chance es fácil: diez mil números entre 0000 y 9999, pero, no siempre es sencillo determinar el número de resultados posibles de un experimento.

Por ejemplo: ¿de cuántas maneras diferentes puede caer el baloto? (el baloto consiste en apostar una cantidad fija, a seis números entre 45 posibles del 01 al 45 y el ganador será aquel que acierte con los seis números sin importar el orden de escogencia). ¿De cuántas maneras posibles puede resultar un campeonato entre ocho equipos? ¿De cuántas maneras posibles se puede elegir un capitán y su suplente de un equipo con once jugadores?

El conocimiento del número de eventos simples de un espacio muestral o un evento es importante para cuantificar la probabilidad de ocurrencia de eventos. En esta sección del capítulo se utilizarán reglas que permiten contar los eventos simples de un espacio muestral o un evento para responder a las preguntas planteadas y a otras que surgirán después.

Regla de la multiplicación

Si un experimento se puede realizar en dos etapas y la primera etapa se consigue de m maneras y para cada una de estas, la segunda de n maneras, entonces, el experimento completo se puede realizar de mn maneras diferentes. Considere el siguiente ejemplo.

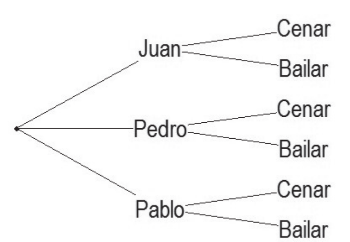
Ejemplo 4.6

Enunciado

Magdalena desea salir el sábado en la tarde y puede escoger entre uno de tres acompañantes: Juan, Pedro y Pablo. Adicionalmente quiere realizar una de las dos siguientes actividades: ir a cenar o a bailar. Considerando las opciones mencionadas, ¿de cuántas maneras diferentes puede escoger?

Solución

Con Juan podría ir a cenar o a bailar; dos posibilidades, e igual con cada uno de los otros dos acompañantes, lo que hace un total de seis posibilidades.



El anterior ejemplo se puede considerar como un experimento en dos etapas: la primera, elección de acompañante y, la segunda, elección de actividad. Las maneras en que se puede dar el experimento es el producto de las opciones en cada etapa, es decir, $3 \times 2 = 6$ resultados posibles. A este método de conteo de eventos simples se le llama **regla de la multiplicación**.▪

Ejemplo 4.7

Enunciado

Se le pide a Juanes que en una presentación interprete dos canciones diferentes de su álbum *Mi sangre*. Si el álbum tiene doce canciones, ¿de cuántas maneras diferentes puede escoger las canciones? Si se le permitiera repetir la interpretación, ¿de cuántas maneras diferentes podría escoger?

Solución

Este es un experimento que se puede realizar en dos etapas, cada etapa consiste en la elección de una canción. Para la primera canción, puede elegir entre cualquiera de las $m = 12$ del álbum. Si las dos canciones que interprete deben ser diferentes, para la segunda tiene $n = 11$ opciones. El número total de eventos simples es:

$mn = (12)(11) = 132$

Si pudiera repetir las canciones, para la primera tendría $m = 12$ posibilidades y para la segunda $n = 12$ posibilidades. En este caso, el número total de eventos simple es:

$mn = (12)(12) = 144$

La diferencia entre los dos conteos es $144 - 132 = 12$ que son las interpretaciones en las que se repite canción.▪

Ejemplo 4.8

Enunciado

¿Cuántas placas diferentes (de vehículo) se pueden construir utilizando tres letras del alfabeto y tres dígitos?

Solución

El experimento se puede realizar en seis etapas, en cada una de las tres primeras etapas es posible escoger entre veintiséis letras y en cada una de las últimas tres etapas es posible escoger entre diez dígitos. Esto es, $26\times26\times26\times10\times10\times10=17.576.000$ maneras diferentes de construir placas con tres letras y tres dígitos.▪

Regla de la multiplicación en dos etapas

Suponga que un experimento se puede realizar en dos etapas. Si la primera etapa se puede realizar de m maneras y la segunda etapa se puede realizar de n maneras, entonces el experimento total se puede realizar de mn maneras.

Regla de la multiplicación en k etapas

Suponga que un experimento se puede realizar en k etapas. Si la primera etapa se puede realizar de n_1 maneras, la segunda de $n_2,\cdots$, la k-ésima de n_k maneras, entonces el experimento total se puede realizar de $n_1\, n_2\cdots n_k$ maneras.

Ejercicio 4.2.1 Reglas de la multiplicación

1. Encuentre el número de resultados posibles en los siguientes experimentos
   a. Lanzamiento de una moneda cinco veces.
   b. Lanzamiento de una moneda y un dado.
   c. Extracción de dos cartas de una baraja de póker.

2. Si en el menú de un restaurante una persona puede elegir entre cinco entradas, cuatro ensaladas y siete postres, ¿cuántos tipos de menús se pueden elegir si cada uno consta de una entrada, una ensalada y un postre?

3. Tres ciudades A, B y C están localizadas geográficamente, formando un triángulo. Para desplazarse entre A y C existen cuatro caminos, entre A y B dos y entre B y C tres. Un turista está en una de estas ciudades y desea pasar por las otras dos. Si el turista desea pasar una vez por cada ciudad, ¿cuántas trayectorias posibles puede escoger? ¿Cuántas si adicionalmente debe escoger la ciudad de partida?

4. Un examen consta de diez preguntas de selección múltiple con cinco opciones cada una. ¿De cuántas formas diferentes se puede contestar el examen?

5. En el menú de llamadas entrantes de un teléfono celular se puede seleccionar tono (de diez diferentes o desactivado), vibración (activada o desactivada) y luces (activadas

o desactivadas). ¿De cuántas maneras diferentes se puede personalizar la entrada de una llamada?

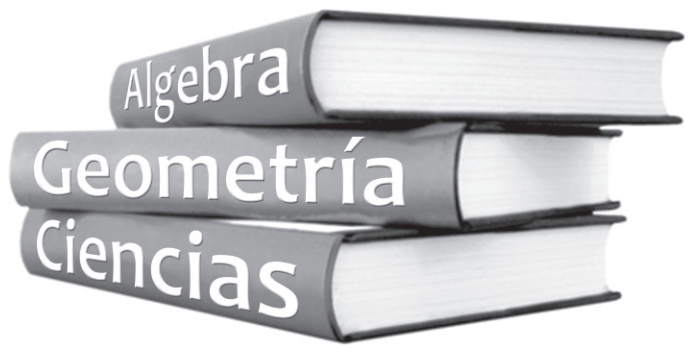
Permutaciones

En muchos experimentos aleatorios es importante contar las ordenaciones posibles de los elementos de un conjunto finito. A cada una de estas ordenaciones se les llama permutación, para explicarlo con detalle considere el siguiente ejemplo.

Ejemplo 4.9

Enunciado

¿De cuántas maneras diferentes es posible organizar tres libros (álgebra, geometría y ciencias) uno encima de otro?



Solución

Si se denota cada libro con su inicial, las posibles ordenaciones de los libros son:

AGC ACG GAC GCA CAG CGA

A cada uno de estos seis arreglos lo llamamos una **permutación**.▪

Permutación
Una permutación es un arreglo o distribución de objetos en un orden determinado.

Ejemplo 4.10

Enunciado

Determine el número de permutaciones de las letras de la palabra *amor*.

Solución

Esta palabra tiene cuatro letras diferentes. Si se comienza a armar la palabra letra por letra, hay cuatro posibilidades para la primera letra, tres para la segunda, dos para la tercera y una para la cuarta. Este es un experimento en cuatro etapas, por tanto, existen $4\times3\times2\times1 = 24$ permutaciones de sus letras. Para simplificar la escritura del producto $4\times3\times2\times1$ se escribe $4!$ y se lee 4 factorial.▪

Ejemplo 4.11

Enunciado

De cuántas maneras diferentes se puede organizar una fila con diez niños.

Solución

Para escoger al primer niño, hay diez posibilidades, al segundo, nueve posibilidades, …, al último, una posibilidad. Por tanto, hay $10\times9\times8\times7\times6\times5\times4\times3\times2\times1 = 10! = 3.628.800$ maneras.▪

Factorial

La escritura del producto $1\times2\times3\times\cdots\times n$ puede ser simplificada con la expresión n! que se lee n factorial. Por definición $0! = 1$.

n! es el número de permutaciones de *n* objetos diferentes.

Ejemplo 4.12

Enunciado

Calcular 3!, 5! y 7!

Solución

$3! = 1\cdot2\cdot3 = 6$

$5! = 1\cdot2\cdot3\cdot4\cdot5 = 120$

$7! = 1\cdot2\cdot3\cdot4\cdot5\cdot6\cdot7 = 5.040$ ▪

Ahora se calcula el número de permutaciones de *n* objetos tomados en grupos de *r* a la vez.

Ejemplo 4.13

Enunciado

¿De cuántas maneras es posible organizar dos libros tomados de un grupo de cinco libros?

Primera solución

Suponga que se cuenta con cinco libros A, B, C, D y E, de los cuales se eligen dos. Considere que este es un experimento en dos etapas. En la primera etapa, se seleccionan cinco de los libros y en la segunda cuatro —pues ya se ha escogido uno—. Por tanto, el número de permutaciones u ordenaciones de los cinco libros en grupos de dos a la vez 5×4 = 20.

Segunda solución

Las permutaciones de cinco objetos son 5! = 120. Es decir, hay 120 maneras diferentes de ordenar los cinco libros. Pero, el ejercicio consiste en seleccionar dos de los cinco libros; para ello, considere solo los dos primeros libros de estas 120 permutaciones. Desde este punto de vista, muchas de las ordenaciones coincidirán, por ejemplo:

ABCDE = ABCED = ABDCE = ABDEC = ABECD = ABEDC

Todos estos arreglos cuentan como una sola permutación, porque solo interesan los dos primeros libros. Por tanto, hay arreglos de seis libros que cuenta como una sola permutación —considerando solo los dos primeros libros—. Son seis, porque corresponden a las permutaciones de los tres elementos no considerados en cada arreglo: 3! = 6.

$$AB\,\underbrace{CDE} = AB\,\underbrace{CED} = AB\,\underbrace{DCE} = AB\,\underbrace{DEC} = AB\,\underbrace{ECD} = AB\,\underbrace{EDC}$$

Permutaciones de las letras CDE

Para encontrar el número de permutaciones de cinco libros, tomados de dos a la vez, se tienen en cuenta las 120 permutaciones de los cinco libros, dividido entre las seis permutaciones de los tres libros que cuentan como la misma. Esto es,120 ÷ 6 =20.▪

Si, en general, interesa contar las permutaciones de *n* objetos tomados de *r* a la vez, se debe tener en cuenta lo siguiente:

1. *n!* permutaciones de *n* elementos.
2. *(n − r)!* permutaciones de los *n*−*r* objetos no considerados.

Con ello, se sabe que las permutaciones de *n* objetos tomados de *r,* a la vez, es *n!* dividido entre *(n − r)!.*

Conteo de permutaciones

El número de permutaciones *nPr* de n objetos distintos tomados de *r* a la vez es:

$$n\,\mathrm{Pr} = \frac{n!}{(n-r)!}$$

Ejemplo 4.14

Enunciado

Un técnico debe armar un equipo con cuatro jugadores de un grupo de diez. Cada jugador puede tener dos roles diferentes: ataque o defensa y dos posiciones: izquierda o derecha. ¿De cuántas maneras diferentes puede armar el equipo?

Solución

Es necesario calcular las permutaciones de diez jugadores tomados de cuatro a la vez, es decir: 10P4:

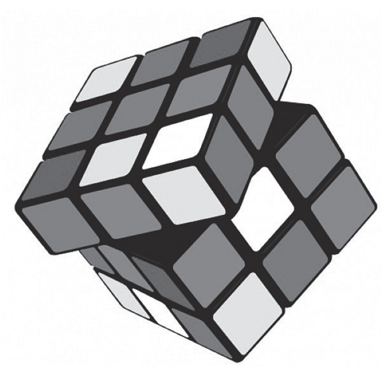
$$10P4 = \frac{10!}{(10-4)!} = \frac{10!}{6!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6!}{6!} = 10 \cdot 9 \cdot 8 \cdot 7 = 5.040 \quad ▪$$

Ejemplo 4.15

Enunciado

Encontrar el número de maneras en que puede organizarse un cubo de Rubik.

Solución



Un cubo de Rubik tiene ocho vértices, por lo tanto hay 8! maneras de ubicarlos, también hay doce aristas y 12! maneras de ubicarlas. Esto hace 12! × 8! posibilidades, pero cada combinación vértice-arista ha sido contada dos veces, por lo que en realidad solo son $\frac{12! \times 8!}{2}$ posibles posiciones. Por otro lado, la ubicación de los vértices debe tener uno de ellos como referencia, esto determina 3^7 posiblidades. De igual manera ocurre con las aristas lo que conduce a 2^{11} posiblidades. Por lo tanto las diferentes maneras de organizar el cubo de Rubik son:

$$\frac{8! \times 12! \times 3^7 \times 2^{11}}{2} = 43.252.003.274.489.856.000 \quad ▪$$

Ejercicio 4.2.2 Permutaciones

1. Resuelva las siguientes operaciones con factoriales:

   a. $3! \times 4!$ b. $\frac{5!}{4!}$ c. $\frac{9!}{10!}$ d. $\frac{3! \times 5!}{(3!)^2}$

2. Calcule e interprete las siguientes permutaciones:

   a. $5P3$ b. $5P4$ c. $6P1$ d. $3P2$ e. $10P5$

3. ¿De cuántas maneras se pueden elegir cuatro personas de un grupo de ocho, si es importante el orden en que se elijan?

4. Ambrosio, Berta, Carlos, Diana y Esteban participarán en una carrera. Si no se consideran empates, ¿cuántas son las diferentes posibilidades para primer y segundo lugar?

5. En un concurso se escogerán primero, segundo y tercer puesto de las tres vitrinas más bonitas de la ciudad de un grupo de cien inscritas. ¿De cuántas maneras diferentes es posible hacer esta elección?

Combinaciones

En el conteo de permutaciones interesa el orden de los objetos. No siempre el orden es importante. Por ejemplo, si una persona escoge dos sabores de helado entre cinco posibles, es igual que escoja vainilla y chocolate que chocolate y vainilla. En general, una situación de conteo en la que el orden no es importante se le denomina *combinación.*

Características de las permutaciones y combinaciones

En el conteo de permutaciones el orden es importante.

En el conteo de combinaciones el orden no es importante.

Combinación

Una combinación de *n* objetos tomados de r a la vez es cualquier arreglo en el que no se tiene en cuenta el orden.

Ejemplo 4.16

Enunciado

Encontrar las combinaciones de las letras A, B, C y D tomadas en grupos de tres.

Solución

Antes de encontrar las combinaciones, primero se calculan las permutaciones de cuatro elementos tomadas de tres, esto es $4P3 = \frac{4!}{1!} = 24$. Hay veinticuatro permutaciones de cuatro letras tomadas en grupos de tres:

ABC	**ABD**	**BCD**	**ACD**
ACB	ADB	BDC	ADC
BAC	BAD	CBD	CAD
BCA	BDA	CDB	CDA
CAB	DAB	DBC	DAC
CBA	DBA	DCB	DCA

Como en las combinaciones el orden no es importante, los seis elementos de la misma columna son el mismo y pueden ser representados por cualquiera de ellos. Las combinaciones de cuatro letras tomadas de tres en tres son solo cuatro:

ABC, ABD, BCD y ACD▪

Hay más permutaciones que combinaciones; en el caso anterior, exactamente 3! = 6 permutaciones por cada combinación. De esta manera, se puede encontrar el número de combinaciones de *n* objetos tomados de *r*, a la vez, denotado *nCr* o $\binom{n}{r}$, calculando primero el número de permutaciones *nPr* dividido entre las posibles permutaciones de sus *r* elementos, es decir, *nPr* dividido entre *r*!:

$$\binom{n}{r} = \frac{nPr}{r!} \neq \frac{\frac{n!}{(n-r)}}{r!} = \frac{n!}{(n-r)!r!}$$

Combinaciones

El número de combinaciones de $\binom{n}{r}$ objetos tomados de *r* a la vez es:

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

Ejemplo 4.17

Enunciado

Encontrar el número de comités diferentes de tres personas que se pueden formar de un grupo de cinco personas.

Solución

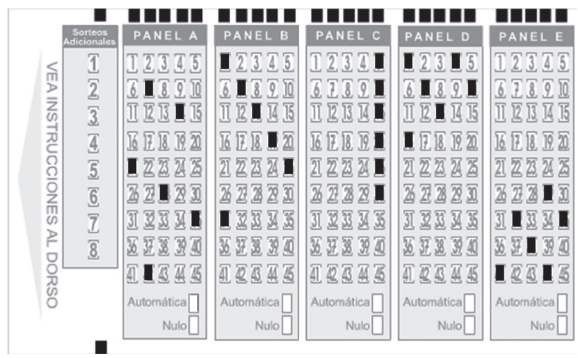
Si el orden en que se elijan los integrantes del comité no es importante, calculamos las combinaciones de tres personas tomadas de un grupo de cinco, esto es, $\binom{5}{3}$.

$$\binom{5}{3} = \frac{5!}{(5-3)!3!} = 10 \quad \blacksquare$$

Ejemplo 4.18

Enunciado

Encontrar el número de resultados posibles del baloto, el cual consiste en apostar una cantidad fija a seis números entre 45 posibles del 01 al 45. El ganador será aquel que acierte los seis números sin importar el orden de escogencia.



Solución

Como el orden de escogencia de los números no es importante, es necesario calcular las combinaciones de 45 números tomados de 6 a la vez, es decir $\binom{45}{6}$.

$$\binom{45}{6} = \frac{45!}{(45-6)!6!} = \frac{45!}{39!6!} = 8'145.060$$

Esto significa que si una persona apuesta una sola vez tendrá una oportunidad entre 8.145.060 para ganar el premio.▪

Ejercicio 4.2.3 Combinaciones y conteo

1. Calcule e interprete las siguientes combinaciones:

   a. $\binom{4}{3}$ b. $\binom{4}{2}$ c. $\binom{6}{2}$ d. $\binom{11}{11}$ e. $\binom{11}{1}$

2. Se desea conformar un comité de tres personas que representen a un grupo de diez, ¿de cuántas maneras diferentes es posible hacer esta elección?

3. Una tienda de ropa otorga un premio a uno de sus clientes que consiste en regalarle tres camisas y dos pantalones. Considerando que la tienda cuenta con diez marcas di-

ferentes de camisas y cinco marcas diferentes de pantalones y que el cliente no puede repetir marca de camisa o pantalón, ¿de cuántas maneras puede elegir el premio?

4. Resuelva las siguientes operaciones con factoriales:

a. $\frac{5!}{1!4!}$ b. $\frac{10!}{7!}$ c. $\frac{69!}{70!}$ d. $\frac{3!\times 3!\times 3!}{2!\times 2!\times 2!}$

5. Evalúe y diga el significado de las siguientes permutaciones:

a. 6P1 b. 6P5 c. 6P6 d. 7P3

6. Evalúe y diga el significado de las siguientes combinaciones:

a. $\binom{5}{1}$ b. $\binom{5}{4}$ c. $\binom{5}{5}$ d. $\binom{10}{8}$

7. Se tienen dos grupos, uno de diez hombres y otro de quince mujeres. ¿De cuántas maneras diferentes puede seleccionar una pareja conformada por un hombre y una mujer? ¿Cuántas si pueden ser del mismo sexo?

8. En un campeonato de fútbol se inscribieron diez equipos. Si en la primera ronda se juega todos contra todos, ¿cuántos partidos se deben jugar?

9. Los conos de helado de una tienda se venden con una o dos bolitas. Si se dispone de ocho sabores de helado y cuatro tipos de cono, ¿de cuántas maneras diferentes es posible armar un cono con una bolita? ¿Cuántos con dos bolitas? ¿Importa el orden de las bolitas?

10. Si no se consideran empates ¿de cuántas maneras diferentes es posible que termine una carrera de diez automóviles?

11. En un comité internacional hay tres cubanos, cuatro franceses y cinco libios. ¿De cuántas maneras diferentes se puede conformar un comité de seis personas, dos de cada país?

12. ¿Cuántas palabras diferentes (sin sentido) se pueden formar con las letras de la palabra amor? ¿Cuántas con la palabra loco?

13. Un hombre tiene cuatro llaves en su llavero una de las cuales es la de su casa. En la noche, en la oscuridad, intenta abrir la puerta seleccionando las llaves en forma aleatoria y descartando las ya probadas. ¿De cuántas maneras diferentes podría abrir la puerta?

14. En un colegio de mil estudiantes se elegirá a un personero y un suplente. ¿De cuántas maneras diferentes es posible hacer esta elección?

15. Para atender un restaurante se necesitan ocho empleados: cinco para atender las mesas, dos para atender la caja y uno para recibir a los clientes a la entrada. Hay tres tareas: mesero, cajero y recepcionista, ¿de cuántas maneras diferentes se les pueden asignar las tareas a ocho empleados?

16. Un grupo de exploradores está dividido en tres patrullas: Águilas, Halcones y Tigres. En cada patrulla hay seis, cinco y cuatro niños, respectivamente. ¿De cuántas maneras diferentes se pueden organizar en una fila los quince niños si cada patrulla debe permanecer junta?

17. De cuántas maneras pueden sentarse en una fila de ocho sillas (en línea recta) cuatro niños y cuatro niñas si:
   a. Pueden sentarse en cualquier orden.
   b. Alternándose niños y niñas

18. De un grupo de doce colombianos y quince españoles, ¿cuántos comités de tres colombianos y dos españoles es posible formar?

19. Un grupo de amigos se saluda con un apretón de manos, todos con todos. Si en total hay 105 apretones de manos, ¿cuántos amigos hay?

Conceptos básicos de probabilidad

Una de las bases teóricas de la estadística es la probabilidad. Una de las primeras motivaciones para la formulación de lo que en la actualidad se conoce como probabilidad lo constituyó el estudio de los juegos de azar. Muchos jugadores encomendaron a matemáticos muy importantes, entre ellos Bernoulli, Laplace y Pascal estrategias para ganar en los juegos de azar. De estos estudios y de otros aportes de matemáticos, así como de otros científicos se derivó la probabilidad. Posteriormente, los estadísticos han usado estos resultados para explicar diversos fenómenos que de alguna manera dependen del azar.

A continuación, se presenta un ejemplo sobre un juego de azar. Con base en este se estudian varios conceptos de probabilidad.

Ejemplo 4.19

Enunciado

En el juego de un casino, se puede apostar a un número entre uno y seis por un valor de $1.000. Luego, se lanzan tres dados. Si el número apostado cae en los tres dados se reciben $3.000 (mil por cada dado); si cae en dos de los tres dados, se reciben $2.000; si cae en uno solo de los tres dados se reciben $1.000 y si no cae en ninguno no se recibe nada. ¿Cuáles son las posibilidades de ganar dinero en este juego? ¿Qué es más probable, ganar o perder?

Solución

Primero, veamos cuáles son los posibles resultados en este juego:

Si el número apostado cae tres veces se reciben $3.000, pero como ya había apostado $1.000, se ganan $2.000.

Si el número apostado cae dos veces, se reciben $2.000, es decir, se ganan $1.000.

Si el número apostado cae una sola vez, se reciben $1.000, es decir, que no se gana nada.

Si el número apostado no cae en ningún dado se pierden $1.000.

Los posibles resultados del juego son cuatro y se pueden resumir así:

Ganar \$2.000
Ganar \$1.000
Ganar \$0
Perder \$1.000

¿Todos estos resultados son igualmente posibles? ¿Uno de estos resultados ocurrirá con más frecuencia? ¿Cómo medir la posibilidad de ocurrencia de cada uno de estos eventos? En la siguiente tabla se intentan esquematizar tanto los resultados posibles del juego como una medida numérica para referirnos a la posibilidad de ocurrencia de cada uno de ellos.

Resultado (ganancia)	2.000	1.000	0	-1.000
Posibilidad de ocurrencia	?	?	?	?

Se utilizará el principio fundamental de conteo para calcular las posibilidades de ocurrencia. Suponga que apostamos al entero k, con $1 \leq k \leq 6$. Por el principio, de la multiplicación, los tres dados pueden caer de $6 \cdot 6 \cdot 6 = 216$ maneras diferentes.

Posibilidad de ganar \$2.000: para que esto ocurra el número k debe aparecer en los tres dados. De estas solo una de ellas tiene el número k tres veces. Por tanto la posibilidad de ganar \$2.000 es de 1 entre 216 posibles.

Posibilidad de ganar \$1.000: en este caso el número k debe aparecer dos veces. Por ejemplo si $k = 1$, se tendrían resultados como $11X$, $1X1$, $X11$ (donde la X significa otro de los cinco números diferentes a 1). Por el principio de multiplicación, esto se puede dar de $3 \cdot (1 \cdot 1 \cdot 5) = 3 \cdot 5 = 15$ maneras diferentes. Por tanto, la posibilidad de ganar \$1.000 es de15 entre 216 posibles.

Posibilidad de ganar \$0: para que esto suceda, el número k debe aparecer una vez y puede aparecer en cualquiera de los tres dados. Por el principio de multiplicación, esto puede suceder de $3 \cdot 1 \cdot 5 \cdot 5 = 75$ maneras diferentes. Por tanto, la posibilidad de ganar \$0 es de 75 entre 216.

Posibilidad de perder \$1.000 (ganar –\$1.000): esto sucederá si el número k no aparece en ninguno de los dados. Otra vez, por el principio de multiplicación, esto puede suceder de $5 \cdot 5 \cdot 5 = 125$ maneras diferentes. Por tanto, la posibilidad de perder \$1.000 es de 125 entre 216.

Note que el conteo anterior es $1 + 15 + 75 + 125$ que corresponde a las 216 posibilidades al lanzar los tres dados. En la siguiente tabla, resumimos los resultados anteriores:

Resultado (ganancia)	2.000	1.000	0	-1.000
Posibilidad de ocurrencia	$\frac{1}{216}$	$\frac{15}{216}$	$\frac{75}{216}$	$\frac{125}{216}$

Después de este análisis, respondamos las preguntas que hicimos al comienzo de la sección:

¿Cuáles son las posibilidades de ganar dinero en este juego? Con dos de los resultados posibles se gana dinero, cuando el número al que se apuesta sale dos o tres veces. En estos casos, se ganan $1.000 o $2.000, respectivamente. Esto ocurre de $1+15=16$ maneras diferentes de 216 posibles.

¿Qué es más probable, ganar o perder? Ya se sabe que la posibilidad de ganar es de 16 entre 216, mientras que la de perder es de 125 entre 216. Por tanto, es más posible perder que ganar.▪

Probabilidad

En el anterior ejemplo se usó el término probabilidad. Una definición informal de probabilidad es la siguiente:

Probabilidad
La probabilidad de un evento es una cuantificación de nuestra creencia sobre la ocurrencia de dicho evento.

Ahora bien, ¿cómo cuantificar la probabilidad? Una manera de hacerlo es mediante el concepto de frecuencia relativa. Supongamos que un experimento se realiza *n* veces y nos fijamos en la frecuencia en que ocurre un evento A. La frecuencia relativa del evento se define como:

$$Frecuencia\,relativa = \frac{frecuencia}{n}$$

La probabilidad del evento A es esta frecuencia pero cuando el experimento se ha repetido un número suficiente de veces, esto es:

$$Probabilidad\,del\,evento\,A = P(A) = \lim_{n\to\infty}\frac{frecuencia}{n}$$

Ejemplo 4.20

Enunciado

Dos jugadores A y B de ajedrez se han enfrentado 35 veces, de estas han empatado cinco veces y A ha ganado solo siete partidas. Si se enfrentaran de nuevo, ¿cuál es la probabilidad de que A gane?

Solución

Es posible que sean varios los factores que influyan sobre el resultado de una partida de ajedrez (la estrategia utilizada por cada uno, las horas de práctica, el estado de ánimo, etc.). Pero, si solo consideramos los resultados anteriores, la probabilidad de que A gane es de 7 sobre 35:

$$\frac{7}{35} = \frac{1}{5} \blacksquare$$

La asignación del valor de probabilidad, por la frecuencia relativa, no siempre es posible. Por ejemplo, considere el experimento del lanzamiento de una moneda. Si se usa la frecuencia relativa para calcular la probabilidad de que el resultado sea cara, se debería lanzar la moneda un número considerable de veces y registrar la frecuencia de caras, con respecto al total de lanzamientos. Si el número de lanzamientos es muy grande y la moneda está balanceada, la frecuencia de caras sobre el número de lanzamientos tenderá a $\frac{1}{2}$. Lanzar una moneda infinitas veces resulta imposible. Por eso, es más razonable pensar que cada uno de los dos eventos posibles (cara y sello) son igualmente probables, es decir, la probabilidad de cara es $\frac{1}{2}$. A este método de asignación de probabilidad se le conoce como concepto clásico de probabilidad.

Probabilidad clásica

Si un experimento puede resultar de n maneras diferentes igualmente probables, la probabilidad de cada evento simple es $\frac{1}{n}$

La probabilidad de un evento A es igual a la suma de probabilidades de los eventos simples contenidos en A (inclusive si los eventos simples no son equiprobables). Tenga en cuenta que si un evento tiene una probabilidad cercana a uno significa que es muy probable que ocurra; si tiene una probabilidad cercana a cero significa que es muy poco probable que ocurra y si tiene una probabilidad de $\frac{1}{2}$ significa que es tan probable que ocurra como que no ocurra.

Dos eventos se dicen **equiprobables** si tienen la misma probabilidad de ocurrencia.

Observe que la probabilidad se ha cuantificado con un número fraccionario en el cual el numerador es menor o igual que el denominador, por consiguiente:

La probabilidad es un número entre cero y uno.
La suma de las probabilidades de todos los eventos simples de un espacio muestral *S* deber ser igual a uno.

Ejemplo 4.21

Enunciado

Suponga que se lanza un dado corriente. ¿Cuál es la probabilidad de obtener seis?

Solución

El espacio muestral de este experimento es $S = \{1, 2, 3, 4, 5, 6\}$. Es posible suponer que cada uno de los seis eventos simples es equiprobable. Por tanto, la probabilidad de obtener un seis es $\frac{1}{6}$.▪

Ejemplo 4.22

Enunciado

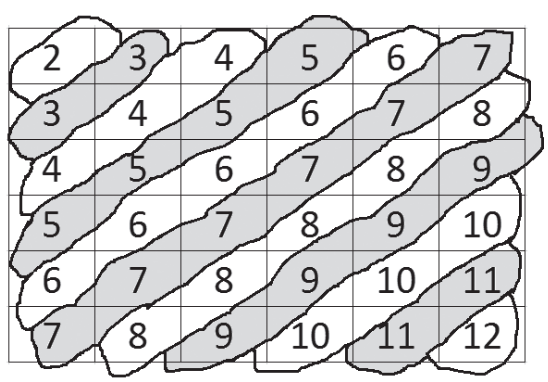
Suponga que se lanzan dos dados corrientes. ¿Cuál es la probabilidad de que la suma de puntos en las caras superiores sea siete?

Solución

El espacio muestral de este experimento consta de 36 resultados posibles igualmente probables.

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

La probabilidad de ocurrencia de cada uno de estos 36 resultados es $\frac{1}{36}$. La suma de las caras superiores de los anteriores resultados se muestra en la siguiente figura:



Los posibles resultados de la suma de las caras de los dos dados van desde dos (doble uno) hasta doce (doble seis). Se puede conseguir el número siete de seis maneras diferentes. Por consiguiente, la probabilidad de que la suma de las caras superiores de dos dados sea siete es:

$$\frac{6}{36} = \frac{1}{6}. \;\blacksquare$$

Ejemplo 4.23

Enunciado

La figura muestra los resultados de un sorteo de baloto. Determinar la probabilidad de que un jugador tenga exactamente 6, 5, 4, 3, 2, 1 o ningún acierto.

Solución

El número de combinaciones de seis números tomados de un conjunto de 45 es $\binom{45}{6}$.

$$\binom{45}{6} = \frac{45!}{(45-6)!6!} = \frac{45!}{39!6!} = 8'145.060$$

Esto significa que el número de resultados posibles del baloto es 8’145.060 y como solo existe una combinación ganadora, la probabilidad de ganar el premio mayor es de 1 entre 8’145.060. Por ejemplo, para calcular el número de maneras en que se pueden obtener exactamente cinco aciertos y, por ende, uno errado, se calcula la combinación $\binom{6}{5}$ y se multiplica por $\binom{39}{1}$ porque cinco serán aciertos y 39 errados. Los cálculos completos se muestran a continuación:

$$P(acertar\,6) = \frac{\binom{6}{6}\times\binom{39}{0}}{\binom{45}{6}} = \frac{1\times 1}{8'145.060} = \frac{1}{8'145.060}$$

$$P(acertar\,5) = \frac{\binom{6}{5}\times\binom{39}{1}}{\binom{45}{6}} = \frac{6\times 39}{8'145.060} = \frac{234}{8'145.060}$$

$$P(acertar\,4) = \frac{\binom{6}{4}\times\binom{39}{2}}{\binom{45}{6}} = \frac{15\times 741}{8'145.060} = \frac{11.115}{8'145.060}$$

$$P(acertar\,3) = \frac{\binom{6}{3}\times\binom{39}{3}}{\binom{45}{6}} = \frac{20\times 9.139}{8'145.060} = \frac{182.780}{8'145.060}$$

$$P(acertar\,2) = \frac{\binom{6}{2}\times\binom{39}{4}}{\binom{45}{6}} = \frac{15\times 82.251}{8'145.060} = \frac{1'233.765}{8'145.060}$$

$$P(acertar\,1)=\frac{\binom{6}{1}\times\binom{39}{5}}{\binom{45}{6}}=\frac{6\times 575.757}{8'145.060}=\frac{3'454.542}{8'145.060}$$

$$P(no\,acertar\,ninguno)=\frac{\binom{6}{0}\times\binom{39}{6}}{\binom{45}{6}}=\frac{1\times 3'262.623}{8'145.060}=\frac{3'262.62}{8'145.060}$$

Es de resaltar que es más probable tener un acierto que ninguno.▪

Ejercicio 4.3 Conceptos básicos de probabilidad

1. Se realizó una encuesta a quinientas madres cabeza de familia de estrato 3 sobre el lugar de la ciudad que escogen para compartir con la familia. Sus respuestas se resumen en la siguiente tabla:

Lugar	**Número de familias**
Museos	78
Cines	101
Parques	75
Clubes	93
Restaurantes	125
Otros	28
Total	**500**

Suponga que se elige al azar una de las quinientas madres y se le pregunta sobre el lugar de la ciudad que prefiere para compartir con la familia.

a. ¿Cuáles son los eventos simples de este experimento?

b. ¿Son equiprobables estos eventos?

c. ¿Cuál es la probabilidad de que la respuesta de esta madre sea "parques"?

d. ¿Cuál es la probabilidad de que la respuesta de esta madre no sea "cines"?

2. Una urna contiene dos monedas de quinientos, dos de doscientos y una de cien. Se extraen al azar y simultáneamente dos monedas. Calcular la probabilidad de:

a. Extraer exactamente $300.

b. Extraer más de $500.

c. Extraer menos de $800.

3. Un vendedor realiza tres llamadas a potenciales clientes. Sabe por experiencia que hace dos ventas cada cinco llamadas:
   a. Describa el espacio muestral.
   b. Calcule la probabilidad de que realice por lo menos una venta.
   c. Calcule la probabilidad de que realice exactamente dos ventas.

4. Una persona revuelve al azar cuatro fichas marcadas con letras del alfabeto para formar una palabra. Dos de ellas tienen la letra "a", una la letra "r" y la última la letra "z". ¿Cuál es la probabilidad de armar la palabra "azar"?

5. Tres niños y dos niñas se sientan aleatoriamente en una banca de parque, uno al lado del otro. ¿Es más probable que las niñas se siente juntas o separadas?

6. Encuentre las probabilidades de los eventos simples de los siguientes experimentos:
   a. Extraer una carta de una baraja de póker.
   b. Lanzar tres dados corrientes.
   c. Lanzar simultáneamente un dado y una moneda.
   d. Lanzar una moneda cinco veces.

7. Una ruleta tiene 38 casillas numeradas del 1 al 36 y las dos restantes con el 0 y el 00. Las casillas impares del 1 al 36 son rojas y las pares negras, el 0 es rojo y el 00 negro. Una vez la ruleta gira, una persona puede apostar a un número, a un grupo de números o a un color; cuando esta se detiene una casilla se identifica como la ganadora.
   a. Liste los eventos simples en un solo giro de la ruleta.
   b. Asigne probabilidades a cada evento simple.
   c. ¿Qué tiene mayor probabilidad de ocurrencia: color rojo o número impar?

8. Cuatro atletas igualmente calificados A, B, C y D apuestan una carrera. Si no se tienen en cuenta empates:
   a. ¿Cuántos eventos simples tiene el espacio muestral?
   b. ¿Qué probabilidad se le puede asignar a cada evento simple?
   c. ¿Cuál es la probabilidad de que A gane y B quede segundo?
   d. ¿Cuál es la probabilidad de que A gane?
   e. ¿Cuál es la probabilidad de que A llegue de último?

9. Se carga un dado de manera que los números pares tienen el doble de posibilidad de salir que los impares. Calcule la probabilidad de:
   a. Obtener número par.
   b. Obtener un número primo.
   c. Obtener un número primo impar.

10. Una pequeña comunidad consta de veinte familias de las cuales cuatro tienen un hijo, ocho tienen dos, cinco tienen tres, dos tienen cuatro y una tiene cinco.
    a. Si una de estas familias es escogida al azar, ¿cuál es la probabilidad de que ella tenga i hijos, $i = 1,2,3,4,5$?
    b. Si uno de los niños es escogido al azar, ¿cuál es la probabilidad de que él provenga de una familia con i hijos, $i = 1,2,3,4,5$?

11. Diez personas están sentadas aleatoriamente en una mesa redonda. ¿Cuál es la probabilidad de que una pareja en particular esté sentada junta?

12. Se mezclan cinco monedas falsas con siete auténticas. Se seleccionan tres monedas al azar con reposición. Calcular la probabilidad de extraer por lo menos una auténtica.

13. Solo una de las cuatro llaves en el llavero de un hombre abre su casa. En la noche, a oscuras, intenta abrir la puerta seleccionando las llaves en forma aleatoria y descartando las ya probadas. ¿Cuál es la probabilidad de que abra la puerta en el *i-ésimo* intento? ($i = 1,2,3,4$).

14. Realice el ejercicio anterior con cinco llaves.

15. En un salón hay veinticinco alumnos, catorce hombres y once mujeres. Cinco de ellos faltaron el jueves. Determine la probabilidad de que:
    a. Exactamente dos de los ausentes sean mujeres.
    b. Exactamente uno de los ausentes sea hombre.
    c. Ninguno sea hombre.

16. Una persona sufre un accidente que le deja afectado el cerebro. Después de esto, se le dan 12 bloques: cuatro cuadrados, cuatro rectángulos y cuatro círculos. Si la persona ubica cuatro de cada tipo en orden: cuatro rectángulos, cuatro cuadrados y luego cua-

tro círculos. ¿Es probable que aún relacione figuras con la misma forma? Ayuda: calcule la probabilidad de este evento suponiendo que no las relaciona.

17. Con el ánimo de indagar sobre el nivel de tolerancia de los jóvenes de una comunidad hacia diferentes grupos minoritarios, se realizó una encuesta a mil estudiantes de colegio. En esta encuesta se les leyeron seis afirmaciones a cada uno de los estudiantes y ellos debían elegir si con respecto a la afirmación estaban: totalmente de acuerdo, moderadamente de acuerdo, moderadamente en desacuerdo o totalmente en desacuerdo.

Los resultados de la encuesta se muestran a continuación:

	Con respecto a la afirmación, usted está:			
	Totalmente de acuerdo	**Moderadamente de acuerdo**	**Moderadamente en desacuerdo**	**Totalmente en desacuerdo**
1. Acepto tener como vecinos de mi casa a un grupo de reinsertados	269	345	311	75
2. Acepto tener como vecinos a enfermos de VIH	158	174	534	134
3. Acepto estudiar o trabajar con un reinsertado	756	244	0	0
4. Acepto estudiar o trabajar con un enfermo de VIH	177	645	123	55
5. Acepto que personas de otras religiones sean mis amigos	811	123	60	6
6. Acepto que personas con orientaciones sexuales diferentes a las mías sean mis amigos	179	437	298	86

Observe que cada una de las preguntas constituye un experimento aleatorio, como los que se han venido estudiando. Con base en este experimento y si se eligiera un estudiante al azar de los mil encuestados:

a. ¿Cuál es la probabilidad de que se encuentre totalmente de acuerdo con la afirmación 1?
b. ¿Cuál es la probabilidad de que se encuentre moderadamente de acuerdo con la afirmación 2?
c. ¿Cuál es la probabilidad de que se encuentre moderadamente en desacuerdo con la afirmación 3?

d. ¿Cuál es la probabilidad de que se encuentre totalmente en desacuerdo con la afirmación 4?
e. ¿Cuál es la probabilidad de que se encuentre totalmente de acuerdo con la afirmación 5?
f. ¿Cuál es la probabilidad de que se encuentre moderadamente de acuerdo con la afirmación 6?

Cálculo de probabilidades

Imagine que una compañía tiene asegurados 2.000 vehículos. De estos, el último año, cuatrocientos reportaron algún tipo de accidente. El gerente de la compañía tiene la creencia de que los automóviles rojos se accidentan con más frecuencia que los de otros colores y estaría tentado a promover un aumento en la póliza de automóviles con este color. Para estudiar esta situación construye la siguiente tabla que resume la información de los vehículos en cuanto a la accidentalidad y a su color:

	Vehículos rojos	**Vehículos de otro color**	**Total**
Vehículos que se accidentaron	300	100	400
Vehículos que no se accidentaron	200	1.400	1.600
Total	500	1.500	2.000

De esta tabla, se puede leer que hay asegurados cuatrocientos vehículos que se accidentaron y 1.600 que no se accidentaron. De los cuatrocientos que se accidentaron trescientos eran rojos y cien de otro color, mientras que de los que no se accidentaron doscientos eran rojos y 1.400 de otro color. ¿Los datos sugieren que los automóviles rojos se accidentan con más frecuencia que los de otros colores? ¿Cómo puede argumentar a favor o en contra de la afirmación del gerente? Más adelante, se darán argumentos probabilísticos para responder estas preguntas.

Adicionalmente, si los datos reflejan lo que sucede en la realidad y se cuenta con una "buena muestra" con información sobre los vehículos, es posible intentar responder preguntas como: ¿cuál es la probabilidad que un automóvil se accidente sabiendo que es de color rojo? ¿Cuál si es de otro color?

Para la compañía sería importante conocer cosas como: ¿cuál es la probabilidad que el automóvil se accidente durante el siguiente año? ¿Existe alguna influencia del color, el género del conductor, o edad en los índices de accidentalidad?

En esta parte del libro se tratan asuntos relacionados con resultados experimentales que se forman con la composición de eventos y con probabilidades asociadas a la ocurrencia de otros eventos.

Ahora supongamos que se selecciona al azar uno de los 2.000 vehículos asegurados por la compañía y definimos los siguientes eventos:

A:= "El vehículo seleccionado es rojo".

B:= "El vehículo seleccionado se ha accidentado".

¿Cuál es la probabilidad de ocurrencia de A? ¿Cuál la de B?

Como el espacio muestral S consta de los 2.000 vehículos asegurados y, de ellos, quinientos son rojos y cuatrocientos se han accidentado se tiene que:

$$P(A) = \frac{500}{2.000} = \frac{1}{4} = 0,25$$

$$P(B) = \frac{400}{2.000} = \frac{1}{5} = 0,2$$

"Intersección" $A \cap B$: el vehículo seleccionado es rojo y se ha accidentado

"Unión" $A \cup B$: el vehículo seleccionado es rojo o se ha accidentado

Se dice que $A \cap B$ ha ocurrido si ocurre A y B y que $A \cup B$ ha ocurrido si ocurre A o B.

Eventos compuestos

Los eventos compuestos son aquellos formados por uniones o intersecciones de eventos o por alguna combinación de los dos.

La intersección de los eventos A y B se denota $A \cap B$ y es el evento de que ocurra A y B.

La unión de dos eventos A y B se denota $A \cup B$ y es el evento de que ocurra A o B.

En diagramas de Venn, los eventos compuestos se pueden representar como:

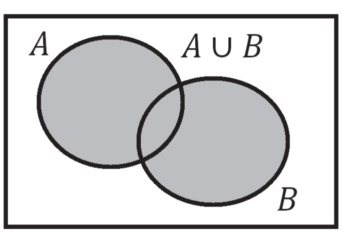


Diagrama de Venn de $A \cup B$

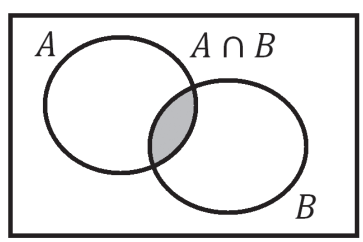
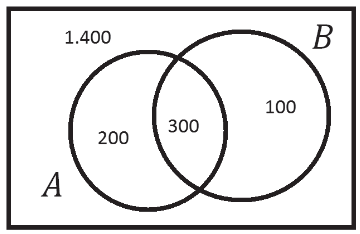


Diagrama de Venn de $A \cap B$

Ahora, ¿cuál es la probabilidad de $A \cap B$? ¿Cuál la de $A \cup B$? Para resolver esta pregunta, se puede construir un diagrama de Venn con base en la información de la tabla inicial:

	Vehículos rojos	**Vehículos de otro color**	**Total**
Vehículos que se accidentaron	300	100	400
Vehículos que no se accidentaron	200	1.400	1.600
Total	500	1.500	2.000

El diagrama de Venn muestra que hay quinientos vehículos rojos (doscientos no accidentados y trescientos accidentados) y cuatrocientos vehículos accidentados (trescientos rojos y cien de otro color). Los 1.400 por fuera de A y B representan los vehículos que no son rojos y que no se accidentaron.



A partir del diagrama, encontramos que:

La probabilidad de que el vehículo seleccionado al azar sea rojo y se haya accidentado es:

$$P(A \cap B) = \frac{300}{2.000} = \frac{3}{20} = 0,15$$

La probabilidad de que el vehículo seleccionado al azar sea rojo o se haya accidentado es:

$$P(A \cap B) = \frac{200 + 300 + 100}{2.000} = \frac{600}{2.000} = 0,3$$

Ejemplo 4.24

Enunciado

Se lanza una moneda corriente dos veces y se definen los siguientes eventos:

A: = "Cara en el primer lanzamiento".

B: = "Sello en el segundo lanzamiento".

Determine los eventos simples de A, B, $A \cap B$, $A \cup B$ y calcule sus probabilidades.

Solución

El espacio muestral del experimento es $S = \{CC, CS, SC, SS\}$, donde C y S significan cara y sello respectivamente. Cada uno de estos eventos es igualmente probable y tiene una probabilidad de ocurrencia de $\frac{1}{4}$. Los eventos simples de A y B son:

$A = \{CC, CS\}$ "cara en el primer lanzamiento".
$B = \{CS, SS\}$ "sello en el segundo lanzamiento".

Como la probabilidad de un evento es igual a la suma de probabilidades de los eventos simples contenidos en él, se tiene que:

$$P(A) = P(B) = \frac{1}{2}$$

La intersección y unión de los eventos A y B son:
$A \cap B = \{CS\}$ "cara en el primer lanzamiento y sello en el segundo lanzamiento".
$A \cup B = \{CC, CS, SS\}$ "cara en el primer lanzamiento o sello en el segundo lanzamiento".

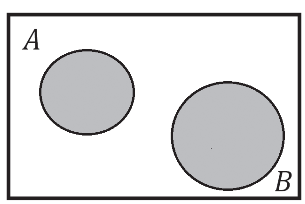
De esta manera $P(A \cap B) = \frac{1}{4}$ y $P(A \cup B) = \frac{3}{4}$. ▪

Eventos mutuamente excluyentes

En el apartado del capítulo se dijo que dos eventos son mutuamente excluyentes si ocurre uno de ellos, el otro no puede ocurrir. Si dos eventos A y B son mutuamente excluyentes, entonces:

$$P(A \cap B) = 0$$
$$P(A \cup B) = P(A) + P(B)$$

La representación de dos eventos mutuamente excluyentes en diagrama de Venn, es la siguiente:



Dos eventos A y B mutuamente excluyentes.

Complemento de un evento

En el ejemplo inicial, el experimento consistía en seleccionar un vehículo entre los 2.000 asegurados por la compañía. Se define el complemento de un evento así: $\overline{A}$:= "El vehículo seleccionado no es rojo" ($\overline{A}$ se lee "A complemento"). El complemento de A consta de los eventos del espacio muestral que no se encuentran en *A*.

Por definición A y $\overline{A}$: son mutuamente excluyentes. Además, como $S = A \cup \overline{A}$, se deduce que:

$$P(S) = P(A) + P(\overline{A})$$

$$1 = P(A) + P(\overline{A})$$

$$P(A) = 1 - P(\overline{A})$$

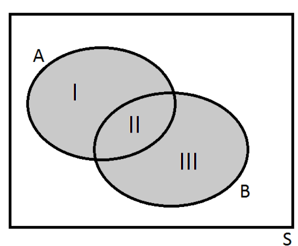
En el ejemplo, se había encontrado que la probabilidad de que el vehículo seleccionado fuera rojo era $P(A) = \frac{1}{4}$, la probabilidad de que el vehículo sea de un color diferente al rojo es:

$$P(\overline{A}) = 1 - P(A) = 1 - \frac{1}{4} = \frac{3}{4}$$

Complemento de un evento

El complemento de un evento *A*, se denota $\overline{A}$, consta de todos los eventos simples del espacio muestral *S* que no están en *A*.

Suponga ahora que dos eventos no son mutuamente excluyentes como en la siguiente figura:



La probabilidad de la unión (parte sombreada de la figura) corresponde al área de las regiones I, II y III. Si para hallar $PA \cup B$ se suma *P*(A) y *P*(B), se habrá contado dos veces el área de II que corresponde a $PA \cap B$. Para solucionar este inconveniente se resta P$(A \cap B)$ de *P*(*A*) + *P*(*B*), esto es:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

La anterior fórmula se puede generalizar para dos o más eventos así:

Sean $A_1, A_2, \cdots, A_n$ donde $n \geq 2$, entonces:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j}^{n} P(A_i \cap A_j) + \sum_{i<j<k}^{n} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} P(A_i \cap A_2 \cap \cdots \cap A_n)$$

Ejemplo 4.25

Enunciado

En el momento de devolver las tarjetas de crédito a n clientes de una tienda, el encargado las confunde y las entrega al azar. ¿Cuál es la probabilidad de que ninguno de los clientes reciba su tarjeta?

Solución

Considere los siguientes eventos:
A: ="Alguno de los cinco clientes recibe su tarjeta".
A_i: ="El cliente i recibe su tarjeta".

Se puede ver que $A = A_i \cup A_2 \cup \cdots \cup A_n$. La probabilidad de que ninguno de los clientes reciba su tarjeta es P($\overline{A}$). Por facilidad se calcula primero P(A), la probabilidad de que por los menos uno de ellos reciba su tarjeta.

Para calcular

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j}^{n} P(A_i \cap A_j) + \sum_{i<j<k}^{n} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} P(A_i \cap A_2 \cap \cdots \cap A_n),$$

se procede término a término así:

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}, \text{ luego } \sum_{i=1}^{n} P(A_i) = 1$$

$$P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}, \text{ pero como } i < j \text{ hay } \binom{n}{2} = \frac{n(n-1)}{2!} \text{ términos, luego}$$

$$\sum_{i<j}^{n} P(A_i \cap A_j) = \frac{1}{n(n-1)} \times \binom{n}{2} = \frac{1}{2!}.$$

De la misma forma:

$$P(A_i \cap A_j \cap A_k) = \frac{(n-3)!}{n!} \text{ y } \sum_{i<j<k}^{n} P(A_i \cap A_j \cap A_k) = \frac{(n-3)!}{n!} \times \binom{n}{3} = \frac{1}{3!} \text{ y así sucesivamente.}$$

Por consiguiente:

$$P(A) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \frac{1}{5!} - \cdots + (-1)^{n+1} \frac{1}{n!} \text{ y } P(\overline{A}) = 1 - P(A)$$

Esta última expresión tiene la forma de la serie de potencias $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} \cdots$. De ella, se deduce que $e^{-1} = 1 - 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} \cdots$ y significa que si el número de clientes es muy grande $(n \to \infty), P(A) = 1 - e^{-1}$ y, por tanto, la probabilidad de que ninguno de los clientes reciba su tarjeta es $P(\overline{A}) = \frac{1}{e} \approx 0,36$. ▪

Ejemplo 4.26

Enunciado

De un grupo de personas se selecciona una al azar y se definen los eventos A, B y C así:

A: = "la persona seleccionada es hombre".
B: = "la persona seleccionada es mayor de edad".
C: = "la persona seleccionada es extranjera".

Calcule las probabilidades y escriba el significado de los complementos $\overline{A}, \overline{B}$ y $\overline{C}$ suponiendo que $P(A) = 0{,}4$, $P(B) = 0{,}7$ y $P(C) = 0{,}05$.

Solución

El complemento de cada uno de estos eventos se forma con los elementos del espacio muestral que no están en A, B o C respectivamente. Por lo tanto:

$\overline{A}$: = "la persona seleccionada no es hombre = la persona seleccionada es mujer".

$\overline{B}$: = "la persona seleccionada no es mayor de edad = la persona seleccionada es menor de edad".

$\overline{C}$: = "la persona seleccionada no es extranjera = la persona seleccionada es del país".

$$P(\overline{A}) = 1 - P(A) = 0{,}6$$

$$P(\overline{B}) = 1 - P(B) = 0{,}3$$

$$P(\overline{C}) = 1 - P(C) = 0{,}95$$

▪

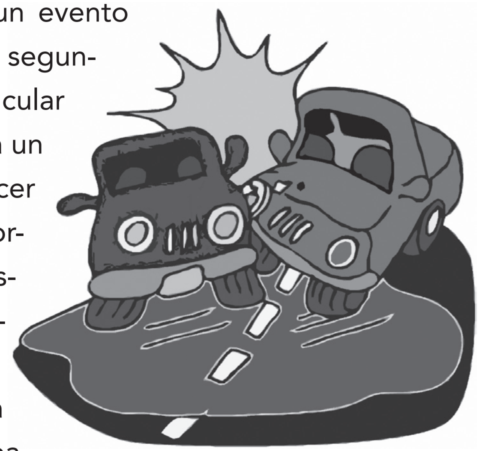
Ejercicio 4.4 Reglas de probabilidad

1. Se lanzan dos dados corrientes uno a uno y se definen los siguientes eventos: A: = "el resultado en el primer lanzamiento es un uno" y B: = "la suma de los puntos de los dos dados es par". Calcule e interprete:

   a. $P(A)$ b. $P(B)$ c. $P(A \cap B)$
   d. $P(A \cup B)$ e. $P(\overline{A})$ f. $P(\overline{B})$

2. Cuando A juega tenis contra B, las probabilidades de que gane A son de dos a una. Suponga que A y B juegan dos partidos, ¿cuál es la probabilidad de que A gane por lo menos un partido?

Probabilidad condicional y Teorema de Bayes



Muchas veces la probabilidad de ocurrencia de un evento está asociada al hecho de que haya ocurrido o no un segundo evento. Por ejemplo, suponga que le interesa calcular la probabilidad de que una persona en particular sufra un accidente de tránsito. Recuerde que esto se puede hacer si se basa en registros históricos para calcular la proporción de personas que han sufrido accidentes, con respecto al total de la población en estudio. Además, suponga que esta persona es hombre. Esta información puede ser útil para encontrar con mayor exactitud la probabilidad deseada, porque es posible que la probabilidad de accidentarse sea diferente entre hombres y mujeres.

En el ejemplo inicial, se dijo que la información de los vehículos asegurados por una compañía, en cuanto a la accidentalidad y a su color, está dado por:

	Vehículos rojos	**Vehículos de otro color**	**Total**
Vehículos que se accidentaron	300	100	400
Vehículos que no se accidentaron	200	1.400	1.600
Total	500	1.500	2.000

También se definieron los eventos:
A: = "el vehículo seleccionado es rojo".
B: = "el vehículo seleccionado se ha accidentado".

Si seleccionamos al azar un vehículo de los asegurados por la compañía, la probabilidad de que se haya accidentado es $P(B)=\frac{400}{2.000}=0,2$. Suponga que el vehículo seleccionado es rojo, ¿cuál es la probabilidad de que se haya accidentado?

El conocimiento de que el vehículo es rojo restringe el espacio muestral de 2.000 a quinientos. De estos quinientos, los accidentados fueron trescientos, luego la probabilidad pedida es $\frac{300}{500}=0,6$. Por tanto, la probabilidad de que el vehículo se haya accidentado dado que se sabe que es rojo es de 0,6. A esta última probabilidad se le llama probabilidad condicional de B dado A, se denota por $P(B|A)$ y se lee "probabilidad de B dado A".

Observe que en el cociente $\frac{300}{500}$, trescientos está asociado al número de vehículos de $A\cap B$ y quinientos al número de vehículos de A.

Ejemplo 4.27

Enunciado

Con respecto al ejemplo inicial, ¿qué significa $P(B|A)$?

Solución

$P(B|A)$ es la probabilidad condicional de A dado que ha ocurrido B, es decir, la probabilidad de que el vehículo seleccionado sea rojo, dado que se sabe que es uno de los que se han accidentado. Como se conoce que el vehículo es uno de los que se han accidentado, el espacio muestral se reduce a cuatrocientos vehículos. De estos, trescientos son rojos, por ende:

$$P(A|B) = \frac{300}{400} = 0{,}75$$

Note que en el cociente $\frac{300}{400}$, trescientos está asociado al número de vehículos de $A \cap B$ y cuatrocientos al número de vehículos de B.▪

Probabilidad condicional

La probabilidad condicional de A dado que ha ocurrido B es:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

siempre y cuando $P(B) \neq 0$.

En el ejemplo inicial, el gerente de la compañía tiene la creencia de que los automóviles rojos se accidentan con más frecuencia que los de otros colores. ¿Tiene razón el gerente? Para responder a esta pregunta, se analiza si es más probable que un vehículo sea de los que se ha accidentado sabiendo que es rojo o si es de los de otro color. Para ello, se compara $P(A|B)$ con $P(A|\overline{B})$, donde $\overline{A}$ significa que el vehículo seleccionado no es rojo, es decir, es de otro color.

Se sabe que:

$$P(B|A) = \frac{300}{500} = \frac{3}{5} = 0{,}6 \text{ y por definición } P(B|\overline{A}) = \frac{P(\overline{A} \cap B)}{P(\overline{A})} = \frac{100}{1.500} = 0{,}0\widehat{6}$$

$\overline{A}$ es el evento en el cual el vehículo seleccionado no es rojo y $\overline{A} \cap B$ es el evento en el cual el vehículo seleccionado no es rojo y se ha accidentado. Como $P(B|\overline{A}) < P(B|A)$, concluimos que si selecciona al azar un automóvil de los 2.000 asegurados, es más probable que se haya accidentado si es de color rojo que si es de otro color.

Independencia

Si dados dos eventos A y B sucede que $P(A|B)=P(A)$, significa que la ocurrencia del evento B no afecta para nada la probabilidad de A. Ya que la probabilidad condicional $P(A|B)$ puede ser expresada como $\frac{P(A\cap B)}{P(B)}$ y se tiene que $\frac{P(A\cap B)}{P(B)}=P(A)$, o lo que es lo mismo:

$$P(A\cap B)=P(A)P(B)$$

Si esto sucede, decimos que los eventos A y B son independientes, en el sentido que si A ocurre, la probabilidad de ocurrencia de B no se ve alterada.

Independencia

Dos eventos A y B son independientes si $P(A\cap B)=P(A)P(B)$.

Ejemplo 4.28

Enunciado

En el ejemplo inicial de esta sección del libro se expuso la siguiente situación. Un experimento consiste en seleccionar al azar uno de los 2.000 vehículos asegurados por una empresa y se definen los eventos:

A: = "el vehículo seleccionado es rojo".
B: = "el vehículo seleccionado se ha accidentado".

Determinar si A y B son independientes.

Solución

A y B son independientes si $P(A\cap B)=P(A)P(B)$. Como $P(A)=\frac{500}{2.000}=\frac{1}{4}$, $P(A)=\frac{400}{2.000}=\frac{1}{5}$ y $P(A\cap B)=\frac{300}{2.000}=\frac{3}{20}$ es fácil ver que $P(A\cap B)\neq P(A)P(B)$, por lo que se puede concluir que A y B no son eventos independientes. Esto significa que la probabilidad de que un vehículo se accidente se ve alterada si se sabe que este es rojo.▪

Ejemplo 4.29

Enunciado

Es posible que en una encuesta las preguntas resulten comprometedoras para los entrevistados. Algunas personas prefieren no contestar sobre asuntos relacionados con su vida privada o dan una información incorrecta. Mantener el anonimato de las respuestas de los entrevistados es fundamental para obtener una información verdadera. ¿Cómo motivar respuestas verdaderas en los entrevistados aun en preguntas comprometedoras?

Solución

Suponga que la pregunta comprometedora es la siguiente:

D: = ¿Ha consumido sustancias alucinógenas durante el último año?

Se puede incluir una segunda pregunta no comprometedora, de tal manera que se conozca de antemano la probabilidad de obtener una respuesta positiva, por ejemplo:

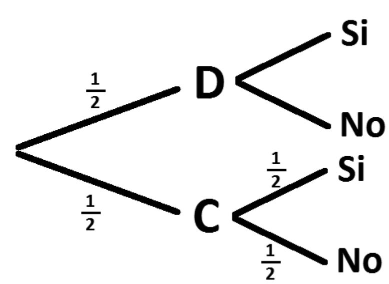
C: = ¿El día de su cumpleaños es un número par?

Ahora se le puede pedir a la persona que de manera privada, lance un moneda al aire y si el resultado es cara debe responder la pregunta D, de lo contario debe responder la pregunta C.

Los entrevistados pueden estar seguros que una respuesta positiva no los compromete de ninguna manera porque se desconoce cuál de las dos preguntas ha respondido.

Note que la probabilidad de obtener un sí como respuesta a la pregunta C, puede ser $\frac{1}{2}$.

En el siguiente diagrama de árbol se resumen las cuatro posibilidades de respuesta de un entrevistado:



Es razonable suponer que la probabilidad de obtener un sí como respuesta cuando la pregunta escogida es la C, es $\frac{1}{2}$, esto es, Como la selección de la pregunta a responder fue hecha al azar mediante una moneda, se sabe que $P(D)=\frac{1}{2}=P(C)$.

Lo que se desea averiguar es la probabilidad de obtener una respuesta positiva cuando se responde a la pregunta D, esto es, $P(\mathrm{Si}|\mathrm{D})$.

Note que obtener un sí como respuesta puede darse de dos maneras:

$$\mathrm{Si}=(\mathrm{Si}\cap\mathrm{D})\cup(\mathrm{Si}\cap\mathrm{C})$$

Como estos eventos son excluyentes, $\mathrm{P}(Si)=\mathrm{P}(\mathrm{Si}\cap\mathrm{D})\cup\mathrm{P}(\mathrm{Si}\cap\mathrm{C})$.

De lo estudiado en la probabilidad condicional se sabe que:

$$\mathrm{P}(\mathrm{Si}|\mathrm{D})=\frac{(\mathrm{Si}\cap\mathrm{D})}{\mathrm{P}(\mathrm{D})}$$

y que:

$$\mathrm{P}(\mathrm{Si}|\mathrm{C})=\frac{(\mathrm{Si}\cap\mathrm{C})}{\mathrm{P}(\mathrm{C})}\rightarrow\mathrm{P}(\mathrm{Si}\cap\mathrm{C})=\mathrm{P}(\mathrm{C})\mathrm{P}(\mathrm{Si}|\mathrm{C}).$$

Por lo tanto:

$$\mathrm{P}(\mathrm{Si}|\mathrm{D})=\frac{(\mathrm{Si}\cap\mathrm{D})}{\mathrm{P}(\mathrm{D})}=\frac{\mathrm{P}(\mathrm{Si})-\mathrm{P}(\mathrm{Si}\cap\mathrm{C})}{\mathrm{P}(\mathrm{D})}=\frac{\mathrm{P}(\mathrm{Si})-\mathrm{P}(\mathrm{C})\mathrm{P}(\mathrm{Si}|\mathrm{C})}{\mathrm{P}(\mathrm{D})}\quad .$$

Observe que la mayoría de las probabilidades del lado derecho de la última ecuación son conocidas salvo $P(S_i)$, pero esta puede ser estimada a través de los resultados de la encuesta. Es una solución bastante razonable a la pregunta planteada, sin embargo el lector notará las restricciones que deben imponerse.

Por ejemplo, si la entrevista se le hace a 200 personas y se obtienen 60 respuestas afirmativas, la proporción de personas que han consumido alguna sustancia alucinógena es:

$$\mathrm{P}(\mathrm{Si}|\mathrm{D})=\frac{\mathrm{P}(\mathrm{Si})-\mathrm{P}(\mathrm{C})\mathrm{P}(\mathrm{Si}|\mathrm{C})}{\mathrm{P}(\mathrm{D})}=\frac{0,3-0,5\times0,5}{0,5}=0,1. \blacksquare$$

Teorema de Bayes

A continuación, se estudiará un resultado que tiene que ver con la probabilidad condicional, el Teorema de Bayes, que se usa para encontrar $P(A|B)$ cuando $P(A\cap B)$ y $P(B)$ no son inmediatas.

Ejemplo 4.30

Enunciado

Investigadores han creado una prueba para detectar si un individuo está infectado con un parásito que se aloja en el aparato digestivo. Con base en un estudio previo se sabe que el 10% de la población está infectada con el parásito. Los investigadores analizan la efectividad de la prueba sobre individuos con diagnóstico confirmado. A partir de este esquema, la prueba resultó positiva en un 99% de los individuos con infección confirmada, sin embargo, cuando se le aplicó a pacientes que se sabe están libres del parásito, el procedimiento reportó erróneamente a 3% como infectados.

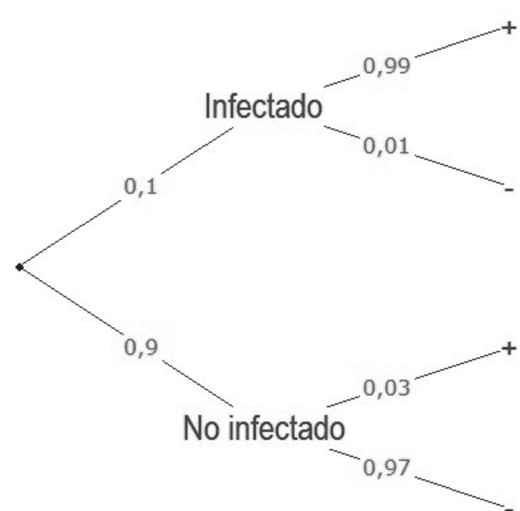
Para que la prueba pueda ser utilizada para detectar la infección por el parásito, se requiere que sea un fuerte indicador de que el parásito está presente. ¿Cuál es entonces la probabilidad de estar infectado por el parásito cuando la prueba así lo reporta?

Solución

De un individuo de la población se sabe que:
I: = "Está infectado con el parásito".
$\bar{I}$: = "No está infectado con el parásito".

Y si se le realiza la prueba los resultados pueden ser:
+: = "La prueba afirma que está infectado".
–: = "la prueba afirma que no está infectado".

Según el estudio previo $P(I) = 0{,}1$ esto es, si se escoge de la población a un individuo al azar, la probabilidad de que este infectado es de 0,1. Mientras que el evento complementario tiene una probabilidad de $P(\bar{I}) = 0{,}9$.



La prueba detecta el parásito en un 99% de los infectados, es decir, $P(+|\,I) = 0{,}99$ y la prueba detecta erróneamente el parásito en un 3% de las personas libres de él, es decir, $P(+|\,\bar{I}) = 0{,}03$. En el diagrama de árbol, se han ubicado las anteriores probabilidades, así como sus probabilidades complementarias.

Para responder la pregunta, se debe calcular $P(I|+)$. Como se ha estudiado, esta probabilidad condicional se puede calcular como:

$$P(I|+) = \frac{P(I \cap +)}{P(+)}$$

Veamos primero cómo calcular la probabilidad del denominador. Cuando se selecciona una persona al azar y luego se le realiza la prueba hay cuatro resultados posibles excluyentes:

$I \cap +$ que la persona este infectada y el resultado sea positivo.

$I \cap -$ que la persona este infectada y el resultado sea negativo.

$\bar{I} \cap +$ que la persona no esté infectada y el resultado sea positivo.

$\bar{I} \cap -$ que la persona no esté infectada y el resultado sea negativo.

Los resultados positivos se pueden dar de dos maneras $+ = (I \cap +) \cup (\bar{I} \cap +)$. Por ser excluyentes $P(+) = P(I \cap +) + P(\bar{I} \cap +)$.

$$P(+) = P(I)P(+|I) + P(\bar{I})P(+|\bar{I}) \qquad (4.1)$$

De esta manera:

$$P(I|+) = \frac{P(I \cap +)}{P(+)} = \frac{P(I)P(+|I)}{P(I)P(+|I) + P(\bar{I})P(+|\bar{I})} \qquad (4.2)$$

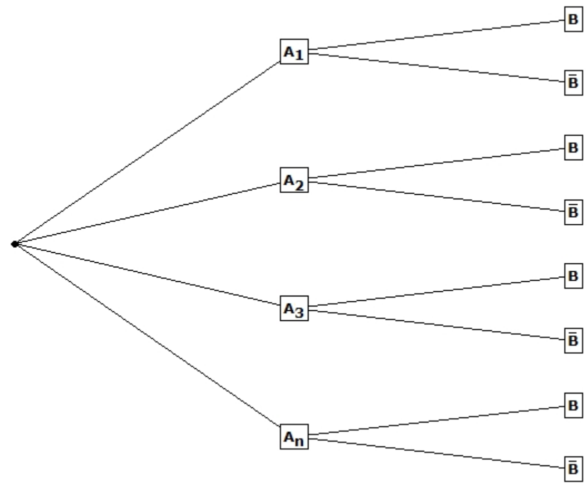
$$P(I|+) = \frac{0,1 \times 0,99}{0,1 \times 0,99 + 0,9 \times 0,03} = 0,7857$$

Esto significa que aunque la prueba dé resultado positivo, la probabilidad de estar infectado por el parásito es apenas de 0,7857.

A la ecuación 4.1 se le llama **probabilidad total** y a la ecuación 4.2 **Teorema de Bayes**, en honor al matemático inglés Thomas Bayes, quién la formuló en 1763.

Estas ideas pueden generalizarse como sigue: suponga que existe un conjunto de eventos mutuamente excluyentes $A_1, A_2, \cdots, A_n$ con probabilidades conocidas y $\sum_{i=1}^{n} A_i = 1$. Si adicionalmente sucede otro evento B y se conoce $P(B|A_i)$ entonces:

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$



Note que en un diagrama de árbol los eventos A_i están en la primer parte del árbol y el evento B en la segunda parte. A $P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$ se le suele llamar la *probabilidad total* de *B*, a $P(B|A_i)$ *probabilidad a priori* y a $P(A_i|B)$ *probabilidad a posteriori.*

Teorema de Bayes

Sea un conjunto de eventos mutuamente excluyentes $A_1, A_2, \cdots, A_n$ con probabilidades conocidas y $\sum_{i=1}^{n} P(A_i) = 1$. Si sucede otro evento B y se conoce $P(B|A_i)$ entonces:

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

Ejemplo 4.31

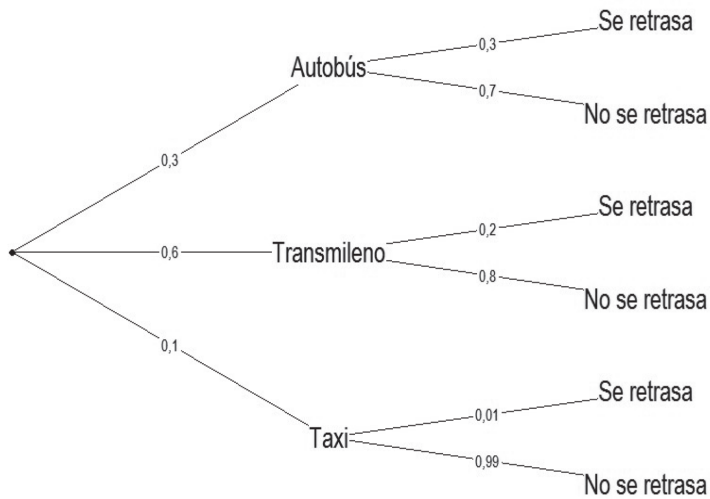
Enunciado

Para llegar a la primera clase de la mañana, un estudiante puede tomar autobús, Transmilenio o taxi, con probabilidades de 0,3, 0,6 y 0,1, respectivamente. En autobús, se retrasa el 30% de las veces, en Transmilenio el 20%, mientras que en taxi solo el 1% de las veces.

a. Determine la probabilidad de que en un día particular el estudiante se retrase.

b. Si un día el estudiante se retrasó, calcule la probabilidad de que haya tomado Transmilenio.

Solución



Consideremos los posibles eventos del estudiante:
A: = "elige viajar en autobús".
T: = "elige viajar en Transmilenio".
X: = "elige viajar en taxi".
R: = "se retrasa".

Según la información del problema $P(A) = 0{,}3$, $P(T) = 0{,}6$, y $P(X) = 0{,}1$. También se conocen algunas probabilidades a priori: $P(R|A) = 0{,}3$, $P(R|T) = 0{,}2$ y $P(R|X) = 0{,}01$.

Entonces, la probabilidad total de retrasarse es:

$$P(R) = P(R \cap A) + P(R \cap T) + P(R \cap X) = P(A)P(R|A) + P(T)P(R|T) + P(X)P(R|X)$$

$$P(R) = 0{,}3 \times 0{,}3 + 0{,}6 \times 0{,}2 + 0{,}1 \times 0{,}01 = 0{,}211$$

Esto significa que antes de la elección del medio de transporte, el estudiante tiene una probabilidad de retrasarse de 0,211.

Por otra parte, la probabilidad de haber tomado Transmilenio, dado que se retrasó es:

$$P(T|R) = \frac{P(T)P(R|T)}{P(A)P(R|A) + P(T)P(R|T) + P(X)P(R|X)}$$

$$P(T|R) = \frac{0{,}6 \times 0{,}2}{0{,}211} \approx 0{,}569 \quad ▪$$

Ejercicio 4.5 Probabilidad condicional y Bayes

1. La tabla recoge la información de un grupo de cien estudiantes sobre su condición de fumador y sexo.

	Fumadores: *F*	**No fumadores: $\overline{F}$**	**Total**
Hombre: H	10	20	30
Mujer: $\overline{H}$	30	40	70
Total	40	60	100

Se selecciona una persona al azar del grupo de estudiantes y se definen los eventos: H: = "la persona seleccionada es hombre" y F: = "la persona seleccionada es un fumador". Calcule e interprete:

a. $P(H)$ b. $P(\overline{H})$ c. $P(F)$ d. $P(\overline{F})$

e. $P(F|H)$ f. $P(F|\overline{H})$ g. $P(H|F)$ h. $P(\overline{H}|\overline{F})$

2. En la siguiente tabla, se muestra la clasificación de un banco realizada a doscientos clientes:

	Deudor moroso	**No es deudor moroso**
Hombre	24	96
Mujer	16	64

Se selecciona un cliente al azar y se definen los siguientes eventos: A: = "el cliente seleccionado es hombre" y B: = "el cliente seleccionado es deudor moroso." ¿Son A y B eventos independientes?

3. Calcule las probabilidades indicadas sobre la siguiente situación: se extrae al azar una carta de una baraja y se definen los siguientes eventos:

A: = "la carta seleccionada es un corazón".
B: = "la carta seleccionada es un as".

a. $P(A)$ b. $P(B)$ c. $P(A \cap B)$

d. $P(A \cup B)$ e. $P(\overline{A})$ f. $P(\overline{B})$

4. La siguiente tabla recoge la información de un grupo de boxeadores en relación con su raza y tipo de sangre:

Raza	**Sangre O+**	**Otro tipo de sangre**
Negra	15	25
Otra raza	20	40

Si se selecciona al azar un boxeador de este grupo de deportistas y se definen los siguientes eventos:

N: = "el deportista seleccionado es de raza negra".

N: = "el tipo de sangre del deportista seleccionado es O+".

Calcule e interprete:

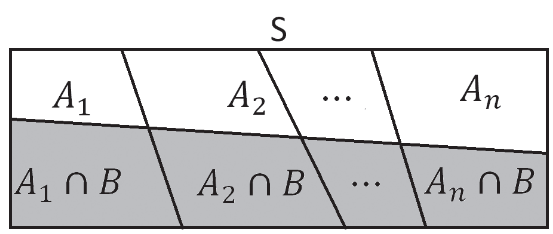
a. $P(N)$ b. $P(\overline{N})$ c. $P(O)$ d. $P(\overline{O})$

e. $P(N|O)$ f. $P(N|\overline{O})$ g. $P(O|N)$ h. $P(\overline{O}|\overline{N})$

5. Una prueba de sangre de laboratorio detecta una enfermedad en el 99,9% de los pacientes que la poseen. No obstante, la prueba da un resultado positivo falso en 1,5% de las personas sanas a las que se les aplica. Si el 0,05% de la población padece la enfermedad, ¿cuál es la probabilidad de que una persona tenga la enfermedad si la prueba dio resultado positivo?

6. El director de tesis de tres estudiantes les anuncia que dos de ellos se han ganado una beca, pero solo se les informará individualmente hasta final de semestre. Uno de los estudiantes le pide a su director que le diga cuál de sus compañeros obtendrá la beca, ya que no importa dar esta información porque él ya sabe que, por lo menos, uno de ellos la obtendrá. El director se niega a responder y afirma que si el estudiante supiera cuál de sus compañeros obtendrá la beca, entonces su propia probabilidad de obtener la beca disminuiría de $\frac{2}{3}$ a $\frac{1}{2}$. ¿Qué opina del razonamiento del director?
7. Explique por qué a $P(B|A_i)$ se le llama probabilidad a priori y $P(A_i|B)$ probabilidad a posteriori.

8. En la figura, la región sombreada corresponde al evento B. Utilice la ilustración para:
   a. Explicar las características de los eventos A_i.
   b. Escribir una expresión para $P(B)$ y para $P(A_i|B)$.



9. Una persona que vuela con frecuencia en avión, selecciona la aerolínea Avianca el 50% de las veces, Lan el 30% y Copa el 20%. Los porcentajes de vuelos con retrasos en estas aerolíneas son 2%, 5% y 4%, respectivamente. Si las condiciones de las aerolíneas se mantienen constantes y la persona decide viajar, calcule:
   a. La probabilidad de que el vuelo se retrase.
   b. La probabilidad de que haya viajado en Avianca sabiendo que el vuelo se retrasó.

10. En Barranquilla el 25% de los carros emiten cantidades excesivas de contaminantes. Cuando se realiza la prueba de control, el 99% de todos los carros que emiten cantidades excesivas de contaminantes no la pasará, pero el 17% de los carros que no emiten cantidades excesivas de contaminantes tampoco pasará. ¿Cuál es la probabilidad de que un carro que no pasa la prueba en realidad, emita cantidades excesivas de contaminantes?

11. Usted le pide a una persona que riegue una planta mientras sale de vacaciones. Si no la riega, la probabilidad de que muera es de 0,7; si la riega la probabilidad de que muera es de 0.05. Si la probabilidad de que la persona riegue la planta es de 0,95:
    a. ¿Cuál es la probabilidad de que la planta aún esté viva cuando usted regrese?
    b. Si la encuentra muerta, ¿cuál es la probabilidad de que la persona no la haya regado?

12. Se tienen dos sobres idénticos, cada uno con dos billetes. Un sobre tiene dos billetes de \$2.000; el otro tiene un billete de \$50.000 y otro de \$2.000. Se escoge al azar un sobre y de él se saca un billete de \$2.000. ¿Cuál es la probabilidad de que en dicho sobre quede un billete de \$50.000?

13. Una prueba diseñada para diagnosticar el cáncer de cuello uterino tiene un coeficiente de falsos negativos y falsos positivos de 0,05 cada uno. De cierta población de mujeres, el 0,4% tiene este tipo de cáncer. ¿Cuál es la probabilidad de que una mujer de la población elegida aleatoriamente tenga cáncer de cuello uterino, dado que el resultado de la prueba dio positivo?

14. Suponga que la prevalencia de la bacteria *Helicobacter pylori* en Colombia es del 30%. Se realiza una prueba que tiene las siguientes características: si una persona tiene la bacteria, la prueba lo indica con una probabilidad del 0,999; si no la tiene, entonces, hay una probabilidad del 0,002 de que la prueba lo reporte erróneamente como infectado. Si una persona es seleccionada al azar, calcular la probabilidad de que:
    a. Tenga la bacteria cuando el resultado de la prueba es positivo.
    b. Obtenga un resultado negativo.

Variables aleatorias

En un experimento, no siempre es de interés detallar todos sus posibles resultados, sino ciertos valores numéricos. Por ejemplo, si una persona compra la boleta de una rifa, no está interesada en todos los posibles resultados de esta, sino en las posibles ganancias, es decir, la persona considera dos escenarios posibles: ganar o perder para luego cuantificar la probabilidad de cada una. En general, a estos resultados numéricos obtenidos de un experimento se les conoce como variables aleatorias.

Ejemplo 4.32

Enunciado

Un estudiante compra por \$20.000 una boleta de una rifa que tiene cien puestos y un premio de \$1’000.000. Determinar las posibilidades de ganancia del estudiante y sus probabilidades.

Solución

Como el estudiante compra una sola boleta, la probabilidad de ganar el premio es $\frac{1}{100}$. Si el número de su boleta es el elegido, la ganancia es de \$980.000 —descontando el valor de la boleta—, de lo contrario tiene una ganancia negativa de –20.000 (pierde \$20.000). Las posibilidades de la ganancia y su respectiva probabilidad se muestran a continuación:

Ganancia	– 20.000	980.000
Probabilidad	$\frac{99}{100}$	$\frac{1}{100}$

La ganancia se puede considerar como una variable aleatoria y, en este caso, solo puede tomar dos valores –20.000 y 980.000.▪

A la fórmula, tabla o gráfica que asigna la probabilidad a cada uno de los valores de una variable aleatoria discreta se le denomina **función de probabilidad** y se denota $P(X = x)$.

Ejemplo 4.33

Enunciado

Se lanzan un par de monedas regulares y se define la variable aleatoria X: = “Número de caras”. Encontrar la función de probabilidad de X.

Solución

El espacio muestral del experimento es $S = \{CC, CS, SC, SS\}$. Cada uno de los eventos simples tiene una probabilidad de ocurrencia de $\frac{1}{4}$. Si el resultado es CC el número de caras es dos, esto se puede escribir $X(CC) = 2$. Igualmente $X(CS) = X(CS) = 1$ y $X(SS) = 0$.

La función de probabilidad se muestra en la siguiente tabla:

x	0	1	2
$P(X=x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

.

En los ejemplos anteriores, las variables aleatorias han tomado un número finito de valores. Si una variable aleatoria toma un número finito o contable de valores, se dice que la variable *aleatoria es* **discreta**. Si la variable puede tomar cualquier valor en algún intervalo (o intervalos) del conjunto de los números reales y la probabilidad de que tome un valor determinado es cero se dice que es **continua**.

Variables aleatorias discretas

Como ya se mencionó, una variable aleatoria es discreta si toma un número finito o contable de valores. Una de las diferencias importantes entre variables aleatorias discretas y continuas es que las discretas no están sujetas a que la probabilidad de que la variable tome un valor particular sea cero.

Función de probabilidad

A la fórmula, tabla o gráfica que asigna la probabilidad a cada uno de los valores de una variable aleatoria discreta se le denomina **función de probabilidad** y se denota $P(X=x)$.

Características de una función de probabilidad discreta

Dos propiedades necesarias para que $P(X=x)$ sea función de probabilidad discreta son:

$$0 \leq P(X=x) \leq 1$$

$\sum P(X=x)=1$ donde la suma recorre los posibles valores de x.

Ejemplo 4.34

Enunciado

Dé ejemplos de variables aleatorias discretas.

Solución

1. Número de comerciales emitidos en un programa de televisión.

2. Cuentas corrientes activas en un banco.
3. Miembros del Senado en la oposición.
4. Número de empresas con errores en su declaración de impuestos en Cundinamarca.
5. Interrupciones de la energía eléctrica durante un mes.▪

Ejemplo 4.35

Enunciado

Se lanza una moneda corriente en repetidas ocasiones y se define la variable aleatoria: $Y:=$ "número de lanzamientos hasta obtener la primera cara".

Determinar la función de probabilidad de Y.

Solución

Los posibles resultados de la variable aleatoria Y son 1, 2, 3, …, n, … En teoría, Y puede tomar valores muy grandes, pero es poco probable que esto ocurra. La variable aleatoria Y toma el valor 1 cuando se obtiene cara en el primer lanzamiento, lo cual ocurre con probabilidad $\frac{1}{2}$, esto es: $P(C) = P(Y=1) = \frac{1}{2}$.

Y toma el valor 2 cuando se obtiene primero un sello y luego una cara:

$$P(SC) = P(Y=2) = \frac{1}{2} \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^2$$

En general, la probabilidad de tener que realizar Y lanzamientos hasta obtener la primera cara es:

$$P(SSS \cdots SC) = P(Y=y) = \left(\frac{1}{2}\right)^y$$

Los valores de la función de probabilidad se pueden poner en una tabla:

y	1	2	3	$\cdots$	n	$\cdots$
$P(Y=y)$	$\left(\frac{1}{2}\right)^1 = \frac{1}{2}$	$\left(\frac{1}{2}\right)^2 = \frac{1}{4}$	$\left(\frac{1}{2}\right)^3 = \frac{1}{8}$	$\cdots$	$\left(\frac{1}{2}\right)^n$	$\cdots$

▪

La **función de distribución** de una variable aleatoria X está definida para todo número real x como: $F(x) = P(X \leq x)$.

La función de distribución $F(x)$ es la probabilidad de que la variable aleatoria tome un valor menor o igual a x. También se le conoce como la función de distribución acumulada de la variable aleatoria X hasta el valor x.

Es frecuente expresar las probabilidades de eventos en términos de la función de distribución. Por ejemplo, si se desea calcular la probabilidad de que una variable aleatoria discreta X tome un valor menor o igual que *b* y mayor que *a*, esto es $P(a < X \leq b)$, se evalúa la función de distribución en *b* y se le resta el valor de la función de distribución en *a*, es decir $P(a < X \leq b) = F(b) - F(a)$.

Ejemplo 4.36

Enunciado

Con respecto al ejemplo 4.33:

a. Determinar la función de distribución.
b. Calcular la probabilidad de tener que lanzar la moneda menos de cuatro veces para obtener la primera cara.
c. Calcular la probabilidad de tener que lanzar la moneda cuatro veces o menos, pero más de dos para obtener la primera cara.

Solución

La función de probabilidad está dada por:

y	1	2	3	4	5	6	…	n	…
$P(Y=y)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$	…	$\left(\frac{1}{2}\right)^n$	…

La función de distribución acumula la probabilidad, por ejemplo:

$$F(3) = P(Y \leq 3) = P(Y = 1) + P(Y = 2) + P(Y = 3) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}.$$

Observe que los eventos que hacen que $X = 1$, $X = 2$ y $X = 3$ son mutuamente excluyentes. Los demás valores de la función de distribución son:

y	1	2	3	4	5	6	…	n	…
$F(y) = P(Y \leq y)$	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{7}{8}$	$\frac{15}{16}$	$\frac{31}{32}$	$\frac{63}{64}$	…	$\frac{2^y - 1}{2^y}$	…

a. Se puede verificar que $F(y) = P(Y \leq y) = \frac{2^y - 1}{2^y}$.
b. La probabilidad de tener que lanzar la moneda menos de cuatro veces para obtener la primera cara es $P(Y < 4) = P(Y = 1) + P(Y = 2) + P(Y = 3) = P(Y \leq 3) = F(3) = \frac{7}{8}$.
c. La probabilidad de tener que lanzar la moneda cuatro veces o menos, pero más de dos para obtener la primera cara es $P(2 < Y \leq 4) = P(Y = 3) + P(Y = 4) = F(4) - F(2) = \frac{15}{16} - \frac{3}{4} = \frac{3}{16}$. ▪

Valor esperado y desviación estándar de una variable aleatoria discreta

En el ejemplo 4.29 se mostraron las opciones de ganancia de un estudiante con sus respectivas probabilidades cuando compra una boleta en una rifa de cien puestos:

Ganancia	-20.000	980.000
Probabilidad	$\frac{99}{100}$	$\frac{1}{100}$

Si el estudiante participa muchas veces en rifas similares, es razonable esperar que la proporción de veces que tendrá una ganancia de 980.000 es $\frac{1}{100}$ y la proporción de veces que ganará – 20.000 es $\frac{1}{100}$. Esto significa que la ganancia promedio por juego es:

$$980.000\frac{1}{100}+(-20.000)\frac{99}{100}=-10.000$$

Este valor se puede interpretar como la ganancia esperada del estudiante en una apuesta. Siguiendo este razonamiento se define:

> **Valor esperado de una variable aleatoria discreta**
> El valor esperado $E(X)$ de una variable aleatoria discreta X se define:
> $E(X)=\sum xP(X=x)$, donde $P(X=x)$ es la función de probabilidad.

El valor esperado de una variable aleatoria X es una media ponderada de los posibles valores que puede tomar X, donde las ponderaciones son las probabilidades de cada valor que X toma.

Ejemplo 4.37

Enunciado

Una compañía de seguros cubre durante un año el 100% del valor de un automóvil, cuyo costo es de $50’000.000. La información estadística muestra que, en el pasado reciente, solo tres de cada mil vehículos de estas características han sido robados cada año. Calcule el valor de la prima (P) para que el valor esperado de la ganancia de la compañía sea cero. Explique con base en qué supuestos se realiza este cálculo.

Solución

Se deben determinar los valores que la ganancia X puede tomar para luego calcular la probabilidad. Si el vehículo no es robado, la compañía tiene una ganancia igual a la prima. Si el vehículo es robado tiene una pérdida igual al costo del vehículo, menos la prima pagada por el tomador. Una pérdida de 50’000.000 – *P* se puede escribir como una ganancia de –(50’000.000 – *P*). Las probabilidades de ocurrencia de los dos eventos se muestran en la siguiente tabla:

Ganancia *X*	***P***	–(50’000.000 – P)
Probabilidad	$\frac{997}{1.000}$	$\frac{3}{1.000}$

Si el valor esperado es cero, es posible encontrar el valor de la *P* de la prima así:

$$E(X) = 0 = P \times \frac{997}{1.000} - (50'000.000 - P) \times \frac{3}{1.000}$$

Al solucionar esta ecuación se obtiene *P*= 150.000. Esto significa que el valor mínimo de la prima deber ser de $150.000 para obtener una ganancia esperada igual a cero. Se debe suponer que los porcentajes de robo y las condiciones de seguridad se mantienen constantes con respecto al tiempo y lugar donde se determinó la probabilidad de robo.▪

Varianza de una variable aleatoria discreta

La varianza *Var*(*X*) de una variable aleatoria discreta *X* se define:

Var(X) = $\sum (x - E(X))^2 P(X = x)$, donde $P(X = x)$ es la función de probabilidad.

Variables aleatorias continuas

Una variable que puede tomar cualquier valor en algún intervalo (o intervalos) del conjunto de los números reales y donde la probabilidad de que tome un valor determinado es cero se dice *continua*. A continuación, se estudian algunas de las características más importantes de las variables aleatorias continuas.

Enunciado

Dé ejemplos de variables aleatorias continuas.

Solución

1. Dosis de un medicamento en centímetros cúbicos que se suministra a un paciente.
2. Peso de los niños de una edad determinada.
3. Duración de los comerciales de televisión.
4. Duración de una pista musical en la radio.
5. Vida útil de un computador portátil.
6. Cantidad de combustible en el tanque de un vehículo.▪

Ejemplo 4.38

Enunciado

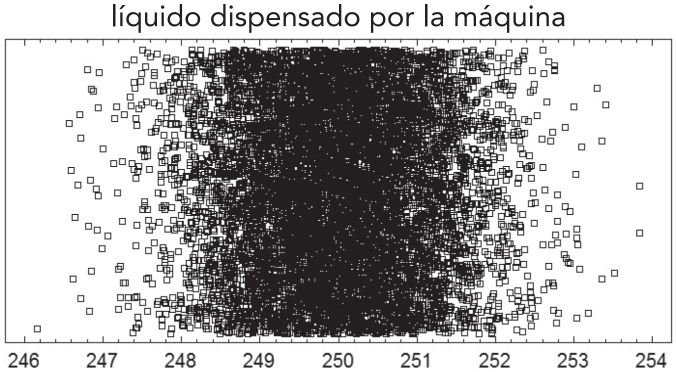
Una máquina está diseñada para llenar una botella de gaseosa de 250 ml. Explicar por qué la variable aleatoria X: = "cantidad de gaseosa dispensada por la máquina" es una variable aleatoria continua.

Solución

Al observar diferentes botellas de gaseosa del mismo tamaño, se puede notar que las alturas del líquido varían ligeramente. La máquina de llenado que dispensa el líquido lo hace con cierta aleatoriedad, es decir, algunas de ellas contendrán una cantidad ligeramente mayor o menor a la especificada en la etiqueta. La cantidad dispensada no se restringe a una colección finita de posibles valores prefijados. Puede tomar cualquier valor en un rango razonable; por ejemplo, entre 246 y 254 ml. Adicionalmente, la probabilidad de que una botella contenga exactamente 250 ml es cero, es decir, es prácticamente imposible encontrar una botella con ese contenido exacto, sin la más mínima desviación. Por las anteriores razones la variable aleatoria se puede suponer continua.▪

Con respecto al ejemplo anterior, suponga que se ha registrado la cantidad de líquido dispensado por la máquina en 10.000 botellas de 250 ml. El gráfico de dispersión de estos valores es semejante al que se muestra en la figura 4.3.

Figura 4.3. Gráfico de dispersión de la cantidad de líquido dispensado por la máquina



Se observa que la mayoría de botellas fueron llenadas con cantidades de líquido cercanas a 250 ml. También que es menos frecuente encontrar botellas que han sido llenadas con cantidades de líquido lejanas a la media.

El histograma de la figura 4.4 muestra un resumen del contenido de las 10.000 botellas. Ahora, asuma que el control de calidad del fabricante solamente permite que salgan a la venta las botellas que encuentran entre los 248 y 252 ml. El histograma de la figura 4.5 muestra la proporción de botellas que estuvieron fuera de los límites de calidad.

Figura 4.4 Histograma de la cantidad de líquido dispensado por la máquina

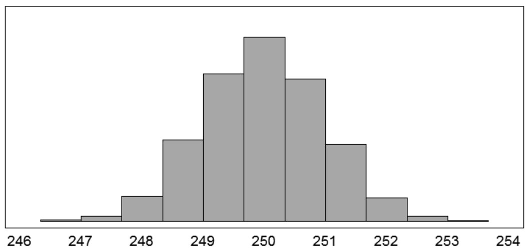
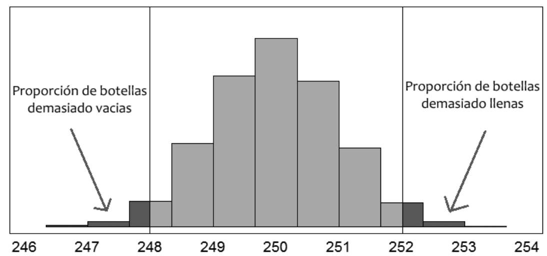


Figura 4.5 Proporciones de botellas fuera de los límites de calidad



Es de resaltar que los histogramas muestran información del pasado, en este caso, de la historia del proceso de llenado de las botellas. Un modelo matemático, para una variable continua, permitirá realizar predicciones sobre lo que podría ocurrir con el proceso en condiciones similares, por ejemplo, determinar la probabilidad de que una botella contenga una cantidad de líquido que esté fuera de las especificaciones de calidad.

La figura 4.6 muestra un modelo matemático que se podría ajustar a la distribución de frecuencias de las cantidades de líquido dispensadas por la máquina, y la figura 4.7 muestra cómo se podría usar este modelo para predecir las probabilidades de estar fuera de las especificaciones de probabilidad cuando las condiciones del proceso permanecen constantes.

Figura 4.6 Modelo matemático para ajustar a la distribución de frecuencias

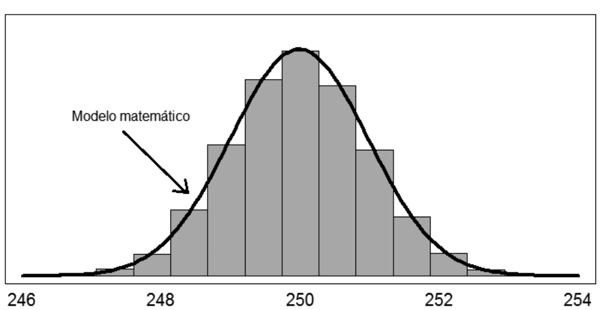
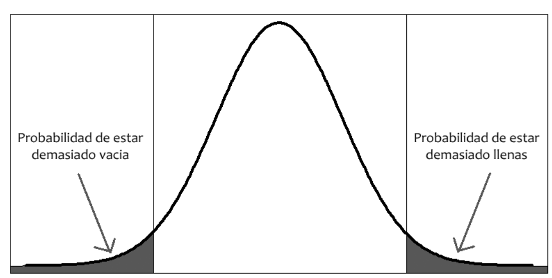


Figura 4.7 Evaluación de la probabilidad de estar fuera de las especificaciones del proceso



En general, un modelo para la distribución de probabilidad de una variable aleatoria continua consiste en una función matemática *f*, llamada *función de densidad de probabilidad*, tal que su gráfica se ajuste, según el criterio del investigador, a la distribución de frecuencias observada con los datos disponibles. Ya que se requiere modelar una probabilidad, es indispensable que el área bajo la función de densidad de probabilidad sea uno.

Función de densidad de probabilidad

Si X es una variable aleatoria continua, la densidad de probabilidad *f* de *X* es una función con las siguientes características:

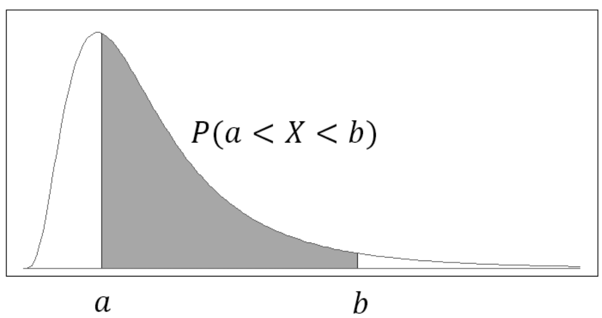
Es positiva.

El área entre la gráfica de *f* y el eje *X* es igual a 1.

La probabilidad de que la variable aleatoria *X* se encuentre entre dos valores *a* y *b*, $P(a < X < b)$, es el área entre la función *f*, el eje *X*, $x = a$ y $x = b$.

La probabilidad de que la variable aleatoria *X* se encuentre entre dos valores *a* y *b* corresponde al área mostrada en la figura 4.8.

Figura 4.8 Una función de densidad *f* y $P(a < X < b)$



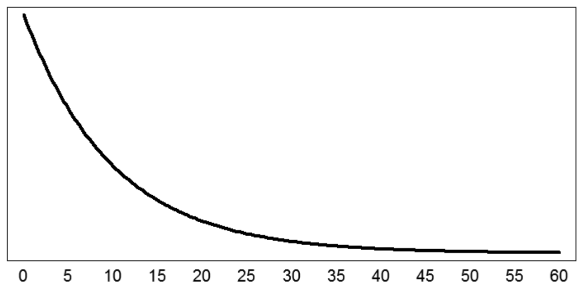
La **función de distribución** de una variable aleatoria continua *X* está definida para todo número real *x* como $F(x) = P(X \leq x)$. En el caso discreto, esta función se evalúa en un punto concreto x_0 sumando la probabilidad de los posibles valores de *X* menores o iguales a x_0. En el caso continuo, la función de distribución evalúa el área delimitada por la gráfica de la función de densidad a la izquierda de un valor particular.

La función de distribución $F(x)$ es la probabilidad de que la variable aleatoria tome un valor menor o igual a *x*. También se le conoce como la función de distribución acumulada de la variable aleatoria *X* hasta el valor *x*.

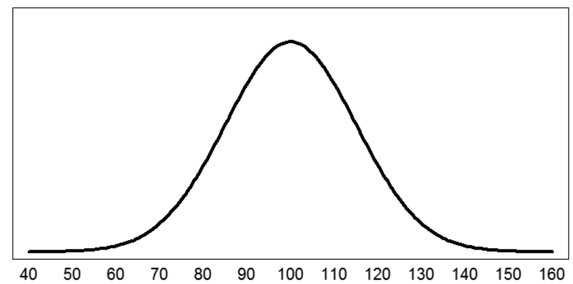
Es frecuente expresar las probabilidades de eventos como la función de distribución. Por ejemplo, si se desea calcular la probabilidad de que una variable aleatoria discreta *X* tome un valor menor o igual que b y mayor que a, esto es $P(a < X \leq b)$, se evalua la función de distribución en b y se le resta el valor de la función de distribución en a, es decir, $P(a < X \leq b) = F(b) - F(a)$.

Como ejemplos de modelos de funciones de probabilidad de variables aleatorias continuas se pueden mencionar los siguientes:

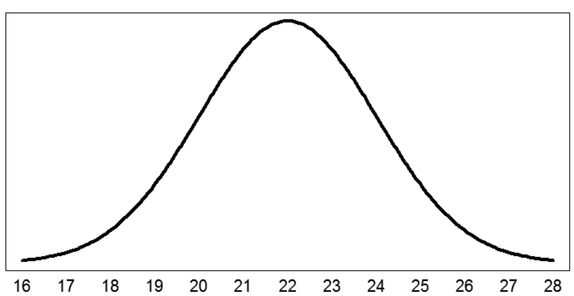
Tiempo de espera en una fila de supermercado (min.)



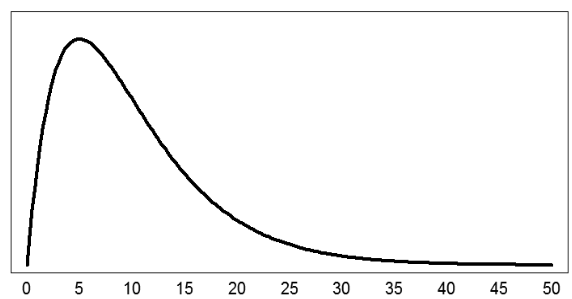
Coeficiente intelectual de una persona



Peso de un niño de 5 años en condiciones normales (kg)



Consumo mensual de agua de una familia en Cali (m^3)



De manera estricta se dice que una *variable aleatoria es* **continua** si su función de distribución F puede ser expresada como:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt\,, x \in \mathbb{R}$$

Donde *f* es una función integrable de dominio real y el rango es los números reales no negativos llamada función de densidad de probabilidad de la variable aleatoria *X* que satisface:

$$\int_{-\infty}^{x} f(t)\,dt = 1$$

Las probabilidades de una variable aleatoria continua se pueden expresar en términos de su función de densidad y de su función de distribución. Por ejemplo, la probabilidad de que la variable aleatoria tome un valor entre *a* y *b* es:

$$P(a \le X \le b) = \int_{a}^{b} f(t)\,dt = F(b) - F(b)$$

Para muchas funciones de densidad de probabilidad, el cálculo de probabilidades requiere cálculos complejos que pueden ser descartados con el uso de tablas de probabilidades acumuladas a las que se hará referencia en el siguiente capítulo.

Valor esperado y desviación estándar de una variable aleatoria continua

De manera informativa, se anotan las definiciones del valor esperado y la varianza de una variable aleatoria continua:

Valor esperado de una variable aleatoria continua

El valor esperado $E(X)$ de una variable aleatoria continua X se define:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Donde $f(x)$ es la función de densidad de probabilidad.

Varianza de una variable aleatoria continua

Si la variable aleatoria es continua su varianza Var (X) se calcula:

$$Var(X) = \int_{-\infty}^{\infty} (x - E(x))^2 f(x)dx$$

Donde $f(x)$ es la función de densidad de probabilidad.

Ejercicio 4.6 Variables aleatorias

1. Determine si las siguientes variables aleatorias son discretas o continuas:
   a. Número de lanzamientos de una moneda necesarios para obtener la primera cara.
   b. Tiempo de vida de una persona.
   c. Altura de una vaca.
   d. Número de hijos de una persona.
   e. Calificación final de un estudiante en estadística.
   f. Número de incendios de edificaciones en un año en Medellín.
   g. Temperatura máxima diaria en Barranquilla.

2. En una rifa, se van a vender 5.000 boletas a \$20.000 cada una. El premio es un carro de \$50’000.000. Si usted compra tres boletas, ¿cuál es la ganancia esperada?

3. Un hombre tiene cuatro llaves en su llavero, una de las cuales es la de su casa. En la noche, en la oscuridad, intenta abrir la puerta seleccionando las llaves en forma aleatoria y descartando las ya probadas.

a. Liste los eventos simples del espacio muestral.
b. Sea la variable aleatoria X: = "número de intentos necesarios para abrir la puerta". Halle la probabilidad de cada valor posible de X.
c. Construya un histograma de la variable aleatoria.
d. Encuentre la media y la desviación estándar de esta variable aleatoria.

4. Una empresa de seguros estima, con base en sus registros, que la probabilidad de que uno de sus asegurados sufra un accidente es de 0,12. Cuando ocurre un accidente, la probabilidad de que el daño alcance el 25% del costo del vehículo es de 0,75; del 50%, del vehículo 0,15 y de pérdida total 0,10.
   a. Determine todos los resultados posibles de los vehículos asegurados y calcule cada una de las probabilidades.
   b. Calcule e interprete el valor de la prima de un vehículo de \$40'000.000 que debería cobrar la empresa de seguros para que la ganancia esperada sea cero.

5. En el juego de un casino se puede apostar a un número entre uno y seis por un valor de \$5.000. Luego, se lanzan tres dados. Si el número apostado cae en los tres dados, se reciben \$10.000; si cae en dos de los tres dados se reciben \$7.000; si cae en uno solo de los tres dados se reciben \$5.000; si no cae en ninguno no se recibe nada. Sea X: = "ganancia del casino cada vez que un jugador apuesta".
   a. Encuentre y grafique la función de probabilidad de X.
   b. Determine e interprete el valor esperado de X.

6. De un grupo de cinco personas, dos mujeres y tres hombres, se eligen dos personas de forma aleatoria. Sea X: = "el número de mujeres en la muestra".
   a. Determine la función de probabilidad de X.
   b. Construya un histograma para la función de probabilidad.
   c. Encuentre e interprete la media de la variable aleatoria X.

7. Computadores PH tiene dos puntos de venta. La probabilidad de que se venda, en un día, un computador en el primer punto de venta es de 0,7 e independientemente, la probabilidad de venta, en un día, de un computador en el segundo punto de venta es de 0,8. Suponga que es igualmente probable vender un computador de escritorio o uno portátil. Los precios de un computador portátil y uno de escritorio, con especificaciones similares, son de \$1'400.000 y \$900.000, respectivamente. Sea X la cantidad de dinero que computadores PH vende en un día. Halle la distribución de X.

8. Un seguro entrega dos millones de pesos por la ocurrencia de un evento que en un periodo de un año se ha presentado una de cada cien veces. Sea Y la ganancia anual de la compañía de seguros por póliza y C la prima de esta.
   a. Calcule C tal que la ganancia esperada por póliza sea igual a cero.
   b. Si la compañía quiere que la ganancia esperada por póliza sea igual a \$50.000, determine la prima que debe cobrar.

CAPÍTULO 5
DISTRIBUCIONES DISCRETAS DE PROBABILIDAD. ALGUNOS CASOS PRÁCTICOS

Muchas situaciones reales pueden ser analizadas por variables aleatorias de tipo discreto. En este capítulo, se estudian tres distribuciones de probabilidad discretas: binomial, Poisson e hipergeométrica.

Distribución de probabilidad binomial

Situación inicial: aprobar un examen sin estudiar. Un estudiante responde al azar las cinco preguntas en un examen de selección múltiple, cada una con un valor de un punto. Si responde al azar y en cada pregunta hay cuatro opciones (a, b, c y d), ¿cuál es la probabilidad de aprobar?

Hoja de respuestas				
Pregunta	**a**	**b**	**c**	**d**
1	⊗	O	O	O
2	O	O	⊗	O
3	O	⊗	O	O
4	O	O	O	⊗
5	O	⊗	O	O

Como en cada pregunta solo hay una respuesta correcta, la probabilidad de acertar cada una es $\frac{1}{4}$ y la de errar $\frac{3}{4}$. En cada pregunta, el estudiante tiene dos posibilidades: acertar o errar. Existen $4 \times 4 \times 4 \times 4 \times 4 = 4^5 = 1.024$ maneras diferentes de responder el examen, pero el estudiante solo está interesado en el número de respuestas correctas. Por ejemplo, las maneras de acertar la primera pregunta y fallar las demás son $1 \times 3 \times 3 \times 3 \times 3 = 81$. El estudiante puede tener una respuesta correcta acertando solo la primera, segunda, tercera, cuarta o quinta pregunta. Esto da un total de $81 \times 5 = 405$ maneras diferentes de responder, con una sola respuesta correcta. Por tanto, la probabilidad de tener una respuesta correcta es:

$$5 \times \frac{1}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} = \frac{405}{1024} = 5\left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^4$$

El cálculo de la probabilidad según el número de respuestas correctas se presenta a continuación:

Respuestas correctas	**Ubicaciones de los A: aciertos y E: equivocaciones**		**Probabilidad**
5	AAAAA	$1 = \binom{5}{5}$ maneras de tener cinco aciertos	$\left(\frac{1}{4}\right)^5$
4	EAAAA AEAAA AAEAA AAAEA AAAAE	$5 = \binom{5}{4}$ maneras de tener 4 aciertos	$5\left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right)^1$
3	EEAAA EAEAA EAAEA EAEAA AEEAA AEAEA AEAAE AAEEA AAEAE AAAEE	$10 = \binom{5}{3}$ maneras de tener 3 aciertos	$10\left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^2$
2	AAEEE AEAEE AEEAE AEEEA EAAEE EAEAE EAEEA EEAAE EEAEA EEEAA	$10 = \binom{5}{2}$ maneras de tener 2 aciertos	$10\left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^3$
1	AEEEE EAEEE EEAEE EEEAE EEEEA	$5 = \binom{5}{1}$ maneras de tener 1 acierto	$5\left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^4$
0	EEEEE	$1 = \binom{5}{0}$ maneras de tener 0 aciertos	$\left(\frac{3}{4}\right)^5$

En resumen, si *X:* = "Número de respuestas correctas":

Respuestas correctas *x*	0	1	2	3	4	5
$P(X = x)$	0,23730	0,39551	0,26367	0,08789	0,01464	0,00098

La probabilidad de aprobar el examen respondiendo al azar es:

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) = 0,08789 + 0,01464 + 0,00098 = 0,10351$$

Se puede ver que $P(X = x) = \binom{5}{x}\left(\frac{1}{4}\right)^{x}\left(\frac{3}{4}\right)^{5-x}$. Este procedimiento se puede generalizar bajo ciertas condiciones, como las que se describen a continuación.

Características de una distribución de probabilidad binomial

Una variable aleatoria que cuenta el número de éxitos en n ensayos se dice que se distribuye o que es binomial si:

1. Consta de n ensayos independientes.
2. En cada ensayo ocurre solamente dos eventos llamados arbitrariamente éxito y fracaso.
3. La probabilidad de éxito p en cada ensayo permanece constante.

Si se satisfacen estas tres condiciones, la probabilidad de cada valor de la variable aleatoria se puede calcular:

$$P(X = x) = \binom{n}{x} \cdot p^{x}(1-p)^{n-x} \qquad x = 0, 1, \ldots, n$$

n es el número de ensayos, $\binom{n}{x}$ es el número de maneras en las que se pueden obtener x éxitos en n ensayos, p es la probabilidad de éxito, $1 - p$ es la probabilidad de fracaso, p^x es la probabilidad de los x éxitos y $(1-p)^{(n-x)}$ la probabilidad de los $n - x$ fracasos. Esta variable aleatoria tiene valor esperado y varianza $E(X) = np$ y $Var(X) = np(1-p)$, respectivamente.

Distribución de probabilidad binomial

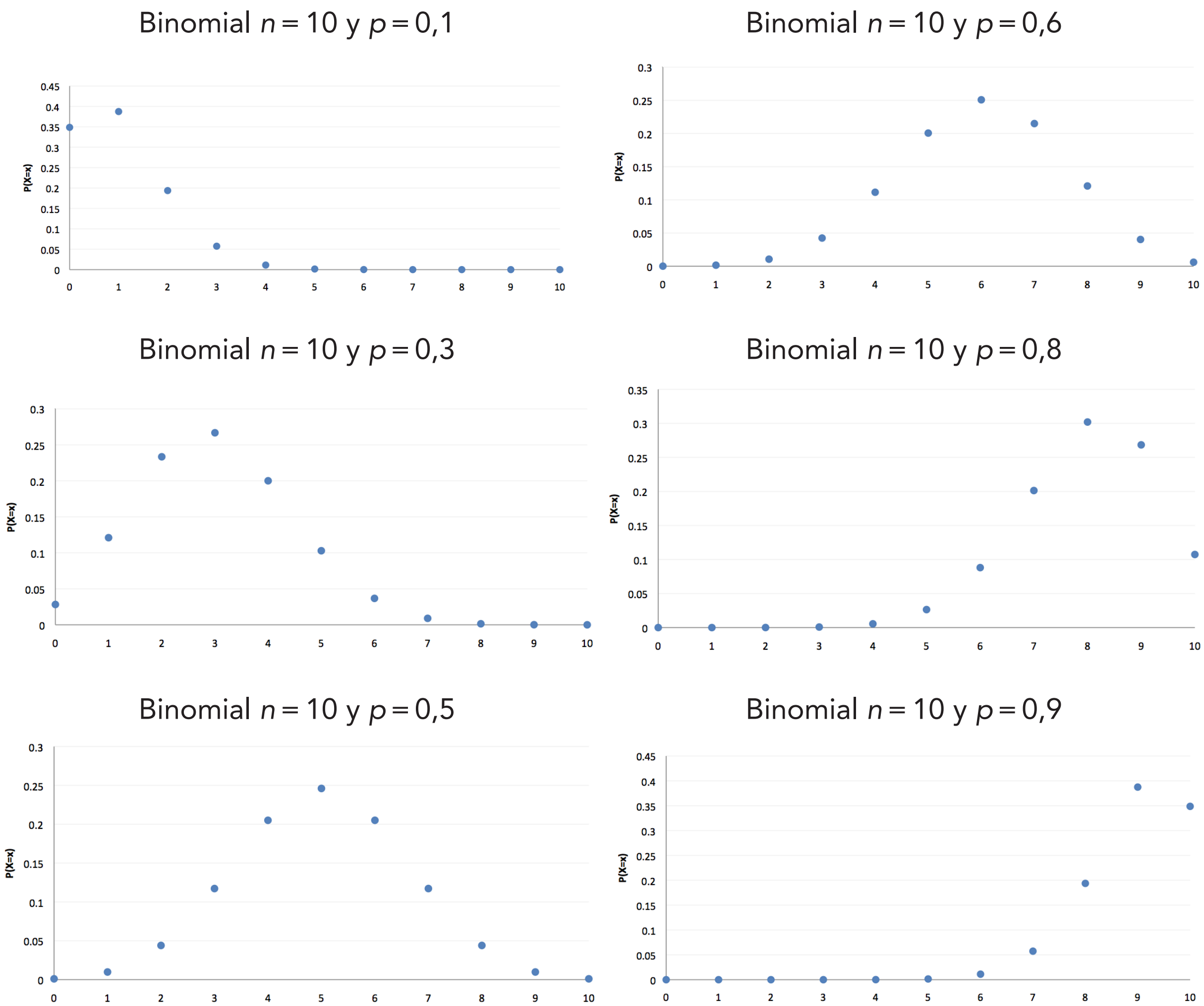
En una variable aleatoria con distribución binomial con n ensayos, la probabilidad de obtener x éxitos es:

$$P(X = x) = \binom{n}{x} \cdot p^{x}(1-p)^{n-x} \qquad x = 0, 1, \ldots, n$$

El valor esperado y varianza de dicha variable aleatoria son $E(X) = np$ *y* $Var(X) = np(1-p)$.

En la siguiente figura se muestran varias variables aleatorias binomiales todas con $n = 10$, y $p = 0,1$, $p = 0,3$, $p = 0,5$, $p = 0,8$ y $p = 0,9$.

Figura 5.1. Funciones de probabilidad binomial con $n = 10$



Observe cómo cambia el sesgo de las gráficas a medida que la probabilidad aumenta. La gráfica de la distribución es simétrica cuando la probabilidad de éxito es igual a la de fracaso.

Por lo general, para el cálculo de la probabilidad no se trabaja con la función de probabilidad $P(X = x)$, sino con la función de distribución de probabilidad $F(x) = P(X \leq x)$, como se verá en los siguientes ejemplos.

Ejemplo 5.1

Enunciado

Calcular la función de distribución de la situación inicial y expresar las siguientes probabilidades en términos de ella.

a. Probabilidad de obtener tres o menos respuestas correctas.
b. Probabilidad de obtener menos de tres respuestas correctas.
c. Probabilidad de obtener exactamente tres correctas.
d. Probabilidad de aprobar el examen.

Solución

Como se mostró, la función de probabilidad es:

x	0	1	2	3	4	5
$P(X=x)$	0,23730	0,39551	0,26367	0,08789	0,01465	0,00098

Y su función de distribución $F(x)=P(X\leq x)$:

x	0	1	2	3	4	5
$F(x)=P(X\leq x)$	0,23730	0,63281	0,89648	0,98438	0,99902	1

a. $P(X\leq 3)=F(3)=0,98438$
b. $P(X<3)=F(2)=0,89648$
c. $P(X=3)=F(3)-F(2)=0,98438-0,89648$
d. $P(X\geq 3)=F(5)-F(2)=1-0.89648$ ▪

Resulta más cómodo trabajar con la función de distribución $F(x)=P(X\leq x)$, porque los cálculos se reducen máximo a una resta. En el apéndice, tabla 2, se incluyen los valores de la función de distribución de probabilidad binomial para diferentes valores de n y p.

Ejemplo 5.2

Enunciado

El 40% de las quejas de los usuarios de Movistar se deben a problemas de facturación. Se seleccionan al azar diez quejas presentadas por usuarios y se determina si son referentes a problemas de facturación o no. Utilice la tabla de probabilidades acumuladas de la distribución binomial para calcular la probabilidad de los siguientes eventos:

a. Todas las quejas se deben a problemas de facturación.
b. Menos de la mitad se deben a problemas de facturación.
c. Entre tres y siete se deben a problemas de facturación.

d. Más de ocho se deben a problemas de facturación.

e. Por lo menos una se debe a problemas de facturación.

Sea X: = "número de quejas debido a problemas de facturación". X tiene una distribución de probabilidad binomial con $n = 10$ y $p = 0{,}4$. La siguiente tabla muestra la función de distribución de dicha variable aleatoria.

n=10

x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
0	0,9044	0,5987	0,3487	0,1074	0,0282	0,0060	0,0010	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0
1	0,9957	0,9139	0,7361	0,3758	0,1493	0,0464	0,0107	0,0017	0,0001	0,0000	0,0000	0,0000	0,0000	1
2	0,9999	0,9885	0,9298	0,6778	0,3828	0,1673	0,0547	0,0123	0,0016	0,0001	0,0000	0,0000	0,0000	2
3	1,0000	0,9990	0,9872	0,8791	0,6496	0,3823	0,1719	0,0548	0,0106	0,0009	0,0000	0,0000	0,0000	3
4	1,0000	0,9999	0,9984	0,9672	0,8497	0,6331	0,3770	0,1662	0,0473	0,0064	0,0001	0,0000	0,0000	4
5	1,0000	1,0000	0,9999	0,9936	0,9527	0,8338	0,6230	0,3669	0,1503	0,0328	0,0016	0,0001	0,0000	5
6	1,0000	1,0000	1,0000	0,9991	0,9894	0,9452	0,8281	0,6177	0,3504	0,1209	0,0128	0,0010	0,0000	6
7	1,0000	1,0000	1,0000	0,9999	0,9984	0,9877	0,9453	0,8327	0,6172	0,3222	0,0702	0,0115	0,0001	7
8	1,0000	1,0000	1,0000	1,0000	0,9999	0,9983	0,9893	0,9536	0,8507	0,6242	0,2639	0,0861	0,0043	8
9	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9990	0,9940	0,9718	0,8926	0,6513	0,4013	0,0956	9
10	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	10

a. $P(X = 10) = F(10) - F(9) = 1 - 0{,}9999 = 0{,}0001$

b. $P(X < 5) = F(4) = 0{,}6331$

c. $P(3 \leq X \leq 7) = F(7) - F(2) = 0{,}9877 - 0{,}1673 = 0{,}8204$

d. $P(X > 8) = F(10) - F(8) = 1 - 0{,}9983 = 0{,}0017$

e. $P(X \geq 1) = F(10) - F(0) = 1 - 0{,}006 = 0{,}994$ ▪

Ejemplo 5.3

Enunciado

En una universidad se ha determinado que el 80% de sus estudiantes se declaran no fumadores. Se selecciona una muestra aleatoria de diez estudiantes. Calcule la probabilidad de que:

a. Ninguno sea fumador.

b. Exactamente la mitad sean fumadores.

c. Más de tres sean fumadores.

d. Todos sean fumadores.

Primera solución

Considere la variable aleatoria X: = "número de fumadores en la muestra de diez estudiantes". X tiene una distribución binomial con $n = 10$, donde la probabilidad de éxito es el porcentaje de fumadores $p = 0{,}2$. Por tanto:

a. $P(X = 0) = F(0) = 0{,}1074$

b. $P(X = 5) = F(5) - F(4) = 0{,}9936 - 0{,}9672 = 0{,}0264$

c. $P(X > 3) = F(10) - F(3) = 1 - 0{,}8791 = 0{,}1209$

d. $P(X = 10) = F(10) - F(9) = 1 - 1 = 0$

Segunda solución

Considere la variable aleatoria Y: = "número de no fumadores en la muestra de diez estudiantes". Y tiene una distribución binomial con $n = 10$, donde la probabilidad de éxito es el porcentaje de no fumadores $p = 0{,}8$. Por tanto:

a. $P(Y = 10) = F(10) - F(9) = 1 - 0{,}8926 = 0{,}1074$

b. $P(Y = 5) = F(5) - F(4) = 0{,}0328 - 0{,}0064 = 0{,}0264$

c. $P(Y \leq 6) = F(6) = 0{,}1209$

d. $P(Y = 0) = F(0) = 0$ ▪

Ejercicio 5.1 Distribución binomial

1. Determine si las variables aleatorias definidas en los literales del a. al i. tienen una distribución binomial, si podrían serlo bajo alguna circunstancia o si definitivamente no pueden serlo. Justifique su respuesta. Intente asignar valores para n y p.

   a. Número de días de lluvia en un periodo de un mes en Quibdó.

   b. Número de accidentes de tránsito en la carrera séptima en un lunes cualquiera.

   c. Número de lanzamientos de una moneda hasta obtener la primera cara.

   d. Número de mujeres en una muestra de diez personas seleccionadas sin reposición en un salón de treinta personas.

   e. Número de mujeres en una muestra de diez personas seleccionadas con reposición en un salón de treinta personas.

f. Número de automóviles que tienen color rojo de los próximos cincuenta que pasen por una vía determinada.

g. Número de semillas que germinan de un total de veinte plantadas.

h. Número de individuos en una clase que se contagian cuando uno de sus compañeros está enfermo de varicela —suponga que está en la etapa de transmisión de la enfermedad—.

i. Número de personas con factor RH sanguíneo positivo en una muestra aleatoria de cinco personas.

2. Sea X una variable aleatoria binomial con $n = 5$ y $p = 0{,}3$.

   a. Determine una expresión para la función de probabilidad.

   b. Trace un gráfico de la función de probabilidad.

   c. Determine la media y la varianza de la distribución.

   d. Halle $P(X < 1)$, $P(X \leq 2)$, $P(X > 3)$ y $P(X > 4)$.

3. Use una tabla de probabilidades binomiales acumuladas para hallar:

   a. $P(X < 2)$ con $n = 2$ y $p = 0{,}01$.

   b. $P(X \leq 4)$ con $n = 5$ y $p = 0{,}10$.

   c. $P(X > 2)$ con $n = 10$ y $p = 0{,}50$.

   d. $P(X \geq 3)$ con $n = 15$ y $p = 0{,}90$.

   e. $P(X < 5)$ con $n = 25$ y $p = 0{,}99$.

4. Las complicaciones de la gripa ocurren más a menudo en pacientes mayores de 64 años de edad. La pulmonía es la complicación más grave de la gripa y es mortal en uno de cada cien casos en este grupo de edad. Suponga que cincuenta personas mayores de 64 años han contraído pulmonía. Encuentre:

   a. Una expresión para la función de probabilidad.

   b. La media y la varianza de la distribución.

   c. La probabilidad de que mueran menos de dos personas.

5. Un estudiante responde al azar un examen que consta de diez preguntas de selección múltiple, cada una con cinco opciones. Si cada punto tiene el mismo valor y se aprueba con seis o más respuestas correctas, ¿cuál es la probabilidad de que el estudiante apruebe el examen?

6. Para poner a prueba un detergente se realizan dos experimentos. El primero consiste en presentar dos prendas de ropa, una nueva y otra idéntica pero lavada cinco veces con el detergente, a diez amas de casa para que decidan cuál es la prenda nueva. En el segundo experimento a una misma ama de casa se le presentan diez pares de prendas, cada par con una prenda nueva y otra idéntica pero lavada cinco veces con el detergente para que decida en cada caso cuál es la prenda nueva. ¿Alguno de estos experimentos es binomial?

7. El 5% de los balances contables de las pymes presentan algún tipo de error. Se selecciona una muestra aleatoria de veinte balances de pymes. Calcule la probabilidad de que el número de balances con errores sea:
   a. Ninguno.
   b. Exactamente tres.
   c. Más de tres.
   d. Entre dos y cuatro.

8. Se sabe que aproximadamente el 90% de las amas de casa prefieren un detergente en presentación de 500 g, contra el mismo producto, pero en una presentación de 1.500 g. Se lleva a cabo un sondeo entre veinte amas de casa seleccionadas al azar y se define la variable aleatoria X: = "número de amas de casa que prefieren la presentación económica". Determine la probabilidad de:
   a. Al menos doce amas de casa prefieren la presentación de 500 g.
   b. No más de diez amas de casa prefieren la presentación 500 g.

9. El contrato de compra de un contenedor de juguetes electrónicos establece que este debe contener un máximo de 5% de estos defectuosos. Para establecer la calidad de los juguetes, se crea un plan de inspección que consiste en probar veinte juguetes del contenedor, seleccionados al azar y sin reposición y se acepta el contenedor si el número de defectuosos es menor o igual a uno.
   a. Determine bajo qué condición este experimento se ajusta a uno de tipo binomial.
   b. Analice cómo varía la probabilidad de aceptar el lote cuando las proporciones reales de juguetes con algún defecto en el contenedor son 0,05 – 0,10 y 0,20.
   c. ¿Es favorable para el vendedor este plan de inspección de la calidad de los juguetes? ¿Lo es para el comparador?
   d. ¿Cómo mejorar el plan de inspección?

10. Un casino confisca el dado de una mesa de juego, porque cree que se ha modificado, de manera que la probabilidad de caer en uno no es $\frac{1}{6}$. El casino realiza una prueba, lanzando el dado sesenta veces y cuenta el número de veces que cae uno. Si usted fuera el investigador del casino, ¿cómo utilizaría esta cuenta para decidir si el dado ha sido modificado?

11. Un experimento binomial tiene una probabilidad de éxito de 0,2. ¿Cuántas veces se debe repetir para que la probabilidad de al menos cinco éxitos sea mínimo 0,579?

12. Un banco reporta que el cobro mediano por manejo de tarjetas de crédito es de \$40.000. Se toma una muestra aleatoria de veinte clientes del banco con tarjeta de crédito y se registra el cobro por dicho concepto, tal como se muestra a continuación.

48.400	43.600	51.200	56.300
55.900	58.600	39.900	48.800
55.400	44.500	46.500	41.500
40.700	45.100	46.100	39.400
47.100	47.900	50.600	48.100

¿Qué tan probable es obtener una muestra como esta si en realidad el cobro mediano es de \$40.000?

Distribución de probabilidad de Poisson

El matemático Simeón-Denis Poisson, en una investigación sobre la probabilidad de los juicios en materias criminales en 1838, propuso una distribución de probabilidad para modelar la ocurrencia de un evento en una unidad de tiempo o espacio.

En la distribución de probabilidad de Poisson, la variable aleatoria cuenta la ocurrencia de eventos puntuales sobre una base continua: unidad de tiempo o espacio. Consiste en un modelo de probabilidad con las siguientes características:

1. Produce a largo plazo un promedio λ constante de eventos.
2. Los eventos ocurren de forma independiente, es decir, el número de ocurrencias en determinado intervalo de tiempo o espacio no altera la probabilidad de ocurrencia posterior.

La variable aleatoria considerada en una distribución de Poisson es $X:=$ "número de eventos en un periodo o espacio".

Distribución de probabilidad de Poisson

Si un evento ocurre en promedio λ veces en cierto periodo o espacio, la probabilidad de que un evento ocurra x veces en el mismo periodo o espacio es:

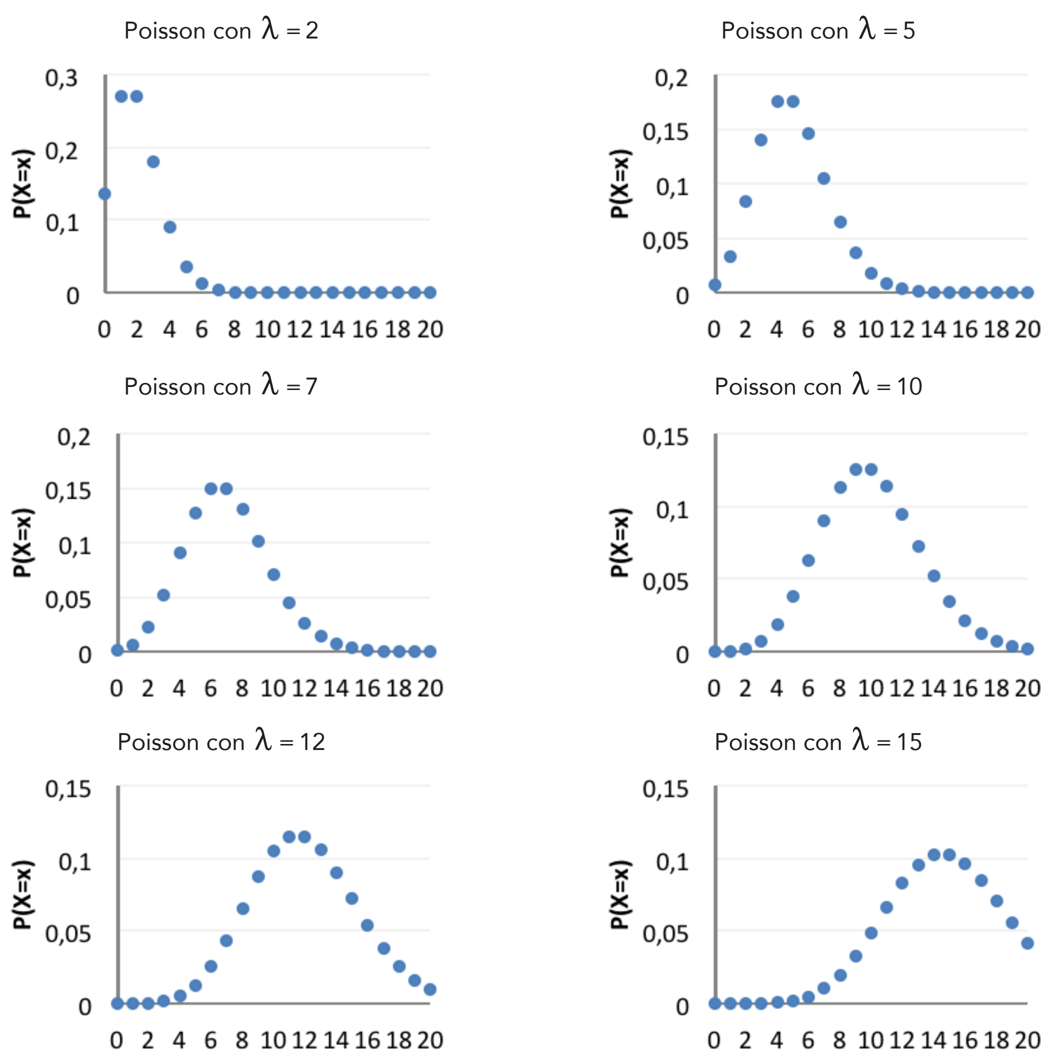
$$P(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!}$$

Para $x=0,1,2,3,\cdots$ y donde $e=2{,}718281828\cdots$ es la base de los logaritmos naturales.

El valor esperado y varianza de dicha variable aleatoria son $E(X)=\lambda$ y $Var(X)=\lambda$.

En la figura 5.2 se muestran varias variables aleatorias de Poisson con diferentes valores de λ.

Figura 5.2 Funciones de probabilidad de Poisson



Observe cómo cambia el sesgo de las gráficas a medida que el promedio aumenta. La gráfica de la distribución tiende a ser simétrica cuando el promedio es muy grande.

Ejemplo 5.4

Son ejemplos de variables aleatorias de Poisson:

a. Número de errores en las ediciones de un periódico.
b. Número de bacterias presentes en muestras de laboratorio.
c. Defectos por metro en la producción de telas.
d. Número de personas que llegan a una caja en un supermercado cada minuto.
e. Número de fallas diarias de una máquina.▪

La distribución de Poisson se aproxima a la distribución binomial cuando $n \to \infty$, de manera que los eventos puedan ocurrir en un conjunto continuo de elementos y $p \to 0$ para que el número promedio de eventos np ocurridos permanezca constante.

Ejemplo 5.5

Enunciado

El número de choques simples por mes en la intersección de la avenida Boyacá con la avenida Suba es una variable aleatoria de Poisson con media $\lambda = 3$. Determinar la probabilidad de los siguientes eventos:

a. Ningún choque simple en un mes.
b. Más de cinco choques simples en un mes.
c. Diez o más choques simples en un semestre.

Solución

La variable aleatoria es X: = "número de accidentes mensuales en la intersección". La función de probabilidad de variable es:

$$P(X = x) = \frac{e^{-3}3^x}{x!}$$

a. $P(X = 0) = \frac{e^{-3}3^0}{0!} = 0,0498$. Este resultado se puede ver en el apéndice, tabla 3 de probabilidades acumuladas de la distribución de probabilidad de Poisson. Para ubicar la probabilidad se busca la columna correspondiente a la media $\lambda = 3$, como $P(X = 0) = F(0)$ se observa la fila correspondiente al valor de $X = 0$.

Es decir, $P(X=0)=0{,}0498$. En conclusión, la probabilidad de que no ocurra ningún choque en un mes es de 0,0498.

b. $P(X>5)=1-F(5)=1-0{,}9161=0{,}0839$. Lo que significa que la probabilidad de que ocurran más de cinco choques en un mes es de 0,0839.

c. En un semestre se espera un promedio de λ = 18, la probabilidad pedida es $P(X\geq 10)=1-F(9)$. En la columna correspondiente, λ = 18 de la tabla de probabilidades acumuladas de Poisson se puede ver que $P(X\geq 10)=1-0{,}0154=0{,}9846$. Lo que significa que este es un evento muy probable.▪

Ejemplo 5.6

Enunciado

El número de ausencias a clase de un estudiante de estadística es una variable aleatoria de Poisson con media dos por corte. Calcule la probabilidad de que un estudiante seleccionado al azar:

a. No falte a clase en el corte.

b. No falte a clase en el semestre.

Solución

Sea X: = "el número de ausencias de un estudiante de estadística por corte". La función de probabilidad de variable es:

$$P(X=x)=\frac{e^{-2}2^{x}}{x!}$$

a. No faltar a clases en un corte es equivalente a tener cero ausencias o $x=0$. $P(X=0)=F(0)=0{,}1353$ se puede ver en el apéndice, tabla 2.

b. Para examinar la probabilidad en un semestre, se considera otra variable de Poisson con promedio λ = 6. $P(X=0)=F(0)=0{,}0025$. La probabilidad de que el estudiante no falte en el semestre es apenas de 0,0025.▪

Ejercicio 5.2 Distribución de Poisson

1. Sea X una variable aleatoria de Poisson con $\lambda=5$

   a. Determine una expresión para la función de probabilidad.

   b. Trace un gráfico para la función de probabilidad y otro para la función de distribución.

c. Determine la media y la varianza de la distribución.

d. Halle $P(X < 1)$, $P(X \leq 5)$, $P(X > 3)$ y $P(X > 4)$.

2. Use una tabla de probabilidades acumuladas de Poisson para hallar:
   a. $P(X < 2)$ con $\lambda = 0{,}3$
   b. $P(X \leq 3)$ con $\lambda = 1$
   c. $P(X > 2)$ con $\lambda = 2$
   d. $P(X \geq 3)$ con $\lambda = 10$
   e. $P(X < 5)$ con $\lambda = 20$

3. En los últimos, dos años se ha observado que en una librería el promedio de ventas mensuales por internet es de diez.
   a. Determine la probabilidad de que durante un mes no se realicen ventas por internet.
   b. De las dos siguientes opciones, ¿cuál es la más probable?
      i. Hacer más de diez ventas por internet en un mes.
      ii. Hacer menos de diez ventas por internet en un mes.
   c. Calcule la probabilidad de que, por lo menos, se realicen cinco ventas por internet en un mes.

4. En la época de admisiones en la Universidad, a la Oficina de Información las llamadas entran con una frecuencia de una cada dos minutos. Calcule la probabilidad de:
   a. Recibir tres llamadas en cinco minutos.
   b. No recibir llamadas en un periodo de cinco minutos.

5. El número de pacientes que llega a un hospital sigue una distribución de Poisson. Si el número promedio es de noventa cada hora, determine la probabilidad de que en un minuto lleguen por lo menos tres pacientes.

6. El número de pacientes que se presentan diariamente a un hospital y que requieren el uso de equipos de rayos *X* sigue una distribución de Poisson con media $\lambda = 15$. Encuentre las probabilidades de los siguientes eventos:
   a. Se presentan más de veinte pacientes durante un día.
   b. Se presentan entre diez y veinte pacientes durante un día.

7. Suponga que el número de veces que una persona se enferma de gripa entre los 15 y 25 años, durante un año, sigue una distribución de Poisson con media $\lambda = 2$. Calcular la probabilidad de que ocurran los siguientes eventos.
   a. La persona no se enferma de gripa durante un año.
   b. La persona se enferma más de dos veces durante un año.

8. En Bogotá, sobre la carrera 7ª entre calles 6 y 32 se presentan en promedio 1,5 choques simples de automóviles cada día, entre semana. Suponiendo que la ocurrencia de choques simples de accidentes diarios en esta vía sigue una distribución de probabilidad de Poisson, calcule la probabilidad de que se presenten:
   a. Tres choques simples un día entre semana.
   b. Al menos, un choque simple un día entre semana.
   c. Al menos, un choque simple en cinco días (de lunes a viernes).

9. La prevalencia de tuberculosis en Colombia en el año 2008 fue de veinticuatro casos por cada 100.000 habitantes. Si se realiza una prueba para detectar la enfermedad a mil individuos colombianos seleccionados aleatoriamente y se supone que la prevalencia de la enfermedad permanece constante, determine la probabilidad de los siguientes eventos:
   a. Ninguna de las personas examinadas tiene tuberculosis.
   b. Al menos una tiene tuberculosis.
   c. Menos de diez tienen tuberculosis.

10. El número promedio de accidentes laborales por mes en una construcción es de 1,8. Determine:
   a. La probabilidad de que en el próximo mes ocurran exactamente dos accidentes.
   b. La probabilidad de que en el próximo mes ocurran tres o más accidentes.
   c. Los supuestos bajo los cuales estos cálculos son acertados.

11. El control de calidad de una empresa de telas ha encontrado que el número promedio de imperfecciones por rollo de veinticinco metros es de 0,2. Las empresas que manufacturan prendas de vestir con estas telas pagan un 25% menos en aquellos rollos de tela que tengan dos o más imperfecciones. Calcule la probabilidad de los siguientes eventos:
   a. A un rollo de tela se le aplica el descuento del 25%.
   b. A un rollo de tela no se aplica el descuento del 25%.

12. Un estudiante fuma en promedio 4,4 cigarrillos por día. Si la distribución de Poisson es un buen modelo para esta situación, calcule las probabilidades de los siguientes eventos:
    a. Fumar diez cigarrillos en un determinado día.
    b. Fumar de cinco a máximo nueve cigarrillos en un día.

Distribución de probabilidad hipergeométrica

En la sección de ejercicios de la distribución de probabilidad binomial, se mencionó el uso de este tipo de distribución para resolver el problema de la inspección de un contenedor con 5% de juguetes con algún defecto con base en una muestra. En la práctica, la inspección de un contenedor se realiza sin reposición, puesto que no tiene sentido volver a probar un elemento dos veces. Si el número de juguetes es muy grande, la probabilidad de defectuoso no cambia significativamente de extracción en extracción. Sin embargo, cuando la población de la cual se extraen las muestras es pequeña, la probabilidad de éxito puede variar mucho a medida que se seleccionan elementos de ella. La distribución de probabilidad hipergeométrica tiene en cuenta que la probabilidad de éxito puede no ser constante en un muestreo sin reposición.

Ejemplo 5.7

Enunciado

Explique por qué en una muestra pequeña las probabilidades de éxito varían mucho en una selección sin reposición.

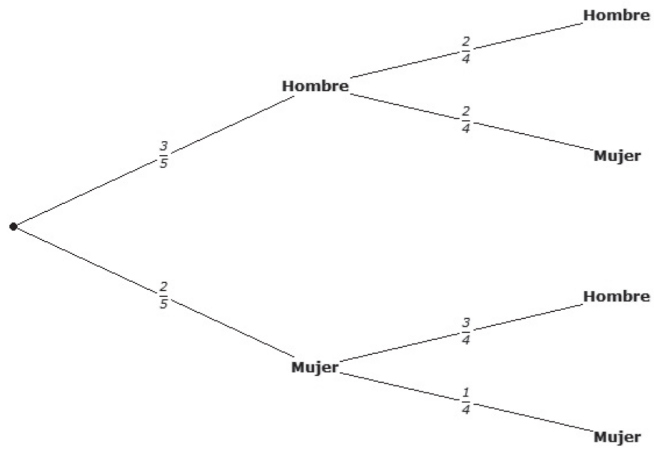
Solución

Suponga que tiene una muestra pequeña compuesta por tres hombres y dos mujeres y que el éxito consiste en seleccionar una mujer. La probabilidad de elegir al azar una mujer es $\frac{2}{5} = 0,4$. La probabilidad de elegir una segunda mujer habiendo seleccionado una previamente es $\frac{1}{4} = 0,25$. Es decir, la probabilidad de éxito pasó de 0,4 a 0,25 (una reducción de 0,15). Si la muestra es grande, por ejemplo, 3.000 hombres y 2.000 mujeres, la probabilidad de éxito en el primer intento es $\frac{2.000}{5.000} = 0,4$, pero la probabilidad de éxito de una segunda mujer habiendo seleccionado una previamente es $\frac{1.999}{4.999} \approx 0,3998$ —lo que implica una reducción insignificante de 0,0002—. Es decir que la probabilidad de éxito cambia muy poco.▪

Ejemplo 5.8

Enunciado

De un grupo de cinco personas, tres hombres y dos mujeres, se seleccionan dos personas para conformar un comité. Determine las probabilidades de los posibles valores de la variable aleatoria X: = "número de mujeres en la muestra".



Primera solución

Las probabilidades en primera y segunda etapa se muestran en el anterior diagrama de árbol. Los cuatro posible resultados HH, HM, MH y MM tienen probabilidades $\frac{6}{20}$, $\frac{6}{20}$, $\frac{6}{20}$ y $\frac{2}{20}$. En estos eventos el número de mujeres es 0, 1, 1 y 2. Por tanto, los posibles valores de la variable aleatoria son 0, 1 y 2 y sus probabilidades son las siguientes:

x	0	1	2
$P(X=x)$	$\frac{6}{20}$	$\frac{12}{20}$	$\frac{2}{20}$

Segunda solución:

Hay $\binom{5}{2} = 10$ maneras de elegir a las dos personas. Las maneras de elegir, por ejemplo, exactamente una mujer son $\binom{2}{1}\binom{3}{1} = 6$. Luego la probabilidad de elegir exactamente una mujer es $\frac{6}{10}$. Siguiendo el mismo razonamiento:

$$P(X=0) = \frac{\binom{2}{0}\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10} \qquad P(X=1) = \frac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}} = \frac{6}{10} \qquad P(X=2) = \frac{\binom{2}{2}\binom{3}{0}}{\binom{5}{2}} = \frac{1}{10} \;\blacksquare$$

Esta estrategia de solución se puede esquematizar así:

Las personas se seleccionan de una población tamaño $N=5$. Hay $C=2$ elementos de la población que son mujeres. N-C=3 elementos de la población no son mujeres. $n=2$ es número de personas que serán seleccionadas. Entonces, la probabilidad de que en la muestra se encuentren x mujeres es:

$$P(X=x) = \frac{\binom{2}{x}\binom{3}{2-x}}{\binom{5}{2}}$$

Distribución de probabilidad hipergeométrica

Si de una población de tamaño N, donde C elementos poseen una característica (y por lo tanto N-C no la poseen) se extrae una muestra tamaño n y se define la variable aleatoria X:= "número de elementos de la muestra con la característica". Entonces, la probabilidad de que la variable aleatoria tome un valor particular x es:

$$P(X=x)=\frac{\binom{C}{x}\binom{N-C}{n-x}}{\binom{N}{n}};\quad x=0,1,2,3,\ldots,\min\{C,n\}$$

N Es el tamaño de la población.
C Es el número de elementos de la población que poseen la característica.
n Es el tamaño de la muestra.
x Es el número de elementos de la muestra que poseen la característica.

El valor esperado y la varianza de esta variable aleatoria son:

$$E(X)=n\frac{C}{N}\text{ y } Var(X)=n\frac{C}{N}\frac{N-C}{N}\frac{N-n}{N-1}$$

Ejemplo 5.9

Enunciado

En una tienda hay veinte celulares Nokia 3050 idénticos exteriormente, pero tres están defectuosos. No se sabe cuáles son. Determinar la probabilidad de vender al menos un celular defectuoso en las dos próximas ventas de esta marca.

Solución

Sea X: = "el número de celulares defectuosos vendidos". Esta variable aleatoria puede tomar los valores 0,1 o 2 en las $n=2$ próximas ventas. Hay $N=20$ celulares en la tienda y $C=3$ son defectuosos. La probabilidad de vender x celulares defectuosos es:

$$P(X=x)=\frac{\binom{3}{x}\binom{17}{2-x}}{\binom{20}{2}}$$

La probabilidad de vender al menos uno defectuoso es:

$$P(X=1)+P(X=2)=\frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}}+\frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}}\approx 0{,}2842\blacksquare$$

Ejemplo 5.10

Enunciado

En una represa hay un número N desconocido de peces. Un biólogo captura, marca y libera una muestra aleatoria de $C = 90$ de ellos. Después de un tiempo prudencial, para que los peces se dispersen libremente, captura nuevamente $n = 10$ de ellos también de forma aleatoria y observa que $X = 2$ de ellos están marcados. Estimar el número N de peces en el estanque.

Solución

Si se asume que la población de peces permanece constante entre las dos capturas y que la probabilidad de atrapar uno marcado es la misma que uno no marcado, la variable aleatoria $X{:}=$ "número de peces marcados en la segunda muestra" sigue una distribución hipergeométrica con función de probabilidad:

$$P(X = x) = \frac{\binom{90}{x}\binom{N-90}{10-x}}{\binom{N}{10}}$$

Los posibles valores de x son $\{0,1,2,\ldots,10\}$. Se puede estudiar la distribución de probabilidad de la anterior familia de variables hipergeométricas para diversos valores de N. La estimación de N es aquella que haga máxima la ocurrencia de $X = 2$ —a partir de la hipótesis de que ha ocurrido lo más probable—.

En la tabla 5.1 se muestran las funciones de probabilidad para diversos valores de N. Se resaltan los valores más probables en cada distribución. Por ejemplo, en la última fila, se observa que si el verdadero valor de N fuera 520, lo más probable debería ser encontrar un solo pez marcado en la segunda muestra de $n = 10$. Observe que a medida que el tamaño de la población N aumenta, cada vez es más probable encontrar menos peces marcados en la segunda muestra.

Aunque en la tabla se muestran valores de N de 20 en 20, se estima que el tamaño de la población de peces debe estar entre 332 y 498. Esta es una estimación que maximiza la probabilidad de ocurrencia del valor de la variable aleatoria, dado por la muestra observada. Es importante resaltar que esta es tan solo una estimación razonable de N. Otros autores, como Ross (2002), proponen tomar como estimación de la proporción de peces marcados $\frac{C}{N}$, a la proporción de peces marcados en la muestra $\frac{x}{n}$, con lo que se obtiene que $N = \frac{nC}{x}$, que para este ejemplo da $N \approx 450$. El lector puede juzgar la conveniencia de cada estimación.▪

Tabla 5.1 Funciones de probabilidad hipergeométricas $R = 90$, $n = 10$

N/x	0	1	2	3	4	5	6	7	8	9	10
100	0,000	0,000	0,000	0,000	0,000	0,001	0,008	0,052	0,202	0,408	0,330
120	0,000	0,000	0,000	0,002	0,013	0,054	0,147	0,261	0,291	0,183	0,049
140	0,000	0,000	0,004	0,020	0,071	0,162	0,250	0,255	0,166	0,062	0,010
160	0,000	0,003	0,017	0,062	0,147	0,234	0,251	0,180	0,082	0,022	0,003
180	0,001	0,008	0,041	0,115	0,209	0,253	0,209	0,115	0,041	0,008	0,001
200	0,002	0,019	0,073	0,167	0,244	0,240	0,160	0,072	0,021	0,003	0,000
220	0,004	0,033	0,109	0,209	0,256	0,212	0,119	0,045	0,011	0,002	0,000
240	0,008	0,052	0,146	0,239	0,253	0,180	0,087	0,029	0,006	0,001	0,000
260	0,013	0,073	0,180	0,258	0,240	0,150	0,064	0,018	0,003	0,000	0,000
280	0,019	0,095	0,209	0,269	0,222	0,124	0,047	0,012	0,002	0,000	0,000
300	0,026	0,118	0,235	0,271	0,203	0,102	0,035	0,008	0,001	0,000	0,000
320	0,035	0,142	0,256	0,269	0,183	0,084	0,026	0,006	0,001	0,000	0,000
340	0,044	0,164	0,272	0,263	0,164	0,069	0,020	0,004	0,000	0,000	0,000
360	0,054	0,186	0,284	0,254	0,146	0,057	0,015	0,003	0,000	0,000	0,000
380	0,065	0,207	0,294	0,243	0,131	0,047	0,012	0,002	0,000	0,000	0,000
400	0,076	0,226	0,300	0,232	0,116	0,039	0,009	0,001	0,000	0,000	0,000
420	0,087	0,244	0,304	0,221	0,104	0,033	0,007	0,001	0,000	0,000	0,000
440	0,099	0,261	0,305	0,209	0,092	0,028	0,006	0,001	0,000	0,000	0,000
460	0,111	0,276	0,305	0,197	0,083	0,023	0,005	0,001	0,000	0,000	0,000
480	0,123	0,290	0,304	0,186	0,074	0,020	0,004	0,000	0,000	0,000	0,000
500	0,135	0,302	0,301	0,175	0,066	0,017	0,003	0,000	0,000	0,000	0,000
520	0,147	0,314	0,298	0,165	0,059	0,014	0,002	0,000	0,000	0,000	0,000

Ejercicio 5.3 Distribución hipergeométrica

1. Explique el significado de los términos “con reposición” y “sin reposición”.

2. Explique las diferencias entre las distribuciones binomial e hipergeométrica.

3. Sea X una variable hipergeométrica con $N=10$, $C=5$ y $n=4$.
   a. Determine una expresión para la función de probabilidad.
   b. Trace un gráfico para la función de probabilidad y otro para la función distribución.
   c. Determine la media y la varianza de la distribución.
   d. Halle $P(X<1)$, $P(X\leq 3)$, $P(X>2)$ y $P(X\geq 3)$.

4. En una urna hay cuatro bolas rojas y tres negras. Se seleccionan al azar tres bolas y se determina el número de bolas negras. Halle la distribución de probabilidad de esta variable aleatoria.
5. En un grupo de diez personas hay cuatro mujeres. Si se seleccionan al azar tres personas, determine la probabilidad de no escoger mujeres.

CAPÍTULO 6
DISTRIBUCIÓN DE PROBABILIDAD NORMAL

Características de la distribución de probabilidad normal

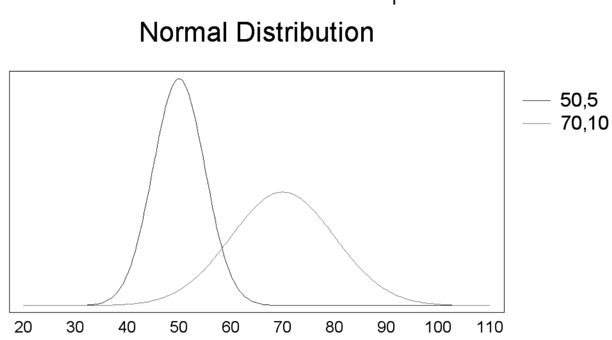
La distribución de probabilidad normal es de gran importancia en muchas técnicas estadísticas, sobre todo, cuando se realizan inferencias. La distribución de probabilidad normal es un modelo de una variable aleatoria continua, es decir, de aquella que puede tomar cualquier valor en algún intervalo (o intervalos) del conjunto de los números reales y cuando la probabilidad de que la variable tome un valor determinado es cero. La normal es una distribución simétrica cuya gráfica tiene forma de campana y es útil para modelar variables continuas si sus distribuciones de frecuencias también son simétricas y acampanadas. El uso de la distribución de probabilidad normal es bastante frecuente, porque las distribuciones de frecuencias de muchas variables continuas adoptan forma de campana. Fue propuesta por el matemático francés Abraham de Moivre, en 1733, para aproximar probabilidades binomiales cuando el parámetro n es grande. También se le conoce como campana de Gauss, en honor al matemático alemán Carl Friedrich Gauss.

En una distribución normal la función de densidad de probabilidad es:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Donde σ es la desviación estándar de la variable aleatoria y μ la media. Esta distribución queda caracterizada por la media y la desviación estándar.

Figura 6.1 Dos distribuciones de probabilidad normal



Por ejemplo, en la figura 6.1, se muestran dos distribuciones normales de pesos de individuos en dos poblaciones. La distribución de la izquierda corresponde a los pesos de un grupo de mujeres con media 50 kg y desviación de 5 kg; la de la derecha a varones con media 70 kg y desviación de 10 kg. Se observan diferencias tanto en los centros de las distribuciones como en sus dispersiones. Casi la totalidad de los pesos de las mujeres están entre los 35 y 65 kg, mientras que los varones, en su mayoría, tienen pesos de entre 40 y 100 kg.

Ejemplo 6.1

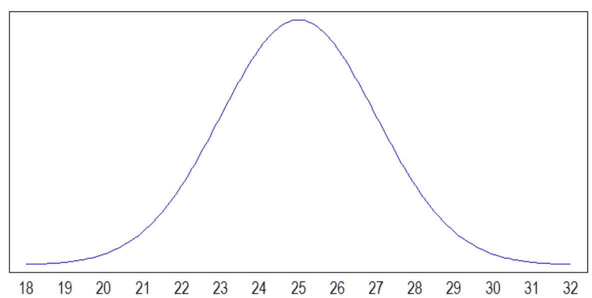
Enunciado

¿Qué significa afirmar que los pesos de los niños de 5 años siguen una distribución normal con media de 25 kg y una desviación de 2 kg? Escriba la función de densidad de esta distribución.

Solución

Significa que los pesos son una variable aleatoria continua y que la distribución es simétrica en forma de campana con centro en 25 kg. Usando la regla de las distribuciones simétricas, es posible establecer que casi la totalidad de los niños de 5 años tienen un peso de entre 19 y 31 kg. Tres desviaciones estándar en torno a la media. La figura 6.2 es la gráfica de la distribución de los pesos de los niños. La función de densidad de la distribución de los pesos de los niños de 5 años es:

Figura 6.2 Distribución del peso de los niños de 5 años



$$f(x) = \frac{1}{\sqrt{2\pi(2)^2}} e^{-\frac{1}{2}\left(\frac{x-25}{2}\right)^2} = \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{8}(x-25)^2}$$

Ejemplo 6.2

Enunciado

Se selecciona al azar un niño de 5 años de una población y se define X: = "peso del niño". Suponga que la distribución de X es normal con media de 25 kg y desviación estándar de 2 kg. Represente gráficamente e interprete las siguientes probabilidades:

a. $P(X \leq 21)$

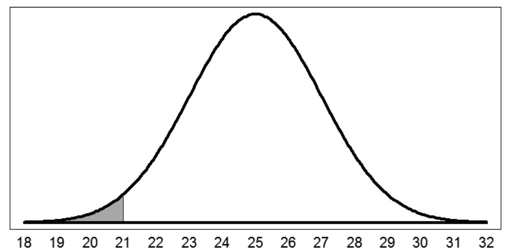
b. $P(X \geq 22)$

c. $P(23 \leq X \leq 27)$

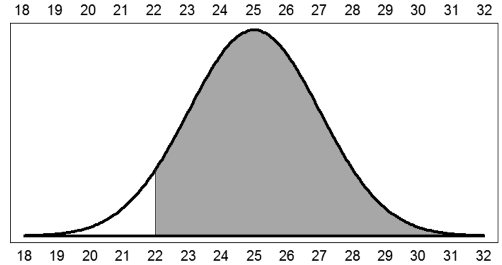
Solución

Contrario a lo que sucede con las distribuciones discretas, en las continuas y, en particular, en la distribución de probabilidad normal, no tiene sentido preocuparse por $P(X = x)$, porque la probabilidad de cualquier valor particular es cero. En este tipo de distribuciones es de interés la probabilidad asociada con intervalos de la variable aleatoria.

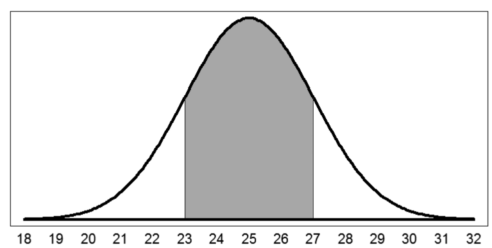
a. $P(X \leq 21)$ es la probabilidad de que el niño pese 21 kg o menos.



b. $P(X \geq 22)$ es la probabilidad de que el niño pese 22 kg o más.



c. $P(23 \leq X \leq 27)$ es la probabilidad de que el niño pese entre 23 y 27 kg.▪



La evaluación de las anteriores probabilidades resulta compleja si se desarrollan a partir de la función de densidad de probabilidad normal. Para superar este escollo, se cuenta con una tabla de probabilidades acumuladas de una distribución normal particular, la distribución de probabilidad normal estándar.

Distribución de probabilidad normal estándar (*Z*)

A una variable normal con media $\mu = 0$ y desviación $\sigma = 1$ se le denomina **normal estándar** o ***Z***. Su función de densidad de probabilidad está dada por la siguiente expresión:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

La importancia de esta distribución de probabilidad radica en que cualquier distribución de probabilidad normal X con media μ y desviación σ puede ser convertida en una normal estándar Z por medio de la transformación $Z = \frac{x - \mu}{\sigma}$. A esta transformación se le llama estandarización o tipificación.

A la función de distribución de probabilidad $F(x)$ de una normal estándar también se le suele denotar $\Phi(x)$. Las probabilidades acumuladas de la función de distribución normal estándar se muestran en la tabla 4 del apéndice.

Estandarización

Si X es una variable aleatoria normal con media μ y desviación σ, la transformación $\frac{X - \mu}{\sigma}$ es una variable aleatoria normal estándar.

$$Z = \frac{X - \mu}{\sigma}$$

Ejemplo 6.3

Enunciado

Utilizar la tabla 4 del apéndice, probabilidades acumuladas de la normal estándar, para calcular:

a. $P(Z \leq 1,64)$.

b. $P(Z > 1,96)$.

c. $P(1,64 < Z \leq 1,96)$.

d. P_{95}.

Se reproduce parcialmente la tabla 4 del apéndice.

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169

Solución

a. La tabla muestra la probabilidad acumulada de la distribución normal estándar en un valor específico z, $F(z)$. $P(Z \leq 1,64) = F(1,64)$, la cual es la probabilidad acumulada hasta el valor 1,64. Para calcularla se ubica en la primera columna el valor 1,6 y se avanza hasta la columna que coincida con el segundo decimal. Allí aparece el valor 0,9495. Lo que quiere decir que $P(Z \leq 1,64) = F(1,644) = 0,9495$.

b. Esta probabilidad es el área bajo la curva normal estándar a la derecha de 1,96. Primero se ubica el valor 1,9 y se avanza hasta que la columna coincida con el segundo decimal. Allí aparece el valor 0,9759. Se debe tener en cuenta que este valor es F (1,96) —área bajo la curva normal estándar a la izquierda de 1,96—. Como el área bajo la curva normal es uno, se concluye que $P(Z > 1,96) = 1 - F(1,96) = 1 - 0,975 = 0,025$.

c. Siguiendo el mismo razonamiento anterior se deduce que:

$$P(1,64 < Z \leq 1,96) = F(1,96) - F(1,64) = 0,975 - 0,9495 = 0,0255\cdot$$

d. Para hallar el percentil 95, P_{95} se busca dentro de la tabla el valor más cercano a 0,95. Este es 0,9495 que se ubica en la fila de 1,6 y la columna 0,04. Por lo tanto $P_{95} = 1,64$.▪

Ejemplo 6.4

Enunciado

En la actualidad, una entidad financiera del Estado otorga créditos a todos sus afiliados con una tasa de interés de 1,5% mensual. Ahora desea crear un sistema de tasas de interés diferenciales según el ingreso, es decir, los afiliados con menores ingresos recibirán una disminución en la tasa, mientras que aquellos con mayores ingresos un aumento. El plan de la entidad financiera se resume en la siguiente tabla:

Situación del afiliado	Acción	Porcentaje de afiliados	Nueva tasa
Con menores ingresos	Disminución de la tasa	5%	1%
Ingresos normales	Sin cambio	90%	1,5%
Con mayores ingresos	Aumento de la tasa	5%	2%

Si la distribución de los ingresos de los afiliados es normal con media $1'500.000 y desviación estándar $400.000, determine el ingreso máximo para tener una disminución de la tasa y el valor del ingreso desde donde se aplicará un aumento.

Solución

Sea X: = “el ingreso de un afiliado”. X sigue una distribución normal con media \$1’500.000 y desviación estándar \$400.000. Suponga que el valor máximo para tener un descuento en la tasa de interés es x_D, por tanto, $P(X < x_D) = 0,05$. El valor correspondiente a x_D en una normal estándar es –1,64485 porque $P(Z < -1,64485) = 0,05$. Esto significa que la estandarización de x_D es –1,64485.

$$\frac{x_D - 1'500.000}{400.000} = -1,64485$$

$$x_D = 842.059$$

Es decir, las personas con un ingreso inferior o igual a \$842.059 podrán tomar un crédito con una tasa de interés del 1%.

Ahora se hallará el valor x_A desde donde se aplicará el aumento de la tasa. Se sigue un razonamiento similar en donde solo basta saber que $P(Z > 1,64485) = 0,05$. Con esto se obtiene que $x_A = 2'157.941$, es decir, personas con ingresos superiores o iguales a \$2’157.941 recibirán créditos con una tasa de interés del 2%.▪

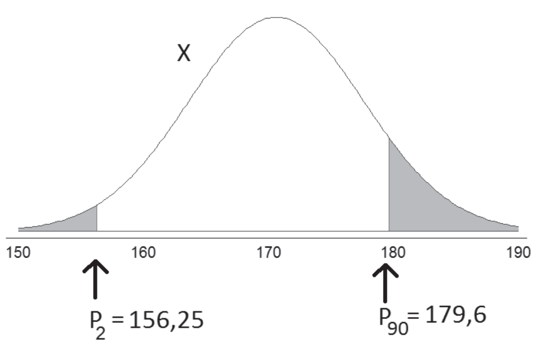
Ejemplo 6.5

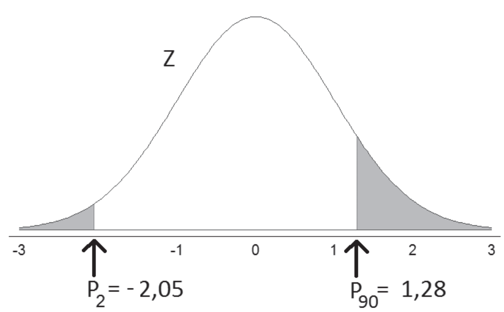
Enunciado

Se sabe que las estaturas de los varones de 18 años en la ciudad de Armenia sigue una distribución normal. Determine la media y la desviación estándar de las estaturas de estos individuos si el 2% de ellos mide menos de 156,25 cm y el 10% más de 179,6 cm.

Solución

Sea X la distribución de las estaturas de los varones de 18 años en la ciudad de Armenia.





Los valores 156,25 cm y 179,6 cm son los percentiles 2 y 90 de la distribución de X (figura de arriba a la izquierda). Estos mismos percentiles en la distribución normal estándar son $P_2 = -2{,}05$ y $P_{90} = 1{,}28$ (figura de arriba a la derecha). Es decir, la estandarización de 156,25 cm y 179,6 cm es –2,05 y 1,28, respectivamente. Así que:

$$\frac{156{,}25-\mu}{\sigma}=-2{,}05$$

$$\frac{179{,}6-\mu}{\sigma}=1{,}28$$

Al resolver estas dos ecuaciones, cada una con dos incógnitas se obtiene que el promedio de las estaturas de los varones de 18 años de la ciudad de Armenia es 170,6 cm con una desviación estándar aproximada de 7 cm.▪

Ejercicio 6. Distribución normal

1. Determine el área bajo la curva normal estándar a la izquierda de los siguientes valores:
   a. $z = 1{,}64$.
   b. $z = 1{,}96$.
   c. $z = -1{,}64$.
   d. $z = -1{,}96$.

2. Determine el área bajo la curva normal estándar a la derecha de los siguientes valores:
   a. $z = 1$.
   b. $z = -1$.
   c. $z = 3{,}29$.
   d. $z = -2{,}53$.

3. Si Z es una variable normal estándar, calcule las siguientes probabilidades:
   a. $P(Z < 1{,}5)$.
   b. $P(Z < -2{,}3)$.
   c. $P(Z > 1{,}55)$.
   d. $P(Z > -0{,}11)$.
   e. $P(-2 < Z < 1{,}5)$.
   f. $P(-1{,}64 < Z < 1{,}64)$.

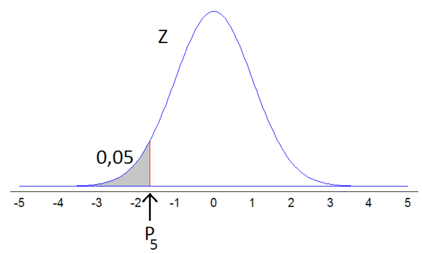
4. Determine el área debajo de la curva normal estándar entre los siguientes valores:

   a. $z = -1{,}64$ y $z = 1{,}64$.

   b. $z = -1{,}96$ y $z = 1{,}96$.

   c. $z = -0{,}5$ y $z = 2{,}3$.

5. Encuentre el percentil de la variable aleatoria normal estándar señalado en la siguiente figura:



6. Encuentre y ubique en una gráfica los siguientes percentiles de una variable aleatoria normal estándar:

   a. Percentil 95.

   b. Percentil 75.

   c. Percentil 50.

   d. Percentil 25.

7. Encuentre z_0 tal que si Z es una variable normal estándar:

   a. $P(Z < z_0) = 0{,}9$.

   b. $P(Z > z_0) = 0{,}97$.

   c. $P(-z_0 < Z < z_0) = 0{,}95$.

   d. $P(-z_0 < Z < z_0) = 0{,}99$.

8. Si X es una variable normal con $\mu = 60$ y $\sigma^2 = 25$, calcule las siguientes probabilidades:

   a. $P(X < 75)$.

   b. $P(X < 30)$.

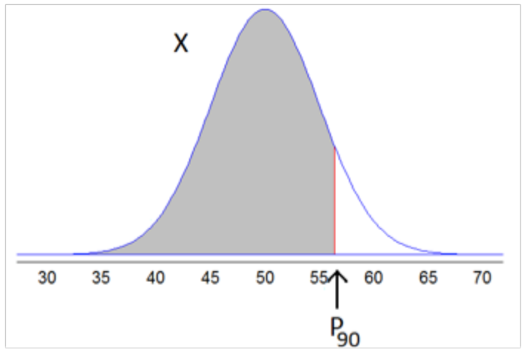
   c. $P(X > 80)$.

   d. $P(X > 40)$.

   e. $P(40 < X < 80)$.

   f. $P(20 < X < 50)$.

9. Encuentre el percentil de la variable aleatoria normal X con $\mu = 50$ y $\sigma = 5$ señalado en la siguiente figura:



10. Si X es una variable normal con $\mu = 60$ y $\sigma^2 = 25$, determine los siguientes percentiles:
   a. Percentil 5.
   b. Percentil 25.
   c. Percentil 50.
   d. Percentil 75.
   e. Percentil 96.

11. X es una variable normal con media μ y varianza $\sigma^2 = 16$ y $(X > 60) = 0{,}0062$. Determine el valor de μ.

12. La ficha técnica de un automóvil de motor 1.200 cc afirma que puede recorrer 50 km con un galón de gasolina. Cuando el operario de una estación de servicio vende un galón de gasolina, la cantidad real de gasolina vendida es una variable aleatoria normal con media un galón y desviación 0,05 galones. Determine la probabilidad de los siguientes eventos:
   a. El automóvil recorre más de 55 km con un galón de gasolina.
   b. El automóvil recorre menos de 40 km con un galón de gasolina

13. Se encontró que el tiempo requerido para completar un examen de matemáticas en la Universidad tiene distribución normal con media de 75 minutos y desviación de 12 minutos. Calcule la probabilidad de los siguientes eventos:
   a. Un estudiante requiere menos de una hora en completar el examen.
   b. Un estudiante requiere entre una hora y una hora y media en completar el examen.

14. Los pesos de un paquete de salchichas están distribuidos normalmente con una media de 1.000 g y una desviación estándar de 50 g. Calcule el porcentaje de paquetes de salchichas que pesan:
    a. Más de 1.100 g.
    b. Entre 1.055 y 1.100 g.
    c. ¿Sería extraño encontrar un paquete con menos de 900 g?

15. El peso de los niños varones de 7 años matriculados en colegios departamentales de Boyacá en el año 2010 sigue una distribución normal con media 22 kg y una desviación de 2 kg, mientras que el de las niñas tiene una media de 21 kg y una desviación de 1,8 kg. En un programa de salud, se desea atender al 5% de los niños con pesos más bajos (posiblemente con desnutrición) y al 1% de los niños con pesos más altos (con sobrepeso o en riesgo). Determine cómo se debe decidir si un niño o una niña debe ser atendido por el programa de salud.

16. Suponga que los puntajes de un examen están distribuidos normalmente con media 76 y desviación estándar 15. El 15% de los estudiantes, los mejores, obtienen una A como nota y el 10%, los peores, pierden el curso y obtienen una P.
    a. Hallar el puntaje mínimo para obtener una A.
    b. Hallar el puntaje mínimo para aprobar.

17. La cantidad de gaseosa depositada por una máquina en una botella es una variable aleatoria normal. Se sabe que el 25% de las botellas sobrepasa los 310 ml y el 5% no alcanza los 290 ml. Determine la media y la desviación de la cantidad de gaseosa depositada por esta máquina.

ESQUEMA PARA UN PROYECTO DE ESTADÍSTICA DESCRIPTIVA

Realizar un proyecto es un buen recurso de aprendizaje, porque amplía la perspectiva desde la cual los estudiantes se han acercado a la estadística. En esta parte de libro se plantea un esquema muy simple y reducido para llevar a cabo un proyecto de estadística descriptiva. La idea básica es recoger, describir y analizar información real. De esta manera, los estudiantes podrán aplicar lo aprendido durante el curso con una información que les sea significativa.

Como las técnicas estadísticas de selección de muestras o el diseño de cuestionarios no han sido abordados aún por los estudiantes y tampoco hacen parte de este curso, el proyecto es apenas un ejercicio de entrenamiento en esta dirección. Una vez terminado el proyecto, se espera que los estudiantes adquieran competencias básicas en el manejo de datos y, por lo menos, se hayan visto envueltos en alguna discusión sobre:

- El planteamiento de preguntas de investigación.
- La formulación de hipótesis.
- La construcción de una base de datos.
- El manejo de software estadístico.

Este proyecto es más una construcción de los estudiantes que un ejercicio completamente descrito y delimitado por el profesor. La creatividad será un elemento importante en su desarrollo.

Una vez se hayan estudiado los tres primeros capítulos del libro, se puede dar comienzo al proyecto con la conformación de equipos de estudiantes. No es una tarea corriente que se hace de un momento a otro, sino que se espera un avance secuencial soportado por el trabajo en equipo y la asesoría del profesor. Al finalizar, cada equipo presentará sus resultados de manera oral y por escrito.

Seguramente, será necesario contar con más apoyo bibliográfico. Para ampliar la visión que se tiene sobre cómo desarrollar un proyecto como este, se recomienda hacer la lectura del capítulo dos del libro *Elementos de muestreo* (Scheafer, 2006) y del capítulo uno de *Muestreo, diseño y análisis* (Lohr, 2000).

Aunque cada proyecto debería tratarse de un modo particular se plantean a continuación unas pautas que pueden ser útiles para comenzar.

Parte 1: elección del tema del proyecto

Escoja un tema que sea de interés para usted y para su grupo y del que se pueda indagar por medio de una encuesta o un experimento sencillo. Plantee una pregunta central que quiera responder y que sea el hilo conductor de su proyecto. Una vez hecho esto:

- Plantee un objetivo sencillo y claro.
- Determine la población sobre la que desea hacer su estudio.
- Piense qué elementos de la población estarán en la muestra y decida el tamaño de esta.
- Decida si va a realizar entrevistas personales, telefónicas, por correo electrónico o si lo que necesita su proyecto es una observación y medición directa.
- Discuta todos estos aspectos con sus compañeros y preséntelos por escrito a su profesor.

Parte 2: diseño de un cuestionario o un instrumento de captura de información

Haga una lista de preguntas que puedan ser incluidas en una encuesta de manera que respondan al objetivo planteado. Una vez decida cuáles serán incluidas, redáctelas como preguntas cerradas. El posterior análisis se facilitará si las opciones de cada pregunta son excluyentes y exhaustivas. Para poder aplicar las técnicas descriptivas estudiadas en el curso, se sugiere contar con el suficiente número de variables, tanto de tipo cuantitativo como cualitativo.

Cada proyecto es diferente y tal vez un cuestionario no sea el medio adecuado para la obtención de la información. Por ejemplo, si va a realizar un experimento sencillo, será necesario construir un instrumento diferente de captura de información. En cualquier caso, pida la opinión y ayuda de su profesor. Por último, haga una prueba piloto con un número pequeño de encuestas y evalúe la eficacia de su instrumento.

Parte 3: trabajo de campo

Planifique con detalle la manera como obtendrá la información y fíjese un límite de tiempo para esta tarea. Recuerde que el tiempo puede ser muy determinante, por ejemplo, en encuestas de opinión. Prolongar la duración de la recolección de la información puede no ser conveniente.

Parte 4: construcción de una base de datos

Una vez haya obtenido toda la información, sintetícela en una hoja de Excel en la cual las filas correspondan a los elementos muestreados y las columnas a las variables que se han medido.

Parte 5: analice los datos

Aplique lo aprendido en los tres primeros capítulos del libro. Si es posible piense en una aplicación relacionada con el capítulo cuatro. Construya diferentes tipos de gráficas y calcule varios estadísticos o plantee regresiones con sus datos. Recuerde que todo lo anterior debe estar en función de la pregunta que ha sido planteada inicialmente y de sus objetivos. Escriba sus conclusiones con referencia a la muestra que analizó.

Parte 6: comunique sus resultados

Prepare una breve exposición con los principales resultados de su proyecto y presente un informe a manera de artículo.

PRUEBAS DE CONOCIMIENTOS

Prueba 1

Preguntas de selección múltiple con única respuesta.

1. De las siguientes, la afirmación **falsa** es:
   a. Una muestra es un subconjunto de los elementos de la población.
   b. Un estadístico es cualquier medida calculada sobre la muestra.
   c. El número de cuentas corrientes con saldo en rojo en un banco es una variable discreta.
   d. Si una variable es discreta es porque es cuantitativa.
   e. Las medidas que se calculan sobre las muestras se denominan parámetros.

2. En una encuesta se preguntó, entre otras cosas, sobre la nacionalidad de los entrevistados. A continuación, se muestra la tabla de frecuencias de la nacionalidad.

Nacionalidad	Frecuencia absoluta
Colombiana	10
Panameña	20
Chilena	7
Boliviana	3

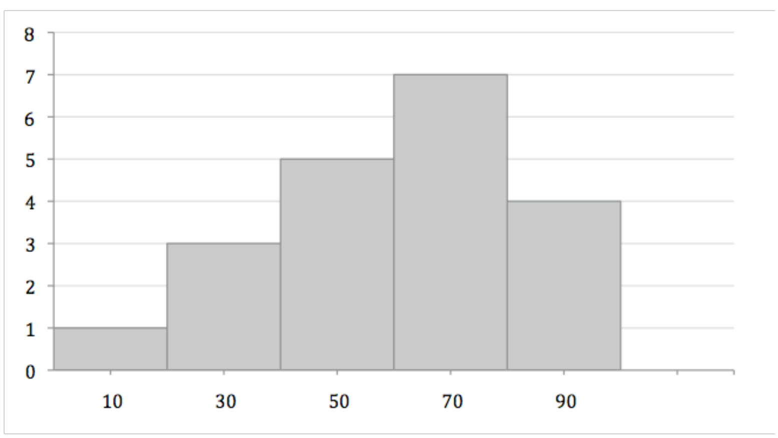
Con respecto a la anterior tabla, la afirmación **verdadera** es:
   a. En la encuesta, el elemento sobre el que se mide la variable nacionalidad es el país de origen.
   b. La variable nacionalidad es cuantitativa.
   c. La representación gráfica para esta tabla se llama histograma.
   d. El porcentaje de personas que no son de Chile es de 82,5%.
   e. La proporción de chilenos y bolivianos es de 10%.

6. El siguiente diagrama de tallo y hojas corresponde a las edades de un grupo de personas.

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

La afirmación **correcta** es:
   a. La moda de los datos es 5.
   b. La media de los datos es aproximadamente 2,68.
   c. La varianza muestral es aproximadamente 9,15238…
   d. La desviación estándar muestral es 83,766…
   e. La distribución de las edades es es más o menos simétrica.

4. Un grupo de estudiantes realizó una prueba física y al finalizarla se les asignó un puntaje entre 00 y 100. El puntaje de aprobación es 60. Los resultados se muestran en el siguiente histograma.



La afirmación que **no es cierta** es:
   a. La gráfica tiene un ligero sesgo a la derecha.
   b. La moda de las calificaciones es setenta.
   c. Veinte estudiantes presentaron la prueba.
   d. El 45% de los estudiantes no aprobó.
   e. La media de las puntuaciones es aproximadamente sesenta.

Prueba 2

Preguntas de selección múltiple con única respuesta.

1. Suponga que es más fácil saber la cantidad de grasa de un yogurt que la cantidad de calorías, y que se utiliza una recta de regresión para estimar las calorías con base en la cantidad de grasa. A partir de esta información **es cierto** que:
   a. La cantidad de calorías es la variable independiente.
   b. La cantidad de grasa es la variable dependiente.
   c. La asociación entre la cantidad de grasa y la cantidad de calorías es lineal negativa.
   d. Se está asumiendo que la cantidad de calorías depende de la cantidad de grasa.
   e. El diagrama de dispersión debe revelar una asociación lineal positiva.

2. Si se supone que existe una asociación lineal negativa fuerte entre el salario y el número de hijos de una persona, entonces **es cierto** que:
   a. Personas con salarios bajos tienden a tener muchos hijos.
   b. Un aumento en el número de hijos produce una disminución en el salario.
   c. El salario depende del número de hijos.
   d. El número de hijos depende del salario.
   e. Personas con salarios altos tienden a tener muchos hijos.

3. Suponga que en una regresión lineal simple a = 4 y b = 2. Si la variable dependiente tiene el valor –1, entonces la variable independiente es:
   a. 2.
   b. −2.
   c. −2,5.
   d. 1,5.
   e. $-\frac{1}{2}$.

4. Si X y Y son dos variables con coeficiente de correlación r y recta de regresión $y = a + bx$, entonces **es cierto** que:
   a. r tiene que ser mayor que cero.
   b. a y b tienen el mismo signo.
   c. El coeficiente de correlación de X sobre Y es $-r$.
   d. $r \to 1$.
   e. r y b tienen el mismo signo.

Prueba 3

Preguntas de selección múltiple con múltiple respuesta.

Si 1 y 2 son correctas marque A.
Si 2 y 3 son correctas marque B.
Si 3 y 4 son correctas marque C.
Si 2 y 4 son correctas marque D.
Si 1 y 3 son correctas marque E.

Considere la siguiente situación:

Una baraja tiene 52 cartas, 13 de cada una de las siguientes figuras: corazones, diamantes, tréboles y picas. Cada figura está compuesta por: As, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K. Adicionalmente, los corazones y diamantes son rojos y los tréboles y picas son negros.

Se extrae al azar una carta de la baraja y se definen los siguientes eventos:

A: la carta extraída es un corazón.
B: la carta extraída es un as.
C: la carta extraída es roja.
D: el número de la carta es par.
E: ocurre A y B.

1. De los siguientes, los enunciados verdaderos son:
    1. El espacio muestral del experimento tiene 52 eventos simples.
    2. El evento A tiene trece eventos simples.
    3. E está conformado por dos eventos simples.
    4. C y D son dos eventos mutuamente excluyentes.

2. De los siguientes, los enunciados falsos son:
    1. Si ocurre C entonces no puede ocurrir A.
    2. Si ocurre E entonces ocurre C.
    3. Si ocurre C entonces ocurre A.
    4. Si ocurre A entonces ocurre C.

3. Si para la extracción de la carta solo contamos con los cuatro ases, es cierto que:
    1. Tiene que ocurrir C.
    2. No puede ocurrir E.
    3. Con seguridad ocurre B.
    4. El espacio muestral se reduce a cuatro eventos simples.

4. Si para la extracción solo contamos con los tréboles, es falso que:
    1. Ahora B es un evento simple.
    2. No puede ocurrir C.
    3. El espacio muestral se reduce a cuatro eventos simples.
    4. Con seguridad ocurre C.

Prueba 4

Preguntas de selección múltiple con múltiple respuesta.

Si 1 y 2 son correctas marque A.
Si 2 y 3 son correctas marque B.
Si 3 y 4 son correctas marque C.
Si 2 y 4 son correctas marque D.
Si 1 y 3 son correctas marque E.

Para responder las preguntas tenga en cuenta la siguiente situación:

Ocho personas son candidatos a ser elegidos para dos puestos en una compañía: un presidente y un vicepresidente. De estas ocho personas hay cinco hombres y tres mujeres.

1. Un experimento consiste en seleccionar al azar una persona de este grupo de ocho para el cargo de presidente. Entonces es cierto que:
   1. El espacio muestral del experimento es $S = \{hombre, mujer\}$.
   2. La probabilidad de que la persona seleccionada sea mujer es $\frac{3}{5}$.
   3. La probabilidad de que la persona seleccionada sea hombre es $\frac{5}{8}$.
   4. La probabilidad de que la persona sea mujer es mayor a que sea hombre.

2. Si se seleccionan dos personas al azar de este grupo, es cierto que:
   1. La probabilidad de que se seleccionen dos hombres es $\frac{25}{64}$.
   2. La probabilidad de que se seleccionen dos mujeres es $\frac{6}{56}$.
   3. Hay 64 posibilidades diferentes de escoger a las dos personas.
   4. El espacio muestral consta de cuatro eventos simples.

3. Si se seleccionan dos personas al azar, pero se conoce que la primera es hombre, es cierto que:
   1. La probabilidad de que la segunda sea hombre es $\frac{4}{7}$.
   2. La probabilidad de que la siguiente sea hombre es mayor a que sea mujer.
   3. La probabilidad de que la siguiente sea hombre es menor a que sea mujer.
   4. La probabilidad de que la segunda sea mujer es $\frac{3}{8}$.

Prueba 5

Preguntas de selección múltiple con única respuesta.

La siguiente tabla muestra los resultados de las proporciones de individuos en una ciudad, según su nivel educativo y su nivel salarial.

	Nivel salarial		
Nivel educativo	Alto S	Bajo $\overline{S}$	Total
Alto E	0,15	0,10	0,25
Bajo $\overline{E}$	0,10	0,65	0,75
Total	0,25	0,75	1

Suponga que se selecciona al azar a uno de los individuos de dicha ciudad y se definen los siguientes eventos:

E: = "el individuo seleccionado tiene un nivel educativo alto".

S: = "el individuo seleccionado tiene un nivel salarial alto".

1. Con base en la información es cierto que:
   a) $P(S) = 0,25$.
   b) $P(E) = P(S)$.
   c) $P(E \cap S) = 0,15$.
   d) $P(\overline{E}) = P(\overline{E} \cap S) + P(\overline{E} \cap \overline{S})$.
   e) $P(E \cup S) = 0,5$.
2. Teniendo en cuenta la tabla, la afirmación correcta es:
   a) $P(E|S) = \frac{3}{5}$.
   b) $P(S|E) = \frac{3}{5}$.
   c) $P(E|\overline{S}) = \frac{13}{15}$.
   d) $P(S|\overline{E}) = \frac{2}{5}$.
   e) $P(E|S) = P(S|E)$.
3. De las siguientes, la afirmación correcta es:
   a) E y S son independientes.
   b) Hay una mayor proporción de individuos con salarios altos.
   c) Hay una menor proporción de individuos con nivel educativo alto.
   d) El 65% de los individuos tiene salarios bajos o nivel educativo bajo.
   e) El 15% de los individuos tiene salarios altos y nivel educativo alto.

Prueba 6

Preguntas de selección múltiple con múltiple respuesta.

Si 1 y 2 son correctas marque A.

Si 2 y 3 son correctas marque B.

Si 3 y 4 son correctas marque C.

Si 2 y 4 son correctas marque D.

Si 1 y 3 son correctas marque E.

1. Sea X una variable aleatoria Poisson con promedio cinco.
   1. La función de probabilidad de X es $f(x) = \frac{e^{-5}5^x}{x!}$.
   2. X es una variable aleatoria continua.
   3. $P(X = 4) = 0,175467$.
   4. La desviación estándar de X es 5.

2. Considere la variable aleatoria binomial X con $n = 1$ y $p = \frac{1}{3}$. Las afirmaciones ciertas son:
   1. El valor esperado de X es $\frac{1}{3}$.
   2. La varianza de X es $\frac{2}{9}$.
   3. Los posibles valores de X son 0,1 y 2.
   4. $P(X > 0,5) = P(X < 0,5)$.

3. Sea X una variable aleatoria hipergeométrica con parámetros $N = 100$, $C = 10$ y $n = 4$.
   1. $P(X = 1) = 0,002754$.
   2. $P(X = 0) = 0,651631$.
   3. $P(X = 5) = 1$.
   4. $P(X = 3) = 0,002754$.

4. Sea X una variable aleatoria normal $\mu = 60$ y $\sigma^2 = 25$.
   1. $\frac{X - 60}{25}$ es una variable aleatoria normal estándar.
   2. $P(X > 70) \approx 0,344578$.
   3. $P(50 < X < 70) \approx 0,9545$.
   4. $P(X < 55) \approx 0,158655$.

5. Sea X una variable aleatoria binomial con parámetros $n=5$ y $p=0{,}35$. Es correcto afirmar:

    1. X toma valores en el intervalo (0,5) de los números reales.
    2. $P(X=0)=0{,}65^5$.
    3. $P(X=0)=P(X=5)$.
    4. El valor de mayor probabilidad de X es 2.

BIBLIOGRAFÍA

BATANERO, C. *Didáctica de la Estadística*, Universidad de Granada, 2001.

BLANCO, L. *Probabilidad*, Bogotá, Universidad Nacional de Colombia, 2000.

GONZÁLEZ, L. *Aciertos Matemáticos*, Educar Editores, 2006.

GRIMMETT, G. y D. STIRZAKER. *Probability and Random Processes*, Oxford Press, 2001.

LOHR, S. *Muestreo, diseño y análisis*, Thomson, 2000.

MARTÍN-PLIEGO, J. *Estadística y probabilidad I*, Editorial AC, 2004.

MEDINA, R. J. *Matemáticas especiales*, Bogotá, Universidad Distrital Francisco José de Caldas, 2011.

MEISEL, A. y M. VEGA. *La estatura de los colombianos: un ensayo de antropometría histórica, 1910-2002*, Banco de la República, 2004.

MENDENHALL, W. *et al. Introducción a la probabilidad y estadística*, Cengage, 2008.

MILTON, S. *Estadística para biología y ciencias de la salud*, Mc Graw Hill, 2001.

MYERS, R. et *al. Generalized Linears Models*, John Wiley, 2002.

MONTGOMERY, D. *et al. Introduction to Linear Regression Analysis*, Mc Graw Hill, 2001.

MOOD, A. *Introduction to the Theory of Statistics*, Mc Graw Hill, 1974.

ROSS, S. *First Course in Probability*, Pearson Prentice Hall, 2010.

ROSS, S. *Probabilidad y estadística para ingenieros*, Mc Graw Hill, 2002.

SHEAFER, R. *et al. Elementos de muestreo*, Thomson, 2006.

APÉNDICE

I. Lectura: buenas prácticas estadísticas

Los miembros de la Comunidad Económica Europea adoptaron el 24 de febrero del año 2005 un código de buenas prácticas estadísticas que rige para los países miembros, en particular, para los institutos nacionales de estadísticas.

A continuación, se reproduce el texto completo del documento con el propósito de divulgar los principios relacionados con la independencia y la responsabilidad de la producción de estadísticas que son deseables, tanto para los institutos de estadísticas como para las personas que producen estadísticas de divulgación pública.

Principio 1. Independencia profesional

La independencia profesional de las autoridades estadísticas de otros departamentos y organismos políticos, reguladores o administrativos, así como de los operadores del sector privado, garantiza la credibilidad de las estadísticas.

Indicadores:

- En la legislación se especifica la independencia de la autoridad estadística de las interferencias políticas y de otras interferencias externas a la hora de elaborar y difundir estadísticas oficiales.
- El director de la autoridad estadística tiene un nivel jerárquico lo suficientemente elevado como para garantizar un acceso de alto nivel a las autoridades políticas y a los

organismos públicos de carácter administrativo. Debe ser una persona de una gran capacidad profesional.

- El director de la autoridad estadística y, cuando proceda, los jefes de sus organismos estadísticos tienen la responsabilidad de garantizar que las estadísticas se elaboran y se difunden de forma independiente.
- El director de la autoridad estadística y, cuando proceda, los jefes de sus organismos estadísticos son los únicos responsables para decidir sobre los métodos, las normas y los procedimientos estadísticos, así como sobre el contenido y el calendario de las comunicaciones estadísticas.
- Se publican los programas de trabajo estadístico y se describen los progresos realizados en informes periódicos.
- Las comunicaciones estadísticas se distinguen claramente y se emiten al margen de las declaraciones políticas.
- Cuando procede, la autoridad estadística realiza comentarios públicos sobre cuestiones estadísticas, en los que incluye críticas y usos inadecuados de las estadísticas oficiales.

Principio 2. Mandato de recogida de datos

Las autoridades estadísticas deben tener un mandato jurídico claro para recoger información destinada a la elaboración de estadísticas. A petición de las autoridades estadísticas, se podrá obligar por ley a las administraciones, las empresas, los hogares y el público, en general, a que permitan el acceso a los datos destinados a la elaboración de estadísticas o a que presenten dichos datos.

Indicadores:

- En la legislación se especifica el mandato de recoger información destinada a la elaboración y a la difusión de estadísticas oficiales.
- La legislación nacional le permite a la autoridad estadística la utilización de expedientes administrativos a efectos estadísticos.
- Sobre la base de un acto jurídico, la autoridad estadística puede obligar a responder encuestas estadísticas.

Principio 3. Adecuación de los recursos

Los recursos a disposición de las autoridades estadísticas deben ser suficientes para cumplir los requisitos de las estadísticas.

Indicadores:

- Se dispone de recursos humanos, financieros e informáticos adecuados, tanto en tamaño como en calidad, para cumplir las necesidades actuales de las estadísticas.
- El alcance, el detalle y el costo de las estadísticas son proporcionados con respecto a las necesidades.
- Existen procedimientos para evaluar y justificar las solicitudes de nuevas estadísticas en relación con su costo.
- Existen procedimientos para evaluar la necesidad continua de todas las estadísticas, para determinar si alguna de ellas se puede realizar de forma discontinua o reducirse y, así, poder liberar recursos.

Principio 4. Compromiso de calidad

Todos los miembros del Sistema Estadístico Europeo (SEE) se comprometen a trabajar y cooperar conforme a los principios establecidos en la declaración sobre la calidad del sistema estadístico.

Indicadores:

- La calidad del producto se controla periódicamente conforme a los componentes de calidad del SEE.
- Existen procesos para controlar la calidad de la recogida, el tratamiento y la difusión de estadísticas.
- Existen procesos para abordar consideraciones de calidad, en los que constan compromisos en este ámbito y para orientar la planificación de las encuestas actuales y futuras.
- Las orientaciones de calidad están documentadas y el personal tiene una formación adecuada. Dichas orientaciones se expresan por escrito y se ponen a disposición del público.
- Existe una revisión periódica y profunda de la producción estadística clave utilizando expertos externos cuando proceda.

Principio 5. Confidencialidad estadística

Se debe garantizar absolutamente la privacidad de los proveedores de datos (hogares, empresas, administraciones y otros encuestados), la confidencialidad de la información que proporcionan y su uso exclusivo a efectos estadísticos.

Indicadores:

- En la legislación, se garantiza la confidencialidad estadística.
- El personal de la autoridad estadística firma un compromiso de confidencialidad jurídica cuando es nombrado.
- Se establecen sanciones importantes por cualquier incumplimiento premeditado de la confidencialidad estadística.
- Se proporcionan instrucciones y orientaciones sobre la protección de la confidencialidad estadística en los procesos de elaboración y difusión. Dichas orientaciones se expresan por escrito y se ponen a disposición del público.
- Existen disposiciones físicas y tecnológicas para proteger la seguridad y la integridad de las bases de datos estadísticas.
- Se aplican protocolos estrictos a los usuarios externos que acceden a microdatos a efectos de investigación.

Principio 6. Imparcialidad y objetividad

Las autoridades estadísticas deben elaborar y difundir estadísticas respetando la independencia científica y hacerlo de forma objetiva, profesional y transparente, de modo que se trate a todos los usuarios por igual.

Indicadores:

- Las estadísticas se recopilan sobre una base objetiva determinada por consideraciones estadísticas.
- La elección de las fuentes y las técnicas dependen de consideraciones estadísticas.
- Los errores descubiertos en las estadísticas publicadas se corrigen y se dan a conocer lo antes posible.
- La información sobre los métodos y los procedimientos utilizados por la autoridad estadística están a disposición del público.
- Se anuncian previamente la fecha y la hora de comunicación de las estadísticas.
- Todos los usuarios tienen acceso al mismo tiempo a las comunicaciones estadísticas, se restringe, se controla y se hace pública toda comunicación previa privilegiada a cualquier usuario externo.
- Las comunicaciones y las declaraciones estadísticas realizadas en ruedas de prensa son objetivas e imparciales.

Principio 7. Metodología sólida

Unas estadísticas de calidad se deben apoyar en una metodología sólida, lo cual exige herramientas, procedimientos y conocimientos especializados adecuados.

Indicadores:

- Existen procedimientos para garantizar que se apliquen coherentemente conceptos, definiciones y clasificaciones estándar en la autoridad estadística.
- El registro de empresas y el marco de las encuestas de población se evalúan periódicamente y, en caso necesario, se ajustan para garantizar una alta calidad.
- Se contratan titulados en las disciplinas académicas pertinentes.
- El personal asiste a cursos de formación y conferencias internacionales pertinentes, y se relaciona con colegas especialistas en estadística en el ámbito internacional para aprender de los mejores profesionales y aumentar sus conocimientos especializados.
- Se establece una cooperación con la comunidad científica para mejorar la metodología, se evalúa, mediante revisiones externas, la calidad y la eficacia de los métodos aplicados y se promueve la adopción de mejores herramientas cuando ello es viable.

Principio 8. Procedimientos estadísticos adecuados

Unas estadísticas de calidad se deben apoyar en unos procedimientos estadísticos adecuados, aplicados desde la recogida de los datos hasta la validación de estos.

Indicadores:

- Cuando las estadísticas se basan en datos administrativos, las definiciones y los conceptos utilizados a efectos administrativos se deben aproximar bastante a los requeridos a efectos estadísticos.
- En el caso de las encuestas estadísticas, se prueban sistemáticamente los cuestionarios antes de la recogida de datos.
- El diseño de las encuestas, así como la selección y la ponderación de las muestras están bien fundamentadas y se revisan o actualizan conforme a lo dispuesto.
- El trabajo de campo, la introducción de los datos y la codificación se controlan y se revisan de forma rutinaria conforme a lo dispuesto.
- Se utilizan sistemas informáticos de edición y de imputación, asimismo, se revisan o se actualizan periódicamente conforme a lo dispuesto.
- Las revisiones siguen procedimientos normalizados, consolidados y transparentes.

Principio 9. Una carga para los encuestados que no sea excesiva

La carga que supone la notificación debería ser proporcionada con respecto a las necesidades de los usuarios y no ser excesiva para los encuestados. La autoridad estadística controla la carga que supone responder a la encuesta y fija objetivos para reducirla progresivamente.

Indicadores:

- El alcance y el detalle de las demandas de estadísticas se limita a lo estrictamente necesario.
- La carga que supone la notificación se reparte lo más ampliamente posible entre la población sobre la que se efectúa la encuesta mediante técnicas de muestreo apropiadas.
- En la medida de lo posible, se puede acceder fácilmente a la información que se solicita de las empresas a partir de sus cuentas y también, cuando es posible, se utilizan medios electrónicos para facilitar su transmisión.
- Se aceptan las estimaciones y las aproximaciones más fiables cuando no se dispone inmediatamente de la información exacta.
- Cuando es posible se utilizan fuentes administrativas para evitar que se dupliquen las solicitudes de información.
- Está generalizada la puesta en común de datos entre las autoridades estadísticas, a fin de evitar la multiplicación de las encuestas.

Principio 10, Relación costo-eficacia

Los recursos se deben utilizar eficazmente.

Indicadores:

- Se controla la utilización de los recursos de la autoridad estadística mediante medidas internas y externas independientes.
- Las operaciones administrativas rutinarias —por ejemplo, toma, codificación y validación de los datos— están automatizadas en la mayor medida posible.
- Se está optimizando el potencial de productividad de la tecnología de la información y la comunicación a efectos de recogida, tratamiento y difusión de los datos.
- Se están realizando esfuerzos proactivos para mejorar el potencial estadístico de los registros administrativos y evitar encuestas directas costosas.

Principio 11. Pertinencia

Las estadísticas deben satisfacer las necesidades de los usuarios.

Indicadores:

- Existen procesos para consultar a los usuarios, controlar la pertinencia y la utilidad práctica de las estadísticas existentes por lo que se refiere a la satisfacción de las necesidades, así como para asesorar sobre las nuevas necesidades y prioridades.
- Se satisfacen las necesidades prioritarias y se reflejan en el programa de trabajo.
- Se realizan periódicamente encuestas para conocer el grado de satisfacción de los usuarios.

Principio 12. Precisión y fiabilidad

Las estadísticas deben reflejar la realidad de forma precisa y fidedigna.

Indicadores:

- Se evalúan y validan los datos originales, los resultados intermedios y la producción estadística.
- Se miden y se documentan sistemáticamente los errores de muestreo y los que no son de muestreo con arreglo al marco de los componentes de calidad del SEE.
- Se realizan de forma rutinaria y se utilizan internamente estudios y análisis de revisiones para moldear los procesos estadísticos.

Principio 13. Oportunidad y puntualidad

Las estadísticas se deben difundir oportuna y puntualmente.

Indicadores:

- Se establece una hora determinada del día para la comunicación de estadísticas.
- Para establecer la periodicidad de las estadísticas se tienen en cuenta los requisitos de los usuarios en la medida de lo posible.
- En caso de que la comunicación no se vaya a producir a la hora establecida, se notifica por adelantado, se dan explicaciones y se fija un nuevo plazo de comunicación.
- Cuando se considere conveniente, se pueden difundir resultados preliminares de una calidad global aceptable.

Principio 14. Coherencia y comparabilidad

Las estadísticas deberían ser coherentes a nivel interno, a lo largo del tiempo y comparables entre regiones y países; debería ser posible combinar y hacer un uso conjunto de los datos relacionados a partir de fuentes distintas.

Indicadores:

- Las estadísticas son coherentes en el ámbito interno; por ejemplo, se observan las identidades aritméticas y contables.
- Las estadísticas son coherentes o conciliables durante un periodo razonable.
- Las estadísticas se recopilan sobre la base de normas comunes con respecto al alcance, las definiciones, las unidades y las clasificaciones en las distintas encuestas y fuentes.
- Se comparan y se concilian las estadísticas de las distintas encuestas y fuentes.
- Se garantiza la comparabilidad transnacional de los datos mediante intercambios periódicos entre el sistema estadístico y otros sistemas estadísticos; se efectúan estudios metodológicos en estrecha colaboración entre los Estados miembros y Eurostat.

Principio 15. Accesibilidad y claridad

Las estadísticas se deberían presentar de forma clara y comprensible, difundirse de forma adecuada y conveniente y estar disponibles; asimismo, se debería permitir el acceso a estas de forma imparcial, con metadatos y orientación de apoyo.

Indicadores:

- Las estadísticas se presentan de forma que facilitan una interpretación adecuada y comparaciones significativas.
- Los servicios de difusión utilizan una tecnología moderna de información y comunicación y, si procede, una copia impresa tradicional.
- Cuando sea posible, se suministran análisis a medida y se hacen públicos.
- Se puede permitir el acceso a los microdatos a efectos de investigación. Dicho acceso está sometido a protocolos estrictos.
- Los metadatos están documentados con arreglo a sistemas de metadatos normalizados.
- Se mantiene informados a los usuarios sobre la metodología de los procesos estadísticos y la calidad de la producción estadística con respecto a los criterios de calidad del SEE.

II. Función de distribución binomial

n=2

x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
0	0,9801	0,9025	0,8100	0,6400	0,4900	0,3600	0,2500	0,1600	0,0900	0,0400	0,0100	0,0025	0,0001	0
1	0,9999	0,9975	0,9900	0,9600	0,9100	0,8400	0,7500	0,6400	0,5100	0,3600	0,1900	0,0975	0,0199	1
2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	2

n=3

x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
0	0,9703	0,8574	0,7290	0,5120	0,3430	0,2160	0,1250	0,0640	0,0270	0,0080	0,0010	0,0001	0,0000	0
1	0,9997	0,9928	0,9720	0,8960	0,7840	0,6480	0,5000	0,3520	0,2160	0,1040	0,0280	0,0073	0,0003	1
2	1,0000	0,9999	0,9990	0,9920	0,9730	0,9360	0,8750	0,7840	0,6570	0,4880	0,2710	0,1426	0,0297	2
3	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	3

n=4

x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
0	0,9606	0,8145	0,6561	0,4096	0,2401	0,1296	0,0625	0,0256	0,0081	0,0016	0,0001	0,0000	0,0000	0
1	0,9994	0,9860	0,9477	0,8192	0,6517	0,4752	0,3125	0,1792	0,0837	0,0272	0,0037	0,0005	0,0000	1
2	1,0000	0,9995	0,9963	0,9728	0,9163	0,8208	0,6875	0,5248	0,3483	0,1808	0,0523	0,0140	0,0006	2
3	1,0000	1,0000	0,9999	0,9984	0,9919	0,9744	0,9375	0,8704	0,7599	0,5904	0,3439	0,1855	0,0394	3
4	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	4

n=5

x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
0	0,9510	0,7738	0,5905	0,3277	0,1681	0,0778	0,0313	0,0102	0,0024	0,0003	0,0000	0,0000	0,0000	0
1	0,9990	0,9774	0,9185	0,7373	0,5282	0,3370	0,1875	0,0870	0,0308	0,0067	0,0005	0,0000	0,0000	1
2	1,0000	0,9988	0,9914	0,9421	0,8369	0,6826	0,5000	0,3174	0,1631	0,0579	0,0086	0,0012	0,0000	2
3	1,0000	1,0000	0,9995	0,9933	0,9692	0,9130	0,8125	0,6630	0,4718	0,2627	0,0815	0,0226	0,0010	3
4	1,0000	1,0000	1,0000	0,9997	0,9976	0,9898	0,9688	0,9222	0,8319	0,6723	0,4095	0,2262	0,0490	4
5	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	5

n=6

x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
0	0,9415	0,7351	0,5314	0,2621	0,1176	0,0467	0,0156	0,0041	0,0007	0,0001	0,0000	0,0000	0,0000	0
1	0,9985	0,9672	0,8857	0,6554	0,4202	0,2333	0,1094	0,0410	0,0109	0,0016	0,0001	0,0000	0,0000	1
2	1,0000	0,9978	0,9842	0,9011	0,7443	0,5443	0,3438	0,1792	0,0705	0,0170	0,0013	0,0001	0,0000	2
3	1,0000	0,9999	0,9987	0,9830	0,9295	0,8208	0,6563	0,4557	0,2557	0,0989	0,0159	0,0022	0,0000	3
4	1,0000	1,0000	0,9999	0,9984	0,9891	0,9590	0,8906	0,7667	0,5798	0,3446	0,1143	0,0328	0,0015	4
5	1,0000	1,0000	1,0000	0,9999	0,9993	0,9959	0,9844	0,9533	0,8824	0,7379	0,4686	0,2649	0,0585	5
6	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	6

n=7

x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
0	0,9321	0,6983	0,4783	0,2097	0,0824	0,0280	0,0078	0,0016	0,0002	0,0000	0,0000	0,0000	0,0000	0
1	0,9980	0,9556	0,8503	0,5767	0,3294	0,1586	0,0625	0,0188	0,0038	0,0004	0,0000	0,0000	0,0000	1
2	1,0000	0,9962	0,9743	0,8520	0,6471	0,4199	0,2266	0,0963	0,0288	0,0047	0,0002	0,0000	0,0000	2
3	1,0000	0,9998	0,9973	0,9667	0,8740	0,7102	0,5000	0,2898	0,1260	0,0333	0,0027	0,0002	0,0000	3
4	1,0000	1,0000	0,9998	0,9953	0,9712	0,9037	0,7734	0,5801	0,3529	0,1480	0,0257	0,0038	0,0000	4
5	1,0000	1,0000	1,0000	0,9996	0,9962	0,9812	0,9375	0,8414	0,6706	0,4233	0,1497	0,0444	0,0020	5
6	1,0000	1,0000	1,0000	1,0000	0,9998	0,9984	0,9922	0,9720	0,9176	0,7903	0,5217	0,3017	0,0679	6
7	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	7

n=8

x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
0	0,9227	0,6634	0,4305	0,1678	0,0576	0,0168	0,0039	0,0007	0,0001	0,0000	0,0000	0,0000	0,0000	0
1	0,9973	0,9428	0,8131	0,5033	0,2553	0,1064	0,0352	0,0085	0,0013	0,0001	0,0000	0,0000	0,0000	1
2	0,9999	0,9942	0,9619	0,7969	0,5518	0,3154	0,1445	0,0498	0,0113	0,0012	0,0000	0,0000	0,0000	2
3	1,0000	0,9996	0,9950	0,9437	0,8059	0,5941	0,3633	0,1737	0,0580	0,0104	0,0004	0,0000	0,0000	3
4	1,0000	1,0000	0,9996	0,9896	0,9420	0,8263	0,6367	0,4059	0,1941	0,0563	0,0050	0,0004	0,0000	4
5	1,0000	1,0000	1,0000	0,9988	0,9887	0,9502	0,8555	0,6846	0,4482	0,2031	0,0381	0,0058	0,0001	5
6	1,0000	1,0000	1,0000	0,9999	0,9987	0,9915	0,9648	0,8936	0,7447	0,4967	0,1869	0,0572	0,0027	6
7	1,0000	1,0000	1,0000	1,0000	0,9999	0,9993	0,9961	0,9832	0,9424	0,8322	0,5695	0,3366	0,0773	7
8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	8

n=9

x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
0	0,9135	0,6302	0,3874	0,1342	0,0404	0,0101	0,0020	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	0
1	0,9966	0,9288	0,7748	0,4362	0,1960	0,0705	0,0195	0,0038	0,0004	0,0000	0,0000	0,0000	0,0000	1
2	0,9999	0,9916	0,9470	0,7382	0,4628	0,2318	0,0898	0,0250	0,0043	0,0003	0,0000	0,0000	0,0000	2
3	1,0000	0,9994	0,9917	0,9144	0,7297	0,4826	0,2539	0,0994	0,0253	0,0031	0,0001	0,0000	0,0000	3
4	1,0000	1,0000	0,9991	0,9804	0,9012	0,7334	0,5000	0,2666	0,0988	0,0196	0,0009	0,0000	0,0000	4
5	1,0000	1,0000	0,9999	0,9969	0,9747	0,9006	0,7461	0,5174	0,2703	0,0856	0,0083	0,0006	0,0000	5
6	1,0000	1,0000	1,0000	0,9997	0,9957	0,9750	0,9102	0,7682	0,5372	0,2618	0,0530	0,0084	0,0001	6
7	1,0000	1,0000	1,0000	1,0000	0,9996	0,9962	0,9805	0,9295	0,8040	0,5638	0,2252	0,0712	0,0034	7
8	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9980	0,9899	0,9596	0,8658	0,6126	0,3698	0,0865	8
9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	9

n=10

x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
0	0,9044	0,5987	0,3487	0,1074	0,0282	0,0060	0,0010	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0
1	0,9957	0,9139	0,7361	0,3758	0,1493	0,0464	0,0107	0,0017	0,0001	0,0000	0,0000	0,0000	0,0000	1
2	0,9999	0,9885	0,9298	0,6778	0,3828	0,1673	0,0547	0,0123	0,0016	0,0001	0,0000	0,0000	0,0000	2
3	1,0000	0,9990	0,9872	0,8791	0,6496	0,3823	0,1719	0,0548	0,0106	0,0009	0,0000	0,0000	0,0000	3
4	1,0000	0,9999	0,9984	0,9672	0,8497	0,6331	0,3770	0,1662	0,0473	0,0064	0,0001	0,0000	0,0000	4
5	1,0000	1,0000	0,9999	0,9936	0,9527	0,8338	0,6230	0,3669	0,1503	0,0328	0,0016	0,0001	0,0000	5
6	1,0000	1,0000	1,0000	0,9991	0,9894	0,9452	0,8281	0,6177	0,3504	0,1209	0,0128	0,0010	0,0000	6
7	1,0000	1,0000	1,0000	0,9999	0,9984	0,9877	0,9453	0,8327	0,6172	0,3222	0,0702	0,0115	0,0001	7
8	1,0000	1,0000	1,0000	1,0000	0,9999	0,9983	0,9893	0,9536	0,8507	0,6242	0,2639	0,0861	0,0043	8
9	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9990	0,9940	0,9718	0,8926	0,6513	0,4013	0,0956	9
10	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	10

n=11

x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
0	0,8953	0,5688	0,3138	0,0859	0,0198	0,0036	0,0005	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0
1	0,9948	0,8981	0,6974	0,3221	0,1130	0,0302	0,0059	0,0007	0,0000	0,0000	0,0000	0,0000	0,0000	1
2	0,9998	0,9848	0,9104	0,6174	0,3127	0,1189	0,0327	0,0059	0,0006	0,0000	0,0000	0,0000	0,0000	2
3	1,0000	0,9984	0,9815	0,8389	0,5696	0,2963	0,1133	0,0293	0,0043	0,0002	0,0000	0,0000	0,0000	3
4	1,0000	0,9999	0,9972	0,9496	0,7897	0,5328	0,2744	0,0994	0,0216	0,0020	0,0000	0,0000	0,0000	4
5	1,0000	1,0000	0,9997	0,9883	0,9218	0,7535	0,5000	0,2465	0,0782	0,0117	0,0003	0,0000	0,0000	5
6	1,0000	1,0000	1,0000	0,9980	0,9784	0,9006	0,7256	0,4672	0,2103	0,0504	0,0028	0,0001	0,0000	6
7	1,0000	1,0000	1,0000	0,9998	0,9957	0,9707	0,8867	0,7037	0,4304	0,1611	0,0185	0,0016	0,0000	7
8	1,0000	1,0000	1,0000	1,0000	0,9994	0,9941	0,9673	0,8811	0,6873	0,3826	0,0896	0,0152	0,0002	8
9	1,0000	1,0000	1,0000	1,0000	1,0000	0,9993	0,9941	0,9698	0,8870	0,6779	0,3026	0,1019	0,0052	9
10	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995	0,9964	0,9802	0,9141	0,6862	0,4312	0,1047	10
11	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	11

n=12

x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
0	0,8864	0,5404	0,2824	0,0687	0,0138	0,0022	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0
1	0,9938	0,8816	0,6590	0,2749	0,0850	0,0196	0,0032	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	1
2	0,9998	0,9804	0,8891	0,5583	0,2528	0,0834	0,0193	0,0028	0,0002	0,0000	0,0000	0,0000	0,0000	2
3	1,0000	0,9978	0,9744	0,7946	0,4925	0,2253	0,0730	0,0153	0,0017	0,0001	0,0000	0,0000	0,0000	3
4	1,0000	0,9998	0,9957	0,9274	0,7237	0,4382	0,1938	0,0573	0,0095	0,0006	0,0000	0,0000	0,0000	4
5	1,0000	1,0000	0,9995	0,9806	0,8822	0,6652	0,3872	0,1582	0,0386	0,0039	0,0001	0,0000	0,0000	5
6	1,0000	1,0000	0,9999	0,9961	0,9614	0,8418	0,6128	0,3348	0,1178	0,0194	0,0005	0,0000	0,0000	6
7	1,0000	1,0000	1,0000	0,9994	0,9905	0,9427	0,8062	0,5618	0,2763	0,0726	0,0043	0,0002	0,0000	7
8	1,0000	1,0000	1,0000	0,9999	0,9983	0,9847	0,9270	0,7747	0,5075	0,2054	0,0256	0,0022	0,0000	8
9	1,0000	1,0000	1,0000	1,0000	0,9998	0,9972	0,9807	0,9166	0,7472	0,4417	0,1109	0,0196	0,0002	9
10	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9968	0,9804	0,9150	0,7251	0,3410	0,1184	0,0062	10
11	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9978	0,9862	0,9313	0,7176	0,4596	0,1136	11
12	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	12

n=15

x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
0	0,8601	0,4633	0,2059	0,0352	0,0047	0,0005	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0
1	0,9904	0,8290	0,5490	0,1671	0,0353	0,0052	0,0005	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1
2	0,9996	0,9638	0,8159	0,3980	0,1268	0,0271	0,0037	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	2
3	1,0000	0,9945	0,9444	0,6482	0,2969	0,0905	0,0176	0,0019	0,0001	0,0000	0,0000	0,0000	0,0000	3
4	1,0000	0,9994	0,9873	0,8358	0,5155	0,2173	0,0592	0,0093	0,0007	0,0000	0,0000	0,0000	0,0000	4
5	1,0000	0,9999	0,9978	0,9389	0,7216	0,4032	0,1509	0,0338	0,0037	0,0001	0,0000	0,0000	0,0000	5
6	1,0000	1,0000	0,9997	0,9819	0,8689	0,6098	0,3036	0,0950	0,0152	0,0008	0,0000	0,0000	0,0000	6
7	1,0000	1,0000	1,0000	0,9958	0,9500	0,7869	0,5000	0,2131	0,0500	0,0042	0,0000	0,0000	0,0000	7

n	x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
n=15	8	1,0000	1,0000	1,0000	0,9992	0,9848	0,9050	0,6964	0,3902	0,1311	0,0181	0,0003	0,0000	0,0000	8
	9	1,0000	1,0000	1,0000	0,9999	0,9963	0,9662	0,8491	0,5968	0,2784	0,0611	0,0022	0,0001	0,0000	9
	10	1,0000	1,0000	1,0000	1,0000	0,9993	0,9907	0,9408	0,7827	0,4845	0,1642	0,0127	0,0006	0,0000	10
	11	1,0000	1,0000	1,0000	1,0000	0,9999	0,9981	0,9824	0,9095	0,7031	0,3518	0,0556	0,0055	0,0000	11
	12	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9963	0,9729	0,8732	0,6020	0,1841	0,0362	0,0004	12
	13	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995	0,9948	0,9647	0,8329	0,4510	0,1710	0,0096	13
	14	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995	0,9953	0,9648	0,7941	0,5367	0,1399	14
	15	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	15

n	x/p	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,99	x/p
n=20	0	0,8179	0,3585	0,1216	0,0115	0,0008	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0
	1	0,9831	0,7358	0,3917	0,0692	0,0076	0,0005	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1
	2	0,9990	0,9245	0,6769	0,2061	0,0355	0,0036	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	2
	3	1,0000	0,9841	0,8670	0,4114	0,1071	0,0160	0,0013	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	3
	4	1,0000	0,9974	0,9568	0,6296	0,2375	0,0510	0,0059	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	4
	5	1,0000	0,9997	0,9887	0,8042	0,4164	0,1256	0,0207	0,0016	0,0000	0,0000	0,0000	0,0000	0,0000	5
	6	1,0000	1,0000	0,9976	0,9133	0,6080	0,2500	0,0577	0,0065	0,0003	0,0000	0,0000	0,0000	0,0000	6
	7	1,0000	1,0000	0,9996	0,9679	0,7723	0,4159	0,1316	0,0210	0,0013	0,0000	0,0000	0,0000	0,0000	7
	8	1,0000	1,0000	0,9999	0,9900	0,8867	0,5956	0,2517	0,0565	0,0051	0,0001	0,0000	0,0000	0,0000	8
	9	1,0000	1,0000	1,0000	0,9974	0,9520	0,7553	0,4119	0,1275	0,0171	0,0006	0,0000	0,0000	0,0000	9
	10	1,0000	1,0000	1,0000	0,9994	0,9829	0,8725	0,5881	0,2447	0,0480	0,0026	0,0000	0,0000	0,0000	10
	11	1,0000	1,0000	1,0000	0,9999	0,9949	0,9435	0,7483	0,4044	0,1133	0,0100	0,0001	0,0000	0,0000	11
	12	1,0000	1,0000	1,0000	1,0000	0,9987	0,9790	0,8684	0,5841	0,2277	0,0321	0,0004	0,0000	0,0000	12
	13	1,0000	1,0000	1,0000	1,0000	0,9997	0,9935	0,9423	0,7500	0,3920	0,0867	0,0024	0,0000	0,0000	13
	14	1,0000	1,0000	1,0000	1,0000	1,0000	0,9984	0,9793	0,8744	0,5836	0,1958	0,0113	0,0003	0,0000	14
	15	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9941	0,9490	0,7625	0,3704	0,0432	0,0026	0,0000	15
	16	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9987	0,9840	0,8929	0,5886	0,1330	0,0159	0,0000	16
	17	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9964	0,9645	0,7939	0,3231	0,0755	0,0010	17
	18	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995	0,9924	0,9308	0,6083	0,2642	0,0169	18
	19	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9992	0,9885	0,8784	0,6415	0,1821	19
	20	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	20

III. Función de distribución de Poisson

x/λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1	1,1	1,2	x/λ
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	0,3679	0,3329	0,3012	0
1	0,9953	0,9825	0,9631	0,9384	0,9098	0,8781	0,8442	0,8088	0,7725	0,7358	0,6990	0,6626	1
2	0,9998	0,9989	0,9964	0,9921	0,9856	0,9769	0,9659	0,9526	0,9371	0,9197	0,9004	0,8795	2
3	1,0000	0,9999	0,9997	0,9992	0,9982	0,9966	0,9942	0,9909	0,9865	0,9810	0,9743	0,9662	3
4		1,0000	1,0000	0,9999	0,9998	0,9996	0,9992	0,9986	0,9977	0,9963	0,9946	0,9923	4
5				1,0000	1,0000	1,0000	0,9999	0,9998	0,9997	0,9994	0,9990	0,9985	5
6							1,0000	1,0000	1,0000	0,9999	0,9999	0,9997	6
7										1,0000	1,0000	1,0000	7

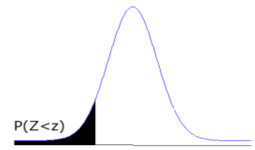
x/λ	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2	2,2	2,4	2,6	2,8	x/λ
0	0,2725	0,2466	0,2231	0,2019	0,1827	0,1653	0,1496	0,1353	0,1108	0,0907	0,0743	0,0608	0
1	0,6268	0,5918	0,5578	0,5249	0,4932	0,4628	0,4337	0,4060	0,3546	0,3084	0,2674	0,2311	1
2	0,8571	0,8335	0,8088	0,7834	0,7572	0,7306	0,7037	0,6767	0,6227	0,5697	0,5184	0,4695	2
3	0,9569	0,9463	0,9344	0,9212	0,9068	0,8913	0,8747	0,8571	0,8194	0,7787	0,7360	0,6919	3
4	0,9893	0,9857	0,9814	0,9763	0,9704	0,9636	0,9559	0,9473	0,9275	0,9041	0,8774	0,8477	4
5	0,9978	0,9968	0,9955	0,9940	0,9920	0,9896	0,9868	0,9834	0,9751	0,9643	0,9510	0,9349	5
6	0,9996	0,9994	0,9991	0,9987	0,9981	0,9974	0,9966	0,9955	0,9925	0,9884	0,9828	0,9756	6
7	0,9999	0,9999	0,9998	0,9997	0,9996	0,9994	0,9992	0,9989	0,9980	0,9967	0,9947	0,9919	7
8	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9998	0,9998	0,9995	0,9991	0,9985	0,9976	8
9					1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9996	0,9993	9
10									1,0000	1,0000	0,9999	0,9998	10
11											1,0000	1,0000	11

x/λ	3	3,2	3,4	3,6	3,8	4	4,2	4,4	4,6	4,8	5	5,2	x/λ
0	0,0498	0,0408	0,0334	0,0273	0,0224	0,0183	0,0150	0,0123	0,0101	0,0082	0,0067	0,0055	0
1	0,1991	0,1712	0,1468	0,1257	0,1074	0,0916	0,0780	0,0663	0,0563	0,0477	0,0404	0,0342	1
2	0,4232	0,3799	0,3397	0,3027	0,2689	0,2381	0,2102	0,1851	0,1626	0,1425	0,1247	0,1088	2
3	0,6472	0,6025	0,5584	0,5152	0,4735	0,4335	0,3954	0,3594	0,3257	0,2942	0,2650	0,2381	3
4	0,8153	0,7806	0,7442	0,7064	0,6678	0,6288	0,5898	0,5512	0,5132	0,4763	0,4405	0,4061	4
5	0,9161	0,8946	0,8705	0,8441	0,8156	0,7851	0,7531	0,7199	0,6858	0,6510	0,6160	0,5809	5
6	0,9665	0,9554	0,9421	0,9267	0,9091	0,8893	0,8675	0,8436	0,8180	0,7908	0,7622	0,7324	6
7	0,9881	0,9832	0,9769	0,9692	0,9599	0,9489	0,9361	0,9214	0,9049	0,8867	0,8666	0,8449	7
8	0,9962	0,9943	0,9917	0,9883	0,9840	0,9786	0,9721	0,9642	0,9549	0,9442	0,9319	0,9181	8
9	0,9989	0,9982	0,9973	0,9960	0,9942	0,9919	0,9889	0,9851	0,9805	0,9749	0,9682	0,9603	9
10	0,9997	0,9995	0,9992	0,9987	0,9981	0,9972	0,9959	0,9943	0,9922	0,9896	0,9863	0,9823	10
11	0,9999	0,9999	0,9998	0,9996	0,9994	0,9991	0,9986	0,9980	0,9971	0,9960	0,9945	0,9927	11
12	1,0000	1,0000	0,9999	0,9999	0,9998	0,9997	0,9996	0,9993	0,9990	0,9986	0,9980	0,9972	12
13			1,0000	1,0000	1,0000	0,9999	0,9999	0,9998	0,9997	0,9995	0,9993	0,9990	13
14						1,0000	1,0000	0,9999	0,9999	0,9999	0,9998	0,9997	14
15								1,0000	1,0000	1,0000	0,9999	0,9999	15
16											1,0000	1,0000	16

x/λ	5,4	5,6	5,8	6	6,5	7	7,5	8	8,5	9	9,5	10	x/λ
0	0,0045	0,0037	0,0030	0,0025	0,0015	0,0009	0,0006	0,0003	0,0002	0,0001	0,0001	0,0000	0
1	0,0289	0,0244	0,0206	0,0174	0,0113	0,0073	0,0047	0,0030	0,0019	0,0012	0,0008	0,0005	1
2	0,0948	0,0824	0,0715	0,0620	0,0430	0,0296	0,0203	0,0138	0,0093	0,0062	0,0042	0,0028	2
3	0,2133	0,1906	0,1700	0,1512	0,1118	0,0818	0,0591	0,0424	0,0301	0,0212	0,0149	0,0103	3
4	0,3733	0,3422	0,3127	0,2851	0,2237	0,1730	0,1321	0,0996	0,0744	0,0550	0,0403	0,0293	4
5	0,5461	0,5119	0,4783	0,4457	0,3690	0,3007	0,2414	0,1912	0,1496	0,1157	0,0885	0,0671	5
6	0,7017	0,6703	0,6384	0,6063	0,5265	0,4497	0,3782	0,3134	0,2562	0,2068	0,1649	0,1301	6
7	0,8217	0,7970	0,7710	0,7440	0,6728	0,5987	0,5246	0,4530	0,3856	0,3239	0,2687	0,2202	7
8	0,9027	0,8857	0,8672	0,8472	0,7916	0,7291	0,6620	0,5925	0,5231	0,4557	0,3918	0,3328	8
9	0,9512	0,9409	0,9292	0,9161	0,8774	0,8305	0,7764	0,7166	0,6530	0,5874	0,5218	0,4579	9
10	0,9775	0,9718	0,9651	0,9574	0,9332	0,9015	0,8622	0,8159	0,7634	0,7060	0,6453	0,5830	10
11	0,9904	0,9875	0,9841	0,9799	0,9661	0,9467	0,9208	0,8881	0,8487	0,8030	0,7520	0,6968	11
12	0,9962	0,9949	0,9932	0,9912	0,9840	0,9730	0,9573	0,9362	0,9091	0,8758	0,8364	0,7916	12
13	0,9986	0,9980	0,9973	0,9964	0,9929	0,9872	0,9784	0,9658	0,9486	0,9261	0,8981	0,8645	13
14	0,9995	0,9993	0,9990	0,9986	0,9970	0,9943	0,9897	0,9827	0,9726	0,9585	0,9400	0,9165	14
15	0,9998	0,9998	0,9996	0,9995	0,9988	0,9976	0,9954	0,9918	0,9862	0,9780	0,9665	0,9513	15
16	0,9999	0,9999	0,9999	0,9998	0,9996	0,9990	0,9980	0,9963	0,9934	0,9889	0,9823	0,9730	16
17	1,0000	1,0000	1,0000	0,9999	0,9998	0,9996	0,9992	0,9984	0,9970	0,9947	0,9911	0,9857	17
18				1,0000	0,9999	0,9999	0,9997	0,9993	0,9987	0,9976	0,9957	0,9928	18
19					1,0000	1,0000	0,9999	0,9997	0,9995	0,9989	0,9980	0,9965	19
20							1,0000	0,9999	0,9998	0,9996	0,9991	0,9984	20
21								1,0000	0,9999	0,9998	0,9996	0,9993	21
22									1,0000	0,9999	0,9999	0,9997	22
23										1,0000	0,9999	0,9999	23
24											1,0000	1,0000	24

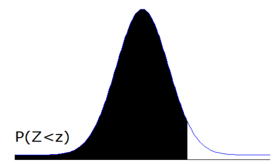
x/λ	11	12	13	14	15	16	17	18	19	20	21	22	x/λ
0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0
1	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1
2	0,0012	0,0005	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	2
3	0,0049	0,0023	0,0011	0,0005	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	3
4	0,0151	0,0076	0,0037	0,0018	0,0009	0,0004	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	4
5	0,0375	0,0203	0,0107	0,0055	0,0028	0,0014	0,0007	0,0003	0,0002	0,0001	0,0000	0,0000	5
6	0,0786	0,0458	0,0259	0,0142	0,0076	0,0040	0,0021	0,0010	0,0005	0,0003	0,0001	0,0001	6
7	0,1432	0,0895	0,0540	0,0316	0,0180	0,0100	0,0054	0,0029	0,0015	0,0008	0,0004	0,0002	7
8	0,2320	0,1550	0,0998	0,0621	0,0374	0,0220	0,0126	0,0071	0,0039	0,0021	0,0011	0,0006	8
9	0,3405	0,2424	0,1658	0,1094	0,0699	0,0433	0,0261	0,0154	0,0089	0,0050	0,0028	0,0015	9
10	0,4599	0,3472	0,2517	0,1757	0,1185	0,0774	0,0491	0,0304	0,0183	0,0108	0,0063	0,0035	10
11	0,5793	0,4616	0,3532	0,2600	0,1848	0,1270	0,0847	0,0549	0,0347	0,0214	0,0129	0,0076	11
12	0,6887	0,5760	0,4631	0,3585	0,2676	0,1931	0,1350	0,0917	0,0606	0,0390	0,0245	0,0151	12
13	0,7813	0,6815	0,5730	0,4644	0,3632	0,2745	0,2009	0,1426	0,0984	0,0661	0,0434	0,0278	13
14	0,8540	0,7720	0,6751	0,5704	0,4657	0,3675	0,2808	0,2081	0,1497	0,1049	0,0716	0,0477	14
15	0,9074	0,8444	0,7636	0,6694	0,5681	0,4667	0,3715	0,2867	0,2148	0,1565	0,1111	0,0769	15
16	0,9441	0,8987	0,8355	0,7559	0,6641	0,5660	0,4677	0,3751	0,2920	0,2211	0,1629	0,1170	16
17	0,9678	0,9370	0,8905	0,8272	0,7489	0,6593	0,5640	0,4686	0,3784	0,2970	0,2270	0,1690	17
18	0,9823	0,9626	0,9302	0,8826	0,8195	0,7423	0,6550	0,5622	0,4695	0,3814	0,3017	0,2325	18
19	0,9907	0,9787	0,9573	0,9235	0,8752	0,8122	0,7363	0,6509	0,5606	0,4703	0,3843	0,3060	19
20	0,9953	0,9884	0,9750	0,9521	0,9170	0,8682	0,8055	0,7307	0,6472	0,5591	0,4710	0,3869	20
21	0,9977	0,9939	0,9859	0,9712	0,9469	0,9108	0,8615	0,7991	0,7255	0,6437	0,5577	0,4716	21
22	0,9990	0,9970	0,9924	0,9833	0,9673	0,9418	0,9047	0,8551	0,7931	0,7206	0,6405	0,5564	22
23	0,9995	0,9985	0,9960	0,9907	0,9805	0,9633	0,9367	0,8989	0,8490	0,7875	0,7160	0,6374	23
24	0,9998	0,9993	0,9980	0,9950	0,9888	0,9777	0,9594	0,9317	0,8933	0,8432	0,7822	0,7117	24
25	0,9999	0,9997	0,9990	0,9974	0,9938	0,9869	0,9748	0,9554	0,9269	0,8878	0,8377	0,7771	25
26	1,0000	0,9999	0,9995	0,9987	0,9967	0,9925	0,9848	0,9718	0,9514	0,9221	0,8826	0,8324	26
27		0,9999	0,9998	0,9994	0,9983	0,9959	0,9912	0,9827	0,9687	0,9475	0,9175	0,8775	27
28		1,0000	0,9999	0,9997	0,9991	0,9978	0,9950	0,9897	0,9805	0,9657	0,9436	0,9129	28
29			1,0000	0,9999	0,9996	0,9989	0,9973	0,9941	0,9882	0,9782	0,9626	0,9398	29
30				0,9999	0,9998	0,9994	0,9986	0,9967	0,9930	0,9865	0,9758	0,9595	30
31				1,0000	0,9999	0,9997	0,9993	0,9982	0,9960	0,9919	0,9848	0,9735	31
32					1,0000	0,9999	0,9996	0,9990	0,9978	0,9953	0,9907	0,9831	32
33						0,9999	0,9998	0,9995	0,9988	0,9973	0,9945	0,9895	33
34						1,0000	0,9999	0,9998	0,9994	0,9985	0,9968	0,9936	34
35							1,0000	0,9999	0,9997	0,9992	0,9982	0,9962	35
36								0,9999	0,9998	0,9996	0,9990	0,9978	36
37								1,0000	0,9999	0,9998	0,9995	0,9988	37
38									1,0000	0,9999	0,9997	0,9993	38
39										0,9999	0,9999	0,9996	39
40										1,0000	0,9999	0,9998	40
41											1,0000	0,9999	41
42												1,0000	42

IV. Función de distribución normal estándar



Segundo decimal del valor *Z*

z	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02	0,01	0,00	z
-4,0	0,00002	0,00002	0,00002	0,00002	0,00003	0,00003	0,00003	0,00003	0,00003	0,00003	-4,0
-3,9	0,00003	0,00003	0,00004	0,00004	0,00004	0,00004	0,00004	0,00004	0,00005	0,00005	-3,9
-3,8	0,00005	0,00005	0,00005	0,00006	0,00006	0,00006	0,00006	0,00007	0,00007	0,00007	-3,8
-3,7	0,00008	0,00008	0,00008	0,00008	0,00009	0,00009	0,00010	0,00010	0,00010	0,00011	-3,7
-3,6	0,00011	0,00012	0,00012	0,00013	0,00013	0,00014	0,00014	0,00015	0,00015	0,00016	-3,6
-3,5	0,00017	0,00017	0,00018	0,00019	0,00019	0,00020	0,00021	0,00022	0,00022	0,00023	-3,5
-3,4	0,00024	0,00025	0,00026	0,00027	0,00028	0,00029	0,00030	0,00031	0,00032	0,00034	-3,4
-3,3	0,00035	0,00036	0,00038	0,00039	0,00040	0,00042	0,00043	0,00045	0,00047	0,00048	-3,3
-3,2	0,00050	0,00052	0,00054	0,00056	0,00058	0,00060	0,00062	0,00064	0,00066	0,00069	-3,2
-3,1	0,00071	0,00074	0,00076	0,00079	0,00082	0,00084	0,00087	0,00090	0,00094	0,00097	-3,1
-3,0	0,00100	0,00104	0,00107	0,00111	0,00114	0,00118	0,00122	0,00126	0,00131	0,00135	-3,0
-2,9	0,00139	0,00144	0,00149	0,00154	0,00159	0,00164	0,00169	0,00175	0,00181	0,00187	-2,9
-2,8	0,00193	0,00199	0,00205	0,00212	0,00219	0,00226	0,00233	0,00240	0,00248	0,00256	-2,8
-2,7	0,00264	0,00272	0,00280	0,00289	0,00298	0,00307	0,00317	0,00326	0,00336	0,00347	-2,7
-2,6	0,00357	0,00368	0,00379	0,00391	0,00402	0,00415	0,00427	0,00440	0,00453	0,00466	-2,6
-2,5	0,00480	0,00494	0,00508	0,00523	0,00539	0,00554	0,00570	0,00587	0,00604	0,00621	-2,5
-2,4	0,00639	0,00657	0,00676	0,00695	0,00714	0,00734	0,00755	0,00776	0,00798	0,00820	-2,4
-2,3	0,00842	0,00866	0,00889	0,00914	0,00939	0,00964	0,00990	0,01017	0,01044	0,01072	-2,3
-2,2	0,01101	0,01130	0,01160	0,01191	0,01222	0,01255	0,01287	0,01321	0,01355	0,01390	-2,2
-2,1	0,01426	0,01463	0,01500	0,01539	0,01578	0,01618	0,01659	0,01700	0,01743	0,01786	-2,1
-2,0	0,01831	0,01876	0,01923	0,01970	0,02018	0,02068	0,02118	0,02169	0,02222	0,02275	-2,0
-1,9	0,02330	0,02385	0,02442	0,02500	0,02559	0,02619	0,02680	0,02743	0,02807	0,02872	-1,9
-1,8	0,02938	0,03005	0,03074	0,03144	0,03216	0,03288	0,03362	0,03438	0,03515	0,03593	-1,8
-1,7	0,03673	0,03754	0,03836	0,03920	0,04006	0,04093	0,04182	0,04272	0,04363	0,04457	-1,7
-1,6	0,04551	0,04648	0,04746	0,04846	0,04947	0,05050	0,05155	0,05262	0,05370	0,05480	-1,6
-1,5	0,05592	0,05705	0,05821	0,05938	0,06057	0,06178	0,06301	0,06426	0,06552	0,06681	-1,5
-1,4	0,06811	0,06944	0,07078	0,07215	0,07353	0,07493	0,07636	0,07780	0,07927	0,08076	-1,4
-1,3	0,08226	0,08379	0,08534	0,08691	0,08851	0,09012	0,09176	0,09342	0,09510	0,09680	-1,3
-1,2	0,09853	0,10027	0,10204	0,10383	0,10565	0,10749	0,10935	0,11123	0,11314	0,11507	-1,2
-1,1	0,11702	0,11900	0,12100	0,12302	0,12507	0,12714	0,12924	0,13136	0,13350	0,13567	-1,1
-1,0	0,13786	0,14007	0,14231	0,14457	0,14686	0,14917	0,15151	0,15386	0,15625	0,15866	-1,0
-0,9	0,16109	0,16354	0,16602	0,16853	0,17106	0,17361	0,17619	0,17879	0,18141	0,18406	-0,9
-0,8	0,18673	0,18943	0,19215	0,19489	0,19766	0,20045	0,20327	0,20611	0,20897	0,21186	-0,8
-0,7	0,21476	0,21770	0,22065	0,22363	0,22663	0,22965	0,23270	0,23576	0,23885	0,24196	-0,7
-0,6	0,24510	0,24825	0,25143	0,25463	0,25785	0,26109	0,26435	0,26763	0,27093	0,27425	-0,6
-0,5	0,27760	0,28096	0,28434	0,28774	0,29116	0,29460	0,29806	0,30153	0,30503	0,30854	-0,5
-0,4	0,31207	0,31561	0,31918	0,32276	0,32636	0,32997	0,33360	0,33724	0,34090	0,34458	-0,4
-0,3	0,34827	0,35197	0,35569	0,35942	0,36317	0,36693	0,37070	0,37448	0,37828	0,38209	-0,3
-0,2	0,38591	0,38974	0,39358	0,39743	0,40129	0,40517	0,40905	0,41294	0,41683	0,42074	-0,2
-0,1	0,42465	0,42858	0,43251	0,43644	0,44038	0,44433	0,44828	0,45224	0,45620	0,46017	-0,1
-0,0	0,46414	0,46812	0,47210	0,47608	0,48006	0,48405	0,48803	0,49202	0,49601	0,50000	-0,0



Segundo decimal del valor *Z*

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	z
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586	**0,0**
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535	**0,1**
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409	**0,2**
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173	**0,3**
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793	**0,4**
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240	**0,5**
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490	**0,6**
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524	**0,7**
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327	**0,8**
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891	**0,9**
1,0	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214	**1,0**
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298	**1,1**
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147	**1,2**
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774	**1,3**
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189	**1,4**
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408	**1,5**
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449	**1,6**
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327	**1,7**
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062	**1,8**
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670	**1,9**
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169	**2,0**
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574	**2,1**
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899	**2,2**
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158	**2,3**
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361	**2,4**
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520	**2,5**
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643	**2,6**
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736	**2,7**
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807	**2,8**
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861	**2,9**
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900	**3,0**
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929	**3,1**
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950	**3,2**
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965	**3,3**
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976	**3,4**
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983	**3,5**
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989	**3,6**
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992	**3,7**
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995	**3,8**
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997	**3,9**
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998	**4,0**

V. Bases de datos incluidas en la plataforma virtual de la Universidad

- Resultados del Mundial 2010.
- Trabajadores.
- Indicadores demográficos.
- Nacimientos por departamento 2007.
- Avalúos de predios y áreas construidas.
- Población mundial por país y sexo.
- Lácteos.
- Crecimiento.
- Tasas mundiales de fertilidad.

www.utadeo.edu.co

ISBN: 978-958-725-114-2

ISBN: 978-958-725-114-2