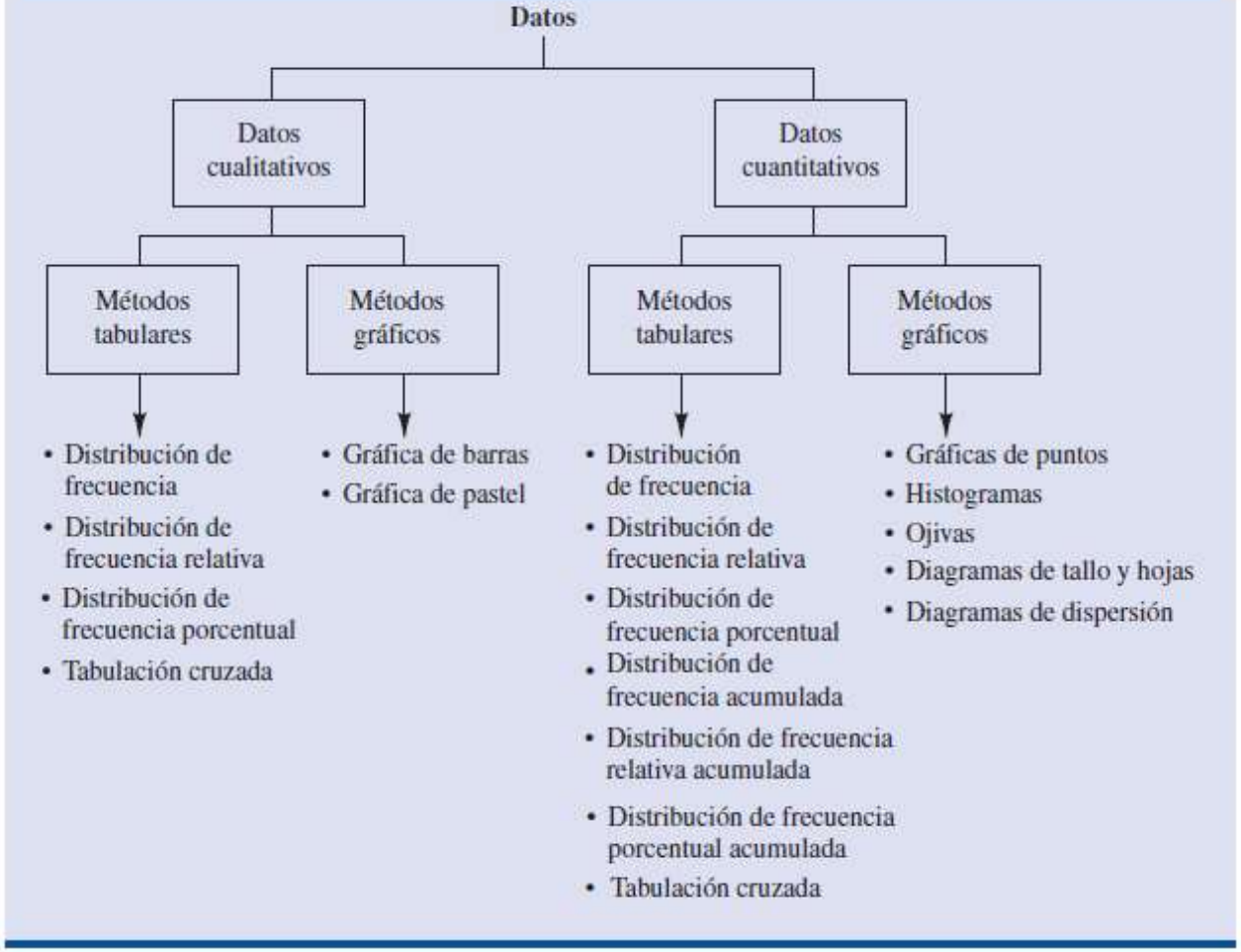
FIGURA 2.9 MÉTODOS TABULARES Y GRÁFICOS PARA RESUMIR DATOS



ESTADÍSTICA DESCRIPTIVA: MEDIDAS NUMÉRICAS

- Se empezará con medidas numéricas para conjuntos de datos que constan de una sola variable.
- Si el conjunto de datos consta de más de una variable, empleará estas mismas medidas numéricas para cada una de las variables por separado.
- En el caso de dos variables, estudiará también medidas de la relación entre dos variables.
- Se presentan medidas numéricas de localización, dispersión, forma, y asociación. Si estas medidas las calcula con los datos de una muestra, se llaman **estadísticos muestrales.** Si estas medidas las calcula con los datos de una población se llaman **parámetros poblacionales.**

Medidas de localización

Media

La medida de localización más importante es la **media**, o valor promedio, de una variable. La media proporciona una medida de localización central de los datos. Si los datos son datos de una muestra, la media se denota $\bar{x}$ si los datos son datos de una población, la media se denota con la letra griega *μ.*

MEDIA MUESTRAL

$$\bar{x} = \frac{\Sigma x_i}{n} \qquad (3.1)$$

Donde:

$\bar{x}$ denota la media muestral

x_1 denota la primera observación de la variable x.

x_2 denota la segunda observación de la variable x.

x_i denota la i-ésima observación de la variable x.

n denota la cantidad total de observaciones.

Ejemplo

Para ilustrar el cálculo de la media muestral, considere los siguientes datos que representan el tamaño de cinco grupos de una universidad.

$$46 \quad 54 \quad 42 \quad 46 \quad 32$$

Se emplea la notación x_1, x_2, x_3, x_4, x_5 para representar el número de estudiantes en cada uno de los cinco grupos.

$$x_1 = 46 \qquad x_2 = 54 \qquad x_3 = 42 \qquad x_4 = 46 \qquad x_5 = 32$$

Por tanto, para calcular la media muestral, escriba

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{46 + 54 + 42 + 46 + 32}{5} = 44$$

La media muestral del tamaño de estos grupos es 44 alumnos.

Ejercicio

Suponga que la bolsa de trabajo de una universidad envía cuestionarios a los recién egresados de la carrera de administración solicitándoles información sobre sus sueldos mensuales iniciales. En la tabla 3.1 se presentan estos datos.

TABLA 3.1 SUELDOS MENSUALES INICIALES EN UNA MUESTRA DE 12 RECIÉN EGRESADOS DE LA CARRERA DE ADMINISTRACIÓN

Egresado	Sueldo mensual inicial ($)	Egresado	Sueldo mensual inicial ($)
1	3450	7	3490
2	3550	8	3730
3	3650	9	3540
4	3480	10	3925
5	3355	11	3520
6	3310	12	3480

Calcular el sueldo mensual promedio de los egresados.

MEDIA POBLACIONAL

$$\mu = \frac{\Sigma x_i}{N} \qquad (3.2)$$

Mediana

Ordenar los datos de menor a mayor (en forma ascendente).

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

Ejemplo

Apliquemos esta definición para calcular la mediana del número de alumnos en un grupo a partir de la muestra de los cinco grupos de universidad. Los datos en orden ascendente son

32 42 46 46 54

Como $n = 5$ es impar, la mediana es el valor de enmedio. De manera que la mediana del tamaño de los grupos es 46. Aun cuando en este conjunto de datos hay dos observaciones cuyo valor es 46, al poner las observaciones en orden ascendente se toman en consideración todas las observaciones.

Ejemplo

Suponga que también desea calcular la mediana del salario inicial de los 12 recién egresados de la carrera de administración de la tabla 3.1. Primero ordena los datos de menor a mayor

3310 3355 3450 3480 3480 3490 3520 3540 3550 3650 3730 3925

Los dos valores de en medio

Como $n = 12$ es par, se localizan los dos valores de enmedio: 3490 y 3520. La mediana es el promedio de estos dos valores.

$$\text{Mediana} = \frac{3490 + 3520}{2} = 3505$$

Moda

La moda es el valor que se presenta con mayor frecuencia.

Ejemplo

32 42 46 46 54

La moda es 46.

3310 3355 3450 3480 3480 3490 3520 3540 3550 3650 3730 3925

La moda es 3480.

Importante

- Hay situaciones en que la frecuencia mayor se presenta con dos o más valores distintos. Cuando esto ocurre hay más de una moda. Si los datos contienen más de una moda se dice que los datos son *bimodales.* Si contienen más de dos modas, son *multimodales.*

- En los casos multimodales casi nunca se da la moda, porque dar tres o más modas no resulta de mucha ayuda para describir la localización de los datos.

Percentiles

El percentil *p* es un valor tal que por lo menos *p* por ciento de las observaciones son menores o iguales que este valor y por lo menos (100 - *p*) por ciento de las observaciones son mayores o iguales que este valor.

CÁLCULO DEL PERCENTIL p

Paso 1. Ordenar los datos de menor a mayor (colocar los datos en orden ascendente).

Paso 2. Calcular el índice i

$$i = \left(\frac{p}{100}\right)n$$

donde p es el percentil deseado y n es el número de observaciones.

Paso 3. (a) Si *i no es un número entero, debe redondearlo.* El primer entero mayor que i denota la posición del percentil p.

(b) Si *i es un número entero,* el percentil p es el promedio de los valores en las posiciones i e $i + 1$.

Ejemplo

Para ilustrar el empleo de este procedimiento, determine el percentil 85 en los sueldos mensuales iniciales de la tabla 3.1.

Paso 1. Ordenar los datos de menor a mayor

3310 3355 3450 3480 3480 3490 3520 3540 3550 3650 3730 3925

Paso 2.

$$i = \left(\frac{p}{100}\right)n = \left(\frac{85}{100}\right)12 = 10.2$$

Paso 3. Como i no es un número entero, se debe *redondear*. La posición del percentil 85 es el primer entero mayor que 10.2, es la posición 11.

Observe ahora los datos, entonces el percentil 85 es el dato en la posición 11, o sea 3730.

Paso 1. Ordenar los datos de menor a mayor

3310 3355 3450 3480 3480 3490 3520 3540 3550 3650 3730 3925

Paso 2.

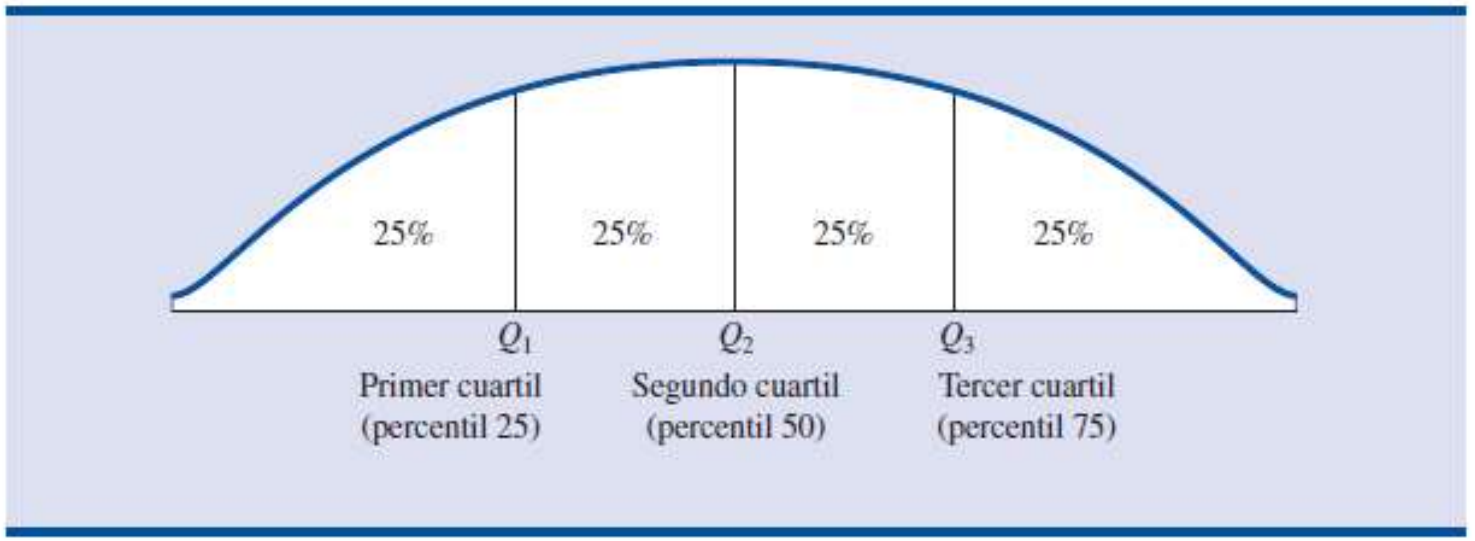
Para ampliar la formación en el uso de este procedimiento, calculará el percentil 50 en los sueldos mensuales iniciales. Al aplicar el paso 2 obtiene.

$$i = \left(\frac{50}{100}\right)12 = 6$$

Como i es un número entero, de acuerdo con el paso 3 b) el percentil 50 es el promedio de los valores de los datos que se encuentran en las posiciones seis y siete; de manera que el percentil 50 es (3490 + 3520)/2 = 3505. Observe que el *percentil 50 coincide con la mediana.*

Cuartiles

FIGURA 3.1 LOCALIZACIÓN DE LOS CUARTILES



Q_1 = primer cuartil, o percentil 25
Q_2 = segundo cuartil, o percentil 50
Q_3 = tercer cuartil, o percentil 75

Una vez más se ordenan los sueldos iniciales de menor a mayor. Q_2, el segundo cuartil (la mediana), ya se tiene identificado, es 3505.

3310 3355 3450 3480 3480 3490 3520 3540 3550 3650 3730 3925

Para calcular los cuartiles Q_1 y Q_3 use la regla para hallar el percentil 25 y el percentil 75. A continuación se presentan estos cálculos.

Para hallar Q_1,

$$i = \left(\frac{p}{100}\right)n = \left(\frac{25}{100}\right)12 = 3$$

Como i es un entero, el paso 3 b) indica que el primer cuartil, o el percentil 25, es el promedio del tercer y cuarto valores de los datos; esto es, $Q_1 = (3450 + 3480)/2 = 3465$.

Para hallar Q_3,

$$i = \left(\frac{p}{100}\right)n = \left(\frac{75}{100}\right)12 = 9$$

Como i es un entero, el paso 3 b) indica que el tercer cuartil, o el percentil 75, es el promedio del noveno y décimo valores de los datos; esto es, $Q_3 = (3550 + 3650)/2 = 3600$.

Los cuartiles dividen los datos de los sueldos iniciales en cuatro partes y cada parte contiene 25% de las observaciones.

3310 3355 3450 | 3480 3480 3490 | 3520 3540 3550 | 3650 3730 3925

$Q_1 = 3465$ $Q_2 = 3505$ (Mediana) $Q_3 = 3600$

Ejercicio

En una prueba sobre consumo de gasolina se examinaron a 13 automóviles en un recorrido de 100 millas, tanto en ciudad como en carretera. Se obtuvieron los datos siguientes de rendimiento en millas por galón.

Ciudad:	16.2	16.7	15.9	14.4	13.2	15.3	16.8	16.0	16.1	15.3	15.2	15.3	16.2
Carretera:	19.4	20.6	18.3	18.6	19.2	17.4	17.2	18.6	19.0	21.1	19.4	18.5	18.7

Use la media, la mediana y la moda para indicar cuál es la diferencia en el consumo entre ciudad y carretera.

Ejercicio

J. D. Powers and Associates hicieron una investigación sobre el número de minutos por mes que los usuarios de teléfonos celulares usan sus teléfonos (Associated Press, junio de 2002). A continuación se muestran los minutos por mes hallados en una muestra de 15 usuarios de teléfonos celulares

615	135	395
430	830	1180
690	250	420
265	245	210
180	380	105

a. ¿Cuál es la media de los minutos de uso por mes?
b. ¿Cuál es la mediana de los minutos de uso por mes?
c. ¿Cuál es el percentil 85?
d. J. D. Powers and Associates informa que los planes promedio para usuarios de celulares permiten hasta 750 minutos de uso por mes. ¿Qué indican los datos acerca de la utilización que hacen los usuarios de teléfonos celulares de sus planes mensuales?

Medidas de localización

Media $\overline{x} = \frac{\sum x_i}{n}$

Mediana $x_1 \quad x_2 \quad \boxed{x_3} \quad x_4 \quad x_5$

$x_1 \quad x_2 \quad \boxed{x_3} \quad \boxed{x_4} \quad x_5 \quad x_6$

Moda Mayor Frecuencia

Percentil $i = \left(\frac{P}{100}\right) n$

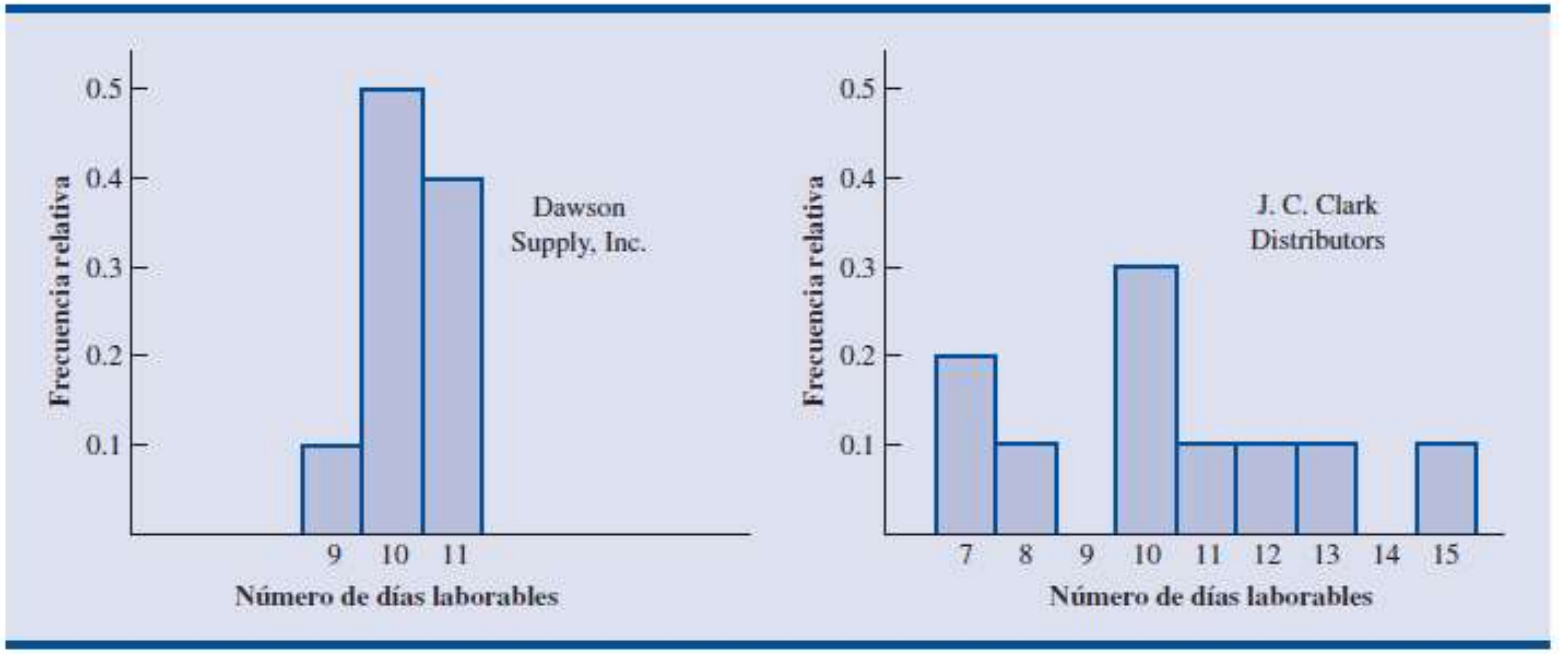
Cuartil Percentil 25% 50% 75%

Medidas de variabilidad o dispersión

Motivación

Después de algunos meses de operación, se percata de que el número promedio de días que ambos proveedores requieren para surtir una orden es 10 días. En la figura 3.2 se presentan los histogramas que muestran el número de días que cada uno de los proveedores necesita para surtir una orden. Aunque en ambos casos este número promedio de días es 10 días, ¿muestran los dos proveedores el mismo grado de confiabilidad en términos de tiempos para surtir los productos? Observe la dispersión, o variabilidad, de estos tiempos en ambos histogramas. ¿Qué proveedor preferiría usted?

FIGURA 3.2 DATOS HISTÓRICOS QUE MUESTRAN EL NÚMERO DE DÍAS REQUERIDOS PARA COMPLETAR UNA ORDER



Rango

RANGO

Rango = Valor mayor − Valor menor

615	135	395
430	830	1180
690	250	420
265	245	210
180	380	105

$$Rango = 1180 - 105 = 1075$$

Rango intercuartílico

RANGO INTERCUARTÍLICO

$$IQR = Q_3 - Q_1 \qquad (3.3)$$

615	135	395
430	830	1180
690	250	420
265	245	210
180	380	105

105 135 180 210 245 250 265 380 395 420 430 615 690 830 1180

Q_1 $$i = \left(\frac{25}{100}\right)15 = 3.75 \approx 4 \qquad Q_1 = 210$$

105 135 180 **210** 245 250 265 380 395 420 430 615 690 830 1180

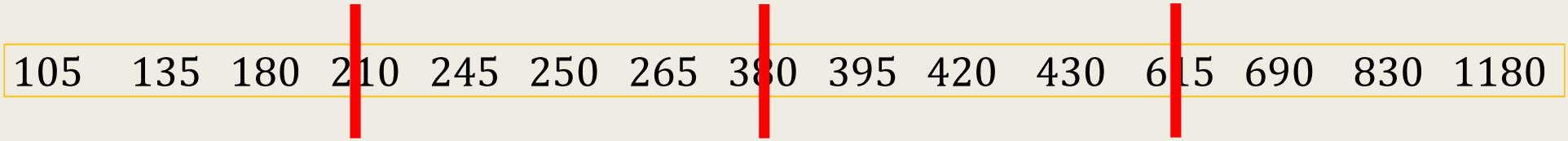
Q_2 $$i = \left(\frac{50}{100}\right)15 = 7.5 \approx 8 \qquad Q_2 = 380$$

105 135 180 210 245 250 265 **380** 395 420 430 615 690 830 1180

Q_3

$$i = \left(\frac{75}{100}\right)15 = 11.25 \approx 12 \qquad Q_3 = 210$$

105 135 180 210 245 250 265 380 395 420 430 615 690 830 1180



$$RIC = Q_3 - Q_1 = 615 - 210 = 405$$

Varianza

VARIANZA POBLACIONAL

$$\sigma^2 = \frac{\Sigma(x_i - \mu)^2}{N} \qquad (3.4)$$

VARIANZA MUESTRAL

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n - 1} \qquad (3.5)$$

TABLA 3.3 CÁLCULO DE LAS DESVIACIONES Y DE LOS CUADRADOS DE LAS DESVIACIONES RESPECTO DE LA MEDIA EMPLEANDO LOS DATOS DE LOS TAMAÑOS DE CINCO GRUPOS DE ESTADOUNIDENSES

Número de estudiantes en un grupo (x_i)	Número promedio de alumnos en un grupo ($\bar{x}$)	Desviación respecto a la media ($x_i - \bar{x}$)	Cuadrado de la desviación respecto de la media $(x_i - \bar{x})^2$
46	44	2	4
54	44	10	100
42	44	−2	4
46	44	2	4
32	44	−12	144
		0	256
		$\Sigma(x_i - \bar{x})$	$\Sigma(x_i - \bar{x})^2$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n - 1} = \frac{256}{4} = 64$$

Estudiantes cuadrados

TABLA 3.4 CÁLCULO DE LA VARIANZA MUESTRAL CON LOS DATOS DE LOS SUELDOS INICIALES

Sueldo mensual (x_i)	Media muestral $(\bar{x})$	Desviación respecto de la media $(x_i - \bar{x})$	Cuadrado de la desviación respecto de la media $(x_i - \bar{x})^2$
3450	3540	−90	8 100
3550	3540	10	100
3650	3540	110	12 100
3480	3540	−60	3 600
3355	3540	−185	34 225
3310	3540	−230	52 900
3490	3540	−50	2 500
3730	3540	190	36 100
3540	3540	0	0
3925	3540	385	148 225
3520	3540	−20	400
3480	3540	−60	3 600
		0	301 850
		$\Sigma(x_i - \bar{x})$	$\Sigma(x_i - \bar{x})^2$

Empleando la ecuación (3.5),

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n - 1} = \frac{301\,850}{11} = 27\,440.91$$

Ejercicio

615	135	395
430	830	1180
690	250	420
265	245	210
180	380	105

Desviación estándar

DESVIACIÓN ESTÁNDAR

$$\text{Desviación estándar muestral} = s = \sqrt{s^2} \quad (3.6)$$

$$\text{Desviación estándar poblacional} = \sigma = \sqrt{\sigma^2} \quad (3.7)$$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{256}{4} = 64$$

Estudiantes cuadrados

$s = 8$ estudiantes

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{301\,850}{11} = 27\,440.91$$

dólares cuadrados

$$s = 165.65\ dólares$$

Coeficiente de variación

Mide que tan grande es la desviación estándar en relación con la media.

COEFICIENTE DE VARIACIÓN

$$\left(\frac{\text{Desviación estándar}}{\text{Media}} \times 100\right)\% \qquad (3.8)$$

$$\frac{8}{44} * 100 = 18.2\%$$

El coeficiente de variación indica que la desviación estándar muestral es 18.2% del valor de la media muestral.

Ejercicio

22. La Asociación Estadounidense de Inversionistas Individuales realiza cada año una investigación sobre los corredores de bolsa con descuento (*AAII Journal*, enero de 2003). En la tabla 3.2 se muestran las comisiones que cobran 24 corredores de bolsa con descuento por dos tipos de transacciones: transacción con ayuda del corredor de 100 acciones a $50 la acción y transacción en línea de 500 acciones a $50 la acción.
    a. Calcule el rango y el rango intercuartílico en cada tipo de transacción.
    b. Calcule la varianza y la desviación estándar en cada tipo de transacción.
    c. Calcule el coeficiente de variación en cada tipo de transacción.
    d. Compare la variabilidad en el costo que hay en los dos tipos de transacciones

24. Las puntuaciones de un jugador de golf en el 2005 y 2006 son las siguientes:

2005	74	78	79	77	75	73	75	77
2006	71	70	75	77	85	80	71	79

    a. Use la media y la desviación estándar para evaluar a este jugador de golf en estos dos años.
    b. ¿Cuál es la principal diferencia en su desempeño en estos dos años? ¿Se puede ver algún progreso en sus puntuaciones del 2006?, ¿cuál?

24. Los siguientes son los tiempos que hicieron los velocistas de los equipos de pista y campo de una universidad en un cuarto de milla y en una milla (los tiempos están en minutos).

Tiempos en un cuarto de milla:	0.92	0.98	1.04	0.90	0.99
Tiempos en una milla:	4.52	4.35	4.60	4.70	4.50

Después de ver estos datos, el entrenador comentó que en un cuarto de milla los tiempos eran más homogéneos. Use la desviación estándar y el coeficiente de variación para resumir la variabilidad en los datos. El uso del coeficiente de variación, ¿indica que la aseveración del entrenador es correcta?

Medidas numéricas

Medidas de localización

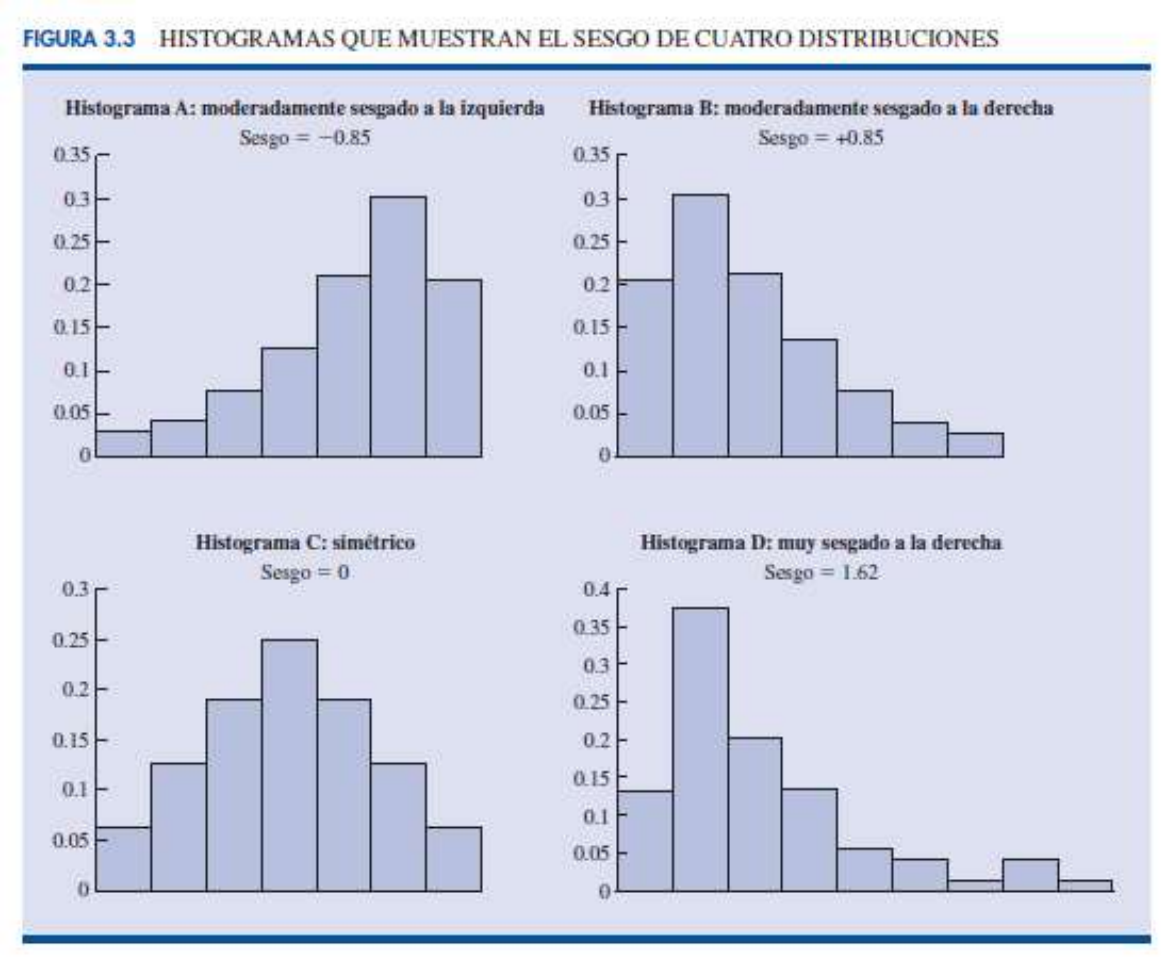
- Media
- Moda
- Mediana
- Percentil
- Cuartil

Medidas de variabilidad

- Rango
- Rango intercuartílico
- Varianza
- Desviación estándar
- Coeficiente de variación

Medidas de la forma de la distribución, de la posición relativa y de la detección de observaciones atípicas

Forma de la distribución



Sesgo

$$Sesgo = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$$

- En una distribución simétrica, la media y la mediana son iguales.
- Si los datos están sesgados a la derecha, la media será mayor que la mediana.
- Si los datos están sesgados a la izquierda, la media será menor que la mediana.
- Cuando los datos están ligeramente sesgados, se prefiere la mediana como medida de localización.

Puntos z

PUNTO z

$$z_i = \frac{x_i - \bar{x}}{s} \qquad (3.9)$$

donde

z_i = punto z para x_i
$\bar{x}$ = media muestral
s = desviación estándar muestral

- Al punto *z* también se le suele llamar *valor estandarizado*.
- El punto z_i puede ser interpretado como el *numero de desviaciones estándar a las que* x_i *se encuentra de la media* .

- Por ejemplo si z_i = 1.2, esto indica que x_i es 1.2 desviaciones estándar mayor que la media muestral. De manera Similar $z_2 = 0.5$ indica que x_2 0.5 o 1/2 desviación estándar menor que la media muestral.
- Puntos *z* mayores a cero corresponden a observaciones cuyo valor es mayor a la media
- Puntos *z* menores que cero corresponden a observaciones cuyo valor es menor a la media.
- Si el punto *z* es cero, el valor de la observación correspondiente es igual a la media.

TABLA 3.5 PUNTOS z CORRESPONDIENTES A LOS DATOS DE LOS TAMAÑOS DE LOS GRUPOS DE ESTUDIANTES

Número de estudiantes en un grupo (x_i)	Desviación respecto de la media ($x_i - \bar{x}$)	Puntos z $\left(\frac{x_i - \bar{x}}{s}\right)$
46	2	2/8 = 0.25
54	10	10/8 = 1.25
42	−2	−2/8 = −0.25
46	2	2/8 = 0.25
32	−12	−12/8 = −1.50

$\bar{x} = 44$ $s = 8$

Teorema de Chebyshev

El **teorema de Chebyshev** permite decir qué proporción de los valores que se tienen en los datos debe estar dentro de un determinado número de desviaciones estándar de la media.

TEOREMA DE CHEBYSHEV

Por lo menos $(1 - 1/z^2)$ de los valores que se tienen en los datos deben encontrarse dentro de z desviaciones estándar de la media, donde z es cualquier valor mayor que 1.

De acuerdo con este teorema para z = 2, 3 y 4 desviaciones estándar se tiene:

- Por lo menos 0.75, o 75%, de los valores de los datos deben estar dentro de z=2 desviaciones estándar de la media.

- Al menos 0.89, o 89%, de los valores deben estar dentro de z=3 desviaciones estándar de la media.

- Por lo menos 0.94, o 94%, de los valores deben estar dentro de z =4 desviaciones estándar de la media.

Ejemplo

En las calificaciones obtenidas por 100 estudiantes en un examen de estadística para la administración, la media es 70 y la desviación estándar es 5.
¿Cuántos estudiantes obtuvieron puntuaciones entre 60 y 80?,

Solución

1. 60 y 80 están a dos desviaciones estándar de la media.

$$\frac{60-70}{5} = -2 \qquad \frac{80-70}{5} = 2$$

2. $1 - \frac{1}{2^2} = \frac{3}{4} = 0{,}75$
3. *El 75% de las calificaciones están entre 60 y 80*

Ejemplo

En las calificaciones obtenidas por 100 estudiantes en un examen de estadística para la administración, la media es 70 y la desviación estándar es 5.
¿Cuántos tuvieron puntuaciones entre 58 y 82?

Solución

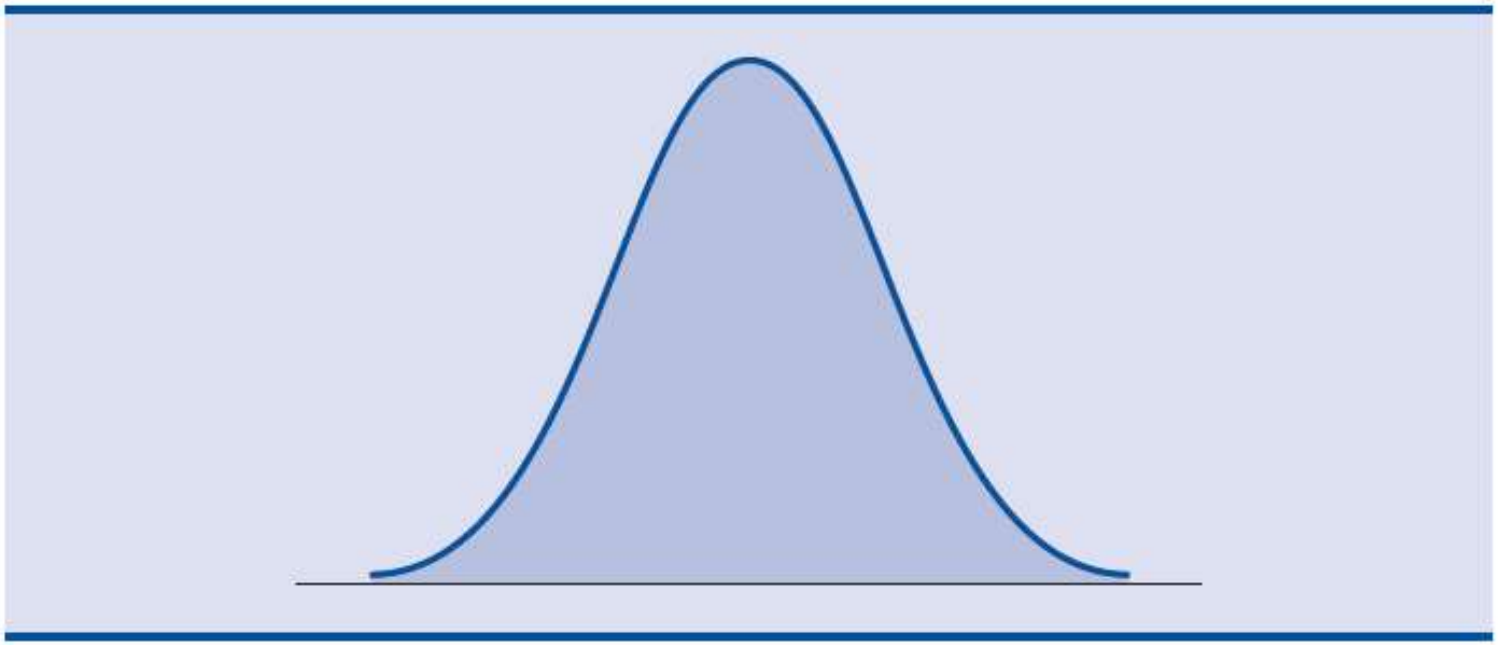
1. 58 y 82 están a 2.4 desviaciones estándar de la media.

$$\frac{58-70}{5} = -2.4 \quad \frac{82-70}{5} = 2.4$$

2. $1 - \frac{1}{2.4^2} = 0.862$
3. *El 86% de las calificaciones están entre* 58 y 82

Regla empírica

FIGURA 3.4 DISTRIBUCIÓN EN FORMA DE MONTAÑA O DE CAMPANA



REGLA EMPÍRICA

Cuando los datos tienen una distribución en forma de campana:

- Cerca de 68% de los valores de los datos se encontrarán a no más de una desviación estándar desde la media.
- Aproximadamente 95% de los valores de los datos se encontrarán a no más de dos desviaciones estándar desde la media.
- Casi todos los valores de los datos estarán a no más de tres desviaciones estándar de la media.

Ejemplo

Los envases con detergente líquido se llenan en forma automática en una línea de producción. Los pesos de llenado suelen tener una distribución en forma de campana. Si el peso medio de llenado es de 16 onzas y la desviación estándar de 0.25 onzas, la regla empírica es aplicada para sacar las conclusiones siguientes:

- Aproximadamente 68% de los envases llenados pesarán entre 15.75 y 16.25 onzas (estarán a no más de una desviación estándar de la media).

- Cerca de 95% de los envases llenados pesarán entre 15.50 y 16.50 onzas (estarán a no más de dos desviaciones estándar de la media).

- Casi todos los envases llenados pesarán entre 15.25 y 16.75 onzas (estarán a no más de tres desviaciones estándar de la media).

Detección de observaciones atípicas

- Para identificar las observaciones atípicas se emplean los valores estandarizados (puntos *z*).
- Cualquier dato cuyo punto *z* sea menor que -3 o mayor que 3 como una observación atípica.
- Debe examinar la exactitud de tales valores y si en realidad pertenecen al conjunto de datos.

Ejercicio

26. Piense en una muestra en que la media es 500 y la desviación estándar es 100. ¿Cuáles son los puntos z de los datos siguientes: 520, 650, 500, 450 y 280?

27. Considere una muestra en que la media es 30 y la desviación estándar es 5. Utilice el teorema de Chebyshev para determinar el porcentaje de los datos que se encuentra dentro de cada uno de los rangos siguientes.
    a. 20 a 40
    b. 15 a 45
    c. 22 a 38
    d. 18 a 42
    e. 12 a 48

28. Suponga datos que tienen una distribución en forma de campana cuya media es 30 y desviación estándar 5. Utilice la regla empírica para determinar el porcentaje de los datos que se encuentra dentro de cada uno de los rangos siguientes.
    a. 20 a 40
    b. 15 a 45
    c. 25 a 35

Análisis exploratorio de datos

Resumen de cinco números

En el **resumen de cinco números** se usan los cinco números siguientes para resumir los datos.

1. El valor menor.
2. El primer cuartil ($Q1$).
3. La mediana ($Q2$).
4. El tercer cuartil ($Q3$).
5. El valor mayor.

Ejemplo

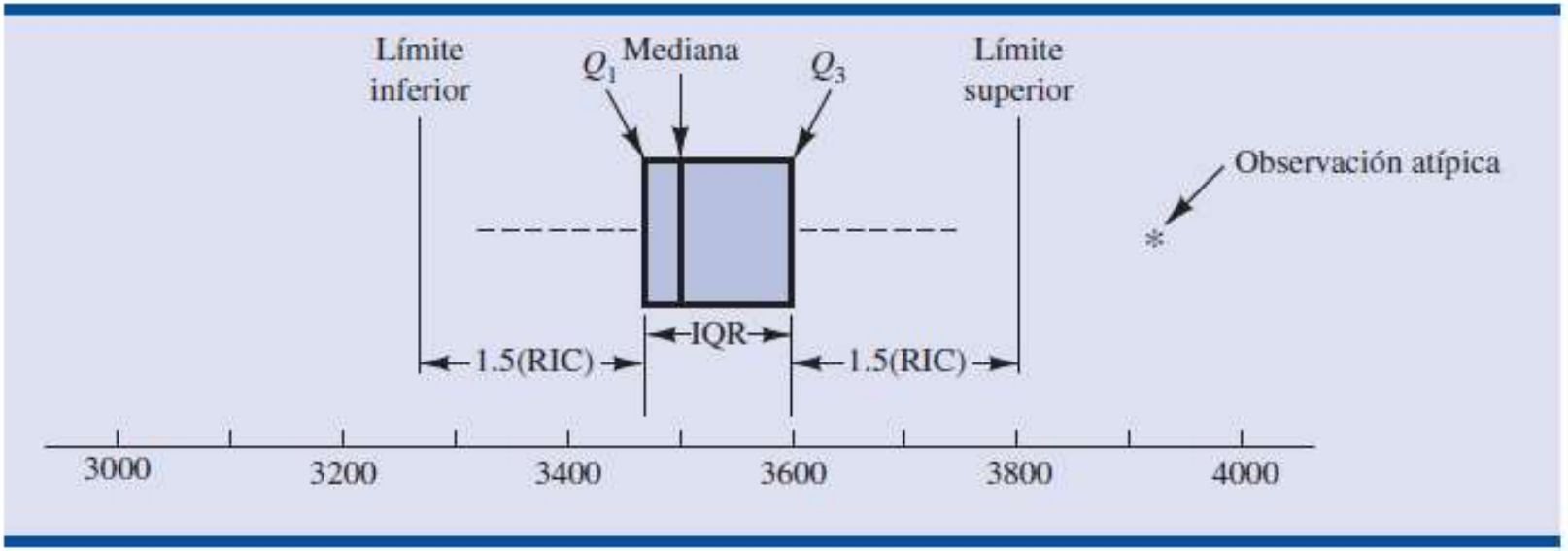
3310 3355 3450 | 3480 3480 3490 | 3520 3540 3550 | 3650 3730 3925

$Q_1 = 3465$ $Q_2 = 3505$ (Mediana) $Q_3 = 3600$

Diagrama de caja

Un **diagrama de caja** es un resumen gráfico de los datos con base en el resumen de cinco números.

FIGURA 3.5 DIAGRAMA DE CAJA DE LOS SALARIOS INICIALES, EN EL QUE SE MUESTRAN LAS LÍNEAS QUE INDICAN LOS LÍMITES INFERIOR Y SUPERIOR



1. Se dibuja una caja cuyos extremos se localicen en el primer y tercer cuartiles.

2. En el punto donde se localiza la mediana se traza una línea vertical.

3. Usando el rango intercuartílico, $RIC = Q3 - Q1$, se localizan los *limites*. En un diagrama de caja los límites se encuentran 1.5(RIC) abajo del $Q1$ y 1.5(RIC) arriba del $Q3$.

$$Límite\ inferior = Q1 - 1.5 * RIC$$
$$Límite\ superior = Q3 + 1.5 * RIC$$

4. Graficar los bigotes. Los bigotes son líneas punteadas que van desde los extremos de la caja hasta los valores menor y mayor *de los limites* calculados en el paso 3.

5. Mediante un asterisco se indica la localización de las observaciones atípicas.

Tarea

41. A continuación se presentan las ventas, en millones de dólares, de 21 empresas farmacéuticas.

8 408	1 374	1872	8879	2459	11 413
608	14 138	6452	1850	2818	1 356
10 498	7 478	4019	4341	739	2 127
3 653	5 794	8305			

a. Proporcione el resumen de cinco números.
b. Calcule los límites superior e inferior.
c. ¿Hay alguna observación atípica en estos datos?
d. Las ventas de Johnson & Johnson son las mayores de la lista, $14 138 millones. Suponga que se comete un error al registrar los datos (un error de transposición) y en lugar del valor dado se registra $41 138 millones. ¿Podría detectar este problema con el método de detección de observaciones atípicas del inciso c, de manera que se pudiera corregir este dato?
e. Dibuje el diagrama de caja.