

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 N° 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.

Prohibida la reproducción total o parcial por cualquier medio sin
autorización escrita de la Universidad.

IMPRESO EN COLOMBIA - PRINTED IN COLOMBIA

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	7
---------------------------	---

CAPÍTULO 1

Estadística descriptiva	9
¿Qué es la estadística?	9
Población, muestra y variables	10
Gráficos para resumir datos cualitativos	13
Gráficos para resumir datos cuantitativos	26

CAPÍTULO 2

Estadísticos de centro y variabilidad	47
Estadísticos de centro	47
Estadísticos de variabilidad	61
La desigualdad de Chevyshev	70
Percentiles muestrales	78

CAPÍTULO 3

Correlación y regresión	87
Regresión lineal simple y correlación	88
El modelo de la regresión lineal simple	94
Modelos alternativos a la regresión lineal simple	98

CAPÍTULO 4

Introducción a la probabilidad	109
Experimentos aleatorios	109
Técnicas de conteo	116

Regla de la multiplicación.....	117
Permutaciones	120
Combinaciones	124
Conceptos básicos de probabilidad.....	129
Cálculo de probabilidades.....	139
Probabilidad condicional y Teorema de Bayes.....	146
Variables aleatorias.....	157

CAPÍTULO 5

Distribuciones discretas de probabilidad. algunos casos

prácticos	171
Distribución de probabilidad binomial.....	171
Distribución de probabilidad de Poisson	180
Distribución de probabilidad hipergeométrica	186

CAPÍTULO 6

Distribución de probabilidad Normal	193
Características de la distribución de probabilidad normal.....	193
Distribución de probabilidad normal estándar (Z)	195

Esquema para un proyecto de estadística descriptiva	203
--	-----

Pruebas de conocimientos	207
---------------------------------------	-----

Bibliografía	217
---------------------------	-----

Apéndice	219
I. Lectura: “Buenas prácticas estadísticas”	219
II. Función de distribución binomial.....	227
III. Función de distribución de Poisson.....	229
IV. Función de distribución normal estándar	232
V. Bases de datos incluidas en la plataforma virtual de la Universidad.....	234

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 **cuantitativa**. 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 contable¹ 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 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** (figura 1.1) 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

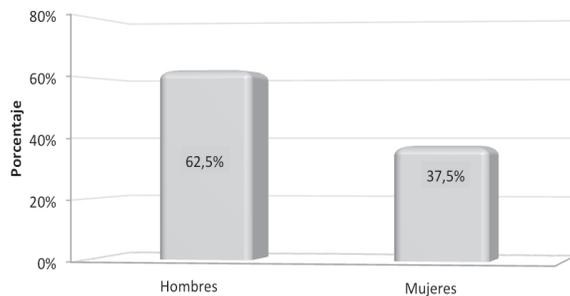
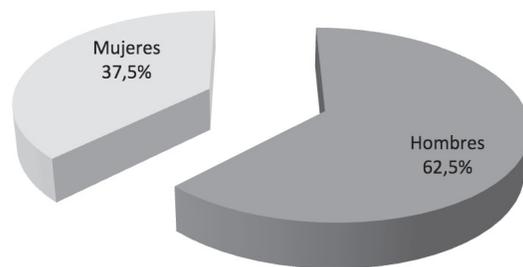


Figura 1.2 Gráfico de sectores.
Distribución del sexo



Ejemplo 1.3

Enunciado

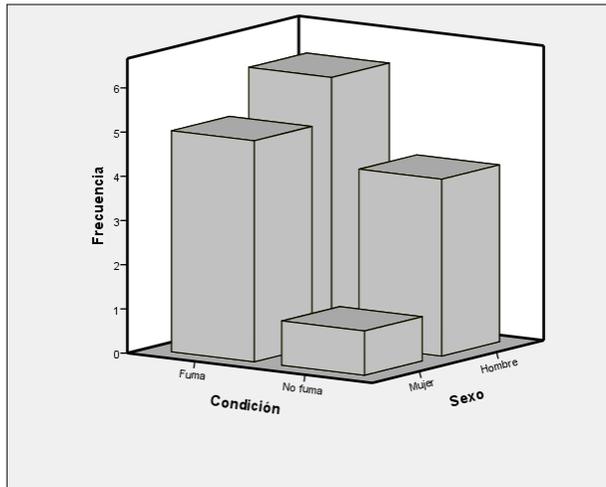
Construya una gráfica para representar las variables sexo y condición de fumador consignadas en la tabla 1.2.

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

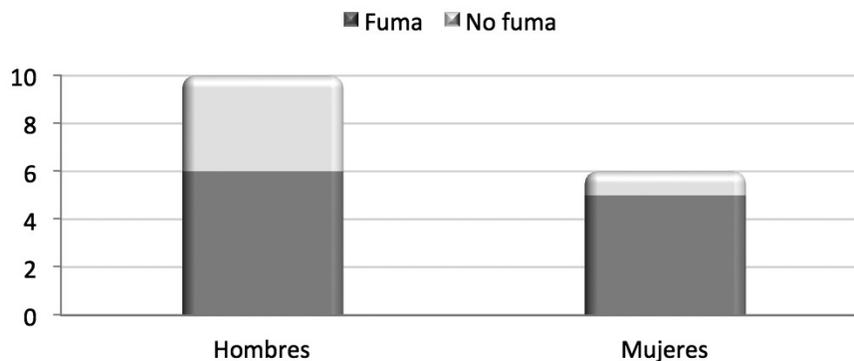


Un gráfico de *rascacielos*, como el de la figura 1.3, es adecuado para estos datos. Las alturas de los prismas rectangulares corresponden a las frecuencias de cada uno de los cuatro grupos formados por las dos variables:

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%

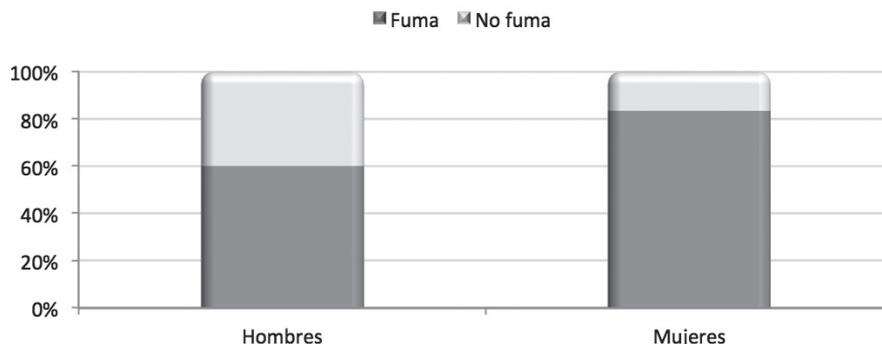
Celda
Frecuencia absoluta
Porcentaje de la tabla
Porcentaje de la fila
Porcentaje de la columna

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

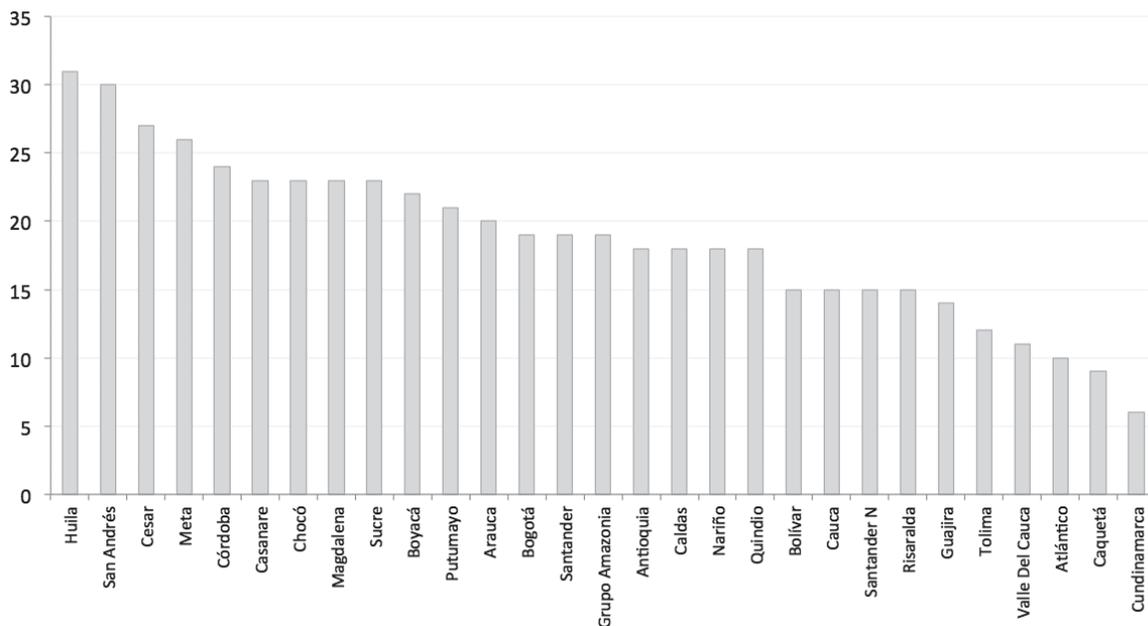
Depto.	Antioquia	Arauca	Atlántico	Bogotá	Bolívar	Boyacá	Caldas	Caquetá	Casanare	Cauca
Casos de MP	1.059	49	231	1.353	292	283	173	41	73	199
Población total	5.988.458	244.507	2284840	7.259.597	1.958.224	1.265.517	976.438	442.033	319.502	1.308.090
MP x 100.000	18	20	10	19	15	22	18	9	23	15

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 m² 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

- ¿Cuáles son los elementos de la población?
- ¿Cuáles son las variables?
- ¿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

- ¿Cuáles son los elementos de la población?
- ¿Estos datos corresponden a una población o a una muestra?
- ¿Cuál es la variable que se mide? ¿Es cualitativa o cuantitativa?
- Construya un gráfico de barras para describir los datos.
- Construya un gráfico de sectores para describir los datos.
- ¿Cuál es la proporción de amas de casa que tienen por preferencia el Éxito?
- ¿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	
<i>Bastante en desacuerdo</i>	93	9	102	
	19,6%	1,9%	21,5%	
	91,2%	8,8%		
	31,5%	5,0%		
<i>Algo en desacuerdo</i>	136	16	152	
	28,6%	3,4%	32,0%	
	89,5%	10,5%		
	46,1%	8,9%		
<i>Neutral</i>	30	26	56	
	6,3%	5,5%	11,8%	
	53,6%	46,4%		
	10,2%	14,4%		
<i>Algo de acuerdo</i>	22	57	79	
	4,6%	12,0%	16,6%	
	27,8%	72,2%		
	7,5%	31,7%		
<i>Bastante de acuerdo</i>	8	67	75	
	1,7%	14,1%	15,8%	
	10,7%	89,3%		
	2,7%	37,2%		
<i>No aplica</i>	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.

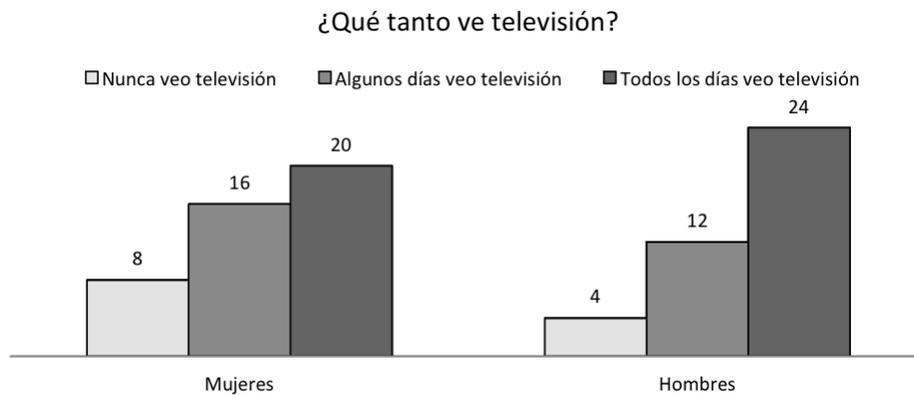
Sexo	Estado	Sexo	Estado	Sexo	Estado
F	Incumplido	F	Cumplido	F	Incumplido
F	Cumplido	F	Incumplido	M	Cumplido
F	Incumplido	M	Cumplido	F	Incumplido
M	Incumplido	F	Incumplido	F	Cumplido
F	Cumplido	F	Incumplido	F	Incumplido
F	Cumplido	M	Cumplido	M	Cumplido
F	Incumplido	F	Incumplido	F	Incumplido
M	Cumplido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



- ¿Cuál es el número de estudiantes en la muestra?
- ¿Qué porcentaje de estudiantes nunca ve televisión?
- ¿Qué porcentaje de mujeres nunca ve televisión?
- 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

- ¿Cuáles son los elementos de la población?
- ¿Estos datos corresponden a una población o a una muestra?
- ¿Cuál es la variable que se mide? ¿Es cualitativa o cuantitativa?
- 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)

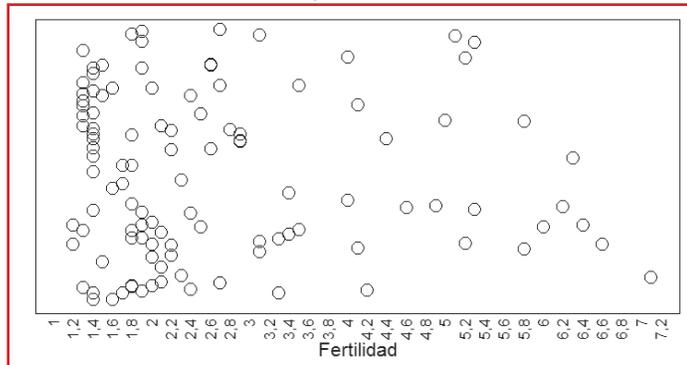
Ubicación	Fer.	Ubicación	Fer.	Ubicación	Fer.	Ubicación	Fer.	Ubicación	Fer.
Bosnia	1,2	Rusia	1,4	Corea del Norte	1,9	Sudáfrica	2,5	Samoa	4,0
Corea del Sur	1,2	España	1,4	Líbano	1,9	Venezuela	2,5	Guatemala	4,1
Andorra	1,3	Barbados	1,5	Noruega	1,9	Ecuador	2,6	Irak	4,1
Alemania	1,3	Cuba	1,5	Suecia	1,9	Panamá	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,

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

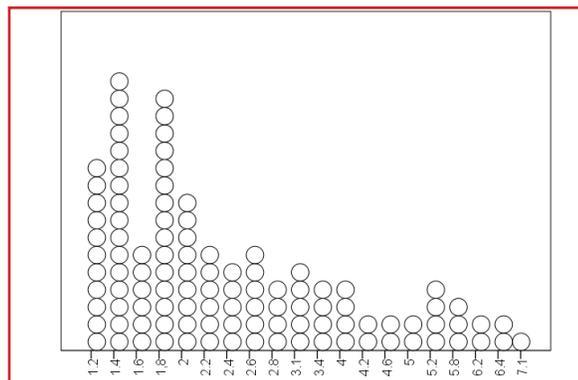
Enunciado

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

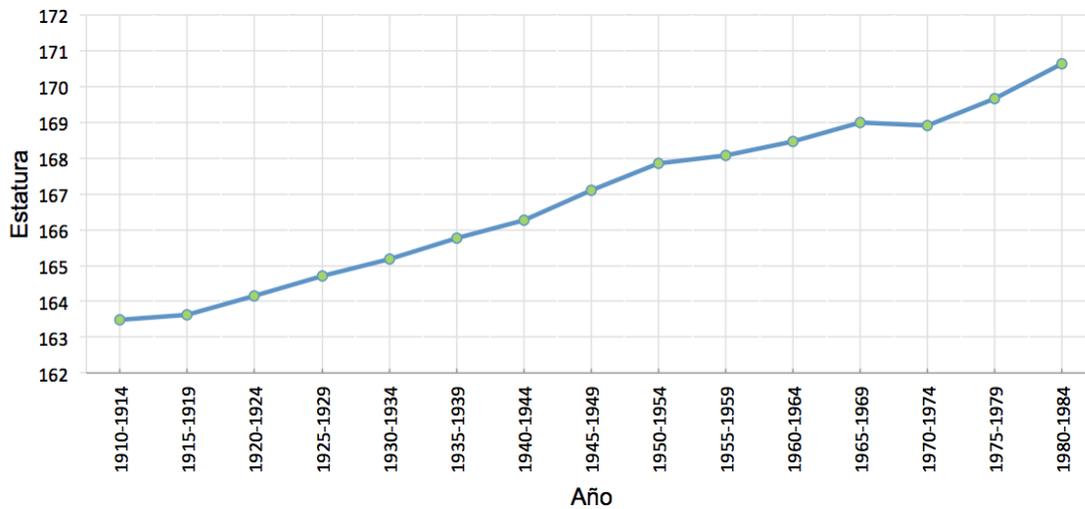
Año de nacimiento	Estatura promedio hombres	Estatura promedio mujeres	Número de hombres	Número de mujeres	Total
1910-1914	163,48	150,78	1.751	2.197	3.948
1915-1919	163,61	151,49	4.582	4.993	9.575
1920-1924	164,16	152,38	9.086	9.779	18.865
1925-1929	164,7	153,06	15.659	16.299	31.958
1930-1934	165,17	153,48	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

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
⋮	⋮
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.	Ubicación	Fer.	Ubicación	Fer.	Ubicación	Fer.	Ubicación	Fer.
Ucrania	1,3	Países Bajos	1,7	Tuquía	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	Reino Unido	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	Paquistán	4,0		

Solución

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

Paso 2:

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

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.

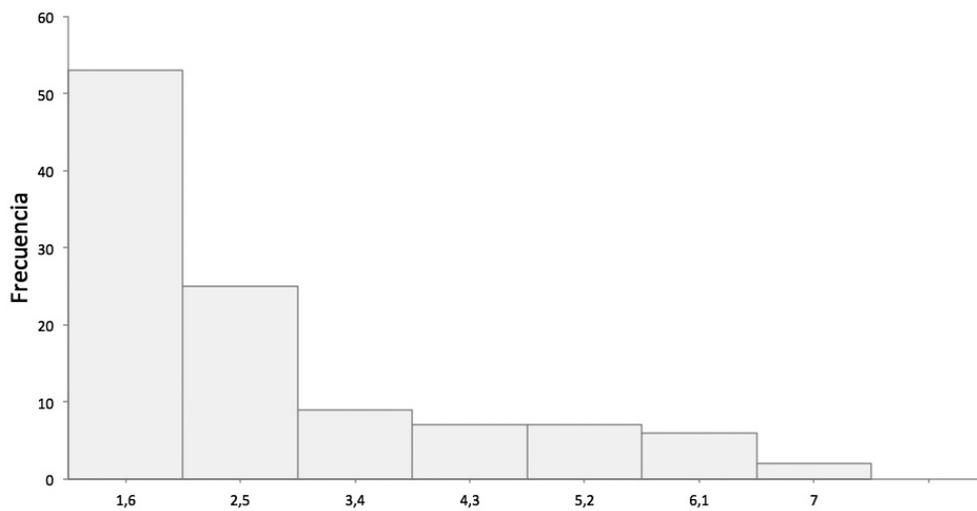
Paso 5: por último, se examina la frecuencia de observaciones en cada una de las clases conformadas y se construye el histograma de la figura 1.10 con las frecuencias absolutas. También se puede construir con las frecuencias relativas.

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

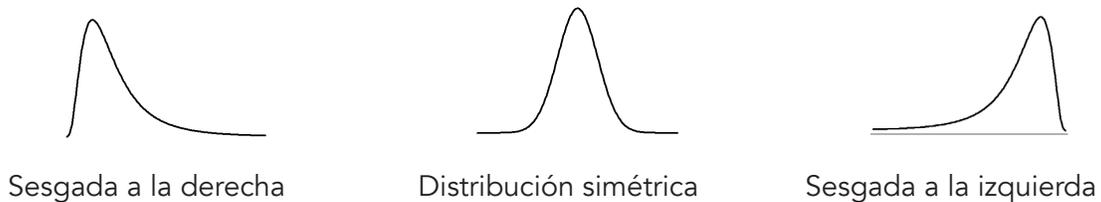
Clase	Límite inferior	Límite superior	Frecuencia	Marca de clase
1	1,15	2,05	53	1,6
2	2,05	2,95	25	2,5
3	2,95	3,85	9	3,4
4	3,85	4,75	7	4,3
5	4,75	5,65	7	5,2
6	5,65	6,55	6	6,1
7	6,55	7,45	2	7,0

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.

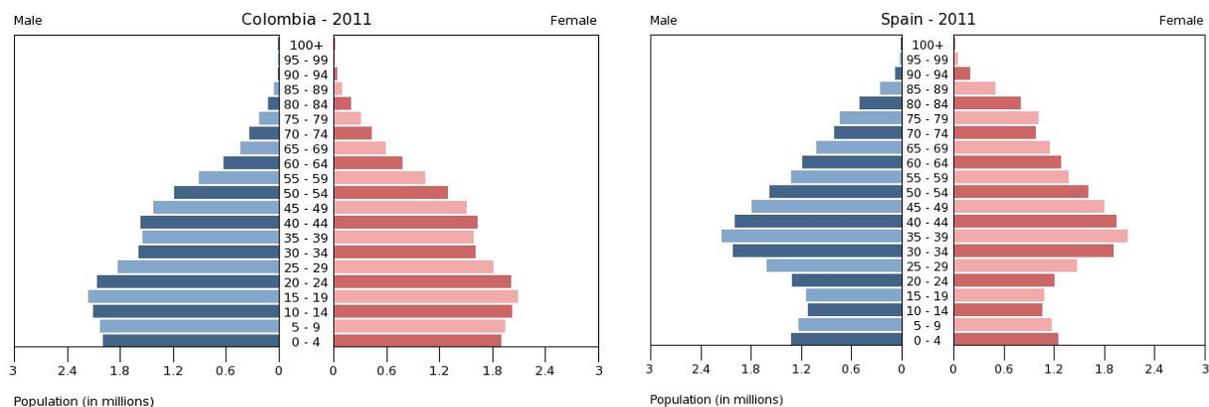
Pirámides de población

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

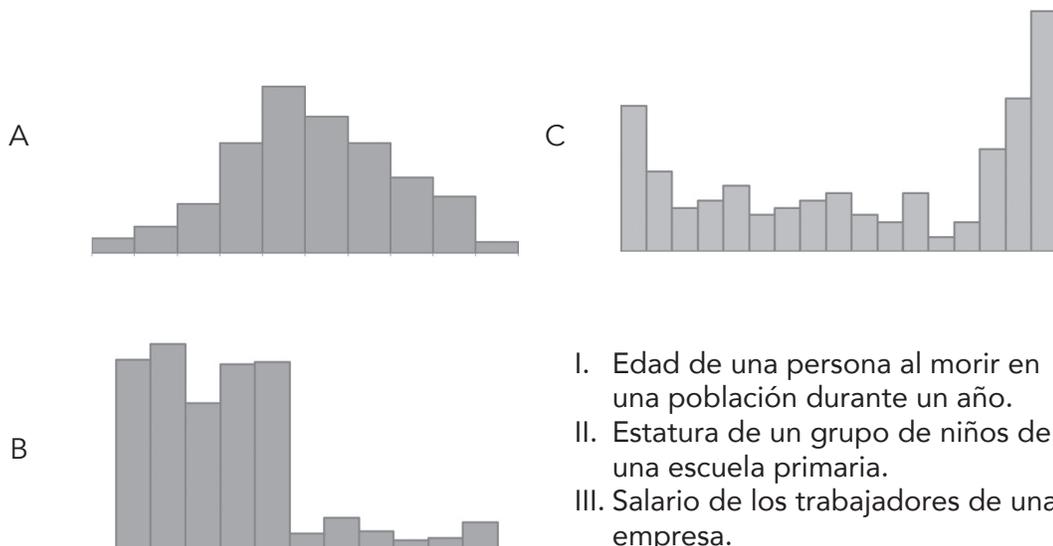
La pirámide de población de Colombia 2011, tiene una base ancha y una cima angosta. Esto significa que hay una gran cantidad de población joven. Sobresale el grupo de colombianos entre 15 y 19 años con la mayor frecuencia y que solo una reducida porción alcanza edades superiores a los 85 años.

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

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



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

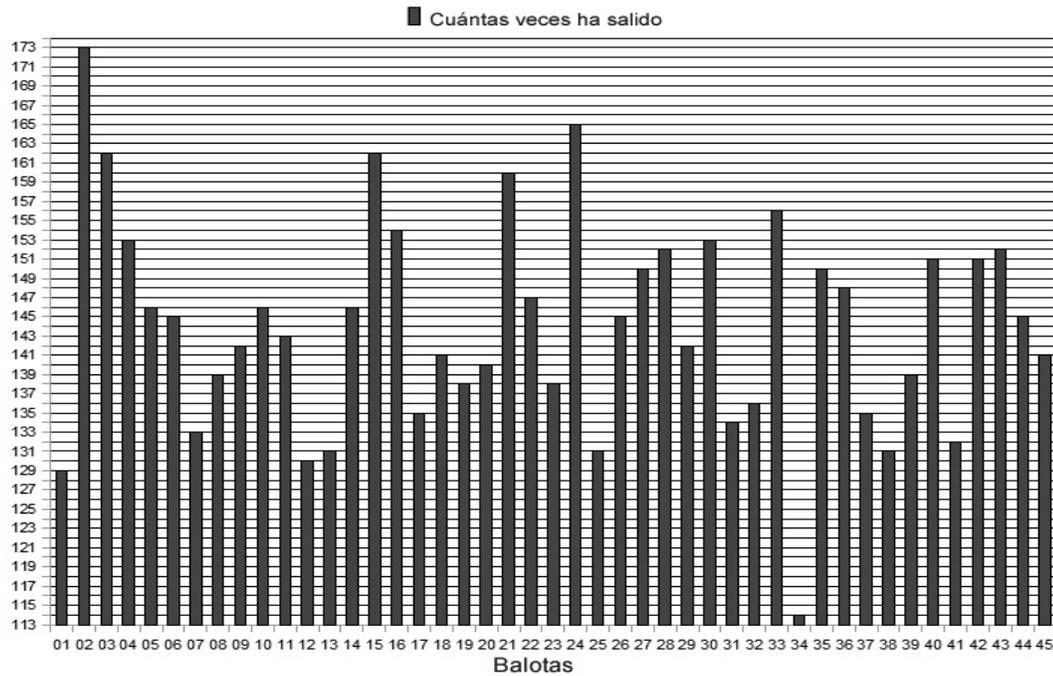
- a. Salarios de trabajadores en una empresa en la cual la mayoría gana poco.
 - b. Tiempos de deportistas que compiten en 100 metros planos.
 - c. Edad al primer parto de las madres chocoanas.
3. Las siguientes son las calificaciones en el examen final de microeconomía de cien estudiantes, pertenecientes 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?

Estadística de números jugados hasta el sorteo 1.081 del sábado 27 de agosto de 2011



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

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

- ¿En cuáles carreras resulta más rentable estudiar un posgrado?
- ¿Qué tipo de gráfica utilizaría para describir los datos? Constrúyala.
- Resuma la información de la gráfica en relación con las carreras, los salarios y el nivel de estudios.

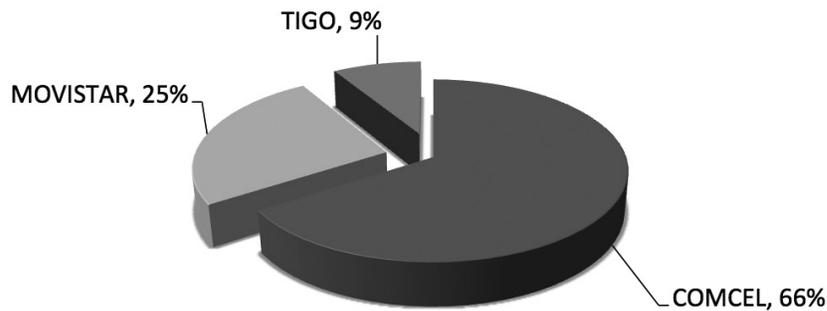
10. Se sabe que se ha arrojado material tóxico a una ciénaga. Los investigadores midieron la cantidad de material tóxico (en partes por millón) en peces capturados en tres lugares diferentes, con los siguientes resultados:

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		

- Use una gráfica apropiada para describir los datos.
- ¿Qué sugieren los datos?
- ¿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: <i>Portafolio</i> , 27 de octubre de 2010	

Quejas de usuarios por operador

- ¿Considera adecuado el anterior gráfico de sectores para describir el número de quejas por empresa?
- Construya un gráfico de barras para los mismos datos.
- ¿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

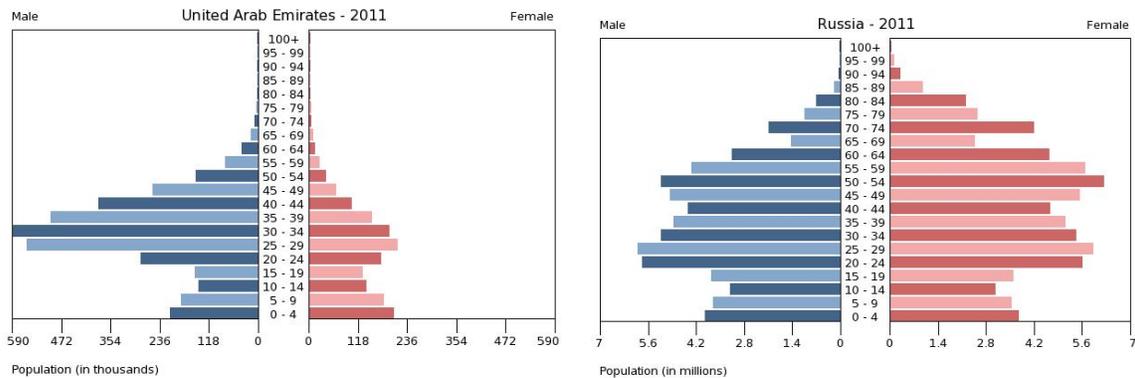
- a. Construya diagramas de tallos y hojas para las variables contenidas en la anterior tabla.
- b. ¿Existen valores inusuales en alguna de ellas?

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:

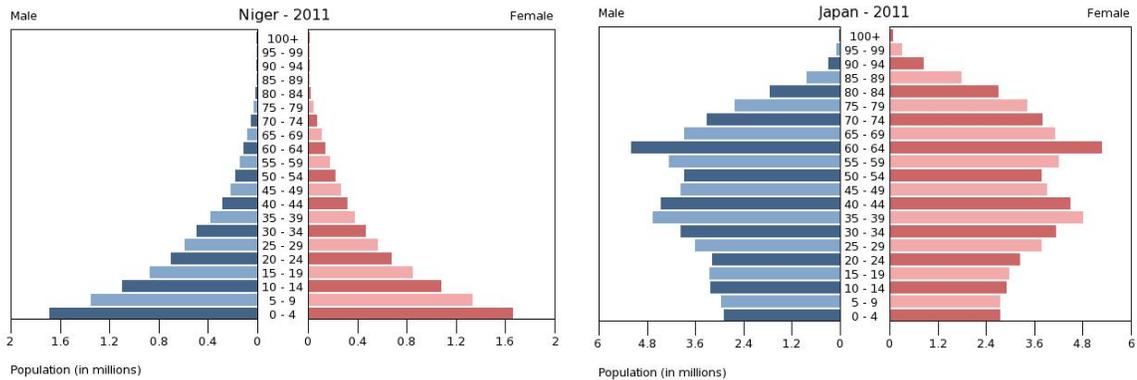
50	52	112	113	114	102	103	103	103	104
55	56	57	34	124	106	106	107	107	84
63	33	33	72	145	68	69	69	79	89
65	65	68	76	91	73	74	74	83	93
71	72	90	81	98	77	78	79	86	99
76	76	97	85	91	81	81	83	92	98
80	80	90	90	99	85	86	86	91	97
85	85								

- 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?

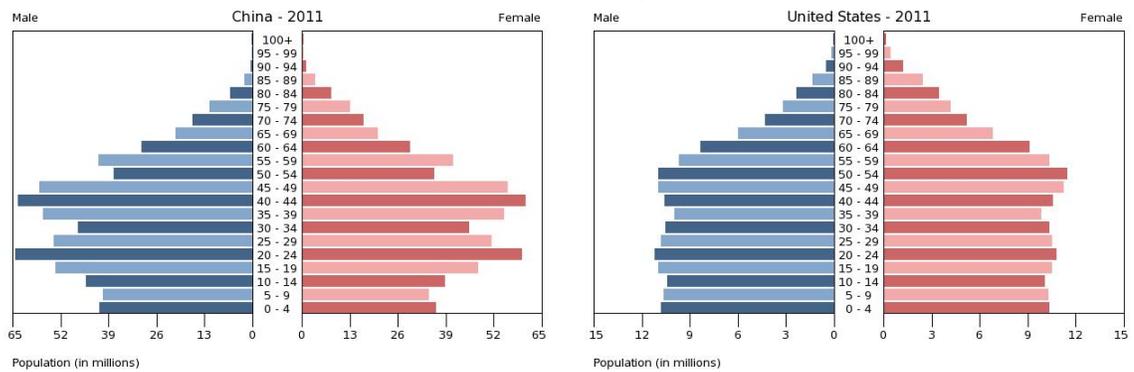
14. Compare las pirámides de población de Emiratos Árabes y Rusia 2011.



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



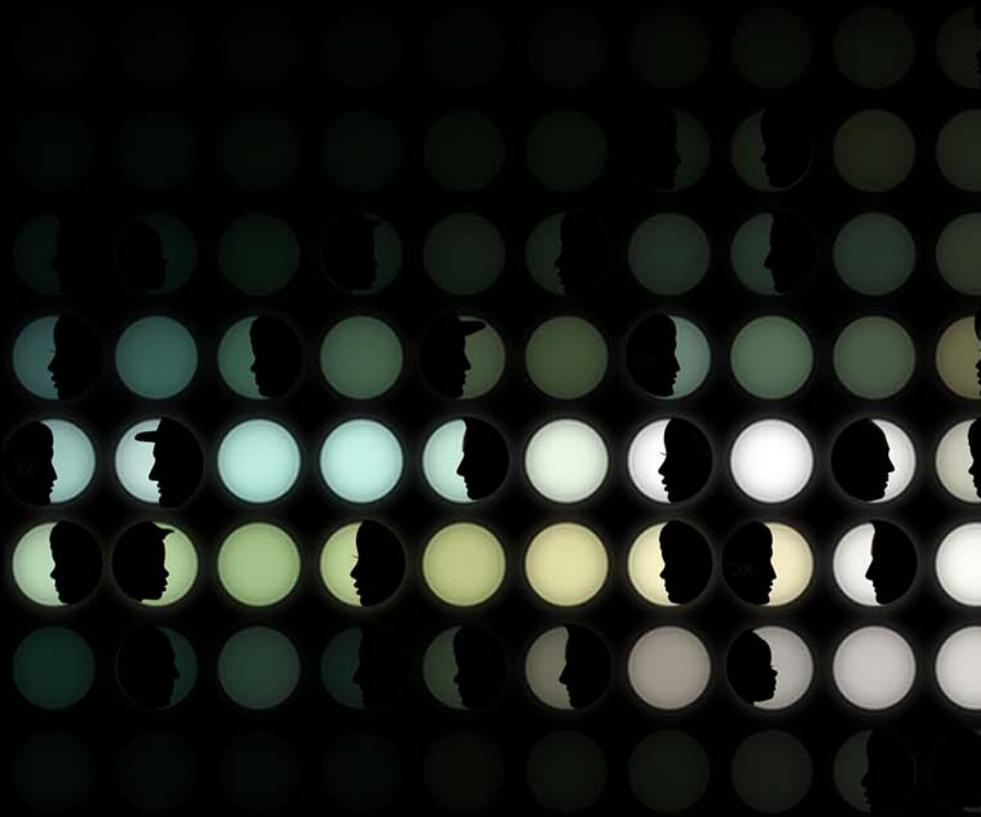
16. Compare las pirámides de población de China y Estados Unidos 2011.



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

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

www.utadeo.edu.co



ISBN: 978-958-725-114-2



ISBN: 978-958-725-114-2