

# Estadística y probabilidades

**Diana Angélica Pulido**

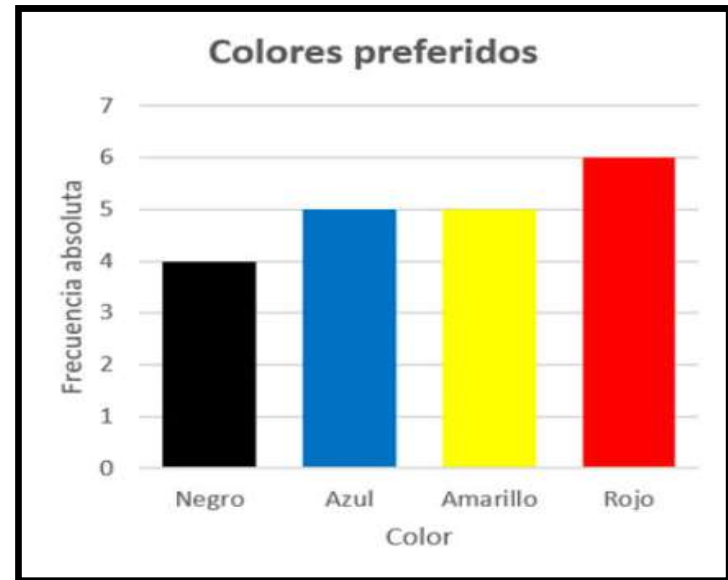
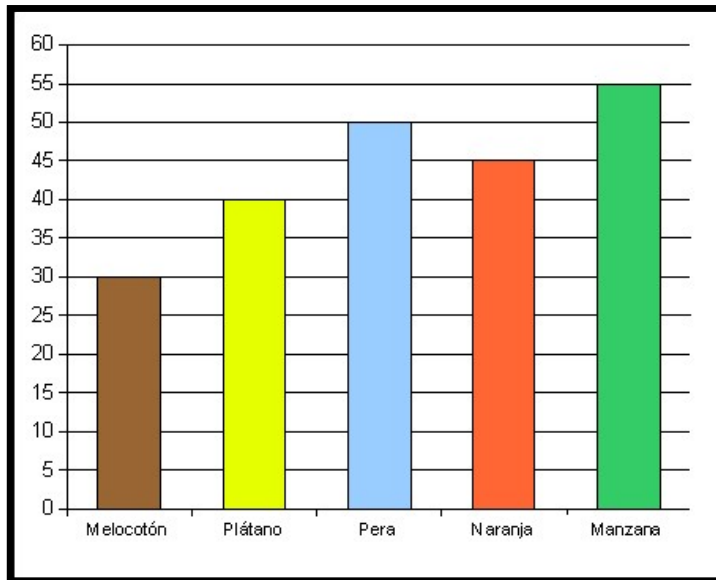
**203 C**

**Lunes – Miércoles 7-9 D309 - D402**

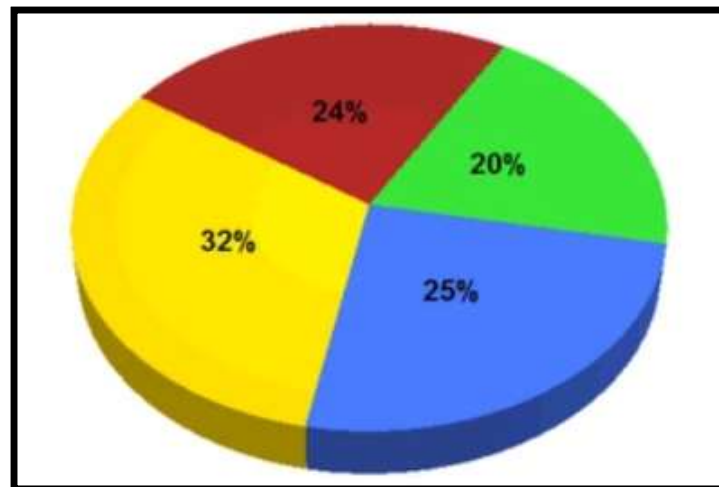
**[Dianaa.pulido@uexternado.edu.co](mailto:Dianaa.pulido@uexternado.edu.co)**



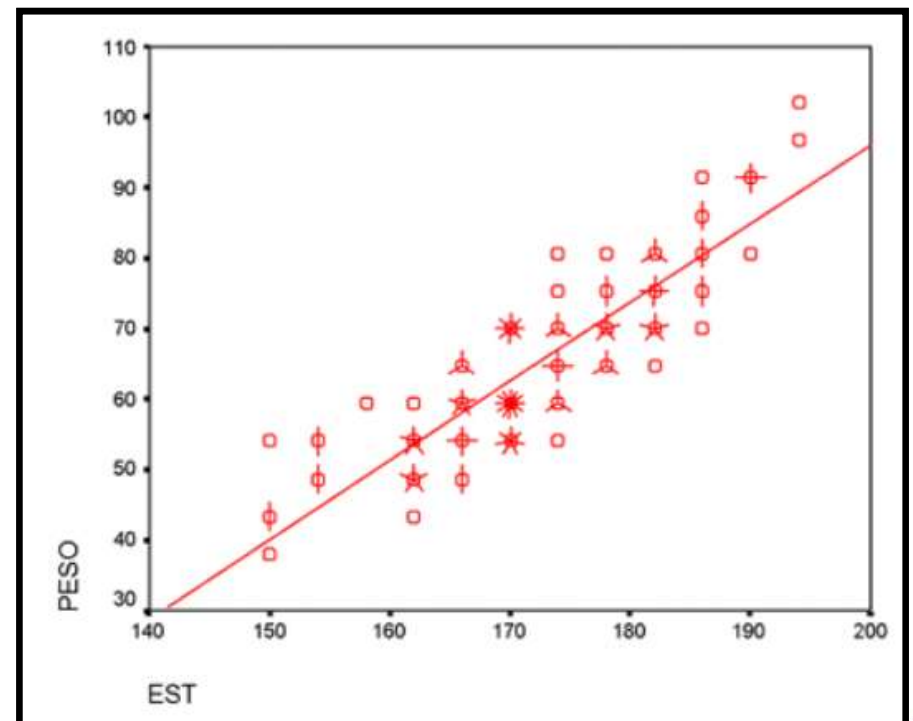
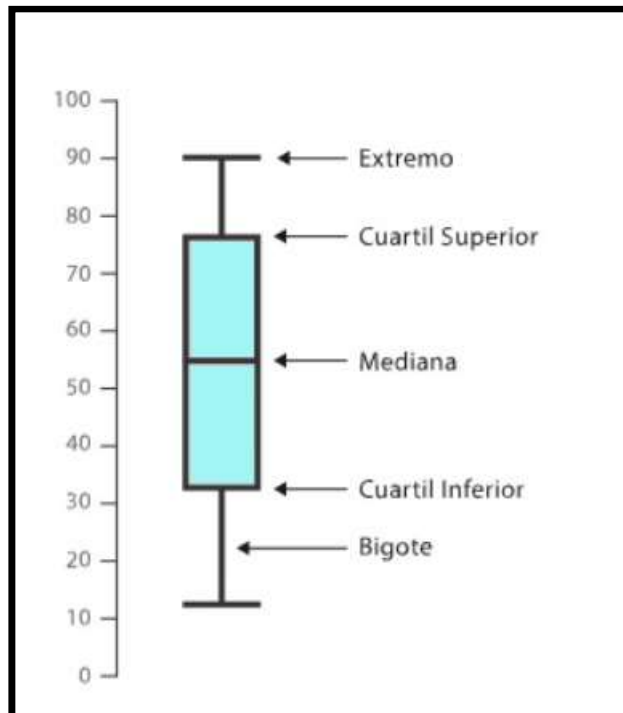
# Estadística en la vida



# ¿En qué situación aplica?



# Otros diagramas



# Datos e información

- La asociación de agentes inmobiliarios informó que la mediana del precio de venta de una casa en Estados Unidos es de \$215 000 (*The Wall Street Journal*, 16 de enero de 2006).
- Durante el Super Bowl de 2006 el costo promedio de un spot publicitario de 30 segundos en televisión fue de \$2.5 millones (*USA Today*, 27 de enero de 2007).
- En una encuesta de Jupiter Media se encontró que 31% de los hombres adultos ven más de 10 horas de televisión a la semana. Entre las mujeres sólo 26% (*The Wall Street Journal*, 26 de enero de 2004).



# Conceptos básicos

- **Datos:** son *hechos/informaciones y cifras* que se recogen, analizan y resumen para su presentación e interpretación.
- **Conjunto de datos:** todos los datos reunidos para un determinado estudio.

## Ejemplo

La tabla 1.1 muestra un conjunto de datos que contiene información sobre 25 empresas que forman parte del S&P 500. El S&P 500 consta de 500 empresas elegidas por Standard & Poor's. Estas empresas representan 76% de la capitalización de mercado de todas las acciones de Estados Unidos. Las acciones de S&P 500 son estrechamente observadas por los inversionistas y por los analistas de Wall Street.



**TABLA 1.1** CONJUNTO DE DATOS DE 25 EMPRESAS S&P 500

<b>Empresa</b>	<b>Bolsa de valores</b>	<b>Denominación abreviada Ticker</b>	<b>Posición en <i>BusinessWeek</i></b>	<b>Precio por acción (\$)</b>	<b>Ganancia por acción (\$)</b>
Abbott Laboratories	N	ABT	90	46	2.02
Altria Group	N	MO	148	66	4.57
Apollo Group	NQ	APOL	174	74	0.90
Bank of New York	N	BK	305	30	1.85
Bristol-Myers Squibb	N	BMJ	346	26	1.21
Cincinnati Financial	NQ	CINF	161	45	2.73
Comcast	NQ	CMCSA	296	32	0.43
Deere	N	DE	36	71	5.77
eBay	NQ	EBAY	19	43	0.57
Federated Dept. Stores	N	FD	353	56	3.86
Hasbro	N	HAS	373	21	0.96
IBM	N	IBM	216	93	4.94
International Paper	N	IP	370	37	0.98
Knight-Ridder	N	KRI	397	66	4.13
Manor Care	N	HCR	285	34	1.90
Medtronic	N	MDT	53	52	1.79
National Semiconductor	N	NSM	155	20	1.03
Novellus Systems	NQ	NVLS	386	30	1.06
Pitney Bowes	N	PBI	339	46	2.05
Pulte Homes	N	PHM	12	78	7.67
SBC Communications	N	SBC	371	24	1.52
St. Paul Travelers	N	STA	264	38	1.53
Teradyne	N	TER	412	15	0.84
UnitedHealth Group	N	UNH	5	91	3.94
Wells Fargo	N	WFC	159	59	4.09

*Fuente: Business Week (4 de abril de 2005).*





# Elementos

**Elementos** son las entidades de las que se obtienen los datos.

En el conjunto de datos de la tabla 1.1, cada acción de una empresa es un **elemento**; los nombres de los elementos aparecen en la primera columna. Como se tienen 25 acciones, el conjunto de datos contiene 25 elementos.

Una **variable** es una característica de los elementos que es de interés.

# Variables

Empresa	Bolsa de valores	Denominación abreviada Ticker	Posición en <i>BusinessWeek</i>	Precio por acción (\$)	Ganancia por acción (\$)
Abbott Laboratories	N	ABT	90	46	2.02

Elementos

Variables





# Observaciones

Los valores encontrados para cada variable en cada uno de los elementos constituyen los datos. *Al conjunto de mediciones obtenidas para un determinado elemento se le llama **observación**.*

## Elementos

## Variables

Empresa	Variables				
	Bolsa de valores	Denominación abreviada Ticker	Posición en <i>BusinessWeek</i>	Precio por acción (\$)	Ganancia por acción (\$)
Abbott Laboratories	N	ABT	90	46	2.02
Altria Group	N	MO	148	66	4.57
Apollo Group	NQ	APOL	174	74	0.90
Bank of New York	N	BK	305	30	1.85
Bristol-Myers Squibb	N	BMJ	346	26	1.21
Cincinnati Financial	NQ	CINF	161	45	2.73
Comcast	NQ	CMCSA	296	32	0.43

Observación para  
el primer elemento

Observación para el  
séptimo elemento



**TABLA 1.1 CONJUNTO DE DATOS DE 25 EMPRESAS S&P 500**

Empresa	Denominación		Posición en <i>BusinessWeek</i>	Precio por acción (\$)	Ganancia por acción (\$)
	Bolsa de valores	abreviada Ticker			
Abbott Laboratories	N	ABT	90	46	2.02
Altria Group	N	MO	148	66	4.57
Apollo Group	NQ	APOL	174	74	0.90
Bank of New York	N	BK	305	30	1.85
Bristol-Myers Squibb	N	BMJ	346	26	1.21
Cincinnati Financial	NQ	CINF	161	45	2.73
Comcast	NQ	CMCSA	296	32	0.43
Deere	N	DE	36	71	5.77
eBay	NQ	EBAY	19	43	0.57
Federated Dept. Stores	N	FD	353	56	3.86
Hasbro	N	HAS	373	21	0.96
IBM	N	IBM	216	93	4.94
International Paper	N	IP	370	37	0.98
Knight-Ridder	N	KRI	397	66	4.13
Manor Care	N	HCR	285	34	1.90
Medtronic	N	MDT	53	52	1.79
National Semiconductor	N	NSM	155	20	1.03
Novellus Systems	NQ	NVLS	386	30	1.06
Pitney Bowes	N	PBI	339	46	2.05
Pulte Homes	N	PHM	12	78	7.67
SBC Communications	N	SBC	371	24	1.52
St. Paul Travelers	N	STA	264	38	1.53
Teradyne	N	TER	412	15	0.84
UnitedHealth Group	N	UNH	5	91	3.94
Wells Fargo	N	WFC	159	59	4.09

*Fuente: Business Week (4 de abril de 2005).*

Variables

Elementos

Observaciones



# Ejercicio

Construya un conjunto de datos que contenga 5 variables, 4 elementos y las observaciones para cada uno de ellos.



# Escalas de medición

La escala de medición determina la cantidad de información contenida en el dato e indica la manera más apropiada de resumir y de analizar estadísticamente los datos.

- **Escala nominal:** cuando las variables se diferencian por sus nombres. Estas variables no tienen orden ni jerarquía asociada a ellas aunque pueden ser representadas por etiquetas numéricas.

✓ **Color.** *Blanco* = 1 *Negro* = 2

✓ **Género.** Hombre = H    Mujer = M

✓ **Comportamiento.** E = Extrovertido    I = introvertido    A = ambas



# ¿Cuáles variables se miden en escala nominal?

Empresa	Bolsa de valores	Denominación abreviada Ticker	Posición en <i>BusinessWeek</i>	Ganancia	
				Precio por acción (\$)	por acción (\$)
Abbott Laboratories	N	ABT	90	46	2.02
Altria Group	N	MO	148	66	4.57
Apollo Group	NQ	APOL	174	74	0.90
Bank of New York	N	BK	305	30	1.85
Bristol-Myers Squibb	N	BMJ	346	26	1.21
Cincinnati Financial	NQ	CINF	161	45	2.73
Comcast	NQ	CMCSA	296	32	0.43



- **Escala ordinal:** Cuando los datos muestran las propiedades de los datos nominales y además *tiene sentido el orden o jerarquía de los datos*.

✓ *Encuesta de satisfacción. Excelente, Bueno, Malo*

Excelente=3      Bueno=2      Malo = 1

✓ *Posición en un ranking. Primero – Segundo – Tercero*



## ■ Escala de intervalo:

- ☐ *Son variables numéricas que representan magnitudes.*
- ☐ *Establecen orden y distancia (significado de la diferencia).*
- ☐ *0 no significa la ausencia de la variable*

✓ Las calificaciones obtenidas por tres alumnos en la prueba de matemáticas con 620, 550 y 470, pueden ser ordenadas en orden de mejor a peor.

✓ La temperatura

✓ Edad





### **Escala de razón:**

- Cuando los datos tienen todas las propiedades de los datos de intervalo y la proporción entre dos valores tiene significado.
- Esta escala requiere que se tenga el valor cero para indicar que en este punto no existe la variable.

✓ Valor de un auto.

✓ *Altura*

✓ *Peso*

✓ *Salario*



# Datos cualitativos y cuantitativos

Los **datos cualitativos** comprenden etiquetas o nombres que se usan para identificar un atributo de cada elemento. Los datos cualitativos emplean la escala nominal o la ordinal y pueden ser numéricos o no. Una **variable cualitativa** es una variable con datos cualitativos.

Los **datos cuantitativos** requieren valores numéricos que indiquen cuánto o cuántos. Los datos cuantitativos se obtienen usando las escalas de medición de intervalo o de razón. Una **variable cuantitativa** es una variable con datos cuantitativos.

Se dividen en datos continuos y discretos.



# Ejercicio

Clasificar las siguientes variables según su escala y su tipo (Cuali/cuanti)

- a. Ventas anuales.
- b. Tamaño de los refrescos (pequeño, mediano, grande).
- c. Clasificación como empleado (GS 1 a GS 18).
- d. Ganancia por acción.
- e. Modo de pago (al contado, cheque, tarjeta de crédito).
- f. Año

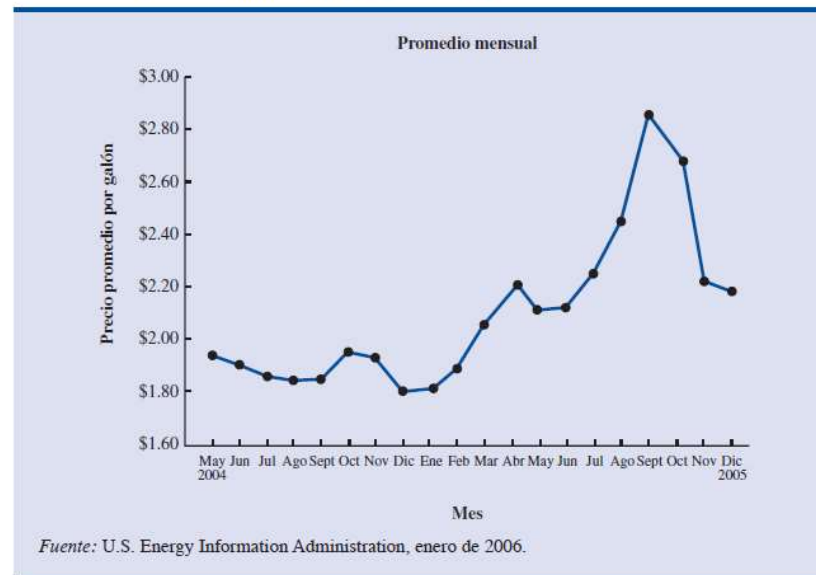


# Sección transversal y series de tiempo

Datos de sección transversal son los obtenidos en el mismo o aproximadamente el mismo momento (punto en el tiempo).

Los datos de series de tiempo son datos obtenidos a lo largo de varios periodos.

**FIGURA 1.1** PRECIO PROMEDIO POR GALÓN DE GASOLINA NORMAL EN LAS CIUDADES DE ESTADOS UNIDOS



# Fuentes de datos

- Fuentes ya existentes
- Encuestas
- Estudios experimentales

**TABLA 1.2** EJEMPLOS DE DATOS DISPONIBLES DE LOS REGISTROS DE EMPRESAS INTERNACIONALES

Fuente	Algunos de los datos disponibles
Registros sobre los empleados	Nombre, dirección, número de seguridad social, salario, días de vacaciones, días de enfermedad y bonos
Registros de producción	Parte o número de producto, cantidad producida, costo de mano de obra y costo de materiales
Registros de inventario	Parte o número de producto, cantidad de unidades disponibles, nivel de reaprovisionamiento, cantidad económica a ordenar y programa de descuento
Registros de ventas	Número del producto, volumen de ventas, volumen de ventas por región y volumen de ventas por tipo de cliente
Registros de créditos	Nombre del cliente, dirección, número de teléfono, crédito límite y cuentas por cobrar
Perfil de clientes	Edad, género, nivel de ingresos, número de miembros en la familia, dirección y preferencias



**TABLA 1.3** EJEMPLO DE LOS DATOS DISPONIBLES DE ALGUNAS DEPENDENCIAS GUBERNAMENTALES

Dependencia gubernamental	Algunos de los datos disponibles
Oficina de Censos <i>www.census.gov</i>	Datos poblacionales, número de hogares e ingresos de los hogares
Junta de la Reserva Federal <i>www.federalreserve.gov</i>	Datos sobre dinero en circulación, créditos a plazos, tasas de cambio y tasas de interés
Oficina de Administración y Presupuesto <i>www.whitehouse.gov/omb</i>	Datos sobre ingresos, gastos y deudas del gobierno federal
Departamento de Comercio <i>www.doc.gov</i>	Datos sobre las actividades comerciales, valor de los embarques por industria, nivel de ganancia por industria e industrias en crecimiento y en decremento
Oficina de Estadística Laboral <i>www.bls.gov</i>	Gasto de los consumidores, salarios por hora, tasa de desempleo y estadísticas internacionales



# Errores en la adquisición

FIGURA 1.4 CUESTIONARIO PARA CONOCER LA OPINIÓN DE LOS CLIENTES EMPLEADO EN EL RESTAURANTE THE LOBSTER POT DE REDINGTON SHORES, FLORIDA

The  
LOBSTER  
Pot  
RESTAURANT

Nos alegramos de su visita al restaurante Lobster Pot y queremos estar seguros de que volverá. De manera que si tiene unos minutos le agradeceríamos mucho que nos llenara esta tarjeta. Sus comentarios y sugerencias son extremadamente importantes para nosotros. Gracias.

Nombre de la persona que lo atendió \_\_\_\_\_

	Excelente	Bueno	Satisfactorio	Insatisfactorio
Calidad de los alimentos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Amabilidad en el servicio	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Prontitud en el servicio	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Limpieza	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gestión	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Comentarios \_\_\_\_\_

¿Qué lo motivó a visitarnos? \_\_\_\_\_

Favor de depositarlo en el buzón de sugerencias que se encuentra a la entrada.





# Población

La población es el conjunto de todos los elementos de interés en un estudio determinado.

# Muestra

La muestra es un subconjunto de la población.

*Al proceso de realizar un estudio para recolectar datos de toda una población se le llama **censo**. Al proceso de efectuar un estudio para recolectar datos de una muestra se le llama **encuesta muestral**. Una de las principales contribuciones de la estadística es emplear datos de una muestra para hacer estimaciones y probar hipótesis acerca de las características de una población mediante un proceso al que se le conoce como **inferencia estadística**.*



# Tarea

La revista *Condé Nast Traveler* realiza una encuesta anual entre sus suscriptores con objeto de determinar los mejores alojamientos del mundo. En la tabla 1.6 se presenta una muestra de nueve hoteles europeos (*Condé Nast Traveler*, enero de 2000) Los precios de una habitación doble estándar van de \$(precio más bajo) a \$\$\$\$ (precio más alto). La calificación general corresponde a la evaluación de habitaciones, servicio, restaurante, ubicación/atmósfera y áreas públicas; cuanto más alta sea la calificación general, mayor es el nivel de satisfacción.

**TABLA 1.6** CALIFICACIONES PARA NUEVE LUGARES DONDE ALOJARSE EN EUROPA

Nombre del lugar	País	Precio de la habitación	Número de habitaciones	Calificación general
Graveteye Manor	Inglaterra	\$	18	83.6
Villa d'Este	Italia	\$\$\$	166	86.3
Hotel Prem	Alemania	\$	54	77.8
Hotel d'Europe	Francia	\$	47	76.8
Palace Luzern	Suiza	\$	326	80.9
Royal Crescent Hotel	Inglaterra	\$\$\$	45	73.7
Hotel Sacher	Austria	\$\$\$	120	85.5
Duc de Bourgogne	Bélgica	\$	10	76.9
Villa Gallici	Francia	\$	22	90.6

Fuente: *Condé Nast Traveler*, enero de 2000.



**TABLA 1.6** CALIFICACIONES PARA NUEVE LUGARES DONDE ALOJARSE EN EUROPA

Nombre del lugar	País	Precio de la habitación	Número de habitaciones	Calificación general
Graveteye Manor	Inglaterra	\$\$	18	83.6
Villa d'Este	Italia	\$\$\$\$	166	86.3
Hotel Prem	Alemania	\$	54	77.8
Hotel d'Europe	Francia	\$\$	47	76.8
Palace Luzern	Suiza	\$\$	326	80.9
Royal Crescent Hotel	Inglaterra	\$\$\$	45	73.7
Hotel Sacher	Austria	\$\$\$	120	85.5
Duc de Bourgogne	Bélgica	\$	10	76.9
Villa Gallici	Francia	\$\$	22	90.6

*Fuente: Condé Nast Traveler, enero de 2000.*

- ¿Cuántos elementos hay en este conjunto de datos?
- ¿Cuántas variables hay en este conjunto de datos?
- ¿Cuáles variables son cualitativas y cuáles cuantitativas?
- ¿Qué tipo de escala de medición se usa para cada variable?



Los equipos de sonido todo en uno, llamados minicomponentes, cuentan con sintonizador AM/FM, casetera doble, cargador para un disco compacto con bocinas separadas. En la tabla 1.7 se muestran los precios de menudeo, calidad de sonido, capacidad para discos compactos, sensibilidad y selectividad de la sintonización y cantidad de caseteras en los artículos de una muestra de 10 minicomponentes (*Consumer Report Buying Guide 2002*).bb

**TABLA 1.7** UNA MUESTRA DE 10 MINICOMPONENTES

Marca y modelo	Precio (\$)	Calidad de sonido	Capacidad para CD	Sintonización FM	Caseteras
Aiwa NSX-AJ800	250	Buena	3	Regular	2
JVC FS-SD1000	500	Buena	1	Muy buena	0
JVC MX-G50	200	Muy buena	3	Excelente	2
Panasonic SC-PM11	170	Regular	5	Muy buena	1
RCA RS 1283	170	Buena	3	Mala	0
Sharp CD-BA2600	150	Buena	3	Buena	2
Sony CHC-CL1	300	Muy buena	3	Muy buena	1
Sony MHC-NX1	500	Buena	5	Excelente	2
Yamaha GX-505	400	Muy buena	3	Excelente	1
Yamaha MCR-E100	500	Muy buena	1	Excelente	0



**TABLA 1.7** UNA MUESTRA DE 10 MINICOMPONENTES

Marca y modelo	Precio (\$)	Calidad de sonido	Capacidad para CD	Sintonización FM	Caseteras
Aiwa NSX-AJ800	250	Buena	3	Regular	2
JVC FS-SD1000	500	Buena	1	Muy buena	0
JVC MX-G50	200	Muy buena	3	Excelente	2
Panasonic SC-PM11	170	Regular	5	Muy buena	1
RCA RS 1283	170	Buena	3	Mala	0
Sharp CD-BA2600	150	Buena	3	Buena	2
Sony CHC-CL1	300	Muy buena	3	Muy buena	1
Sony MHC-NX1	500	Buena	5	Excelente	2
Yamaha GX-505	400	Muy buena	3	Excelente	1
Yamaha MCR-E100	500	Muy buena	1	Excelente	0

- ¿Cuántos elementos contiene este conjunto de datos?
- ¿Cuál es la población?
- ¿Cuántas variables hay en este conjunto de datos?
- De estas variables, ¿cuáles son cualitativas y cuáles son cuantitativas?



*La Columbia House vende discos compactos a los miembros de su club de venta por correo. En una encuesta sobre música se les pidió a los nuevos miembros del club que llenaran un cuestionario con 11 preguntas. Algunas de las preguntas eran:*

- a. ¿Cuántos discos compactos has comprado en los últimos 12 meses?
- b. ¿Eres miembro de algún club de venta de libros por correo (Sí o No)?
- c. ¿Cuál es tu edad?
- d. Incluyéndote a ti, de cuántas personas (adultos y niños) consta tu familia.
- e. ¿Qué tipo de música te interesa comprar? Se presentaban quince categorías entre las que se encontraban rock pesado, rock ligero, música contemporánea para adultos, rap y rancheras.

Responde si los datos que se obtienen con cada pregunta son cualitativos o cuantitativos.



*El hotel Ritz Carlton emplea un cuestionario de opinión del cliente para obtener datos sobre la calidad de sus servicios de restaurante y entretenimiento (The Ritz-Carlton Hotel, Naples, Florida, febrero de 2006). Se les pidió a los clientes que evaluaran seis puntos: recibimiento, servicio, alimentos, menú, atención y atmósfera. Los datos registrados para cada factor fueron 1 para Pasadero, 2 Regular, 3 Bueno y 4 Excelente.*

- a. Las respuestas de los clientes proporcionan datos para seis variables. ¿Son estas variables cualitativas o cuantitativas?
- b. ¿Qué escala de medición se usa?

