

Estadística aplicada a los negocios y la economía

Tercera edición

- Incluye 850 ejercicios
- Disquete de datos
- Ejercicios y proyectos por internet

Allen L. Webster

**Mc
Graw
Hill**

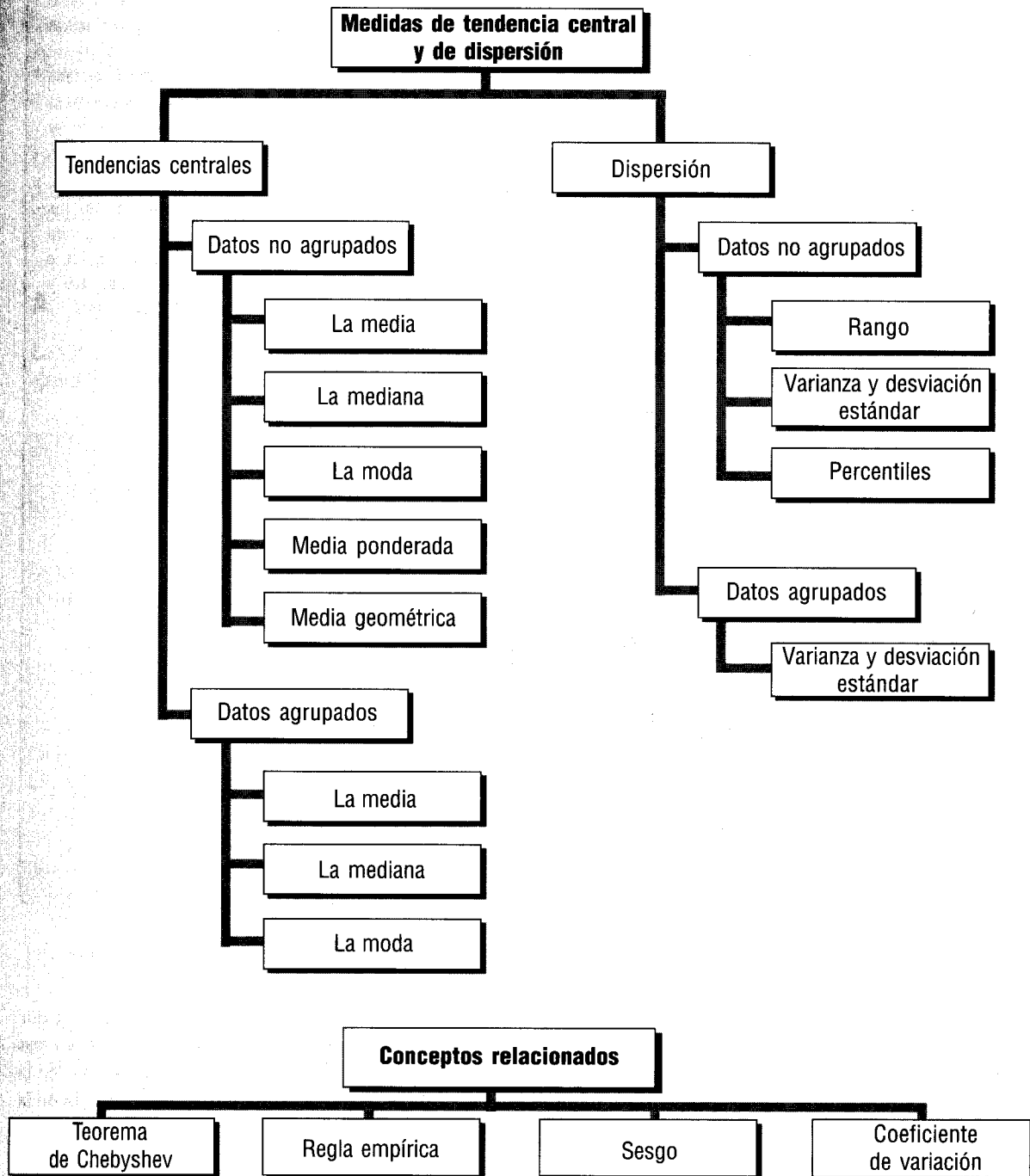
**Irwin
McGraw-Hill**

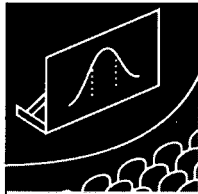
3

Medidas de tendencia central y de dispersión

Plan del capítulo

Este capítulo ilustra cómo puede describirse con unos cuantos números un conjunto de datos completo. Se observará lo útil que pueden ser la tendencia central y las medidas de dispersión en el análisis estadístico.





ESCENARIO

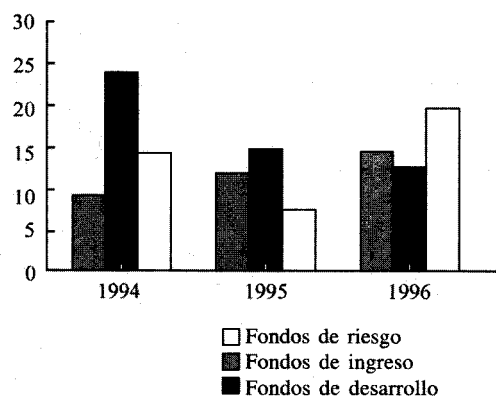
La revista *Fortune* (febrero 17 de 1997), reportó el rendimiento promedio anual de tres fondos (fondos de riesgo, de títulos y de desarrollo) que ofrecía Janus Funds, una gran compañía de fondos mutuos de inversión. Como analista de inversiones se le asigna la tarea de analizar las operaciones de Janus. Para efectuar esta asignación laboral se debe preparar un estudio descriptivo del desempeño de los tres fondos que aparecen en el gráfico.

Como lo demuestra el gráfico, los rendimientos en cada uno de estos fondos han variado considerablemente con los años. Hay muchas consideraciones que se deben tener en cuenta al realizar el análisis, y el supervisor inmediato ha enfatizado la importancia de una presentación minuciosa y precisa.

Surgen muchas preguntas sobre la forma como se debe proceder. Como lo enseña este capítulo, existen muchas herramientas descriptivas del análisis estadístico que se pueden utilizar para desarrollar el análisis. Cada herramienta juega un papel importante en el suministro de valiosa información. Los estadísticos descriptivos tales como rendimientos promedio y medidas de dispersión, serán muy útiles.

En este capítulo se analizarán muchas de las herramientas descriptivas que se utilizan usualmente para transmitir información valiosa de manera concisa e informativa. Tales prácticas reducen la incertidumbre y facilitan ampliamente el proceso de toma de decisiones.

Rendimientos porcentuales – Janus Funds



3.1 Introducción

Los datos, al igual que los estudiantes, se congregan alrededor de sus puntos de encuentro favoritos. Parece que los estudiantes acuden en masa a sitios tales como partidos de fútbol, fraternidades, bares populares y otros sitios de reunión y en raras ocasiones hasta la biblioteca. De igual forma, los números parecen disfrutar de la compañía de otros números y están propensos a reunirse alrededor de un punto central denominado **medida de la tendencia central** o, más comúnmente, *media*. Una medida de tendencia central ubica e identifica el punto alrededor del cual se centran los datos.

Un conjunto grande de datos puede ser rápidamente descrito de manera sucinta con un solo número. Si el profesor dice que el promedio de la clase en el último examen de estadística fue de 95, esto indica algo. Si se dice que el promedio fue de 35, esto indica algo totalmente diferente.

Además, las **medidas de dispersión** indican el punto hasta el cual las observaciones individuales se esparcen alrededor de su punto central. Miden la dispersión o la variabilidad de los datos y reflejan la tendencia de las observaciones individuales a desviarse de dicho punto central.

3.2 Medidas de la tendencia central a partir de datos no agrupados

Existen tres métodos comunes para identificar el centro de un conjunto de datos: la **media**, la **mediana** y la **moda**. Su cálculo e interpretación dependerá de la definición de “centro”. En cada caso, se ubican en el punto alrededor del cual se aglomeran los datos.

A. La media

La media, o media aritmética, es la medida de la tendencia central que usualmente se le llamaba promedio. Por ejemplo, para calcular la media de los últimos 10 exámenes de estadística, simplemente se suman y se divide por 10.

Media aritmética La medida de la tendencia central que normalmente era considerada como el promedio.

La media de una población es el parámetro μ (que se pronuncia miu). Si hay N observaciones en el conjunto de datos de la población, la media se calcula así:

Media poblacional
$$\mu = \frac{X_1 + X_2 + X_3 + \cdots + X_N}{N} = \frac{\sum_{i=1}^N X_i}{N} \quad [3.1]$$

La letra griega mayúscula Σ es el signo de sumatoria que indica que se suman todas las observaciones de 1 a N . Las X_i denotan las observaciones individuales. En aras de la simplicidad, el subíndice y el superíndice se reducirán y aparecerá el signo de la sumatoria como simplemente Σ .

La media de una muestra es un estadístico $\bar{X}$ (que se lee X – barra). Con n observaciones en el conjunto de datos de la muestra, la media se determina así:

Media muestral
$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n} \quad [3.2]$$

Se supone que una muestra de los ingresos por ventas mensuales en miles de dólares para cinco meses es de 56, 67, 52, 45, y 67. La media se calcula así:

$$\bar{X} = \frac{56 + 67 + 52 + 45 + 67}{5} = 57.4$$

B. La mediana

La mediana algunas veces es llamada *media posicional*, porque queda exactamente en la mitad del conjunto de datos después de que las observaciones se han colocado en serie ordenada. La mitad de las observaciones estará por encima de la mediana, la otra mitad estará por *debajo* de ella.

La mediana La observación de la mitad después de que se han colocado los datos en una serie ordenada.

Si el conjunto de datos tiene un número impar de observaciones, la *posición* de la mediana es

$$\text{Posición de la mediana} = \frac{n + 1}{2} \quad [3.3]$$

Dado el ejemplo anterior sobre los ingresos por ventas, se debe colocar primero los datos en serie ordenada:

45, 52, 56, 67, 67

La posición del valor de la mediana se vuelve

$$\text{Posición de la mediana} = \frac{5 + 1}{2} = 3$$

o la tercera posición. De la serie ordenada anterior se puede observar que este valor es 56.

Con un conjunto de datos que contiene un número par de observaciones es necesario promediar los dos valores medios. Si los ingresos para un sexto mes son de 35 y se adiciona al conjunto de datos, entonces la serie ordenada se vuelve 35, 45, 52, 56, 67, 67. La posición de la mediana es

$$\text{Posición de la mediana} = \frac{6 + 1}{2} = 3.5$$

es decir, la posición tres y medio. Los dos valores de las posiciones tercera y cuarta se promedian para producir una mediana de $52 + 56 = 54$. Esto significa que en la mitad de los meses las ventas estuvieron por debajo de US\$54,000, y en la mitad de los meses los ingresos excedieron dicha suma.

C. La moda

La observación modal es la observación que ocurre con mayor frecuencia. Utilizando las seis observaciones anteriores de 35, 45, 52, 56, 67, 67, la moda es 67. Si la séptima observación de 56 se agregara, el conjunto de datos sería *bimodal*, con modas de 56 y 67.

Ejemplo 3.1 Medidas de tendencia central

La emisión de la revista *Fortune* del 17 de febrero de 1997 reportó que en 1996 las utilidades en millones de dólares de varias de las 500 mejores compañías que aparecen en la revista incluían:

Exxon	US\$7,510	General Electric	US\$7,280
Philip Morris	6,246	IBM	5,429
Intel	5,157	General Motors	4,289

Calcule las tres medidas de tendencia central.

Solución

Media: Tratando los datos como una muestra se tiene que

$$\bar{X} = \frac{7,510 + 7,280 + 6,246 + 5,429 + 5,157 + 4,289}{6} = 5,985$$

Mediana: Primero deben colocarse los datos en una serie ordenada. Si no están ordenados, la posición que se encuentre utilizando la fórmula 3.3 no tiene sentido.

$$4,289, 5,157, 5,429, 6,246, 7,280, 7,510$$

La posición de la mediana es $(n + 1)/2 = 3.5$. La mediana es el promedio de los valores tercero y cuarto: $5,429 + 6,246 = 5,837.5$

Moda: Este conjunto de datos no tiene moda debido a que todas las observaciones ocurrieron con igual frecuencia.

Interpretación

La media y la mediana identifican el centro del conjunto de datos alrededor del cual se encuentran las observaciones.

D. La media ponderada

En la discusión sobre la media, se asume que cada observación era de igual importancia. Sin embargo, en ciertos casos, puede querer darse mayor peso a algunas de las observaciones. Por ejemplo, si el profesor de estadística amenaza que el examen final valdrá el doble de los otros exámenes para determinar la nota final, entonces al puntaje que se obtenga en el examen final debe dársele el doble de peso. Es decir, que debe contarse doble al calcular la nota. Esto es exactamente lo que hace la media ponderada al utilizar la fórmula 3.4

$$\bar{X}_w = \frac{\sum XW}{\sum W} \quad [3.4]$$

en donde $\bar{X}_w$ es la media ponderada

X es la observación individual

W es el peso o ponderación asignada a cada observación.

Se asume que se obtuvo un puntaje de 89, 92 y 79 en los exámenes parciales y 94 en el examen final. Estos puntajes y sus respectivas ponderaciones están reflejados en la tabla 3.1 La fórmula (3.4) da

$$\bar{X}_w = \frac{\sum XW}{\sum W} = \frac{448}{5} = 89.6$$

Tabla 3.1

Cálculo de la media ponderada

Nota (X)	Peso (W)	XW
89	1	89
92	1	92
79	1	79
94	2	188
	5	448

Este método es igual que sumar la nota del examen final dos veces al calcular la media:

$$\bar{X}_w = \frac{89 + 92 + 79 + 94 + 94}{5} = 89.6$$

Media ponderada La media ponderada toma en cuenta la importancia relativa de las observaciones.

Ejemplo 3.2 Cálculo de la media ponderada

Paul el fontanero vende cinco tipos de limpiadores para desagües. En la tabla se muestra cada tipo junto con la utilidad por lata y el número de latas vendidas.

Limpiador	Utilidad por lata (X)	Volumen de ventas en latas (W)	XW
Glunk Out	US\$ 2.00	3	US\$ 6.00
Bubble Up	3.50	7	24.50
Dream Drain	5.00	15	75.00
Clear More	7.50	12	90.00
Main Drain	6.00	15	90.00
	US\$ 24.00	52	US\$ 285.50

Se puede calcular la media aritmética simple de la utilidad de Paul como $\text{US\$}24/5 = \text{US\$}4.80$ por lata.

Solución

Sin embargo, probablemente este no sea un buen estimado de la utilidad promedio de Paul, debido a que vende más de algunos tipos de limpiadores que de otros. Para obtener un estado financiero más representativo del desempeño real de su negocio, Paul debe dar más peso a los tipos más populares de limpiadores. Por tanto, el cálculo apropiado sería la media ponderada. La medida de peso apropiada sería las cantidades vendidas. La media ponderada es entonces

$$\bar{X}_w = \frac{\sum XW}{\sum W} = \frac{\text{US\$}285.50}{52} = \text{US\$}5.49$$

Interpretación

La media ponderada es mayor que la media aritmética simple porque Paul vende más limpiadores de los tipos que tienen un margen de utilidad mayor.

E. La media geométrica

La media geométrica puede utilizarse para mostrar los cambios porcentuales en una serie de números positivos. Como tal, tiene una amplia aplicación en los negocios y en la economía, debido a que con frecuencia se está interesado en establecer el cambio porcentual en las ventas, en el producto nacional bruto o en cualquier serie económica.

Media geométrica La media geométrica proporciona una medida precisa de un cambio porcentual promedio en una serie de números.

La media geométrica (MG) se halla tomando la raíz enésima del producto de n números. Así,

$$MG = \sqrt[n]{X_1 X_2 X_3 \cdots X_n} \quad [3.5]$$

La mayoría de las calculadoras manuales pueden calcular la raíz enésima de cualquier número.

La media geométrica se utiliza con más frecuencia para calcular la tasa de crecimiento porcentual promedio de algunas series dadas, a través del tiempo. Para ilustrar esta aplicación en un escenario de negocios, se consideran las cifras sobre ingresos del ejemplo 3.3 para White-Knuckle Airlines, la principal competencia de P&P durante los últimos cinco años.

Ejemplo 3.3 La media geométrica

El CEO (director ejecutivo) de White-Knuckle Airlines desea determinar la tasa de crecimiento promedio en los ingresos con base en las cifras dadas en la tabla. Si la tasa de crecimiento promedio es menor que el promedio industrial del 10%, se asumirá una nueva campaña publicitaria.

Ingresos para White-Knuckle Airlines		
Año	Ingreso	Porcentaje del año anterior
1992	US\$ 50,000	-- --
1993	55,000	55/50 = 1.10
1994	66,000	66/55 = 1.20
1995	60,000	60/66 = 0.91
1996	78,000	78/60 = 1.30

Solución

Primero es necesario determinar el porcentaje que los ingresos de cada año representan respecto de los obtenidos el año anterior. En otras palabras, ¿qué porcentaje del ingreso de 1992 es el ingreso en 1993? Esto se encuentra dividiendo los ingresos de 1992 entre los de 1993. El resultado, 1.10 revela que los ingresos de 1993 son 110% de los ingresos de 1992. También se calculan los porcentajes para los tres años restantes. Tomando la Media Geométrica (MG) de estos porcentajes da

$$MG = \sqrt[4]{(1.10)(1.2)(0.91)(1.3)} = 1.1179$$

Restando 1 para convertirlo a un incremento anual promedio da 0.1179, o un incremento promedio de 11.79% para el período de cinco años.

Por otro lado, la media aritmética simple es

$$\bar{X} = \frac{1.1 + 1.2 + 0.91 + 1.3}{4} = \frac{4.51}{4} = 1.1275$$

o un cambio promedio de 12.75%. Se divide por 4 ya que se presentaron cuatro cambios durante el período de cinco años.

Sin embargo, si un incremento promedio de 12.75%, basado en la media aritmética simple, se aplica a la serie que comienza con US\$50,000, los resultados son

$$\text{US\$ } 50,000 \times 1.1275 = \text{US\$ } 56,375$$

$$\text{US\$ } 56,375 \times 1.1275 = \text{US\$ } 63,563$$

$$\text{US\$ } 63,563 \times 1.1275 = \text{US\$ } 71,667$$

$$\text{US\$ } 71,667 \times 1.1275 = \text{US\$ } 80,805$$

Ya que US\$80,805 excede los US\$78,000 que White-Knuckle Airlines en realidad ganó, el incremento del 12.75% es obviamente muy alto. Si se utiliza la tasa de crecimiento de la media geométrica del 11.79%, se obtiene

$$\text{US\$ } 50,000 \times 1.1179 = \text{US\$ } 55,895$$

$$\text{US\$ } 55,895 \times 1.1179 = \text{US\$ } 62,485$$

$$\text{US\$ } 62,485 \times 1.1179 = \text{US\$ } 69,852$$

$$\text{US\$ } 69,852 \times 1.1179 = \text{US\$ } 78,088 \approx \text{US\$ } 78,000$$

Esto da un valor de US\$78,088, lo que está mucho más cerca al ingreso real de US\$78,000.

Interpretación

La media geométrica representa el cambio promedio con el tiempo. Debido a que la tasa de crecimiento supera el promedio de la industria del 10%, la nueva campaña publicitaria no se llevará a cabo.

La media geométrica siempre será menor que la media aritmética salvo en el extraño caso en el que todos los incrementos porcentuales sean iguales. Entonces las dos medias serán iguales.

3.3 Comparación entre media, mediana y moda

La media es la medida más común de tendencia central. Se presta para mayor manipulación e interpretación algebraica. Desafortunadamente, la media se ve afectada por valores extremos, o valores atípicos, y a diferencia de la mediana, puede ser sesgada por las observaciones que están muy por encima o muy por debajo de ésta.

Por ejemplo, para los datos 4, 5, 6, 6, 7, 8 la media y la mediana son ambas 6 y representa una excelente medida del punto central de los datos. Si la última observación fuera 80 en lugar de 8, la media sería 18, pero la mediana todavía sería 6. Debido a que la mediana no se ve afectada por este valor extremo, representa mejor las seis observaciones.

La moda también es menos afectada por unas pocas observaciones atípicas y sigue siendo 6 aun cuando el último valor sea 80. Sin embargo, si no hay moda, o si el conjunto de datos es bimodal, su uso puede ser confuso.

Esto no implica que una medida sea necesariamente mejor que las otras. La medida que se seleccione depende de la naturaleza de los datos o de la forma como se utilicen los datos. Por ejemplo Land's End un vendedor minorista popular de equipos para acampar, se beneficiaría muy poco del hecho de saber que la talla promedio de las botas de excursionismo que vendió fue de 7.3492. De mayor utilidad para las decisiones futuras del negocio sería conocer el tamaño *modal* - reconociendo que vendió más botas de talla 8 que de cualquier otra talla.

Sin embargo, se asume que Land's End desea comercializar una nueva tienda de acampar. Las dimensiones de la tienda dependerán, entre otras cosas, de la estatura promedio de los adultos. La experiencia ha demostrado que la media sirve muy bien como medida de tendencia central cuando se trata de productos que están hechos para acomodarse a la estatura de las personas. El tamaño de los marcos de las puertas de las entradas de los hogares y negocios minoristas, y gran parte del mobiliario se fabrica con base en la estatura promedio.

Ejercicios de la sección

1. Su firma está introduciendo un nuevo chip de computador del cual se promociona que realiza cálculos estadísticos mucho más rápidamente que los que actualmente se encuentran en el mercado. Se hacen veinte cálculos diferentes, produciendo los tiempos en segundos que se ven más adelante. Aunque usted no puede

tergiversar su producto, usted desea presentar los resultados de la manera más favorable para su empresa. Determine la media, la mediana y la moda. Comente los beneficios relativos de utilizar cada estadístico.

3.2	4.1	6.3	1.9	0.6
5.4	5.2	3.2	4.9	6.2
1.8	1.7	3.6	1.5	2.6
4.3	6.1	2.4	2.2	3.3

2. Como gerente de ventas de Acme, Inc., usted desea calcular las medidas de tendencia central para los niveles de utilidad de su firma, durante los últimos doce meses, ya que las siguientes utilidades mensuales están dadas en miles de dólares:

US\$12.3	US\$14.3	US\$25.7
21.6	21.6	-12.8
22.3	18.0	23.1
-3.4	17.9	22.3

3. El director de la planta de Intel desea que usted compare los salarios promedio en su planta de Palo Alto con las de la competencia que queda ubicada en las cercanías de San José. De los 6,012 empleados que supervisa 1,212 ganan US\$12.30 la hora; a 650 se les paga US\$15.50; 3,098 ganan US\$23.50, y al resto se les paga US\$17.12. De los 5,634 empleados que hay en la otra planta 1,654 ganan US\$12.75; 815 reciben US\$17.80 y los otros US\$20.10. Escriba un breve informe para que el director obtenga detalladamente la información que desea.
4. Una empresa grande de equipos deportivos está probando el efecto de dos planes publicitarios sobre las ventas de los últimos 4 meses. Dadas las ventas que se ven aquí, ¿cuál programa de publicidad parece producir el crecimiento promedio más alto en ventas mensuales?

Mes	Plan 1	Plan 2
Enero	US\$ 1,657	US\$ 4,735
Febrero	1,998	5,012
Marzo	2,267	5,479
Abril	3,432	5,589

3.4 Medidas de dispersión

En nuestro esfuerzo por describir un conjunto de números hemos visto que es de utilidad ubicar el centro del conjunto de datos. Pero identificar una medida de tendencia central rara vez es suficiente. Una descripción más completa del conjunto de datos puede obtenerse si se mide qué tan dispersos están los datos alrededor de dicho punto central. Esto es precisamente lo que hacen las medidas de dispersión. Indican cuánto se desvían las observaciones alrededor de su media.

Medidas de dispersión Miden qué tanto se dispersan las observaciones alrededor de su media.

Se toman por ejemplo los tres conjuntos pequeños de datos que se observan aquí.

Conjunto de datos 1	Conjunto de datos 2	Conjunto de datos 3
0,5,10	4,5,6	5,5,5

Los tres tienen una media de cinco. ¿Se debe por tanto concluir que los conjuntos de datos son similares? Claro que no. Sin embargo, si se informa sólo sus medias, sin ver las observaciones, se puede concluir que hay similitud. Una imprecisión más notoria de los conjuntos de datos resultaría si se compara el grado en el cual se dispersaron las observaciones individuales en cada conjunto de datos o se expandieron alrededor de la media cinco. Las observaciones en el primer conjunto de datos están muy dispersas por encima y por debajo de la media, mientras que aquellas del segundo grupo de datos están comparativamente cerca de ésta. El primer conjunto de datos tiene una medida de dispersión mayor que la segunda. El tercer conjunto de datos no tiene dispersión, todas las observaciones son iguales a la media. Sabiendo esto, sería poco probable asumir de manera errónea cualquier similitud en los conjuntos de datos simplemente con base en su media. En este sentido, las medidas de dispersión son muy útiles e informativas.

A. El rango

La medida de dispersión más simple (y menos útil) es el rango o recorrido. El **rango** es simplemente la diferencia entre la observación más alta y la más baja. Su ventaja es que es fácil de calcular. Su desventaja es que considera sólo dos de los cientos de observaciones que hay en un conjunto de datos. El resto de las observaciones se ignoran. Los rangos de los tres conjuntos de datos anteriores son 10, 2 y 0 respectivamente.

B. Varianza y desviación estándar de una población

La varianza y su raíz cuadrada, y la desviación estándar son medidas de dispersión mucho más útiles. Proporcionan una medida más significativa sobre el punto hasta el cual se dispersan las observaciones alrededor de su media.

La **varianza** es el “promedio de las desviaciones respecto a su media elevadas al cuadrado”. ¿Qué significa esto? Significa que (1) se encuentra la cantidad por la cual cada observación se desvía de la media, (2) se elevan al cuadrado tales desviaciones, y (3) se halla la media de tales desviaciones elevadas al cuadrado. Así, se tiene el promedio de las desviaciones de la media elevadas al cuadrado.

Varianza El promedio de las observaciones respecto a su media elevadas al cuadrado.

La varianza para una población σ^2 (se lee sigma al cuadrado) es

Varianza poblacional	$\sigma^2 = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + (X_3 - \mu)^2 + \cdots + (X_N - \mu)^2}{N}$	[3.6]
	$= \frac{\sum (X_i - \mu)^2}{N}$	

En donde $X_1, X_2, X_3, \dots, X_N$ son las observaciones individuales
 μ es la media poblacional
 N es el número de observaciones

La desviación estándar σ es

Desviación estándar poblacional	$\sigma = \sqrt{\sigma^2}$	[3.7]
---------------------------------	----------------------------	-------

Vale la pena notar que debido a que se está trabajando con una población, la media es μ , no $\bar{X}$, como para una muestra, y el número de observaciones es N y no n , como para una muestra.

Desviación estándar Es la raíz cuadrada de la varianza. Es una medida importante de la dispersión de los datos.

A manera de ilustración, Chuckie Chambers vende cinco pólizas de seguro diferentes del baúl de su Plymouth modelo 1973. Sus respectivas primas mensuales son de US\$110, US\$145, US\$125, US\$95 y US\$150. La prima promedio es

$$\mu = \frac{110 + 145 + 125 + 95 + 150}{5} = \text{US\$125}$$

La varianza se halla: (1) restando la media de US\$125 de cada una de las observaciones, (2) elevando al cuadrado estas desviaciones, y (3) hallando el promedio de estas desviaciones al cuadrado. Al seguir estos tres pasos resulta:

$$\begin{aligned}\sigma^2 &= \frac{(110 - 125)^2 + (145 - 125)^2 + (125 - 125)^2 + (95 - 125)^2 + (150 - 125)^2}{5} \\ &= 430\end{aligned}$$

A pesar del uso común de la varianza, ésta presenta dos problemas: es un número muy grande con respecto a las observaciones. Como se puede ver, es varias veces mayor incluso que la observación más grande. Debido a su gran tamaño, con frecuencia la varianza se vuelve difícil para trabajar.

Un problema aún más angustioso es que debido a que las desviaciones son elevadas al cuadrado, la varianza siempre se expresa en términos de los datos originales elevados al cuadrado. En el caso de Chuckie, debido a que elevó al cuadrado las desviaciones de la media, entonces se convierte en 430 dólares al cuadrado – una unidad de medida que no tiene sentido. En la mayoría de los casos la varianza se expresa en términos que no tienen significado o interpretación lógica.

Sin embargo, ambas complicaciones pueden resolverse rápidamente. Tan sólo con hallar la desviación estándar σ , sacando la raíz cuadrada de la varianza:

$$\sigma = \sqrt{430} = \text{US\$20.74}$$

Así de fácil, se solucionan ambos problemas. Ahora se tiene un número más pequeño con el cual es más fácil trabajar, y más importante aún, ahora está expresado en dólares ya que se tomó la raíz cuadrada de los dólares elevados al cuadrado.

El concepto de desviación estándar es muy importante en los negocios y en la economía. Por ejemplo, en finanzas la desviación estándar se utiliza como medida de riesgo relacionada con varias oportunidades de inversión. Mediante el uso de la desviación estándar para medir la variabilidad en las tasas de rendimiento ofrecidas por diferentes inversiones, el analista financiero puede medir el nivel de riesgo que tiene cada activo financiero. Generalmente, entre mayor sea la desviación estándar de la tasa de rendimiento de una inversión en particular, mayor será el grado de riesgo. Veamos el siguiente ejemplo:

Ejemplo 3.4 Varianza y desviación estándar de una población

Markus Boggs es gerente de Nest Egg Investments. Recientemente, Markus estaba interesado en las tasas de rendimiento de los últimos cinco años de dos diferentes fondos mutuos. Megabucks, Inc. mostró, durante un período de cinco años, tasas de rendimiento del 12, 10, 13, 9 y 11%, mientras

que Dynamics Corporation arrojó 13, 12, 14, 10 y 6%. Un cliente se acercó a Boggs y expresó su interés en uno de estos fondos mutuos. ¿Cuál debería escoger Boggs para su cliente?

Solución

Vale la pena destacar que ambos fondos ofrecen un rendimiento promedio del 11%. Debido a que ambos ofrecen el mismo rendimiento en promedio, una inversión más segura es la que tiene un grado menor de riesgo, tal como se midió mediante la desviación estándar. Para Megabucks, Boggs halla

$$\sigma^2 = \frac{(12 - 11)^2 + (10 - 11)^2 + (13 - 11)^2 + (9 - 11)^2 + (11 - 11)^2}{5} = 2$$

La desviación estándar es

$$\sigma = \sqrt{2} = 1.41\%$$

Para Dynamics,

$$\sigma^2 = \frac{(13 - 11)^2 + (12 - 11)^2 + (14 - 11)^2 + (10 - 11)^2 + (6 - 11)^2}{5} = 8$$

La desviación estándar por tanto es

$$\sigma = \sqrt{8} = 2.83\%$$

Interpretación

Debido a que Megabucks presenta menos variabilidad en sus rendimientos y ofrece la misma tasa de rendimiento promedio que ofrece Dynamics, Megabucks representa la más segura de las dos inversiones y por ende es la oportunidad de inversión preferida.

C. Varianza y desviación estándar para una muestra

Tome nota: Los ejemplos anteriores se relacionan con la varianza y la desviación estándar para una *población*. Los símbolos σ^2 y σ son letras griegas típicas de los parámetros.

Rara vez se pueden calcular parámetros. En la mayoría de los casos más bien se estimarán tomando una muestra y calculando los estadísticos correspondientes. Teniendo esto presente, esta sección analiza la forma como se calculan estas importantes medidas de dispersión en la medida en que se relacionan con las muestras.

La varianza y la desviación estándar para una muestra representan medidas de dispersión alrededor de la media. Se calculan de manera parecida a aquellas para una población. La varianza de la muestra s^2 es

Varianza
de la muestra

$$s^2 = \frac{\sum(X_i - \bar{X})^2}{n - 1}$$

[3.8]

y la desviación estándar de la muestra es

Desviación estándar
de la muestra

$$s = \sqrt{s^2}$$

[3.9]

Vale la pena destacar que la media en la fórmula (3.8) se expresa como $\bar{X}$, y no μ , ya que se está trabajando con muestras. Además, se divide por $n - 1$ en lugar de N , debido a que se tiene $n - 1$ *grados de libertad*, $gl = n - 1$. El número de grados de libertad en toda operación estadística es igual al número de observaciones menos toda restricción impuesta en tales observaciones. Una restricción es cualquier valor que deba calcularse de dichas observaciones.

Por ejemplo, se asume que se deben seleccionar $n = 4$ observaciones que deben promediar $\bar{X} = 10$. Bajo estas condiciones, se está en libertad de escoger tres números que se deseen – digamos el tamaño del sombrero, la edad o el coeficiente intelectual-CI. Aunque una vez que esos tres números se seleccionan, el cuarto está predeterminado si los anteriores deben promediar $\bar{X} = 10$. Como lo ilustra la fórmula (3.8), la varianza utiliza el valor para $\bar{X}$ que funciona como una restricción y por tanto reduce los grados de libertad en 1. Por consiguiente, se disminuye el número de observaciones n , en 1.

Otra razón por la cual se toma $n - 1$ es que una muestra generalmente está un poco menos dispersa que la población de la cual se tomó. Por tanto, existe la tendencia a que la desviación estándar de la muestra s sea un poco menor que la desviación estándar de la población σ . Este es un hecho infortunado. Vale la pena recordar que se intenta utilizar el valor de s como una estimación de σ . Sin embargo, s subestimaré σ de manera consistente. Se debe compensar esta condición inflando artificialmente s dividiendo por un número levemente menor $n - 1$, en lugar de n .

Para ilustrar la técnica para establecer estas medidas de dispersión para una muestra, se considera otro problema que tiene Boggs en su esfuerzo por ayudar a sus clientes a tomar decisiones sobre inversión.

Ejemplo 3.5 Varianza y desviación estándar para una muestra

El Sr. Boggs desea determinar la estabilidad del precio de una acción en particular. Decide basar su juicio en la estabilidad de la desviación estándar del precio de cierre diario de dicha acción. Al revisar las páginas financieras, Boggs sabe que la acción ha sido transada en la bolsa durante algún tiempo y que hay muchos precios de cierre desde hace varios meses. En lugar de utilizar todos estos precios, Boggs decide simplificar su aritmética y seleccionar una muestra aleatoria de $n = 7$ días. (Aunque 7 probablemente es una muestra muy pequeña, servirá por el momento para los propósitos que se tienen.) Boggs nota que los precios de cierre son de

US\$87, US\$120, US\$54, US\$92, US\$73, US\$80, y US\$63

Solución

Entonces, $\bar{X} = \text{US\$81.29}$ y

$$\begin{aligned} s^2 &= \frac{\sum (X_i - \bar{X})^2}{n - 1} \\ &= \frac{(87 - 81.29)^2 + (120 - 81.29)^2 + \dots + (63 - 81.29)^2}{7 - 1} \\ s^2 &= 465.9 \text{ cuadrado de dólares} \\ s &= \sqrt{465.9} = \text{US\$21.58} \end{aligned}$$

Interpretación

Boggs ha estimado que la media del precio de cierre de la acción es de US\$81.29, con una tendencia a variar por encima o por debajo de dicho precio en US\$21.58. Una explicación adicional del uso e interpretación de la desviación estándar se proporciona más tarde en este capítulo. Sin embargo, se tiene en mente que Boggs puede interpretar siempre la desviación estándar de US\$21.58 como una

medida de la tendencia de los precios de cierre que tienden a fluctuar alrededor de su media de \$81.29 dólares.

Todos estos estadísticos importantes pueden obtenerse en los computadores utilizando varios paquetes. La pantalla 3.1 muestra una copia impresa en Excel para los datos de P&P. Es de destacar que las medidas de tendencia central y de dispersión se dan junto con otros estadísticos que se discutirán más adelante.

Pantalla 3.1

Estadísticos descriptivos para P&P

Mean (Media)	78.36
Standard Error (Error estándar)	1.599938774
Median (Mediana)	78.5
Mode (Moda)	83
Standard Deviation (Desviación estándar)	11.31327557
Sample Variance (Varianza de la muestra)	127.9902041
Kurtosis (Curtosis)	-0.036918787
Skewness (Sesgo)	-0.05225955
Range (Rango)	52
Minimum (Mínimo)	50
Maximum (Máximo)	102
Sum (Suma)	3918
Count (Conteo)	50

Ejercicios de la sección

5. Se utilizan dos procesos para producir discos de computador. Han surgido problemas respecto a las variaciones en los tamaños de tales discos. Con base en los datos de muestra aquí observados, de ocho tamaños de discos en pulgadas para cada proceso, explique cuál proceso aconsejaría usted si su objetivo es minimizar la desviación en el tamaño alrededor de la media.

Proceso 1		Proceso 2	
3.41	3.22	3.81	3.26
3.74	3.06	3.26	3.79
3.89	3.65	3.07	3.14
3.65	3.33	3.35	3.51

6. Explique con sus propias palabras qué miden la varianza y la desviación estándar. ¿Por qué su cálculo es algo diferente para las poblaciones y las muestras?
7. Un analista de inversiones sugiere que usted invierta en Boomer Securities en lugar de Reliable Stocks. Dadas las tasas anuales de rendimiento que se muestran a continuación para una muestra de cada inversión, ¿qué le dice al analista si usted desea minimizar su exposición al riesgo?

Boomer		Reliable	
15.5%	3.6%	4.5%	6.2%
21.7	27.2	5.5	7.2
-7.8	2.2	3.5	4.2
-5.0	12.2	4.1	

8. Curly, Moe, y Larry venden seguros de vida para la Shemp Insurance Company. El Sr. Shemp ascenderá a uno de sus vendedores a un cargo administrativo con base en su desempeño en ventas. Su decisión depende de cuál miembro de su equipo de ventas tiene (1) el promedio más alto en ventas y (2) el registro de ventas más consistente. Dados los siguientes datos de muestra semanales en ventas, ¿cuál vendedor obtendrá el ascenso?

Curly		Moe		Larry	
US\$ 986	US\$ 1,265	US\$645	US\$893	US\$534	US\$534
1,337	734	645	230	534	534
2,745	245	734	415	534	534
2,645	5,344	822	723	534	534
3,658	4,867				

3.5 Medidas de tendencia central y de dispersión para datos agrupados

Al trabajar con datos que han sido agrupados en una distribución de frecuencia, no se conoce cuáles son las observaciones individuales. En una tabla de frecuencia para los pasajeros de P&P, la tabla 2.2, se sabe solamente que en tres días, entre 50 y 59 pasajeros abordaron un avión. No se tienen las cifras exactas de esos tres días. Sin los valores específicos, los procedimientos mostrados anteriormente para calcular las medidas descriptivas, simplemente no se aplican. Deben encontrarse métodos alternativos. Debe tenerse en mente que los cálculos hechos utilizando datos agrupados son sólo aproximaciones. Por tanto, las observaciones individuales no agrupadas deberían utilizarse cuando sea posible.

A. La media

Al calcular la media de datos agrupados, se supone que las observaciones en cada clase son iguales al punto medio de la clase. Aunque esto puede ser una suposición heroica, probablemente se compensa debido a que posiblemente algunas de las observaciones superen el punto medio, mientras que otros queden por debajo de él. Dada esta suposición, se debe tener en cuenta la frecuencia y los puntos medios de cada clase cuando se calcule la media utilizando datos agrupados. La fórmula 3.10 hace justo esto.

Media con datos agrupados

$$\bar{X}_g = \frac{\sum fM}{n} = \frac{\sum fM}{\sum f} \quad [3.10]$$

en donde f es la frecuencia o número de observaciones en cada clase
 M es el punto medio de cada clase
 n es el tamaño de la muestra y es igual a las frecuencias sumadas en todas las clases.

La tabla de frecuencia para Pigs & People Airlines del capítulo 2 se repite en la tabla 3.2 por conveniencia, junto con los puntos medios para cada clase, los cuales, como se debe recordar, se determinaron promediando los límites superior e inferior.

Utilizando la fórmula (3.10), se puede ver que P&P transportó un promedio diario de 78.7 pasajeros.

$$\bar{X}_g = \frac{\sum fM}{n} = \frac{3935}{50} = 78.7 \text{ pasajeros}$$

Tabla 3.2
Distribución
de frecuencia
para pasajeros

Clase (pasajeros)	Frecuencia (<i>f</i>) (días)	<i>M</i>	<i>fM</i>
50-59	3	54.5	163.5
60-69	7	64.5	451.5
70-79	18	74.5	1341.0
80-89	12	84.5	1014.0
90-99	8	94.5	756.0
100-109	2	104.5	209.0
	50		3935.0

B. La mediana

Si se han registrado datos en una tabla de frecuencia, no pueden colocarse en un arreglo ordenado para calcular la mediana. A manera de ilustración, se suministra la tabla de frecuencia de P&P Airlines, en la tabla 3.3.

Tabla 3.3
Distribución
de frecuencia
para pasajeros

Clase	<i>f</i>	Frecuencia acumulada
50-59	3	3
60-69	7	10
70-79	18	28
80-89	12	40
90-99	8	48
100-109	2	50

Primero se debe hallar la clase de la mediana de la distribución de frecuencia. La **clase mediana** es la clase cuya frecuencia acumulada es mayor que o igual a $n/2$.

Debido a que n es 50, se necesita localizar la primera clase con una frecuencia acumulada de 25 o más. La tercera clase tiene una frecuencia acumulada de 28. La mediana puede determinarse entonces como

Mediana para datos agrupados	$\text{Mediana} = L_{md} + \left[\frac{n/2 - F}{f_{md}} \right] (C)$	[3.11]
------------------------------	---	--------

en donde L_{md} es el límite inferior de la clase de la mediana (70)
 F es la frecuencia acumulada de la clase que antecede a la clase de la mediana (10)
 f_{md} es la frecuencia de la clase de la mediana (18)
 C es el intervalo de clase de la clase de la mediana (10)

Utilizando la fórmula (3.11), se obtiene la mediana:

$$\text{Mediana} = 70 + \left[\frac{50/2 - 10}{18} \right] 10 = 78.33 \text{ pasajeros}$$

Se puede concluir que en 25 días – la mitad de los 50 días estudiados – menos de 78.33 pasajeros volaron en P&P Airlines, y en los otros 25 días, más de 78.33 pasajeros volaron los agradables cielos de P&P.

C. La moda

Ya que por definición la moda es la observación que ocurre con mayor frecuencia, se hallará en la clase que tenga la frecuencia más alta, llamada la **clase modal**. Para estimar la moda en el caso de datos agrupados, se utiliza la fórmula (3.12).

Moda para datos agrupados	$\text{Moda} = L_{mo} + \left[\frac{D_a}{D_b + D_a} \right] (C) \quad [3.12]$
---------------------------	--

en donde L_{mo} es el límite inferior de la clase modal
 D_a es la diferencia entre la frecuencia de la clase modal y la clase que la antecede
 D_b es la diferencia entre la frecuencia de la clase modal y la clase que le sigue
 C es el intervalo de clase de la clase modal

De la tabla 3.3, la moda es

$$\text{Moda} = 70 + \left[\frac{18 - 7}{(18 - 12) + (18 - 7)} \right] (10) = 76.47 \text{ pasajeros}$$

D. Varianza y desviación estándar

Si los datos están agrupados en una tabla de frecuencia, la varianza y la desviación estándar pueden calcularse como

Varianza de la muestra de datos agrupados	$s^2 = \frac{\sum fM^2 - n\bar{X}^2}{n - 1} \quad [3.13]$
---	---

y

Desviación estándar muestral para datos agrupados	$s = \sqrt{s^2} \quad [3.14]$
---	-------------------------------

Los datos de pasajeros para P&P Airlines se utilizarán como ilustración.

Ejemplo 3.6 Medidas de dispersión con datos agrupados

El director de vuelo de P&P requiere información respecto a la dispersión del número de pasajeros. Las decisiones que se tomen respecto a la programación y al tamaño más eficiente de los aviones,

dependerá de la fluctuación en el transporte de pasajeros. Si esta variación en número de pasajeros es grande, se pueden necesitar aviones más grandes para evitar el sobrecupo en los días en los que el transporte de pasajeros es más solicitado. La tabla de frecuencia para P&P es la siguiente:

Clase (pasajeros)	f (días)	M	fM	M^2	fM^2
50-59	3	54.5	163.5	2,970.25	8,910.75
60-69	7	64.5	451.5	4,160.25	29,121.75
70-79	18	74.5	1,341.0	5,550.25	99,904.50
80-89	12	84.5	1,014.0	7,140.25	85,683.00
90-99	8	94.5	756.0	8,930.25	71,442.00
100-109	2	104.5	209.0	10,920.25	21,840.50
$n = 50$			$\Sigma fM = 3,935.0$		$\Sigma fM^2 = 316,902.50$

Solución

Se calculó $\bar{X}_g$ en un ejemplo anterior

$$\bar{X}_g = \frac{\Sigma fM}{n} = \frac{3935}{50} = 78.7$$

Por tanto, la fórmula (3.13) da

$$s^2 = \frac{\Sigma fM^2 - n\bar{X}^2}{n - 1}$$

$$s^2 = \frac{316,902.50 - 50(78.7)^2}{49} = 147.31 \text{ pasajeros elevado al cuadrado}$$

$$s = \sqrt{147.31} = 12.14 \text{ pasajeros}$$

Interpretación

El director de vuelo ahora puede decidir si los aviones que se están utilizando actualmente pueden acomodar fluctuaciones en los niveles de pasajeros tal y como lo mide una desviación estándar de 12.14. Si no, quizá se utilizarán aviones más grandes para acomodar cualquier excedente que pueda ocurrir en esos días de tráfico pesado.

Ejercicios de la sección

9. El ausentismo diario en su oficina parece ir en aumento. El año pasado un promedio de 47.8 empleados estuvo ausente algunos días, con una desviación estándar de 14.7. Se recolectó una muestra de datos para el año en curso y se ubicaron en la tabla de frecuencias que se muestra a continuación. Calcule la media, la mediana, la moda, y la desviación estándar para estos datos y compárelos con los del año anterior. ¿A qué conclusiones llega?

Número de empleados ausentes	Días en los que ese número estuvo ausente
20-29	5
30-39	9
40-49	8
50-59	10
60-69	12
70-79	11
80-89	8
90-99	3

10. Anteriormente, el tiempo para completar un trabajo determinado en las oficinas de Harmon electronics había arrojado las siguientes estadísticas en horas: una media de 12.2, una mediana de 13.2, y una moda de 14.5. La varianza fue de 8.21. Se reflejan datos más recientes en la siguiente tabla de frecuencias. El Sr. Harmon lo contrata como consultor externo para evaluar los cambios en la eficiencia de los empleados. Calcule los estadísticos correspondientes con base en estos datos y prepare un breve informe. ¿Qué conclusiones obtiene?

Horas tomadas hasta la finalización del trabajo	Número de veces que dicho trabajo se tomó este tiempo
5 y menos de 7	4
7 y menos de 9	8
9 y menos de 11	12
11 y menos de 13	8
13 y menos de 15	5
15 y menos de 17	2

3.6 Otras medidas de dispersión

Aunque la varianza y la desviación estándar son las medidas de dispersión más útiles en análisis estadístico, existen otras técnicas con las cuales puede medirse la dispersión de un conjunto de datos. Estas medidas adicionales de dispersión son los **cuartiles**, los **deciles** y los **percentiles**.

Cada conjunto de datos tiene tres cuartiles que lo dividen en cuatro partes iguales. El primer cuartil es ese valor debajo del cual clasifica el 25% de las observaciones, y sobre el cual puede encontrarse el 75% restante. El segundo cuartil es justo la mitad. La mitad de las observaciones están por debajo y la mitad por encima; en este sentido, es lo mismo que la mediana. El tercer cuartil es el valor debajo del cual está el 75% de las observaciones y encima del cual puede encontrarse el 25% restante.

La determinación de cuartiles con frecuencia es de utilidad. Por ejemplo muchas escuelas de postgrados admitirán sólo a aquellos estudiantes que estén en el 25% superior (tercer cuartil) de los candidatos. Las empresas, con frecuencia, desean señalar las plantas cuyos deficientes registros de producción los colocan por debajo del cuartil inferior. Con un poco de imaginación es posible prever numerosos ejemplos en los cuales la determinación de cuartiles puede ser de gran beneficio.

Los deciles separan un conjunto de datos en 10 subconjuntos iguales, y los percentiles en 100 partes. El primer decil es la observación debajo de la cual se encuentra el 10% de las observaciones, mientras que el 90% restante se encuentra encima de éste. El primer percentil es el valor debajo del cual se encuentra el 1% de las observaciones, y el resto están encima de éste. Puede aplicarse una interpretación similar al resto de deciles y percentiles. Todo conjunto de datos tiene 9 deciles y 99 percentiles.

Un percentil y su ubicación en un arreglo ordenado se identifica mediante los subíndices. Por ejemplo, el decimoquinto percentil se indica como P_{15} , y su ubicación en la serie ordenada es L_{15} .

Para ilustrar el cálculo de percentiles, se asume que se tienen observaciones para el número de acciones correspondientes a 50 acciones transados en la Bolsa de Valores de Nueva York, como se muestra en la tabla 3.4. Vale la pena destacar que los datos han sido puestos en una serie ordenada. El lugar del P ésimo percentil se halla

Ubicación de un percentil

$$L_p = (n + 1) \frac{P}{100}$$

[3.15]

en donde L_p es el sitio del percentil deseado en una serie ordenada
 n es el número de observaciones
 P es el percentil deseado

Se asume que se desea calcular el percentil 25, P_{25} , para las acciones de la tabla 3.4. Se debe hallar primero su ubicación en la serie ordenada.

$$L_{25} = (50 + 1) \frac{25}{100} \\ = 12.75$$

Tabla 3.4

Números de acciones
transadas en la Bolsa
de Valores de Nueva
York (en 100's)

3	10	19	27	34	38	48	56	67	74
4	12	20	29	34	39	48	59	67	74
7	14	21	31	36	43	52	62	69	76
9	15	25	31	37	45	53	63	72	79
10	17	27	34	38	47	56	64	73	80

El valor resultante de 12.75 dice que el percentil 25 está ubicado al 75% del trayecto comprendido entre la doceava observación, que es 20 y la treceava observación que es 21, es decir, $P_{25} = 20 + 0.75(21 - 20) = 20.75$.

Si se desea calcular el percentil 35, se halla

$$L_{35} = (50 + 1) \frac{35}{100} \\ = 17.85$$

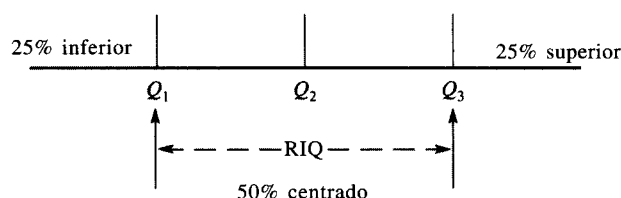
El percentil 35 está al 85% del trayecto comprendido entre la observación 17, que es 29 y la observación 18 que es 31, es decir, $P_{35} = 29 + (0.85)(31 - 29) = 30.7$. Por tanto el 35% de las observaciones está por debajo de 30.7 y el 65% restante por encima de 30.7.

Regresando a los deciles y cuartiles por un momento, se nota que el primer decil es igual a P_{10} , el segundo decil es igual a P_{20} , y así sucesivamente. Adicionalmente, el primer cuartil es igual a P_{25} , el segundo cuartil es igual a P_{50} , y P_{75} se encuentra en el tercer cuartil. Teniendo esto en mente, el cálculo de deciles y cuartiles se vuelve simplemente un asunto de determinación de los percentiles apropiados de acuerdo con las reglas que se acaban de establecer.

Una medida única de dispersión es el **rango o recorrido intercuartílico** (*Interquartile range - RIQ*). El RIQ es la diferencia entre el tercer cuartil y el primer cuartil. Es decir, $P_{75} - P_{25}$. La mitad de las observaciones se clasifican dentro de este rango. Consta del 50% de la mitad de las observaciones y corta el 25% inferior y el 25% superior de los puntos de datos. Como resultado, el RIQ proporciona una medida de dispersión que no está muy influenciada por unas cuantas observaciones extremas. El rango intercuartil se ilustra en la figura 3.1.

Figura 3.1

Recorrido
intercuartílico



Ejercicios de la sección

11. Defina en sus propias palabras:
 - a. El percentil 80
 - b. El cuarto decil
 - c. El tercer cuartil
12. Utilizando los datos de la tabla 3.4, calcule e interprete:
 - a. El rango o recorrido intercuartílico.
 - b. La mediana
 - c. El percentil 50
 - d. El percentil 70
 - e. El tercer decil

3.7 Usos frecuentes de la desviación estándar

Como se ha enfatizado, la desviación estándar es útil para describir un conjunto de datos midiendo el grado de dispersión de las observaciones individuales alrededor de su media. Existen dos aplicaciones adicionales para la desviación estándar: (1) Teorema de Chebyshev y (2) la Regla empírica.

A. Teorema de Chebyshev

El teorema de Chebyshev (algunas veces escrito teorema de Tchebysheff) fue formulado por el matemático ruso P.L. Chebyshev (1821 – 1894). Establece que para *todo* conjunto de datos, por lo menos $1 - 1/K^2$ % de las observaciones están dentro de K desviaciones estándar de la media, en donde K es cualquier número mayor que 1. El teorema de Chebyshev se expresa como

Teorema de Chebyshev	$1 - \left[\frac{1}{K^2} \right]$	[3.16]
----------------------	------------------------------------	--------

Así, por ejemplo, si se forma un intervalo de $K =$ tres desviaciones estándar por encima de la media hasta tres desviaciones estándar por debajo de la media, entonces por lo menos

$$1 - \frac{1}{3^2} = 88.89\%$$

de todas las observaciones estarán dentro de dicho intervalo.

Ejemplo 3.7 Teorema de Chebyshev aplicado a Pigs & People Airlines

El trabajo anterior en este capítulo con datos agrupados para P&P Airlines, reveló una media de 78.7 pasajeros por día, con una desviación estándar de 12.14. Para programar los tiempos para una nueva ruta que abrió P&P, la gerencia desea saber con qué frecuencia los pasajeros están dentro de $K =$ dos desviaciones estándar de la media, y cuál es dicho intervalo.

Solución

Si se transportan dos desviaciones estándar (2×12.14) = 24.28 pasajeros por encima y por debajo de la media de 78.7, se tendrá un intervalo de $(78.7 - 24.28) = 54.42$ a $(78.7 + 24.28) = 102.98$ pasajeros. Se puede estar seguro de que por lo menos