



#vivaFCA

# Proceso de Selección 2024

## LICENCIATURA EN ACTUARÍA





#vivaFCA

## SESIÓN 2

# Introducción a la estadística



# Introducción

- ▶ 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).



# Introducción

- ▶ A los datos numéricos de las frases anteriores se les llama estadísticas. En este sentido el término estadística se refiere a datos numéricos, tales como promedios, medianas, porcentajes y números índices que ayudan a entender una gran variedad de negocios y situaciones económicas.



# Introducción

- ▶ En un sentido amplio, la estadística se define como el arte y la ciencia de reunir datos, analizarlos, presentarlos e interpretarlos.



# Aplicaciones en los negocios y la economía

- ▶ En el entorno mundial actual de los negocios y de la economía, todo mundo tiene acceso a enormes cantidades de información estadística. Los directivos y los encargados de tomar decisiones que tienen éxito entienden la información y saben usarla de manera eficiente. En esta sección se proporcionan ejemplos que ilustran algunos de los usos de la estadística en los negocios y en la economía.



# Aplicaciones en los negocios y la economía

- ▶ Contaduría
- ▶ Finanzas
- ▶ Marketing
- ▶ Producción
- ▶ Economía



# Datos

- ▶ Datos son hechos/informaciones y cifras que se recogen, analizan y resumen para su presentación e interpretación. A todos los datos reunidos para un determinado estudio se les llama conjunto de datos para el estudio.



# Elementos, variables y observaciones

- ▶ Elementos son las entidades de las que se obtienen los datos.
- ▶ Una variable es una característica de los elementos que es de interés.
- ▶ 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 .



# Escalas de medición

- ▶ La recolección de datos requiere alguna de las escalas de medición siguientes: nominal, ordinal, de intervalo o de razón. 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.



# Escalas de medición

- ▶ Cuando el dato de una variable es una etiqueta o un nombre que identifica un atributo de un elemento, se considera que la escala de medición es una **escala nominal**.



# Escalas de medición

- ▶ Una escala de medición para una **variable es ordinal** si los datos muestran las propiedades de los datos nominales y además tiene sentido el orden o jerarquía de los datos.



# Escalas de medición

- ▶ Una escala de medición para una variable es una **escala de intervalo** si los datos tienen las características de los datos ordinales y el intervalo entre valores se expresa en términos de una unidad de medición fija. Los datos de intervalo siempre son numéricos.



# Escala de medición

- ▶ Una variable tiene una **escala de razón** si los datos tienen todas las propiedades de los datos de intervalo y la proporción entre dos valores tiene significado.



# Datos Cualitativos y cuantitativos

- ▶ 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.



# 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.



# Datos de sección transversal y de series de tiempo

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

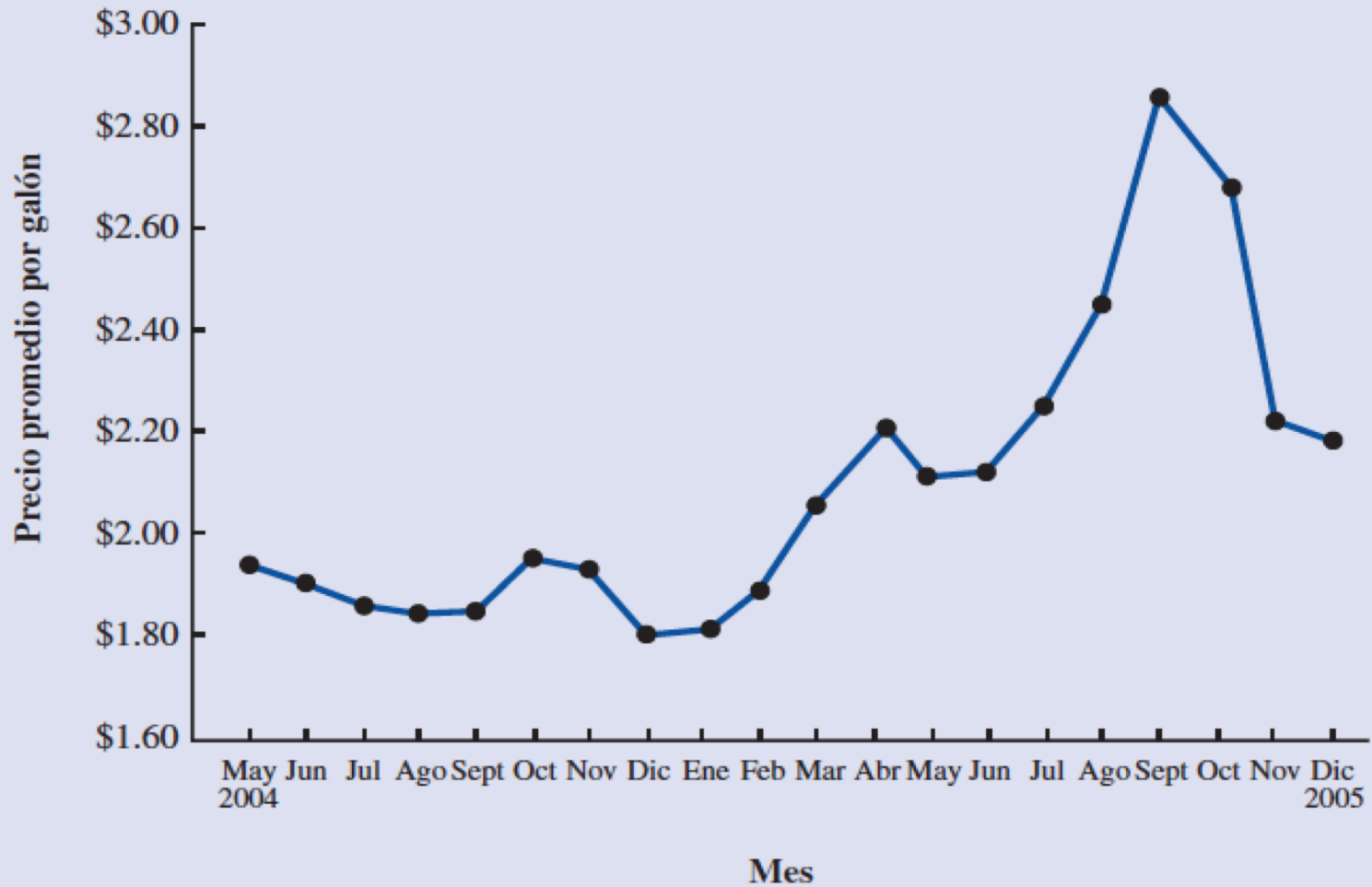


# Datos de sección transversal y de series de tiempo

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



### Promedio mensual



Fuente: U.S. Energy Information Administration, enero de 2006.



# Fuentes de datos

- ▶ Los datos se obtienen de fuentes ya existentes o por medio de encuestas y estudios experimentales realizados con objeto de recolectar nuevos datos.



# Fuentes Existentes

- ▶ En algunos casos los datos que se necesitan para una determinada aplicación ya existen. Las empresas cuentan con diversas bases de datos sobre sus empleados, clientes y operaciones de negocios.
- ▶ La importancia de Internet como fuente de datos y de información estadística sigue creciendo.
- ▶ Las dependencias de los gobiernos son otra fuente importante de datos.



# Estudios estadísticos

- ▶ Algunas veces, los datos necesarios para una aplicación particular no se pueden obtener de las fuentes existentes. En tales casos los datos suelen conseguirse realizando un estudio estadístico. Dichos estudios se clasifican como experimentales u observacionales.



# Estudios experimentales

- ▶ En los estudios experimentales se identifica primero la variable de interés. Después se ubica otra u otras variables que son controladas para lograr datos de cómo ésta influye sobre la variable de interés.



# Estudios observacionales

- ▶ En los estudios estadísticos no experimentales y observacionales, no se controlan las variables de interés. El tipo más usual de estudio observacional es quizá una encuesta.



# Errores en la adquisición de datos

- ▶ Usar datos erróneos es peor que no usar ningún dato. Un error en la adquisición de datos se tiene siempre que el valor del dato obtenido no es igual al verdadero valor o al valor real que se hubiera obtenido con un procedimiento correcto.



# Errores en la adquisición de datos

- ▶ Por ejemplo, un entrevistador puede cometer un error de escritura, como una transposición al escribir la edad de una persona y en lugar de 24 años escribir 42 años, o en una entrevista, el entrevistado puede malinterpretar una pregunta y dar una respuesta incorrecta.



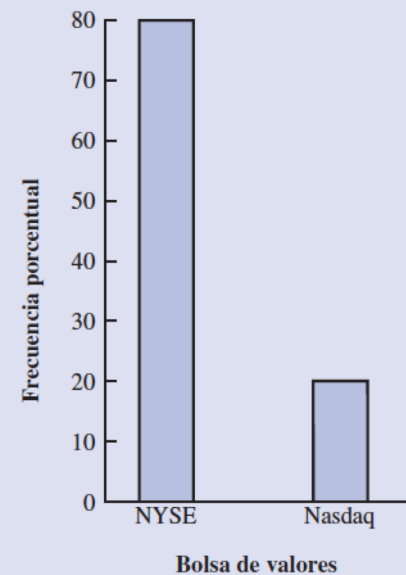
# Estadística descriptiva

- ▶ La mayor parte de la información estadística en periódicos, revistas, informes de empresas y otras publicaciones consta de datos que se resumen y presentan en una forma fácil de leer y de entender. A estos resúmenes de datos, que pueden ser tabulares, gráficos o numéricos se les conoce como estadística descriptiva .



# Estadística descriptiva

- ▶ Los métodos de la estadística descriptiva pueden emplearse para resumir la información en este conjunto de datos.





# Inferencia estadística

- ▶ En muchas situaciones se requiere información acerca de grupos grandes de elementos (individuos, empresas, votantes, hogares, productos, clientes, etc.). Pero, debido al tiempo, costo y a otras consideraciones, sólo es posible recolectar los datos de una pequeña parte de este grupo



# Inferencia estadística

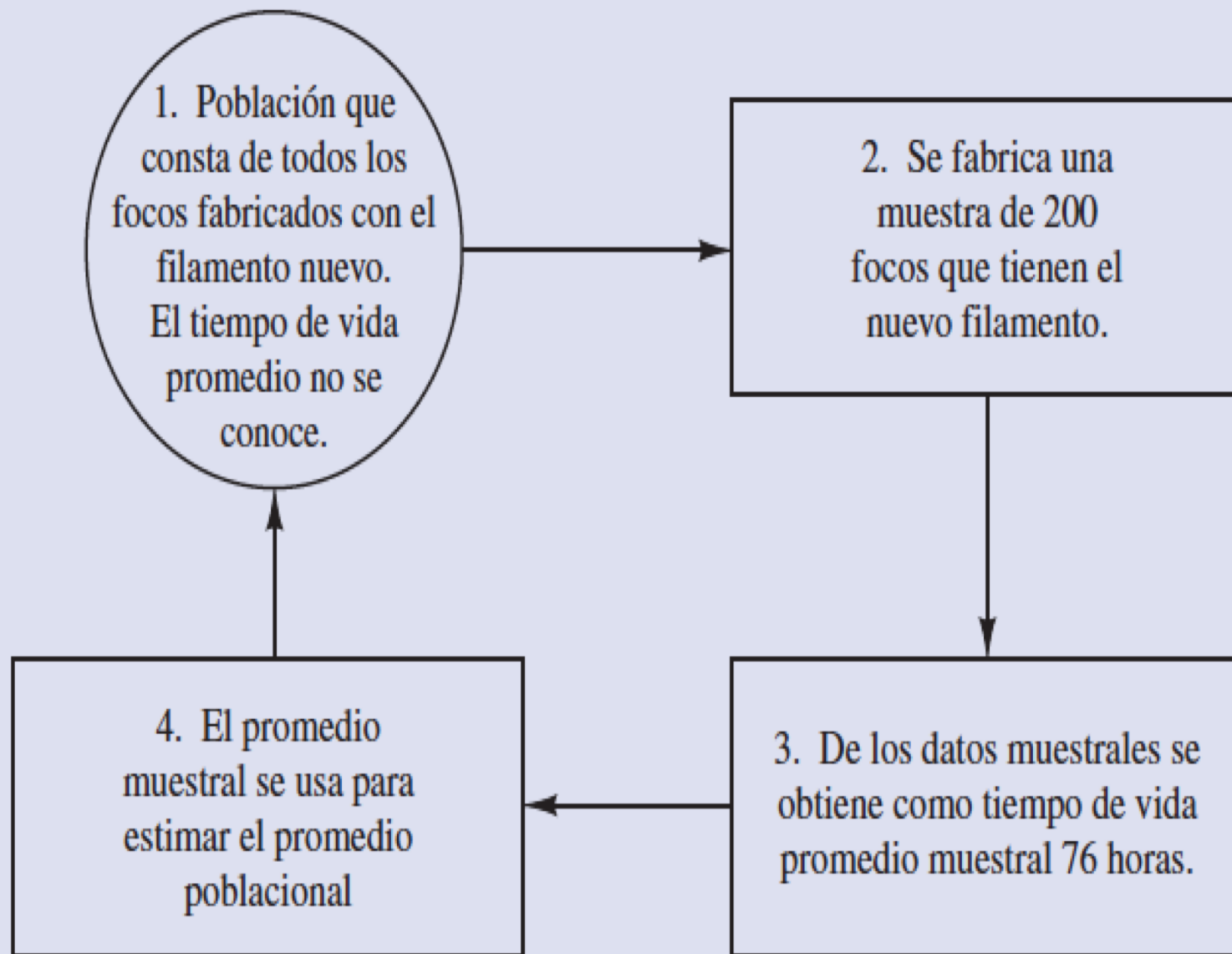
- ▶ Al grupo grande de elementos en un determinado estudio se le llama población y al grupo pequeño muestra
- ▶ **POBLACIÓN:** La población es el conjunto de todos los elementos de interés en un estudio determinado.
- ▶ **MUESTRA:** La muestra es un subconjunto de la población.



# Inferencia estadística

- ▶ 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 .









Datos

Cualitativos



# Distribución de frecuencia

- ▶ Una distribución de frecuencia es un resumen tabular de datos que muestra el número (frecuencia) de elementos en cada una de las diferentes clases disyuntas (que no se sobreponen).




Coke Classic  
Diet Coke  
Pepsi  
Diet Coke  
Coke Classic  
Coke Classic  
Dr. Pepper  
Diet Coke  
Pepsi  
Pepsi  
Coke Classic  
Dr. Pepper  
Sprite  
Coke Classic  
Diet Coke  
Coke Classic  
Coke Classic

Sprite  
Coke Classic  
Diet Coke  
Coke Classic  
Diet Coke  
Coke Classic  
Sprite  
Pepsi  
Coke Classic  
Coke Classic  
Pepsi  
Coke Classic  
Sprite  
Dr. Pepper  
Pepsi  
Diet Coke

Pepsi  
Coke Classic  
Coke Classic  
Coke Classic  
Pepsi  
Dr. Pepper  
Coke Classic  
Diet Coke  
Pepsi  
Pepsi  
Pepsi  
Coke Classic  
Dr. Pepper  
Pepsi  
Sprite





Refresco	Frecuencia
Coke Classic	19
Diet Coke	8
Dr. Peper	5
Pepsi	13
Sprite	5
Total	50



# Distribución de frecuencia relativa

- ▶ La frecuencia relativa de una clase es igual a la parte o proporción de los elementos que pertenecen a cada clase. En un conjunto de datos, en el que hay  $n$  observaciones.

## FRECUENCIA RELATIVA

$$\text{Frecuencia relativa de una clase} = \frac{\text{Frecuencia de la clase}}{n}$$




# Distribución de frecuencia porcentual

- ▶ Una distribución de frecuencia porcentual da la frecuencia porcentual de los datos de cada clase.

$$\text{Frecuencia porcentual de una clase} = \left( \frac{\text{Frecuencia de la clase}}{n} \right) 100$$





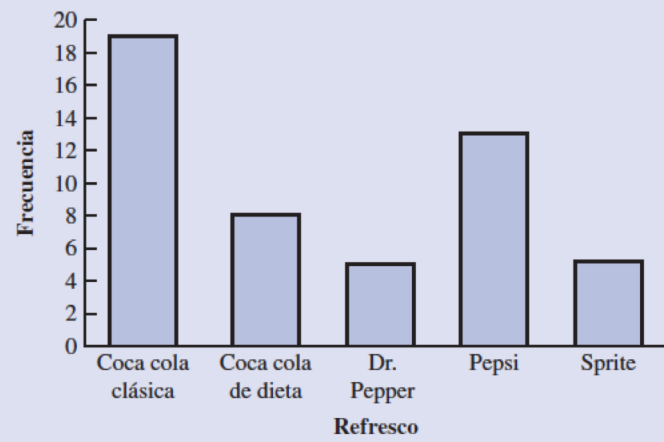
<b>Refresco</b>	<b>Frecuencia relativa</b>	<b>Frecuencia porcentual</b>
Coke Classic	0.38	38
Diet Coke	0.16	16
Dr. Pepper	0.10	10
Pepsi	0.26	26
Sprite	<u>0.10</u>	<u>10</u>
Total	1.00	100



# Gráfica de barras

- ▶ Una gráfica de barras o un diagrama de barras, es una gráfica para representar los datos cualitativos de una distribución de frecuencia, de frecuencia relativa o de frecuencia porcentual. En uno de los ejes de la gráfica (por lo general en el horizontal), se especifican las etiquetas empleadas para las clases (categorías). Para el otro eje de la gráfica (el vertical) se usa una escala para frecuencia, frecuencia relativa o frecuencia porcentual.
- ▶ Cuando se tienen datos cualitativos, las barras deben estar separadas para hacer énfasis en que cada clase está separada.





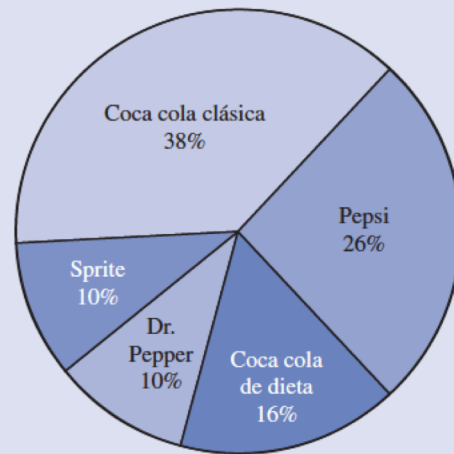


# Gráfica de pastel

- ▶ La gráfica de pastel proporciona otra gráfica para presentar distribuciones de frecuencia relativa y de frecuencia porcentual de datos cualitativos.

*Frecuencia relativa ( $360^\circ$ )*







# Ejercicio

Clase	Frecuencia relativa
A	0.22
B	0.18
C	0.40
D	

- ▶ a. ¿Cuál es la frecuencia relativa de la clase D?
- ▶ b. El tamaño de la muestra es 200. ¿Cuál es la frecuencia de la clase D?
- ▶ c. Muestre la distribución de frecuencia.
- ▶ d. Dé la distribución de frecuencia porcentual.



Los cinco apellidos más comunes en Estados Unidos, en orden alfabético son, Brown, Davis, Johnson, Jones, Smith y Williams (*The World Almanac, 2006*). Suponga que en una muestra de 50 personas con uno de estos apellidos se obtienen los datos siguientes.

Brown	Williams	Williams	Williams	Brown
Smith	Jones	Smith	Johnson	Smith
Davis	Smith	Brown	Williams	Johnson
Johnson	Smith	Smith	Johnson	Brown
Williams	Davis	Johnson	Williams	Johnson
Williams	Johnson	Jones	Smith	Brown
Johnson	Smith	Smith	Brown	Jones
Jones	Jones	Smith	Smith	Davis
Davis	Jones	Williams	Davis	Smith
Jones	Johnson	Brown	Johnson	Davis

Resuma estos datos construyendo:

- Distribuciones de frecuencia relativa y porcentual.
- Una gráfica de barras.
- Una gráfica de pastel.
- De acuerdo con estos datos, ¿cuáles son los tres apellidos más comunes?





# Datos cuantitativos



# Distribución de frecuencia

- ▶ Cuando se trata de datos cuantitativos se debe tener más cuidado al definir las clases disyuntas que se van a usar en la distribución de frecuencia.
- ▶ Los tres pasos necesarios para definir las clases de una distribución de frecuencia con datos cuantitativos son
  1. Determinar el número de clases disyuntas.
  2. Determinar el ancho de cada clase
  3. Determinar los límites de clase.



# Numero de clases

- ▶ Las clases se forman especificando los intervalos que se usarán para agrupar los datos. Se recomienda emplear entre 5 y 20 clases. Cuando los datos son pocos, cinco o seis clases bastan para resumirlos. Si son muchos, se suele requerir más clases. La idea es tener las clases suficientes para que se muestre la variación en los datos, pero no deben ser demasiadas si algunas de ellas contienen sólo unos cuantos datos.



# Ancho de clase

- ▶ Como regla general es recomendable que el ancho sea el mismo para todas las clases. Así, el ancho y el número de clases no son decisiones independientes. Entre mayor sea el número de clases menor es el ancho de las clases y viceversa.


$$\text{Ancho aproximada de clase} = \frac{\text{Valor mayor en los datos} - \text{Valor menor en los datos}}{\text{Número de clase}}$$



# Limite de clase

- ▶ Los límites de clase deben elegirse de manera que cada dato pertenezca a una y sólo una de las clases. El límite de clase inferior indica el menor valor de los datos a que pertenece esa clase. El límite de clase superior indica el mayor valor de los datos a que pertenece esa clase.





AUDITORÍA ANUAL (DÍAS DE DURACIÓN)			
12	14	19	18
15	15	18	17
20	27	22	23
22	21	33	28
14	18	16	13



- ▶ Número de clases = 5

$$\frac{33 - 12}{5} = 4.2 = 5$$

Límites de las clases	Frecuencia	Frecuencia relativa	Frecuencia porcentual
12-16	7	0.35	35%
17-21	7	0.35	35%
22-26	3	0.15	15%
27-31	2	0.10	10%
32-36	1	0.05	5%
Sumas	20	1	100



DISTRIBUCIÓN DE FRECUENCIA DE LAS AUDITORÍAS	
Duración de las audito- rías (días)	Frecuencia
10-14	4
15-19	8
20-24	5
25-29	2
30-34	1
Total	20



# Frecuencia relativa y porcentual

- ▶ Las distribuciones de frecuencia relativa y de frecuencia porcentual para datos cuantitativos se definen de la misma forma que para datos cualitativos.

$$\text{Frecuencia relativa de la clase} = \frac{\text{Frecuencia de la clase}}{n}$$

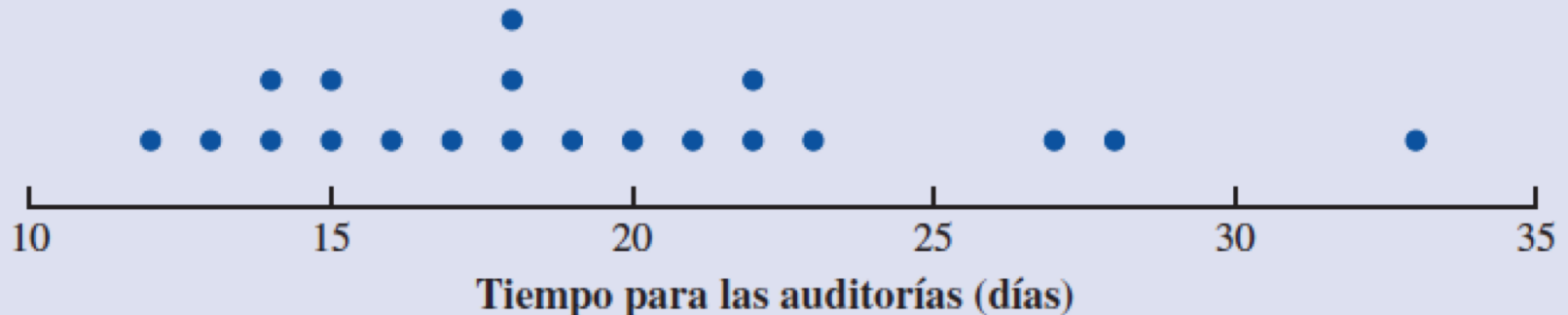


<b>Duración de las auditorías (días)</b>	<b>Frecuencia relativa</b>	<b>Frecuencia porcentual</b>
10–14	0.20	20
15–19	0.40	40
20–24	0.25	25
25–29	0.10	10
30–34	0.05	5
Total	1.00	100



# Gráfico de puntos

- ▶ Uno de los más sencillos resúmenes gráficos de datos son las gráficas de puntos. En el eje horizontal se presenta el intervalo de los datos. Cada dato se representa por un punto colocado sobre este eje.

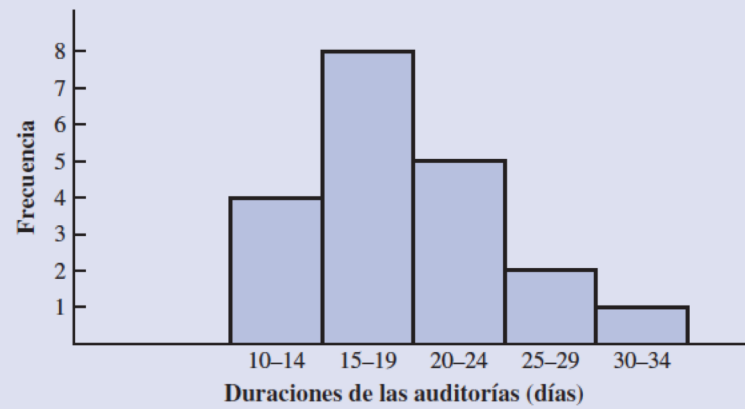




# Histograma

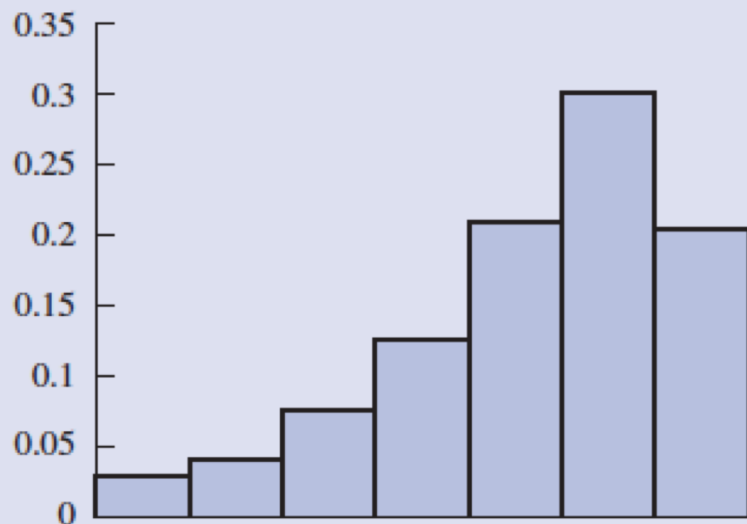
- ▶ Un histograma se construye colocando la variable de interés en el eje horizontal y la frecuencia, la frecuencia relativa o la frecuencia porcentual en el eje vertical.
- ▶ Uno de los usos más importantes de un histograma es proveer información acerca de la forma de la distribución.



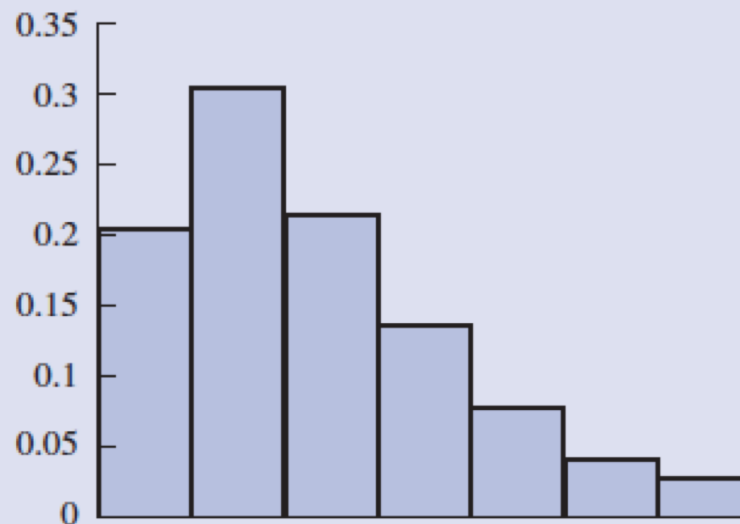




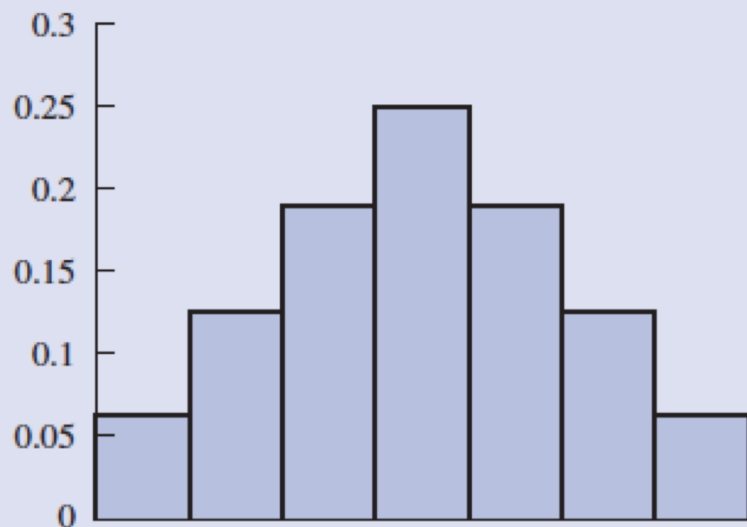
**Histograma A: Moderadamente sesgado a la izquierda**



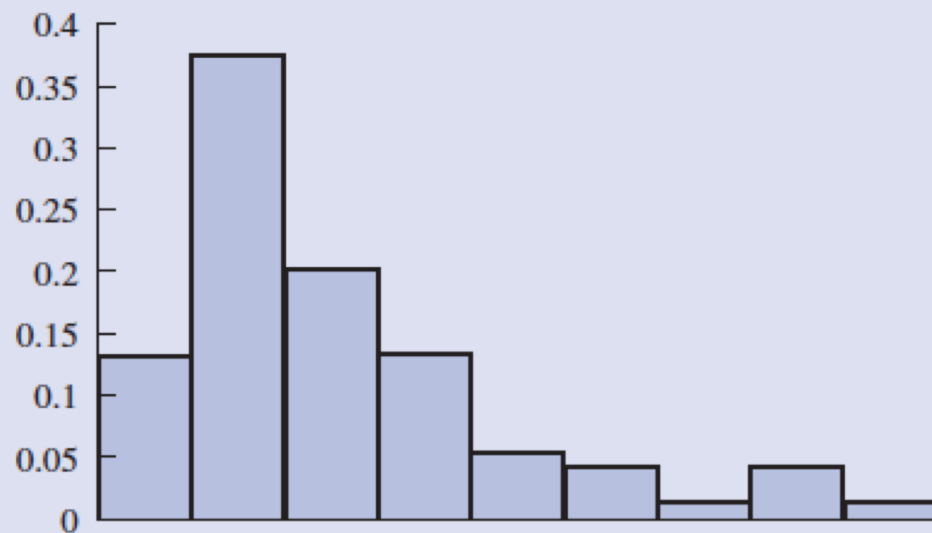
**Histograma B: Moderadamente sesgado a la derecha**



**Histograma C: Simétrico**



**Histograma D: Muy sesgado a la derecha**





# Frecuencia acumulada

- ▶ La distribución de frecuencia acumulada usa la cantidad, las amplitudes y los límites de las clases de la distribución de frecuencia. Sin embargo, en lugar de mostrar la frecuencia de cada clase, la distribución de frecuencia acumulada muestra la cantidad de datos que tienen un valor menor o igual al límite superior de cada clase.



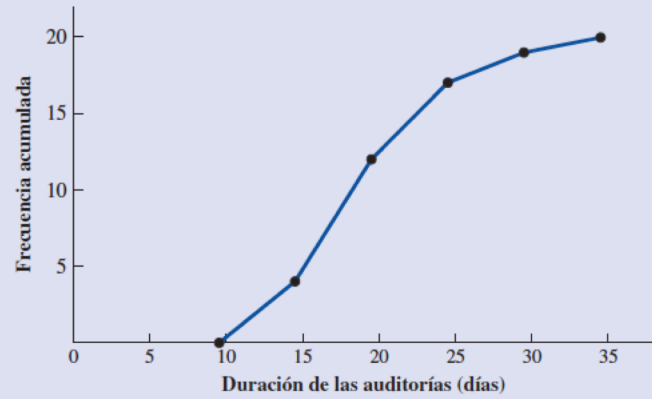
<b>Duración de la auditoría en días</b>	<b>Frecuencia acumulada</b>	<b>Frecuencia relativa acumulada</b>	<b>Frecuencia porcentual acumulada</b>
Menor o igual que 14	4	0.20	20
Menor o igual que 19	12	0.60	60
Menor o igual que 24	17	0.85	85
Menor o igual que 29	19	0.95	95
Menor o igual que 34	20	1.00	100



# Ojiva

- ▶ La gráfica de una distribución acumulada, llamada ojiva , es una gráfica que muestra los valores de los datos en el eje horizontal y las frecuencias acumuladas, las frecuencias relativas acumuladas o las frecuencias porcentuales acumuladas en el eje vertical.







NRF/BIG proporciona los resultados de una investigación sobre las cantidades que gastan en vacaciones los consumidores (*USA Today*, 20 de diciembre de 2005). Los datos siguientes son las cantidades gastadas en vacaciones por los 25 consumidores de una muestra.

1200	850	740	590	340
450	890	260	610	350
1780	180	850	2050	770
800	1090	510	520	220
1450	280	1120	200	350

- ¿Cuál es la menor cantidad gastada en vacaciones? ¿Cuál la mayor?
- Use \$250 como amplitud de clase para elaborar con estos datos una distribución de frecuencia y una distribución de frecuencia porcentual.
- Elabore un histograma y comente la forma de la distribución.
- ¿Qué observaciones le permiten hacer las cantidades gastadas en vacaciones?