## 3.4. Rango, varianza y desviación estándar

### 3.4.1. Rango

También conocido con el nombre de amplitud o recorrido, el *rango* se define como la diferencia que existe entre el valor máximo y el valor mínimo de un conjunto de datos. Es la medida de dispersión más fácil de calcular, y es especialmente útil en aquellas situaciones en que el objetivo de la investigación sólo consiste en averiguar el alcance de las variaciones extremas.

Por ejemplo, el desempeño del precio de las acciones en el mercado bursátil se suele reconocer por los rangos, al citar los precios máximos y mínimos de cada sesión. Es decir, la variación en el precio de una acción puede medirse obteniendo el rango existente entre los dos valores más extremos y así interpretar qué tanta volatilidad manifestó la acción en una jornada o periodo. Si se comparan dos acciones, se puede interpretar que la acción que tiene mayor variación es aquella que tiene mayor rango.

### Ejemplo 20

Una compañía de seguros desea conocer la variación que existe en las ventas de sus ocho vendedores y de esa manera determinar la productividad de cada uno de ellos. Calcula el rango empleando la siguiente información de seguros vendidos durante un mes: 8, 11, 5, 14, 11, 8, 11, 16.

Si se desea hallar el *rango* de tales observaciones sólo hay que identificar el valor máximo (16) y el valor mínimo (5) y obtener la diferencia entre ellos.

*Rango* = *Valor máximo* – *Valor mínimo* = 16 – 5 = 11

El *rango* es 11, lo cual quiere decir que la diferencia entre el número de seguros vendidos por dos vendedores distintos, el mejor vendedor y el peor vendedor, es de 11, indicando una gran dispersión o variabilidad, ya que sería ilógico que si un vendedor logra vender 16 seguros, el otro sólo venda 5 si se trata de los mismos seguros. Lo anterior puede atribuirse a la experiencia, a la capacitación o a la cartera de clientes que cada vendedor tiene.

### Ejemplo 21

Un analista desea comparar el desempeño de la Bolsa Mexicana de Valores de dos meses: septiembre y octubre de 2001. Para esto toma su principal indicador, el Índice de Precios y Cotizaciones (IPC), y obtiene las siguientes gráficas.

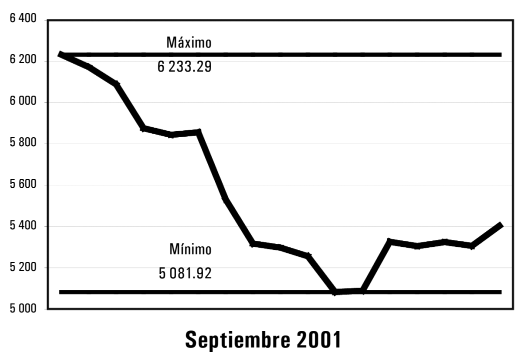


**Figura 3.6. Bolsa Mexicana de Valores en septiembre de 2001.**

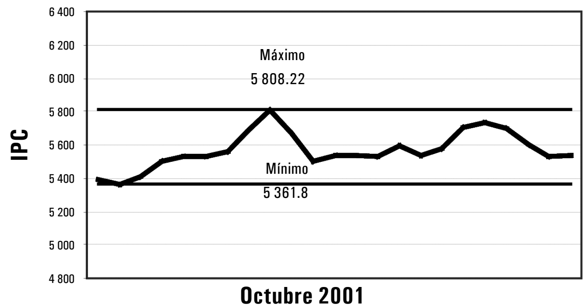


**Figura 3.7. Bolsa Mexicana de Valores en octubre de 2001.**

Si se desea conocer en cuál de los dos meses se presentó mayor volatilidad en el mercado de valores encontramos los rangos del IPC en cada uno de ellos:

Rango en septiembre 2001 = 6 233.29 – 5 081.92 = 1 151.37

Rango en octubre 2001 = 5 808.22 – 5 361.8 = 446.42

Se puede decir que en el mes de septiembre de 2001, la Bolsa Mexicana de Valores registró mayor volatilidad que en el mes de octubre, pues su rango de 1 151.37 fue superior al observado durante el mes de octubre de 446.42.

Este resultado también puede apreciarse de manera visual en las figuras 3.6. y 3.7., donde los rangos se representan por el diferencial existente entre el nivel máximo y el nivel mínimo del IPC. En el mes de septiembre se observa un rango mucho más ancho que el del mes de octubre, el cual se atribuyó al nerviosismo generado por los ataques terroristas del día 11 de septiembre en el Pentágono y en el World Trade Center de Nueva York.

## Ventajas y desventajas del rango

La principal ventaja del rango radica en que es la medida de dispersión más fácil de obtener, pues únicamente se toman los dos valores extremos y se diferencian entre sí. Además, al medirse la amplitud entre los dos valores más extremos en una serie de datos, esta medida de dispersión suele ser muy útil cuando se desea conocer qué tan extremos son los límites máximos y mínimos de una variable; por ejemplo, las temperaturas de ciertas ciudades del país o la ganancia de las casas de cambio que se obtienen diferenciando los precios de compra y los precios de venta para cada divisa.

Sin embargo, el hecho de que se tomen en cuenta únicamente los dos valores más extremos de un conjunto de datos, el rango puede ser una medida de dispersión que resulta afectada ante la presencia de datos atípicos.

EJERCICIOS

## Ejercicio 6

**1.** El rango se define como:

a) La amplitud entre el valor más grande y el valor más pequeño de la serie de datos.
b) La suma del valor más grande y el valor más pequeño de la serie de datos.
c) La diferencia entre los valores extremos y el valor central de la serie de datos.
d) La diferencia entre los valores centrales de la serie de datos.

**2.** El rango presenta fallas como medida de dispersión cuando:

a) Se tiene la presencia de medias desproporcionadas.
b) Se realiza un muestreo aleatorio.
c) Los datos emanan de una muestra y no de una población.
d) Se tiene la presencia de datos atípicos.

**3.** Es una de las ventajas de utilizar el rango:

a) Es una medida que señala hacia dónde se concentran los datos.
b) Es la medida de dispersión más fácil de calcular.
c) Es la medida de dispersión más exacta que existe en una serie.
d) Señala cómo se dispersan los datos de la media.

**4.** Si tenemos los siguientes datos: 0, 1, 1, 3, y 5, entonces el rango es:

a) 5
b) 4
c) 2
d) 6

**5.** El departamento de crédito y cobranza de una empresa quiere conocer la variación que existe en una muestra de 15 datos, correspondientes a los próximos cobros (en pesos) que debe hacer. Calcula el rango para los datos siguientes:

| | | | | |
|---|---|---|---|---|
| 10 000 | 12 000 | 15 000 | 16 000 | 15 000 |
| 9 000 | 13 500 | 12 700 | 9 700 | 18 000 |
| 13 200 | 12 600 | 14 000 | 18 700 | 16 500 |

### 3.4.2. Varianza

Es una medida de variabilidad que toma en cuenta la dispersión que los valores de los datos tienen respecto a su media. Es decir, aquellos conjuntos de datos que tengan valores más alejados de la media, sea muestral o poblacional, tendrán una mayor varianza. Su resultado se expresa en unidades al cuadrado.

Existen dos símbolos para representar la varianza ($\sigma^2$ y $S^2$). La $S^2$ se refiere a un estadístico, es decir, a la varianza de una muestra; mientras que $\sigma^2$ se refiere a un parámetro, es decir, a la varianza de una población. A la $S^2$se le conoce como la varianza muestral mientras que a $\sigma^2$ se le conoce como la varianza poblacional.

La manera de obtener la varianza de un conjunto de datos depende de la forma como se encuentren organizados los datos, ya sea que estén agrupados o no agrupados, así como del tipo de información con la que se trabaje, ya sea que provenga de una muestra o de una población.

#### a) La varianza para datos no agrupados

Cuando tenemos una variable cuya serie de datos no se encuentra agrupada, $X_1$, $X_2$, $X_3$,..., $X_n$, la *varianza poblacional* se calcula mediante la siguiente fórmula:

$$V(X) = \sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$$

Donde:

$\Sigma(X_i - \mu)^2$= Suma de los cuadrados de las desviaciones del valor de cada dato de la serie respecto a la media poblacional.

$X_i$ = El valor de cada dato de la serie.

$\mu$ = La media poblacional.

$N$ = Tamaño de la población.

Es decir, la varianza de una población para datos no agrupados es el promedio del cuadrado de las desviaciones respecto a su media $\mu$.

Cuando tenemos una variable cuya serie de datos no se encuentra agrupada, $X_1$, $X_2$, $X_3$,..., $X_n$, la *varianza muestral* se calcula mediante la siguiente fórmula:

$$S^2 = \frac{\Sigma(X - \overline{X})^2}{n - 1}$$

Donde:

$\Sigma(X_i - \overline{X})^2$ = Suma de los cuadrados de las desviaciones del valor de cada dato de la serie respecto a la media muestral.

$X_i$ = El valor de cada dato de la serie.

$\overline{X}$ = La media muestral.

$N$ = Tamaño de la muestra.

A diferencia de lo que ocurre con otras fórmulas, la varianza de una muestra no equivale exactamente, en términos de cálculo, a la varianza de una población. El denominador de la fórmula de la varianza poblacional es el total de la población $N$, mientras que en la varianza muestral se incluye un factor de corrección $n - 1$.

Los pasos para obtener la varianza muestral o poblacional para datos no agrupados son los siguientes:

1. Encuentra la media muestral o poblacional, según sea el caso.

2. Obtén cada una de las desviaciones respecto a la media, es decir, a cada uno de los datos $X_1, X_2, \ldots, X_n$ se le resta la media obtenida en el paso anterior para quedar los siguientes valores:

   $(X_1 - \mu), (X_2 - \mu), \ldots, (X_n - \mu)$ en caso de una población.
   $(X_1 - \overline{X}), (X_2 - \overline{X}), \ldots, (X_n - \overline{X})$ en caso de una muestra.

3. Eleva al cuadrado cada una de las desviaciones obtenidas en el paso anterior y súmalas entre sí, para obtener la suma del cuadrado de las desviaciones:

   $\Sigma (X - \mu)^2 = (X_1 - \mu)^2 + (X_2 - \mu)^2 + \ldots + (X_n - \mu)^2$ en caso de una población.
   $\Sigma (X - \overline{X})^2 = (X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \ldots + (X_n - \overline{X})^2$ en caso de una muestra.

4. La suma del cuadrado de las desviaciones respecto a su media se divide entre $N$, en caso de una población; o entre $n - 1$, en caso de una muestra.

Tanto para una población como para una muestra, la fórmula de la varianza puede ser transformada en las siguientes expresiones, las cuales son conocidas como el **método corto de la varianza**:

**Varianza poblacional** $$V(X) = \sigma^2 = \frac{\sum X_i^2}{N} - \mu^2$$

**Varianza muestral** $$S^2 = \frac{\sum X_i^2 - n\overline{X}^2}{n-1}$$

Estas fórmulas tienen la ventaja de simplificar las operaciones que se deben realizar cuando se calcula la varianza, sea poblacional o muestral. Cabe señalar que las fórmulas establecidas por el método corto nos conducen al mismo resultado que si se hubieran empleado las fórmulas anteriores, siempre y cuando no se hayan omitido algunos dígitos en las distintas operaciones. La conveniencia de utilizar una u otra fórmula queda sujeta a la libre elección del lector, según la comodidad que le produzca cada una de ellas para realizar las operaciones.

## Ejemplo 22

Emplea los datos de las ventas de seguros del ejemplo 20 y calcula la varianza, suponiendo que los datos constituyen la población total de los agentes de seguro de la compañía.

Se tiene que la media es:

$$\mu = \frac{\sum X}{N} = \frac{(8+11+5+14+11+8+11+16)}{8} = \frac{84}{8} = 10.5$$

Para calcular la varianza se requiere obtener cada una de las diferencias o desviaciones de los datos respecto a la media $(X - \mu)$, elevarlas al cuadrado $(X - \mu)^2$ y sumar estos resultados:

| X | $(X - \mu)$ | $(X - \mu)^2$ |
|---|---|---|
| 8 | –2.5 | 6.25 |
| 11 | 0.5 | 0.25 |
| 5 | –5.5 | 30.25 |
| 14 | 3.5 | 12.25 |
| 11 | 0.5 | 0.25 |
| 8 | –2.5 | 6.25 |
| 11 | 0.5 | 0.25 |
| 16 | 5.5 | 30.25 |
| Σ | 0 | 86 |

**Tabla 3.20. Desviaciones de la venta de seguros.**

Ahora aplicamos la fórmula de varianza poblacional para datos no agrupados y obtenemos:

$$V(X) = \sigma^2 = \frac{\Sigma(X_i - \mu)^2}{N} = \frac{86}{8} = 10.75$$

Puede apreciarse que la varianza es de 10.75. Sin embargo, esta medida de variación no tiene un significado práctico debido a que el resultado obtenido está expresado en términos cuadrados, es decir, la variabilidad de seguros vendidos es de 10.75 seguros cuadrados.

Por esa razón, la varianza sólo tiene sentido lógico cuando comparamos diferentes conjuntos de datos con la misma unidad de medida, es decir, su interpretación es una medida relativa en el sentido de que aquel conjunto que tenga la mayor varianza será el de mayor grado de dispersión.

Por otra parte, si el lector hubiera optado por el método corto para estimar la varianza poblacional, el resultado hubiera sido el mismo. Para ello debemos estimar $\Sigma X_i^2$ y $\mu^2$:

$$\Sigma X_i^2 = 8^2 + 11^2 + 5^2 + 14^2 + 11^2 + 8^2 + 11^2 + 16^2$$

$$= 64 + 121 + 25 + 196 + 121 + 64 + 121 +256 = 968$$

$$\mu^2 = 10.5^2 = 110.25$$

$$V(X) = \sigma^2 = \frac{\Sigma X_i^2}{N} = \mu^2 = \frac{968}{8} = -110.25 = 121 - 110.25 = 10.75$$

Si se compara este resultado mediante el método corto con el primer método, se puede apreciar que los resultados no fueron distintos.

## Ejemplo 23

En las tablas 3.21 y 3.22 se exponen las cotizaciones mensuales del tipo de cambio entre el peso mexicano y el dólar estadounidense para los años de 1995 y 2000. Observa cuidadosamente la información contenida en cada tabla.

a) Realizando una inspección visual, ¿en cuál de los dos años se observa una mayor estabilidad en el tipo de cambio?

b) Encuentra la varianza para el tipo de cambio entre el peso y el dólar estadounidense en cada uno de los dos años.

| Mes | Tipo de cambio en 1995 |
|---|---|
| Enero | 5.69 |
| Febrero | 5.83 |
| Marzo | 6.81 |
| Abril | 5.78 |
| Mayo | 6.17 |
| Junio | 6.30 |
| Julio | 6.08 |
| Agosto | 6.31 |
| Septiembre | 6.41 |
| Octubre | 7.17 |
| Noviembre | 7.65 |
| Diciembre | 7.64 |

*Fuente*: Banco de México: www.banxico.org.mx

**Tabla 3.21. Tipo de cambio mensual peso-dólar en el año 1995.**

| Mes | Tipo de cambio en el 2000 |
|---|---|
| Enero | 9.47 |
| Febrero | 9.44 |
| Marzo | 9.29 |
| Abril | 9.37 |
| Mayo | 9.50 |
| Junio | 9.79 |
| Julio | 9.46 |
| Agosto | 9.28 |
| Septiembre | 9.33 |
| Octubre | 9.51 |
| Noviembre | 9.51 |
| Diciembre | 9.44 |

*Fuente*: Banco de México: www.banxico.org.mx

**Tabla 3.22. Tipo de cambio mensual peso-dólar en el año 2000**

Se observa que los valores del tipo de cambio en el año de 1995 se encuentran muy dispersos entre sí, lo que indica una gran variabilidad o inestabilidad en el mercado cambiario. En contraste, en el año 2000 se puede observar que los valores de la divisa estadounidense se encuentran poco dispersos por lo que se esperaría que la varianza en este año sea menor a la de 1995.

Como los datos no se encuentran organizados mediante tablas de frecuencias, procedemos a encontrar la varianza muestral para datos no agrupados, obteniendo en primer lugar sus medias respectivas:

La media de 1995 es: $\overline{X} = \dfrac{\sum X}{N} = \dfrac{(5.69+5.83+6.81+\ldots+7.67)}{12} = \dfrac{77.84}{12} = 6.48$

La media de 2000 es: $\overline{X} = \dfrac{\sum X}{N} = \dfrac{(9.47+9.44+9.29+\ldots+9.44)}{12} = \dfrac{113.39}{12} = 9.44$

Procedemos a encontrar la suma del cuadrado de las desviaciones del tipo de cambio respecto a la media, de acuerdo con las siguientes tablas:

| Mes | $(X-\overline{X})$ | $(X-\overline{X})^2$ |
|---|---|---|
| Enero | −0.79 | 0.6241 |
| Febrero | −0.65 | 0.4225 |
| Marzo | 0.33 | 0.1089 |
| Abril | −0.70 | 0.49 |
| Mayo | −0.31 | 0.0961 |
| Junio | −0.18 | 0.0324 |
| Julio | −0.40 | 0.16 |
| Agosto | −0.17 | 0.0289 |
| Septiembre | −0.07 | 0.0049 |
| Octubre | 0.69 | 0.4761 |
| Noviembre | 1.17 | 1.3689 |
| Diciembre | 1.16 | 1.3456 |
| Suma | | 5.1584 |

**Tabla 3.23. Desviaciones del tipo de cambio en el año 1995.**

| Mes | $(X-\overline{X})$ | $(X-\overline{X})^2$ |
|---|---|---|
| Enero | 0.03 | 0.0009 |
| Febrero | 0 | 0 |
| Marzo | −0.15 | 0.0225 |
| Abril | −0.07 | 0.0049 |
| Mayo | 0.06 | 0.0036 |
| Junio | 0.35 | 0.1225 |
| Julio | 0.02 | 0.0004 |
| Agosto | −0.16 | 0.0256 |
| Septiembre | −0.11 | 0.0121 |
| Octubre | 0.07 | 0.0049 |
| Noviembre | 0.07 | 0.0049 |
| Diciembre | 0 | 0 |
| Suma | | 0.2023 |

**Tabla 3.24. Desviaciones del tipo de cambio en el en el año 2000.**

De los resultados obtenidos en las tablas 3.23. y 3.24., se divide la suma del cuadrado de las desviaciones entre $n - 1$ y así se obtiene la varianza muestral del tipo de cambio para los años de 1995 y 2000.

Para el año de 1995 $$S^2 = \frac{\Sigma(X_i - \overline{X})^2}{n-1} = \frac{5.1584}{11} = 0.4689$$ pesos al cuadrado

Para el año 2000 $$S^2 = \frac{\Sigma(X_i - \overline{X})^2}{n-1} = \frac{0.2023}{11} = 0.0183$$ pesos al cuadrado

Si bien los pesos al cuadrado continúan siendo una idea abstracta, ambas varianzas tienen sentido lógico cuando son comparadas entre sí, pues se encuentran expresadas en la misma unidad de medida. En este caso, el tipo de cambio en el año de 1995 tiene una mayor dispersión que el observado en el año 2000, tal como lo señalan ambas varianzas y tal como lo apreciamos de manera visual en el inciso anterior.

Este contraste se debe a la diferencia en los escenarios macroeconómicos que se vivieron durante esos años. Al ser mayor la varianza del año 1995, se refleja una gran volatilidad y nerviosismo en el mercado cambiario producido por una fuerte crisis económica que se vivía en ese año. En el año 2000 podemos observar que el peso mexicano gozó de una gran fortaleza, pues su cotización se mantuvo muy estable en el transcurso de los 12 meses, incluso en el mes de junio, cuando se presentaba la recta final de un proceso electoral en el país.

### b) La varianza para datos agrupados

En el caso de datos agrupados, para encontrar la varianza es necesario conocer el punto medio de cada clase. El método se basa en la suposición de que el punto medio de cada clase es aproximadamente igual a la media aritmética de las medidas contenidas en un intervalo. El punto medio de la clase *j* se denota por $m_j$.

**i) La *varianza poblacional* para datos agrupados se define como:**

$$\sigma^2 = \frac{\Sigma[(m_j - \mu)^2 f_j]}{N}$$

Donde:
$\sigma^2$ = Varianza de la población.
$m_j$ = Punto medio de clase.
$\mu$ = Media de la población.
$N$ = Tamaño de la población.
$f$ = Frecuencia de la clase.

**ii) La fórmula para calcular la *varianza muestral* es:**

$$S^2 = \frac{\Sigma[(m_j - \overline{X})^2 f_j]}{n-1}$$

Donde:
$S^2$ = Varianza de la muestra.
$m_j$ = Punto medio de clase.

$\overline{X}$ = Media de la muestra.
$n$ = Tamaño de la muestra.
$f$ = Frecuencia de la clase.

Para obtener la varianza para datos agrupados, sea muestral o poblacional, se tienen que realizar los siguientes pasos:

1. Se obtiene la media muestral o poblacional para datos agrupados, según corresponda. Por ejemplo, si se pretende obtener la varianza muestral, entonces procedemos a encontrar la media a través de la siguiente fórmula:

$$\overline{X} = \frac{\sum m_j f}{\sum f}$$

2. Se encuentran los puntos medios para cada una de las clases $m_1, m_2, \ldots, m_n$ y a cada uno se resta la media muestral o poblacional según corresponda. Por ejemplo, para el caso de la varianza muestral se encontrarían las siguientes desviaciones:

$$(m_1 - \overline{X}),\ (m_2 - \overline{X}), \ldots,\ (m_n - \overline{X})$$

3. Se eleva al cuadrado cada una de las desviaciones de los puntos medios de clases respecto a la media. Por ejemplo, en caso de una población:

$$(X_1 - \mu)^2,\ (X_2 - \mu)^2, \ldots,\ (X_n - \mu)^2$$

4. Cada uno de los cuadrados se multiplica por su respectiva frecuencia de clase. Por ejemplo, en el caso de una población:

$$(X_1 - \mu)^2 f_1,\ (X_2 - \mu)^2 f_2, \ldots,\ (X_n - \mu)^2 f_n$$

5. Se suma cada uno de estos resultados y se divide, en el caso de la varianza poblacional, entre el número total de datos de la población ($N$), y en el caso de una muestra entre el $n$ –1.

## Ejemplo 24

Una gran empresa de ventas por teléfono quiere conocer la variación existente en las ventas realizadas (en miles de pesos) por sus operadores. Para esto realiza una muestra de 25 operadores telefónicos, obteniendo los resultados de la siguiente tabla. Calcula la varianza muestral.

| **Ventas (miles $)** | $f$ |
|---|---|
| 5.00 – 8.99 | 3 |
| 9.00 – 12.99 | 5 |
| 13.00 – 16.99 | 7 |
| 17.00 – 20.99 | 6 |
| 21.00 – 24.99 | 3 |
| 25.00 – 28.99 | 1 |
| $\Sigma$ | 25 |

**Tabla 3.25 Distribución de las ventas por teléfono.**

Las clases denotan las ventas realizadas en miles de pesos y la frecuencia del número de operadores telefónicos.

| Ventas (miles de $) | $F$ | $m_j$ | $(m_j \cdot f)$ | $(m_j - \overline{X})$ | $(m_j - \overline{X})^2$ | $[(m_j - \overline{X})^2]\, f$ |
|---|---|---|---|---|---|---|
| 5.00 – 8.99 | 3 | 6.995 | 20.985 | –8.64 | 74.6496 | 223.9488 |
| 9.00 – 12.99 | 5 | 10.995 | 54.975 | –4.64 | 21.5296 | 107.648 |
| 13.00 – 16.99 | 7 | 14.995 | 104.965 | –0.64 | 0.4096 | 2.8672 |
| 17.00 – 20.99 | 6 | 18.995 | 113.97 | 3.36 | 11.2896 | 67.7376 |
| 21.00 – 24.99 | 3 | 22.995 | 68.985 | 7.36 | 54.1696 | 162.5088 |
| 25.00 – 28.99 | 1 | 26.995 | 26.995 | 11.36 | 129.0496 | 129.0496 |
| Σ | 25 | | 390.875 | | | 693.76 |

**Tabla 3.26 Distribución de las ventas por teléfono.**

Para obtener la varianza, en primer lugar se debe calcular la media muestral para datos agrupados, encontrando el punto medio de clase, multiplicarlo por su frecuencia de la clase correspondiente, y sus resultados se suman para obtener la media, tal y como se muestra a continuación:

$$\overline{X} = \frac{\Sigma m_j \cdot f}{n} = \frac{390.875}{25} = 15.635$$

Se obtiene la varianza restándole a cada punto medio de clase la media muestral, elevando cada una de estas diferencias al cuadrado y multiplicando cada diferencia cuadrática por la frecuencia respectiva de clase de la manera siguiente:

$$S^2 = \frac{\Sigma[(m_j - \overline{X})^2 f_j]}{n-1} = \frac{693.76}{(25-1)} = \frac{693.76}{24} = 28.90666667 \quad \text{pesos al cuadrado}$$

La varianza obtenida señala que la dispersión existente entre las ventas entre

$(n - 1)$ es de 28.90666667 miles de pesos al cuadrado.

### Ventajas y desventajas de la varianza

La varianza mide la variabilidad tomando en cuenta la dispersión que los valores de los datos tienen respecto a su media. Es decir, aquellos conjuntos que tengan valores más alejados de la media, sea muestral o poblacional, tendrán una mayor varianza, mientras que aquellos conjuntos con valores más cercanos a la media mostrarán una mayor uniformidad al contar con una varianza menor.

La varianza únicamente adquiere valores mayores o iguales a cero, nunca valores negativos, y se utiliza para comparar la dispersión de dos o más conjuntos de datos que se encuentren expresados en la misma unidad de medida; por ejemplo, para observar la variación existente entre dos líneas de producción, la tasa de interés de dos instrumentos financieros, las ventas de productos expresados en la misma moneda, etcétera.

La principal desventaja de la varianza es que su resultado se expresa en unidades al cuadrado, resultando darle una interpretación lógica. Además, la varianza no puede comparar la dispersión de dos conjuntos de datos expresados en diferentes unidades de medida; por ejemplo, chamarras con coches, diferentes divisas, el IPC de la Bolsa Mexicana de Valores con el índice Dow Jones de la Bolsa de Nueva York, etcétera.

## Ejercicio 7

**1.** Grandes varianzas implican:

a) Que los datos no varían.
b) Que hay gran variación en los datos.
c) Que hay poca variación en los datos.
d) Que las medias son desproporcionadas.

**2.** Uno de los inconvenientes de utilizar la varianza como medida de dispersión es que:

a) La varianza muestral es sesgada y la poblacional no.
b) La varianza se ve afectada por el tipo de dato que estamos utilizando.
c) Las varianzas poblacionales y muestrales son distintas.
d) Los resultados se expresan en unidades al cuadrado.

**3.** Si tenemos cinco datos cuyos valores son las constantes: 2, 2, 2, 2 y 2; entonces la varianza es:

a) Cualquier valor.
b) Un valor mayor o igual a cero.
c) Un valor igual a cero.
d) Tanto valores positivos como negativos, excepto el cero.

**4.** Una serie compuesta con los siguientes datos: 0, 1, 1, 3 y 5, su varianza será:

a) 2
b) 4
c) 0
d) 1

**5.** Con los siguientes datos de crédito y cobranza, calcula la varianza para datos no agrupados, con el fin de determinar la variabilidad de los datos de los próximos cobros (en pesos).

| | | | | |
|---|---|---|---|---|
| 10 000 | 12 000 | 15 000 | 16 000 | 15 000 |
| 9 000 | 13 500 | 12 700 | 9 700 | 18 000 |
| 13 200 | 12 600 | 14 000 | 18 700 | 16 500 |

**6.** Un despacho de consultoría en cuestiones de mercado hace una encuesta de los ingresos anuales (en miles de pesos) de 300 familias para clasificarlas por nivel de ingreso y con esto establecer qué artículos son susceptibles de promocionarse y posicionar en el mercado, considerando las variaciones existentes. Con la información siguiente calcula la varianza:

| Ingreso (miles de $) | $f$ |
|---|---|
| 1.50 – 2.999 | 25 |
| 3.00 – 4.999 | 31 |
| 5.00 – 6.999 | 42 |
| 7.00 – 8.999 | 45 |
| 9.00 – 10.999 | 52 |
| 11.00 – 12.999 | 42 |
| 13.00 – 14.999 | 35 |
| 15.00 – 16.999 | 28 |
| $\Sigma$ | 300 |

**Distribución de salarios.**

### 3.4.3. Desviación estándar

Al igual que la varianza, la desviación estándar es una medida de variabilidad que también toma en cuenta la dispersión de los valores de los datos respecto a su media. Sin embargo, su significado es más valioso que el de la varianza, pues su resultado se encuentra expresado en las mismas unidades de la variable que se examina y no en valores elevados al cuadrado como lo hace la varianza.

La desviación estándar se representa mediante la letra griega $\sigma$ para el caso de una población, o por *S* en el caso de una muestra. Se obtiene sacando la raíz cuadrada al resultado de la varianza, no importa si ésta se trata de una varianza para datos no agrupados o para datos agrupados, o provenientes de una muestra o de una población. Al proporcionar sus resultados en unidades no cuadradas, la desviación estándar es muy fácil de interpretar y su resultado tiene mayor significado en el análisis de un fenómeno.

Las fórmulas para la *desviación estándar para datos no agrupados* son:

$$\sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{N}} \qquad \text{o} \qquad S = \sqrt{\frac{\Sigma(X-\overline{X})^2}{n-1}}$$

Cuando se trabaja con *datos agrupados*, la desviación estándar también se calcula sacando la raíz cuadrada, pero empleando las fórmulas respectivas de la varianza para datos agrupados:

$$\sigma = \sqrt{\frac{\Sigma[(m_j-\mu)^2 f_j]}{N}} \qquad \text{o} \qquad S = \sqrt{\frac{\Sigma[(m_j-\overline{X})^2 f_j]}{n-1}}$$

Tanto en datos no agrupados como en datos agrupados, $\sigma$ indica la desviación estándar para una población, mientras que la *S* representa la desviación estándar para una muestra.

## Ejemplo 25

Una casa de bolsa desea realizar un comparativo entre los rendimientos anuales y los riesgos de dos instrumentos financieros que han estado operando durante los últimos siete años. Sus rendimientos anuales, expresados en porcentajes, son los siguientes:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Instrumento A: | 4.0% | 14.3% | 19.5% | 14.7% | 26.5% | 37.2% | 23.8% |
| Instrumento B: | 6.5% | 4.4% | 4.8% | 6.9% | 8.5% | 5.8% | 5.1% |

Obtener la media y la desviación estándar de los rendimientos observados por los dos instrumentos financieros.

En primer lugar se obtiene el rendimiento promedio por instrumento:

$$\mu_A = \frac{\sum X}{N} = \frac{(4+14.3+19.5+14.7+26.5+37.2+23.8)}{7} = \frac{140}{7} = 20\%$$

$$\mu_B = \frac{\sum X}{N} = \frac{(6.5+4.4+4.8+6.9+8.5+5.8+5.1)}{7} = \frac{42}{7} = 6\%$$

Como puede observarse, el instrumento que presenta el mayor rendimiento promedio es *A* con 20%, mientras que el instrumento *B* tiene un rendimiento promedio de 6%. En ese sentido, resultaría más atractivo invertir en el fondo A que en el fondo *B*.

Para medir el riesgo de cada uno de los fondos encontramos sus desviaciones estándar; para esto, primero se deben obtener las varianzas poblacionales y posteriormente se les saca la raíz cuadrada:

**Acción A**

$$\Sigma(X-\mu)^2_A = (4-20)^2 + (14.3-20)^2 + \ldots + (23.8-20)^2 = 669.36$$

$$V(X)_A = \sigma^2 = \frac{\Sigma(X-\mu)^2}{N} = \frac{669.36}{7} = 95.62285714$$

$$\sigma_A = \sqrt{\frac{\Sigma(X-\mu)^2}{N}} = \sqrt{95.62285714} = 9.778694041$$

**Acción B**

$$\Sigma(X-\mu)^2_B = (6.5-6)^2 + (4.4-6)^2 + \ldots + (5.1-6)^2 = 12.16$$

$$V(X)_B = \sigma^2 = \frac{\Sigma(X-\mu)^2}{N} = \frac{12.16}{7} = 1.737142857$$

$$\sigma_B = \sqrt{\frac{\Sigma(X-\mu)^2}{N}} = \sqrt{1.737142857} = 1.318007154$$

Puede observarse que el instrumento *A* tiene una variabilidad de 9.778694041%, mientras que el instrumento *B* tuvo una variabilidad de 1.318007154%. Esto indica que los rendimientos del instrumento A tienen una mayor dispersión que los rendimientos del instrumento *B*.

En el contexto de este ejemplo puede pensarse en la desviación estándar como una medida de la incertidumbre o riesgo de la rentabilidad de una inversión. Es decir, la rentabilidad promedio fue mayor para el instrumento A, pero su riesgo en términos de la desviación estándar de la rentabilidad también fue mayor.

Por otra parte, para obtener la desviación estándar cuando se trabaja con datos agrupados se utiliza la misma metodología que en el caso de los datos no agrupados. En primer lugar se encuentra la varianza a través de su respectiva fórmula y posteriormente se le saca la raíz cuadrada.

## Ejemplo 26

Con los datos del ejemplo 5 calcula la desviación estándar.

$$S = \sqrt{\frac{\Sigma(m_j - \overline{X})^2 f_j}{n-1}} = \sqrt{\frac{693.76}{24}} = \sqrt{28.90666667} = 5.376492041$$

Con este resultado se deduce que la variación promedio que existe en las ventas realizadas por teléfono es de 5.38 miles de pesos. Esto puede ayudar a la empresa a analizar las ventas que realizan los operadores de una manera más sencilla que utilizando ventas al cuadrado.

### Ventajas y desventajas de la desviación estándar

La principal ventaja de la desviación estándar es que indica la manera en que se dispersan los datos respecto a la media en las mismas unidades de la variable que se examina y no en valores elevados al cuadrado. Al igual que la varianza, la desviación estándar únicamente adquiere valores mayores o iguales a cero, nunca valores negativos.

Es utilizada para comparar la dispersión entre distintos conjuntos de datos. Aquellos conjuntos que tengan valores más alejados de la media tendrán una mayor desviación estándar, mientras que aquellos conjuntos con valores más cercanos a la media mostrarán una menor desviación estándar.

Al igual que la varianza, una desventaja de la desviación estándar es que tampoco puede comparar la dispersión de dos conjuntos de datos que se expresan en diferentes unidades de medida.

## Ejercicio 8

1. Con los datos de crédito y cobranza que se presentan a continuación, calcula la desviación estándar de los próximos cobros.

| | | | | |
|---|---|---|---|---|
| 10 000 | 12 000 | 15 000 | 16 000 | 15 000 |
| 9 000 | 13 500 | 12 700 | 9 700 | 18 000 |
| 13 200 | 12 600 | 14 000 | 18 700 | 16 500 |

2. Con los siguientes datos de los ingresos anuales (en miles de pesos) de 300 familias, calcula la desviación estándar.

| Ingreso (miles de $) | $f$ |
|---|---|
| 1.50 – 2.999 | 25 |
| 3.00 – 4.999 | 31 |
| 5.00 – 6.999 | 42 |
| 7.00 – 8.999 | 45 |
| 9.00 – 10.999 | 52 |
| 11.00 – 12.999 | 42 |
| 13.00 – 14.999 | 35 |
| 15.00 – 16.999 | 28 |
| Σ | 300 |

**Distribución de salarios.**

EJERCICIOS

## 3.6. Coeficiente de variación

Es una medida de dispersión que señala qué tan grande es la magnitud de la desviación estándar respecto a la media del conjunto de datos que se examina. A diferencia de otras medidas de variabilidad, el coeficiente de variación mide la dispersión en términos de porcentaje y no en unidades de medida. De esta manera, este coeficiente se utiliza para comparar la dispersión entre dos conjuntos de datos expresados en diferentes unidades de medidas.

Por ejemplo, si los analistas de un despacho de bienes raíces están interesados en determinar si el valor de un avalúo tiene mayor variabilidad que el tamaño del lote, resultaría imposible comparar en forma directa la dispersión mediante el rango, la varianza o la desviación estándar, pues el valor del avalúo se mide en unidades monetarias, por ejemplo en miles de pesos, mientras que el tamaño del lote se mide en metros cuadrados. En este caso, los analistas pueden utilizar el coeficiente de variación, expresado en porcentajes, y así comparar la dispersión de dos variables expresadas en distintas unidades de medida.

El coeficiente de variación se representa mediante la expresión *CV* y se obtiene dividiendo la desviación estándar entre la media, multiplicando este resultado por 100, no importando si se trata de datos no agrupados o de datos agrupados, o que provengan de una muestra o de una población.

El coeficiente de variación se puede calcular mediante la fórmula siguiente:

$$CV = \left(\frac{S}{\overline{X}}\right)100\% \qquad \text{En caso de una muestra}$$

$$CV = \left(\frac{\sigma}{\mu}\right)100\% \qquad \text{En caso de una población}$$

Donde:
$CV$ = Coeficiente de variación.
$S$ = Desviación estándar de la muestra.
$\overline{X}$ = Media de los datos.
$\sigma$ = Desviación estándar de la población.
$\mu$ = Media poblacional.

### Ejemplo 31

Con los datos del ejemplo 25, calcula el coeficiente de variación con el fin de hacer una comparación de los rendimientos de las acciones:

$$CV_A = \left(\frac{S_A}{\mu_A}\right)100\% = \left(\frac{9.778694041}{20}\right)100 = (0.488934702)\ (100) = 48.8934702\%$$

$$CV_B = \left(\frac{S_B}{\mu_B}\right)100\% = \left(\frac{1.318007154}{6}\right)100 = (0.219667859)\ (100) = 21.9667859\%$$

La acción que presenta la menor variabilidad es la *B*, que como ya se había mencionado es la que presenta un menor rendimiento promedio y una menor desviación estándar (menor riesgo), con lo que se concluye que la acción más conveniente para invertir sin incurrir en un gran riesgo es la acción *B*.

## Ejemplo 32

Los analistas de un centro financiero desean comparar el desempeño del tipo de cambio y el porcentaje de la participación extranjera en el mercado accionario de la Bolsa Mexicana de Valores durante el año 2000. Para esto se calcula el coeficiente de variación para cada uno de los mercados.

| Mes | Tipo de cambio en el 2000 |
|---|---|
| Enero | 9.47 |
| Febrero | 9.44 |
| Marzo | 9.29 |
| Abril | 9.37 |
| Mayo | 9.50 |
| Junio | 9.79 |
| Julio | 9.46 |
| Agosto | 9.28 |
| Septiembre | 9.33 |
| Octubre | 9.51 |
| Noviembre | 9.51 |
| Diciembre | 9.44 |

*Fuente*: Banco de México: www.banxico.org.mx

**Tabla 3.28. Tipo de cambio mensual peso-dólar en el año 2000.**

| Mes | Inversión extranjera en el 2000 |
|---|---|
| Enero | 44.01 |
| Febrero | 46.58 |
| Marzo | 44.78 |
| Abril | 47.25 |
| Mayo | 45.07 |
| Junio | 46.69 |
| Julio | 44.07 |
| Agosto | 44.96 |
| Septiembre | 44.72 |
| Octubre | 44.62 |
| Noviembre | 43.03 |
| Diciembre | 41.31 |

*Fuente*: Banco de México: www.banxico.org.mx

**Tabla 3.29. Participación extranjera en la bolsa en el año 2000.**

Las variables que se desean comparar vienen expresadas en diferentes unidades de medida; el tipo de cambio se expresa en pesos mientras que la inversión extranjera se representa en proporciones. Por tal razón, se calculan los coeficientes de variación para cada una de las variables y así se compara la variabilidad de ambos mercados. Para ello tomamos las medias y las desviaciones estándar de los ejemplos 4 y 11.

$$CV_{\textit{Tipo de cambio}} = \left(\frac{S}{\overline{X}}\right)100\% = \left(\frac{0.1352}{9.44}\right)100 = (0.0143)(100) = 1.43\%$$

$$CV_{\textit{Inv. extranjera}} = \left(\frac{S}{\overline{X}}\right)100\% = \left(\frac{1.6328}{44.7575}\right)100 = (0.0364)\ (100) = 3.64\%$$

Los analistas de este centro financiero pueden concluir que el mercado cambiario durante el año 2000 tuvo mayor estabilidad que la participación extranjera en el mercado accionario, pues el coeficiente de variación del primero fue de 1.43%, mientras que el del segundo fue de 3.64%. De esta forma, los analistas comparan la variación de dos mercados que tienen distintas unidades de medición.

### Ventajas y desventajas del coeficiente de variación

El coeficiente de variación es útil cuando pretende comparar la variabilidad de dos o más conjuntos de datos expresados en diferentes unidades de medición, pues el resultado será señalado en porcentajes.

La única desventaja que adolece el coeficiente de variación es cuando se tienen que comparar dos conjuntos de datos donde uno tiene una media con valores negativos y el otro tiene una media positiva. Para el primer conjunto, el coeficiente de variación será negativo; mientras que para el segundo, el coeficiente de variación será positivo, haciendo difícil la comparación entre ambos. Esto puede solucionarse tomando los valores absolutos del resultado que se obtenga en ambos coeficientes.

EJERCICIOS

## Ejercicio 10

1. El coeficiente de variación es una medida de dispersión que expresa sus resultados como:

   a) Unidades métricas.
   b) Desviaciones estándar.
   c) Porcentajes.
   d) Desviaciones respecto a la media.

2. El coeficiente de variación tiene la ventaja de:

   a) Comparar conjuntos de datos expresados en diferentes unidades de medición.
   b) Comparar conjuntos de datos expresados en diferentes unidades cuadradas.
   c) Comparar conjuntos de datos expresados en desviaciones.
   d) Comparar conjuntos de datos expresados en porcentajes.

3. Si tenemos tres diferentes acciones *A*, *B* y *C*, y el coeficiente de variación de sus precios son $CV_A$=13%, $CV_B$=15% y $CV_C$=7%, entonces:

   a) La acción *A* es la de mayor variabilidad y la acción *B* es la de menor variabilidad.
   b) La acción *B* es la de mayor variabilidad y la acción *A* es la de menor variabilidad.
   c) La acción *C* es la de mayor variabilidad y la acción *A* es la de menor variabilidad.
   d) La acción *B* es la de mayor variabilidad y la acción *C* es la de menor variabilidad.

4. Una casa de cambio desea conocer la variación existente entre el valor de dos monedas (pesos/dólar y pesos/libra ) en las transacciones de 10 días para determinar qué moneda es la que representa una mayor estabilidad. Con los siguientes datos, calcula el coeficiente de variación.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Dólar | 150 | 125 | 120 | 200 | 250 | 175 | 200 | 250 | 180 | 140 |
| Libra | 200 | 275 | 180 | 195 | 280 | 250 | 240 | 200 | 300 | 290 |

## 3.7. Índice de asimetría y kurtosis

Cuando se estudiaron las medidas de tendencia central se analizó la relación que existe entre la media, la mediana y la moda, señalando que esta relación depende de la forma en que se distribuyen los datos. Se dijo que el posicionamiento de las medidas de tendencia central estaba en función del tipo de sesgo que se observaba en la distribución de frecuencias.

Por ejemplo, cuando se tiene sesgo positivo o derecho, la media es mayor que la mediana y que la moda, es decir, la media se encuentra más a la derecha de la moda, dando así una distribución con una cola alargada que se extiende hacia el lado derecho. En el caso contrario, cuando la media es menor que la moda, la distribución tiene un sesgo negativo o izquierdo, dando así una distribución con una cola alargada que se extiende hacia el lado izquierdo. También se señaló que cuando las tres medidas de tendencia central coinciden con el mismo valor, la distribución de frecuencias era acampanada y simétrica.

En esta sección se analizará de manera más formal la asimetría de una distribución de frecuencias a través del índice de asimetría. Este aspecto es sumamente importante en el análisis de datos, pues dependiendo del tipo y de la magnitud del sesgo que se observe en una distribución de frecuencias se conocerá con más detalle la forma en que se dispersan los datos de una serie, detectando con mayor facilidad la presencia de datos atípicos.

El índice de asimetría es una medida de dispersión mediante la cual se conozce el tipo y la magnitud de sesgo en una distribución de frecuencias. Se representa mediante la expresión $\alpha_3$.

Para el caso de **datos no agrupados**, las fórmulas del índice de asimetría son:

$$\alpha_{3\ Poblacional} = \frac{\left(\frac{\Sigma(X_j-\mu)^3}{N}\right)}{(\sigma)^3} \qquad \alpha_{3\ Muestral} = \frac{\left(\frac{\Sigma(X_j-\overline{X})^3}{n-1}\right)}{(S)^3}$$

Para el caso de **datos agrupados**, las fórmulas del índice de asimetría son:

$$\alpha_{3\ Poblacional} = \frac{\left(\frac{\Sigma[(m_j-\mu)^3]\,f}{N}\right)}{(\sigma)^3} \qquad \alpha_{3\ Muestral} = \frac{\left(\frac{[\Sigma(\ m_j-\overline{X})^3]\,f}{n-1}\right)}{(S)^3}$$

Donde:

| | | | |
|---|---|---|---|
| $\alpha_3$ = | Coeficiente de asimetría. | $f$ = | Frecuencia de clase. |
| $m_j$ = | Punto medio de clase. | $\sigma$ = | Desviación estándar de la población. |
| $\mu$ = | Media poblacional. | $S$ = | Desviación estándar de la muestra. |
| $\overline{X}$ = | Media muestral | $N$ = | Tamaño de la población. |
| $n$ = | Tamaño de la muestra. | | |

La interpretación del índice de asimetría se define según el caso que se trate:

1. Si el índice de asimetría es igual o cercano a cero ($\alpha_3$ = 0), la distribución es simétrica o insesgada; es decir, si la distribución es dividida exactamente a la mitad, y la figura de la primera mitad es idéntica a la otra mitad.

2. Si el índice de asimetría es mayor que cero ($\alpha_3$ > 0), la distribución es asimétricamente positiva o sesgada hacia la derecha, es decir, si la distribución es dividida exactamente a la mitad, se observará que la cola de la figura se extiende hacia la derecha de la distribución, mientras que su cima o valor más alto de la distribución se ubicará en la parte izquierda.

3. Si el índice de asimetría es menor que cero ($\alpha_3 < 0$), la distribución es asimétrica negativa o sesgada hacia la izquierda. Es decir, si la distribución es dividida exactamente a la mitad, se observará que la cola de la figura se encuentra hacia la izquierda de la distribución, mientras que su cima o valor más alto de la distribución se ubicará en la parte derecha.

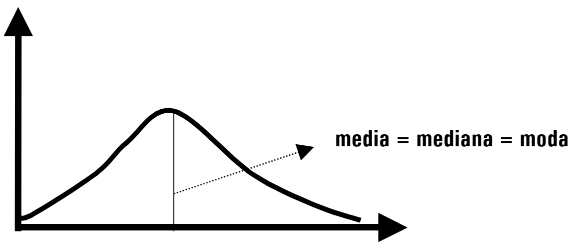


**Figura 3.12. Distribución simétrica $\alpha_3 = 0$.**

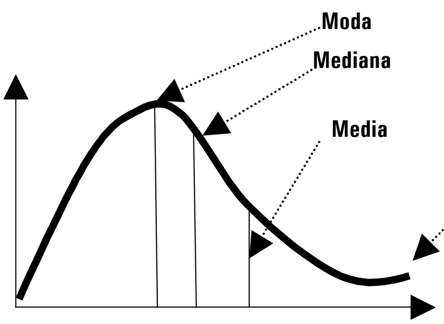


**Figura 3.13. Distribución sesgada a la derecha $\alpha_3 > 0$.**

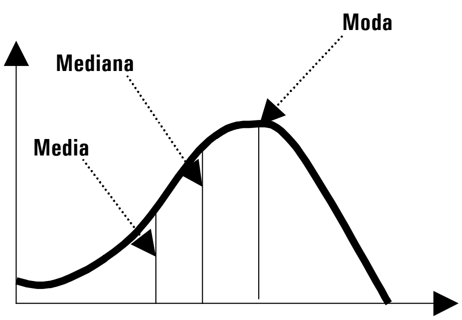


**Figura 3.14. Distribución sesgada a la izquierda $\alpha_3 < 0$.**

## Ejemplo 33

Calcula el índice de asimetría para determinar qué tipo de sesgo tiene la siguiente serie de datos de una población: 1, 1, 2, 2, 2, 3, 3, 4, 5 y 6.

Para obtener el índice de asimetría, primero debemos encontrar cada uno de los elementos de su fórmula.

Se encuentra la media poblacional:

$$\mu = \frac{\Sigma X}{N} = \frac{(1+1+2+\ldots+6)}{10} = \frac{29}{10} = 2.9$$

Se encuentra la varianza poblacional:

$$V(X) = \sigma^2 = \frac{\Sigma(X-\mu)^2}{N} = \frac{24.9}{10} = 2.49$$

Se encuentra la desviación estándar:

$$\sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{N}} = 1.57$$

Se eleva al cubo la desviación estándar:

$$\sigma^3 = 3.86$$

Se obtiene la suma del cubo de las desviaciones con respecto a la media:

$$\Sigma(X - \mu)^3 = 24.9$$

Finalmente, se sustituyen estos resultados en la fórmula del índice de asimetría:

$$\alpha_3 = \frac{\left(\frac{\Sigma(X_j-\mu)^3}{n}\right)}{(\sigma)^3} = \frac{\left(\frac{24.48}{10}\right)}{3.86} = \frac{2.448}{3.86} = 0.6341$$

Se obtiene un índice de asimetría positivo, por lo que se puede decir que la distribución tiene un pequeño sesgo positivo o derecho. Si se observa la figura de la distribución de frecuencias, se notará que tiene una cola que se alarga hacia el lado derecho de la distribución:

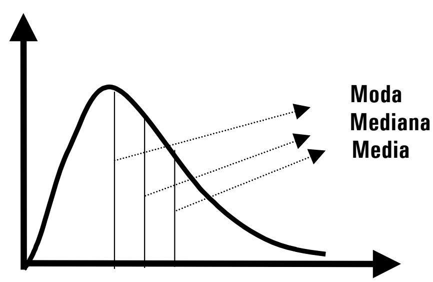


**Figura 3.15. Distribución asimétrica positiva.**

## Ejemplo 34

Con la información del ejemplo 5, calcula el coeficiente de asimetría para saber hacia qué lado se carga la cola de la curva de estos datos.

| **Tiempo de servicio** | $f$ | $m_j$ | $(m_j \cdot f)$ | $(m_j - \overline{X})$ | $(m_j - \overline{X})^3$ | $[(m_j - \overline{X})^3] f$ |
|---|---|---|---|---|---|---|
| 5.00 – 8.99 | 3 | 6.995 | 20.985 | –8.64 | –644.972544 | –1 934.917632 |
| 9.00 – 12.99 | 5 | 10.995 | 54.975 | –4.64 | –99.897344 | –499.48672 |
| 13.00 – 16.99 | 7 | 14.995 | 104.965 | –0.64 | –0.262144 | –1.835008 |
| 17.00 – 20.99 | 6 | 18.995 | 113.97 | 3.36 | 37.933056 | 227.598336 |
| 21.00 – 24.99 | 3 | 22.995 | 68.985 | 7.36 | 398.688256 | 1 196.064768 |
| 25.00 – 28.99 | 1 | 26.995 | 26.995 | 11.36 | 1 466.003456 | 1 466.003456 |
| Σ | 25 | | 390.875 | | | 453.4272 |

**Tabla 3.30. Distribución de las ventas por teléfono.**

Los datos obtenidos son:

$$\overline{X} = \frac{\Sigma(m_j \cdot f)}{n} = \frac{390.875}{25} = 15.635$$

$$S^2 = \frac{\Sigma(m_j - \overline{X})^2 f}{n-1} = \frac{693.76}{(25-1)} = \frac{693.76}{24} = 28.90666667$$

$$S = \sqrt{\frac{\Sigma(m_j - \overline{X})^2 f}{n-1}} = \sqrt{28.90666667} = 5.376492041$$

El numerador de la fórmula empleada para calcular el coeficiente se denota por:

$$\frac{\Sigma(m_j - \overline{X})^3 f}{n-1} = \frac{453.4272}{24} = 18.8928$$

Con los datos anteriores, el coeficiente de asimetría es:

$$\alpha_3 = \frac{\left(\frac{\Sigma(m_j - \overline{X})^3 f}{n-1}\right)}{(S)^3} = \frac{18.8928}{(5.376492041)^3} = \frac{18.8928}{155.4164633} = 0.121562411$$

Con el resultado se puede observar que el coeficiente es cercano a cero, así la distribución se caracteriza por ser insesgada, es decir, que la curva tiene una forma simétrica tal que las colas tienden a ser iguales.

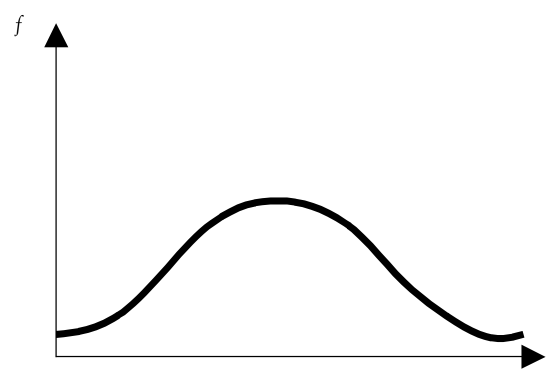


**Figura 3.16. Distribución insesgada de las ventas por teléfono.**

EJERCICIOS

## Ejercicio 11

1. En una distribución con $\alpha_3 = 0$:

   a) Media, mediana y moda son diferentes.
   b) Media, mediana y moda coinciden en el mismo valor.
   c) La media es mayor que la mediana y la moda.
   d) La moda es mayor que la media y la mediana.

2. En una distribución con $\alpha_3 > 0$:

   a) La mediana es mayor que la media y la moda.
   b) Media, mediana y moda coinciden en el mismo valor.
   c) La media es mayor que la mediana y la moda.
   d) La moda es mayor que la mediana y la moda.

3. En una distribución $\alpha_3 < 0$:

   a) La mediana es mayor que la media y la moda.
   b) Media, mediana y moda coinciden en el mismo valor.
   c) La media es mayor que la mediana y la moda.
   d) La moda es mayor que la mediana y la moda.

4. Encuentra el índice de asimetría para una serie conformada por los siguientes datos provenientes de una muestra: 0, 1, 1, 3 y 5, y señala qué tipo de distribución es.

5. Con los datos de los ingresos anuales (en miles) de 300 familias que se presentan a continuación, calcula el coeficiente de asimetría para saber cómo es el sesgo de la distribución.

| Ingreso (miles de $) | $f$ |
|---|---|
| 1.50 – 2.999 | 25 |
| 3.00 – 4.999 | 31 |
| 5.00 – 6.999 | 42 |
| 7.00 – 8.999 | 45 |
| 9.00 – 10.999 | 52 |
| 11.00 – 12.999 | 42 |
| 13.00 – 14.999 | 35 |
| 15.00 – 16.999 | 28 |
| $\Sigma$ | 300 |

**Distribución de salarios.**

### 3.7.1. Kurtosis

El índice de kurtosis es una medida de dispersión mediante la cual se conoce qué tan concentrados o qué tan dispersos se encuentran los datos alrededor de la media. Su resultado representa el grado de apuntamiento de una distribución, es decir, qué tan puntiaguda o qué tan aplanada es la curva de una distribución. Cuando es muy puntiaguda se dice que los datos se encuentran muy concentrados alrededor de la media, mientras que si es muy chata o aplanada, se dice que existe una gran dispersión de los datos alrededor de la media.

Para encontrar el índice de kurtosis, las fórmulas dependen de la información con la que se trabaje y de la manera en que se encuentren organizados los datos, ya sea que se trate de una muestra o de una población, o que los datos se encuentren no agrupados o agrupados. Se representa mediante la expresión $\alpha_4$.

Para el caso de **datos no agrupados**, la kurtosis poblacional y muestral se expresan mediante las siguientes fórmulas:

$$\alpha_{4\,Poblacional} = \frac{\left(\frac{\Sigma(X_j - \mu)^4}{N}\right)}{(\sigma)^4} \qquad \alpha_{4\,Muestral} = \frac{\left(\frac{\Sigma(X_j - \overline{X})^4}{n-1}\right)}{(S)^4}$$

Para el caso de **datos agrupados**, la kurtosis poblacional y muestral se obtienen utilizando las siguientes fórmulas:

$$\alpha_{4\,Poblacional} = \frac{\left(\frac{\Sigma(m_j - \mu)^4 f}{N}\right)}{(\sigma)^4} \qquad \alpha_{4\,Muestral} = \frac{\left(\frac{\Sigma(m_j - \overline{X})^4 f}{n-1}\right)}{(S)^4}$$

Donde:

$\alpha_4$ = Coeficiente de kurtosis.
$m_j$ = Punto medio de clase.
$\overline{X}$ = Media de la muestra.
$f$ = frecuencia de la clase.
$\mu$ = Media poblacional.
$n$ = Tamaño de la muestra.
$N$ = Tamaño de la población.
$\sigma$ = Desviación estándar poblacional.
$S$ = Desviación estándar de la muestra.

La interpretación del índice de kurtosis se define según el caso que se trate:

1. Si el índice de kurtosis es igual a tres ($\alpha_4$ = 3), la distribución no es ni tan puntiaguda ni tan plana. A este tipo de distribución se le conoce como **distribución mesocúrtica**.

2. Si el índice de kurtosis es mayor a tres ($\alpha_4$ > 3), la distribución es muy puntiaguda, es decir, los datos se encuentran muy concentrados alrededor de la media. A este tipo de distribución se le conoce como **distribución leptocúrtica**.

3. Si el índice de kurtosis es menor a tres ($\alpha_4$ < 3), la distribución es muy plana, es decir, los datos se encuentran muy dispersos del valor de la media. A este tipo de distribución se le conoce como **distribución platicúrtica**.

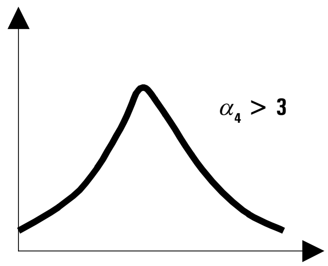


**Figura 3.17. Distribución leptocúrtica.**

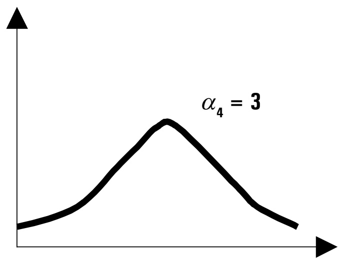


**Figura 3.18. Distribución mesocúrtica.**

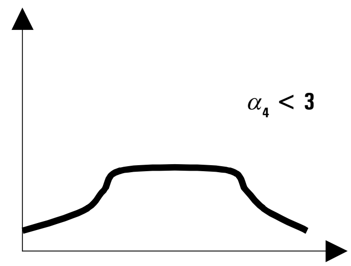


**Figura 3.19. Distribución platicúrtica.**

## Ejemplo 35

Empleando los datos del ejemplo 5, calcula el coeficiente de kurtosis para saber cómo es la forma de la curva de estos datos.

| Tiempo de servicio | $f$ | $m_j$ | $(m_j \cdot f)$ | $(m_j - \overline{X})$ | $(m_j - \overline{X})^4$ | $[(m_j - \overline{X})^4]\, f$ |
|---|---|---|---|---|---|---|
| 5.00 – 8.99 | 3 | 6.995 | 20.985 | –8.64 | 5 572.56278 | 1 6717.68834 |
| 9.00 – 12.99 | 5 | 10.995 | 54.975 | –4.64 | 463.5236762 | 2 317.618381 |
| 13.00 – 16.99 | 7 | 14.995 | 104.965 | –0.64 | 0.16777216 | 1.17440512 |
| 17.00 – 20.99 | 6 | 18.995 | 113.97 | 3.36 | 127.4550682 | 764.7304092 |
| 21.00 – 24.99 | 3 | 22.995 | 68.985 | 7.36 | 2 934.345564 | 8 803.036692 |
| 25.00 – 28.99 | 1 | 26.995 | 26.995 | 11.36 | 16 653.79926 | 16 653.79926 |
| Σ | 25 | | 390.875 | | | 45 258.04749 |

**Tabla 3.31. Distribución de las ventas por teléfono.**

Obtenemos la información necesaria para encontrar la kurtosis:

$$\overline{X} = \frac{\Sigma(m_j \cdot f)}{n} = \frac{390.875}{25} = 15.635$$

$$S^2 = \frac{\Sigma(m_j - \overline{X})^2 f}{n-1} = \frac{693.76}{(25-1)} = \frac{693.76}{24} = 28.90666667$$

$$S = \sqrt{\frac{\Sigma(m_j - \overline{X})^2 f}{n-1}} = \sqrt{28.90666667} = 5.376492041$$

El numerador de la fórmula empleada para calcular el coeficiente se denota por:

$$\frac{\Sigma(m_j - \overline{X})^4 f}{n-1} = \frac{45258.049}{24} = 1\,885.751979$$

Con los datos anteriores, el coeficiente de kurtosis es:

$$\alpha_4 = \frac{\left(\frac{\Sigma(m_j - \overline{X})^4 f}{n-1}\right)}{(S)^4} = \frac{1\,885.751979}{(5.376492041)^4} = \frac{1\,885.751979}{835.5953778} = 2.25677646$$

Con el resultado se puede observar que el coeficiente es menor a tres, por lo que la distribución se caracteriza por ser platicúrtica, es decir, que la curva tiene una forma tal que su apuntamiento es achatado, tal y como se muestra a continuación:

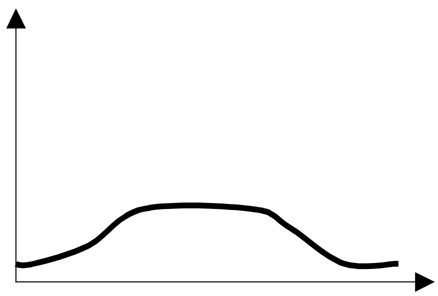

**Figura 3.20. Distribución de las ventas por teléfono.**

EJERCICIOS

## Ejercicio 12

**1.** El índice de kurtosis mide:

a) La simetría de una distribución.
b) Un valor típico o representativo de la distribución.
c) La dispersión existente entre el valor mayor y el menor.
d) El grado de apuntamiento de una distribución.

**2.** Si el índice de kurtosis $\alpha_4$ es igual a tres, entonces:

a) La distribución es asimétrica.
b) La distribución es mesocúrtica.
c) La distribución es leptocúrtica.
d) La distribución es platicúrtica.

**3.** Si el índice de kurtosis $\alpha_4$ es menor a tres, entonces:

a) La distribución es asimétrica.
b) La distribución es mesocúrtica.
c) La distribución es leptocúrtica.
d) La distribución es platicúrtica.

**4.** Si el índice de kurtosis $\alpha_4$ es mayor a tres, entonces:

a) La distribución es asimétrica.
b) La distribución es mesocúrtica.
c) La distribución es leptocúrtica.
d) La distribución es platicúrtica.

**5.** Con los siguientes datos de los ingresos anuales (en miles) de 300 familias, calcula el coeficiente de kurtosis para conocer cómo es la forma de la curva de distribución:

| Ingreso (miles de $) | $f$ |
|---|---|
| 1.50 – 2.999 | 25 |
| 3.00 – 4.999 | 31 |
| 5.00 – 6.999 | 42 |
| 7.00 – 8.999 | 45 |
| 9.00 – 10.999 | 52 |
| 11.00 – 12.999 | 42 |
| 13.00 – 14.999 | 35 |
| 15.00 – 16.999 | 28 |
| $\Sigma$ | 300 |

**Distribución de salarios.**