El tercer principio es una de las razones y objetivos principales de la mejora: reducir la variabilidad hasta lograr niveles de excelencia en calidad, como el nivel de calidad Seis Sigma (vea el capítulo 15). El reto es que en una organización se logre profundizar en la filosofía del pensamiento estadístico para conocer la realidad tal como es (con variación), pero también le permitirá direccionar mejor sus esfuerzos de mejora. En la figura 8.3 se muestra la forma en la que el pensamiento estadístico puede ayudar en los diferentes niveles de una organización.

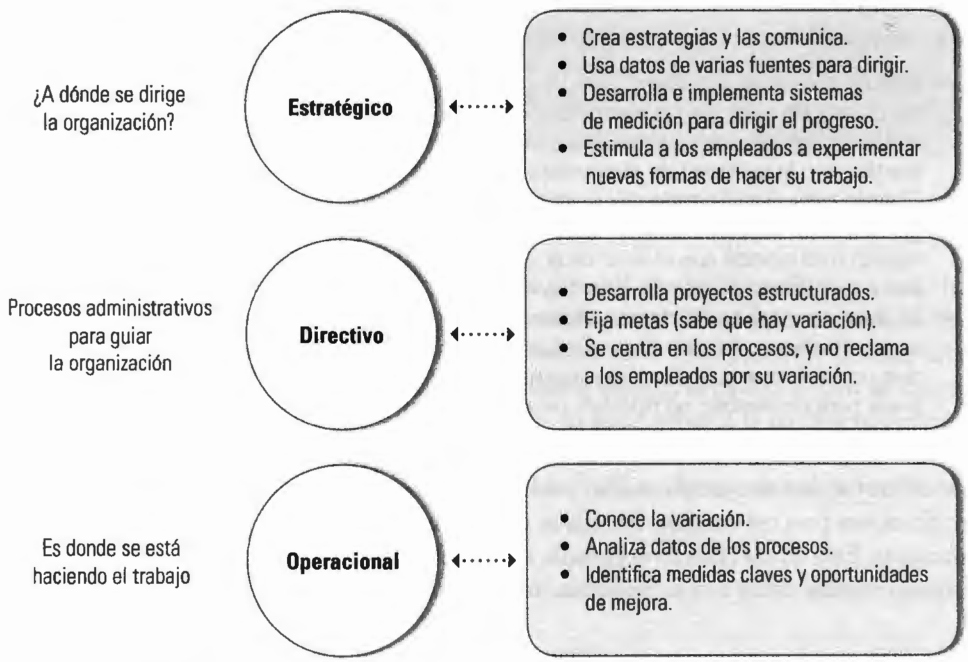


**Figura 8.3** El pensamiento estadístico en los tres niveles de la organización.

## Tipos de variables

En función de los valores que pueden tomar, las variables se clasifican en cualitativas y cuantitativas. Las **variables cualitativas** o **de atributos** asumen valores que representan categorías o atributos de las cosas y que no tienen de base una escala numérica; por ejemplo, tipo de producto, si este funciona o no, etc. Existen varios métodos estadísticos a analizar este tipo de variables (vea, por ejemplo, los del capítulo 10). Mientras que las **variables cuantitativas** son aquellas cuyos valores proceden de mediciones o conteos referidos a escalas numéricas. Por ejemplo, peso de un lote, número de clientes atendidos, número de productos defectuosos.

Las variables cuantitativas se clasifican como discretas y continuas. Las **discretas** con frecuencia proceden de conteos, y solo pueden tomar valores dentro de un conjunto numerable. Por ejemplo: número de clientes atendidos (0, 1, 2, 3, ..., etc.), número de artículos defectuosos por lote, número de quejas, número de servicios de mantenimiento. Por su parte, las **variables continuas** pueden tomar cualquier valor dentro de un intervalo específico. Por ejemplo, el tiempo en el que un cliente es atendido. Intuitivamente las variables de tipo continuo son aquellas que requieren un instrumento de medición para cuantificarse, como peso, volumen, voltaje, longitud, resistencia, temperatura, humedad, tiempo, dimensiones varias, etcétera.

Por otro lado, por su función en un proceso se pueden tener **variables de entrada** y de salida. Las primeras, también llamadas variables independientes (las $X$), por lo general son variables de control del proceso, como temperatura, velocidad, presión, cantidad y/o características de algún insumo o material, etc. Además entre las variables de entrada se consideran aquellas que, aunque normalmente no están controladas, influyen en los resultados de un proceso, como la humedad relativa en el medio ambiente, la habilidad de un operario, el método de trabajo, etcétera.

**variables cualitativas o de atributos**
Son aquellas cuyos valores representan categorías o atributos que no tienen de base una escala numérica.

**variables cuantitativas**
Sus valores proceden de mediciones o conteos referidas a escalas numéricas.

**variables discretas**
Variables cuantitativas que, con frecuencia, proceden de conteos y que sólo pueden tomar valores dentro de un conjunto numerable.

**variables continuas**
Variables cuantitativas que pueden tomar cualquier valor en un intervalo de números reales.

**variables de entrada**
Son variables que reflejan las condiciones de operación de un proceso o de sus insumos (las $X$).

**variables de salida**
Son aquellas que reflejan los resultados de un proceso (las *Y*).

Las **variables de salida**, también llamadas *variables de respuesta* o *dependientes* (las *Y*), reflejan los resultados obtenidos por un proceso, como pueden ser las características de calidad del producto; de tal forma que a través de los valores que toman estas variables se evalúa el desempeño del proceso. Por lo general, las variables de salida tienen especificaciones o tolerancias, porque los valores que deben tener están especificados. Por ejemplo, en un proceso que produce piezas metálicas que se van a ensamblar, las dimensiones de estas deben caer dentro de cierto rango o especificaciones, de lo contrario no embonarán. Cuando se satisfacen estos requerimientos se dice que el proceso cumple las especificaciones de calidad. Existen tres tipos de variables de salida o características de calidad, de acuerdo con el tipo de especificaciones que deben cumplir:

- *Cuanto más pequeño mejor.* Son variables o características de calidad cuya única exigencia es que no excedan un cierto valor máximo tolerado o una especificación superior (ES), y cuanto más pequeño sea su valor, mejor. Por ejemplo, el porcentaje de impurezas en una sustancia o la cantidad de sustancias tóxicas en un producto alimenticio.
- *Cuanto más grande mejor.* Son variables o características de calidad a las que se les exige que sean mayores que un valor mínimo o que una cierta especificación inferior (EI), y cuanto más grande sea el valor de la variable, será mejor. Por ejemplo, la resistencia de una pieza de plástico inyectado o la "blancura" de una tela de color blanco.
- *El valor nominal es el mejor.* Variables que deben tener un valor específico y que, por lo tanto, no deben ser menores que una especificación inferior (EI), pero tampoco mayores que una superior (ES). Por ejemplo, el diámetro interior de una tuerca o la longitud de una pieza para ensamble; no pueden ser ni muy chicas ni muy grandes.

Una tarea primordial del control de calidad es conocer qué tanto los valores de una variable de salida de un proceso son satisfactorios, y saber de esa manera si el proceso es capaz de cumplir las especificaciones para esa variable. Por ello es necesario tomar datos de esta variable y analizarlos adecuadamente. Esto queda claro en el ejemplo 8.1, sobre el azúcar, donde se requiere decidir cuál de las dos marcas cumple mejor con la especificación: peso más cercano a 500 gramos.

## Ejemplo 8.1 La variación

En un restaurante se tiene una fórmula específica para elaborar una cantidad determinada de "agua fresca", la cual contempla agregar 500 gramos de azúcar. Es claro que resulta de suma importancia añadir exactamente esa cantidad de azúcar para la calidad del agua, de lo contrario, esta queda muy dulce o desabrida. Aunque a los cocineros se les ha insistido sobre lo anterior, es frecuente que no pesen el azúcar y la agreguen al tanteo. Al considerar la calidad del agua como un aspecto clave, se decide diseñar un procedimiento a prueba de olvidos: comprar bolsas que contengan 500 gramos de azúcar. Suponga dos marcas de azúcar que cuentan con la presentación de 500 gramos; ahora es necesario decidir qué marca comprar. Con este propósito se pesan 40 bolsas de ambas marcas, y se obtienen los datos de la tabla 8.1.

Es claro que los datos son variables, por lo que se requieren analizar con técnicas estadísticas. En estos casos, para saber el comportamiento (distribución) de un conjunto de datos es necesario estudiar tres de sus aspectos: tendencia central, variabilidad y forma de su distribución. En las secciones siguientes veremos diferentes métodos de la estadística descriptiva especializadas en estudiar uno o más de estos aspectos.

**Tabla 8.1** Datos para el ejemplo 8.1

| Marca | Peso de las bolsas de azúcar (g) | Media | Mediana |
|---|---|---|---|
| A | 503 507 492 499 498 506 502<br>502 506 502 505 493 500 489<br>500 492 500 515 510 502 508<br>499 510 494 503 499 508 513<br>502 515 514 507 510 498 507<br>491 507 502 484 500 | 502.3 | 502 |

*(continúa)*

**Ejemplo 8.1** La variación (*continuación*)

**Tabla 8.1** Datos para el ejemplo 8.1 (*continuación*)

| Marca | Peso de las bolsas de azúcar (g) | Media | Mediana |
|---|---|---|---|
| B | 505 492 502 499 496 499 496<br>495 498 501 504 501 498 498<br>499 495 501 500 497 495 500<br>491 493 507 496 492 499 492<br>501 500 497 500 498 496 494<br>497 504 496 500 499 | 498.1 | 498.5 |

## Medidas de tendencia central

Con las mediciones de una característica de calidad o variable de tipo cuantitativo, como las del ejemplo 8.1, el primer aspecto a investigar es la **tendencia central** de los datos para identificar un valor en torno al cual los datos tienden a agruparse o concentrarse. Esto permitirá conocer tal valor y hacer algún tipo de evaluación con relación a lo adecuado de este, lo cual es el caso en control de calidad donde se busca saber si el proceso está centrado; es decir, saber si la tendencia central de la variable de salida es igual o está muy próxima a un valor nominal deseado (en el ejemplo, el valor nominal es 500 gramos). Enseguida veremos tres medidas de la tendencia central: la media, mediana y moda.

**tendencia central**
Valor en torno al cual los datos o mediciones de una variable tienden a agruparse o concentrarse.

### Media muestral

Suponga que $x_1, x_2, x_3, \dots, x_n$ son las observaciones numéricas de una muestra, entonces la medida más usual de tendencia central es la **media** o **promedio muestral**, que es igual a la media aritmética de todos los datos:

**media** o **promedio muestral**
Medida de tendencia central que es igual al promedio aritmético de un conjunto de datos que se obtiene al sumarlos y el resultado se divide entre el número de datos.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

es decir, la media muestral se obtiene sumando todos los datos, y el resultado se divide entre el número de datos ($n$). En Excel, la función PROMEDIO() se puede utilizar para calcular la media. Por ejemplo, los siguientes datos representan el sueldo semanal de siete trabajadores de cierta área de una empresa:

1 100; 1 300; 1 000; 1 500; 800; 1 600; 1 100;

entonces $\bar{x} = 1\,200$, por lo que el sueldo promedio de los trabajadores es de 1 200 pesos. Es claro que esto no quiere decir que todos o la mayoría de los trabajadores ganen 1 200 pesos, es más, en el ejemplo nadie gana tal cantidad.

Si para calcular la media se utilizaron todos los elementos de la población (el universo sobre el que se quiere tomar decisiones, por ejemplo, el sueldo de todos los trabajadores), entonces el promedio calculado es la media poblacional y se denota con la letra griega mu, $\mu$. Cabe destacar que la media del proceso $\mu$ es igual a cierto valor, aunque no siempre se conoce; mientras que el valor de $\bar{x}$ se obtiene para cada muestra y es diferente (variable) de una muestra a otra, ya que su valor depende de los elementos que se seleccionan en la muestra ($\bar{x}$ es una variable aleatoria). Por lo anterior, el valor que se observa de la media muestral, $\bar{x}$, es por lo general diferente de la media poblacional o del proceso, $\mu$. Luego hay que tener cuidado con las afirmaciones que se hacen con base en $\bar{x}$ sobre la media del proceso o población.

## Mediana o percentil 50

**mediana**
Medida de tendencia central que es igual al valor central que divide los datos a la mitad cuando son ordenados de menor a mayor.

Otra medida de tendencia central de un conjunto de datos es la **mediana**, $\tilde{x}$, que es igual al valor que divide a la mitad los datos cuando se ordenan de menor a mayor. Para calcular la mediana cuando el número de datos es impar, los datos se ordenan de manera creciente y el que quede en medio de dicho ordenamiento será la mediana; y si el número de datos es par, entonces la mediana se calcula dividiendo entre dos la suma de los números que están en el centro del ordenamiento. Así, los datos que son menores o, a lo más, igual que la mediana $\tilde{x}$ constituyen 50% de los datos; mientras que los que son mayores o iguales son el restante 50%. Por ello la mediana también se conoce como *percentil 50.* En Excel, la función MEDIANA() se puede utilizar para calcular la mediana.

Por ejemplo, en el caso de los sueldos de los trabajadores, para calcular la mediana se ordenan los datos:

800; 1 000; 1 100; 1 100; 1 300; 1 500 y 1 600;

Entonces, como el número de datos es impar ($n = 7$), la mediana es $\tilde{x} = 1\,100$, con lo que es posible asegurar que la mitad de los trabajadores de la muestra gana 1 100 pesos semanales o menos.

## Moda

**moda**
Medida de tendencia central de un conjunto de datos que es igual al dato que se repite más veces.

Otra medida tradicional de la tendencia central de un conjunto de datos es la **moda**, que es igual al dato que se repite con más frecuencia. Su cálculo se puede hace en Excel por medio de la función MODA( ). En el caso de los sueldos de los trabajadores la moda es 1 100. De aquí se desprende que no siempre la media es el dato más frecuente, ya que en este caso la media es 1 200, y no 1 100.

Cuando en un grupo de datos hay algunos valores bastante diferentes del resto, ya sean muy pequeños, o bien, muy grandes, entonces la media no es una buena medida de tendencia central, ya que a esta la "jalan" los datos atípicos o raros. Por ejemplo, en el caso de los sueldos de los trabajadores, suponga que en esa área hay un trabajador más que gana 7 600 pesos por mes, con lo que ahora los sueldos son:

800; 1 000; 1 100; 1 100; 1 300; 1 500; 1 600 y 7 600.

En este caso la media es 2 000. Resulta evidente que el 2 000 no refleja la tendencia central de los sueldos, ya que solo un trabajador gana más que la media. En estos casos la *mediana es mejor medida de tendencia central,* ya que no resulta afectada por datos "raros". Por ejemplo, en el caso de los trabajadores, cuando uno gana mucho más que el resto, la mediana es igual a $(1\,100 + 1\,300)/2 = 1\,200$. Este valor sí refleja la tendencia central de la mayoría de los sueldos.

De lo anterior se desprende que, para describir la tendencia central de los datos, es importante apoyarse tanto en la media como en la mediana. Y en caso de que la media sea mucho más grande que la mediana, es señal de que existen datos más grandes que el resto, los que hacen que la media esté "inflada". Por el contrario, si la media es significativamente menor que la mediana, entonces eso indica la presencia de datos mucho más pequeños que el resto, los cuales hacen que la media esté "subestimada". Tomar en cuenta lo anterior es de importancia primordial en la toma de decisiones, ya que no siempre la media refleja la verdadera tendencia central.

**Ejemplo 8.1** **Continuación del ejemplo 8.1**

En la tabla 8.1 se ve que la media y la mediana para la muestra de la marca A son 502.3 y 502, respectivamen[te], mientras que para la marca B son 498.1 y 498.5. Además, la moda es 502 y 498 para las marcas A y B, respecti[va]mente. Por lo que la tendencia central para cada marca se aleja aproximadamente lo mismo respecto al peso requ[e]rido de 500 gramos. De aquí que al tomar en cuenta únicamente las medidas de tendencia central, no se pue[de] decidir cuál de las dos marcas satisface mejor el requerimiento de peso. Por lo general, como en este caso, dec[idir] con base en el promedio equivale a "lanzar un volado". A continuación se analizará cómo al considerar también la va]riabilidad de los datos se pueden tomar decisiones más acertadas.

## Medidas de dispersión o variabilidad

Además de la tendencia central de un conjunto de datos, es necesario conocer qué tan diferentes son entre sí, es decir, es importante saber su **variabilidad** o **dispersión**. Esto es un elemento vital en el análisis estadístico de un conjunto de datos, particularmente cuando se quiere hacer un estudio de capacidad de un proceso. Enseguida veremos cuatro formas de medir la variabilidad.

La **desviación estándar muestral** es la medida más usual de variabilidad e indica qué tan esparcidos están los datos respecto a la media; se denota con $S$ y se calcula mediante la siguiente expresión:

$$S = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}}$$

donde $x_1, x_2, \ldots, x_n$, son las observaciones numéricas de la muestra y $\bar{x}$ es la media muestral. Como se puede apreciar, $S$ mide la distancia que en "promedio" hay entre los datos y la media; por ello, cuanto más grande sea el valor de $S$, mayor variabilidad habrá en los datos. La desviación estándar está expresada en las mismas unidades de medición (gramos, milímetros, etc.) que los datos. Además, $S$ no refleja la magnitud de los datos, solo lo retirado que están los datos de la media, y al igual que esta, se ve afectada por datos atípicos. Su cálculo en Excel se puede hacer con la función DESVESTA().

**Desviación estándar poblacional** o **del proceso**, $\sigma$. Si para calcular la desviación estándar se utilizan todos los elementos de la población o proceso, entonces se obtiene la desviación estándar poblacional y se denota por la letra griega sigma, $\sigma$.

Por otra parte, el cuadrado de la desviación estándar, $S^2$, se conoce como **varianza muestral**, que es muy importante para propósitos de inferencia estadística. Y en forma equivalente, $\sigma^2$ es la varianza (o variancia) poblacional.

Otra medida de dispersión es el **rango** o **recorrido**, $R$, que es igual a la diferencia entre el dato mayor y el dato menor de un conjunto de datos, por lo que $R$ mide la amplitud de la variación de un grupo de datos y también es independiente de la magnitud de los datos. En Excel, se puede obtener con el auxilio de las funciones Max( ) – Min( ). Por ejemplo, sean los dos conjuntos de datos:

$$A = \{10, 12, 14\} \text{ y } B = \{159, 161, 163\},$$

entonces se ve que la magnitud de los datos es diferente, y eso lo refleja la media, que es de 12 y 161, respectivamente. Pero en cuanto a la variabilidad, los datos de ambos conjuntos están igualmente dispersos, como lo indica la desviación estándar que es igual a 2 en ambos casos, y el rango que es de 4 para los dos conjuntos.

El **coeficiente de variación (CV)** es una medida de variación relativa a la magnitud de los datos, que es igual a la desviación estándar entre la media de los datos:

$$CV = \frac{S}{\bar{x}}(100)$$

El CV es útil para comparar la variación de dos o más variables que están medidas en diferentes escalas o unidades de medición (por ejemplo, metro contra centímetro o metro contra kilogramo). Este coeficiente suele interpretarse como una medición en términos porcentuales de la variación de una variable. Por ejemplo, en el caso de los conjuntos de datos A y B que se presentaron en la definición del rango, se tiene que sus correspondientes CV son:

$$CV_A = \frac{2}{12}x100 = 16.66, \quad CV_B = \frac{2}{161}x100 = 1.242$$

respectivamente. Por lo que la variabilidad en los términos relativos del CV para el conjunto A es de 16.66%, mientras que para el conjunto B es solo de 1.242%.

**variabilidad o dispersión**
Se refiere a las diferencias que se hallan entre los datos de un conjunto.

**desviación estándar muestral**
Medida de la variabilidad que indica qué tan esparcidos están los datos de la muestra con respecto a su media.

**desviación estándar poblacional o del proceso, $\sigma$**
Medida de la variabilidad de un proceso. Para su cálculo se debe utilizar un número grande de datos que hayan sido obtenidos en el transcurso de un lapso de tiempo amplio. Se denota con la letra griega sigma $\sigma$.

**varianza muestral**
Medida de variación que es igual al cuadrado de la desviación estándar, $S^2$.

**rango o recorrido**
Medición de la variabilidad de un conjunto de datos que es resultado de la diferencia entre el dato mayor y el dato menor de tal conjunto.

**coeficiente de variación (CV)**
Medida relativa de variabilidad que se obtiene al dividir la desviación estándar entre la media. Es útil para contrastar la variación de dos o más variables que están medidas en diferentes escalas o unidades.

**Ejemplo 8.1** **Continuación del ejemplo 8.1**

Por medio de las medidas de tendencia central no se pudo decir cuál de las dos marcas de azúcar satisface mejor la exigencia de que las bolsas pesen 500 gramos. Ahora, con el uso de las medidas de variabilidad, sí se podrá decidir. A partir de la tabla 8.1 se obtiene:

$$\text{Marca A: } S = 7.23,\ R = 31$$
$$\text{Marca B: } S = 3.68,\ R = 16$$

De aquí se desprende que la muestra de bolsas de azúcar de la marca A tienen una dispersión dos veces que la de la marca B, tanto en términos de $S$ como de $R$. Por lo tanto, como las muestras son representativas de ambas marcas y tienen un tamaño moderado, esto es una evidencia a favor de la marca B, ya que estas proporcionan un peso más cercano al peso deseado (500 gramos).

Si la discrepancia entre las desviaciones estándar hubiera sido menor, entonces para decidir si tal discrepancia es significativa o si podría atribuirse a variaciones debido al muestreo, se hace una prueba de hipótesis (Gutiérrez y de la Vara, 2013).

## Relación entre $\bar{x}$ y $S$

**desigualdad de Chebyshev**
Resultado teórico que relaciona $\bar{x}$ y $S$, y establece el porcentaje mínimo de datos que caen en el intervalo $(\bar{x} - kS, \bar{x} + kS)$, con $k > 1$.

**regla empírica**
Resultado práctico que relaciona $\bar{x}$ y $S$, y establece el porcentaje de datos de la muestra que caen dentro del intervalo $(\bar{x} - kS, \bar{x} + kS)$ con $k = 1, 2, 3$.

Una forma de apreciar más claramente el significado de la desviación estándar como medida de dispersión en torno a la media es a través de la relación entre la media y la desviación estándar, la cual está dada por la **desigualdad de Chebyshev** y la **regla empírica**. Dos hechos particulares que afirman la desigualdad de Chebyshev[2] es que entre $\bar{x} - 2S$ y $\bar{x} + 2S$ está por lo menos 75% de los datos de la muestra, y que entre $\bar{x} \pm 3S$ está por lo menos 89%.

En cuanto a la regla empírica, afirma que en muchos de los datos que surgen en la práctica se ha observado por la experiencia que:

- entre $\bar{x} - S$ y $\bar{x} + S$ está 68% de los datos de la muestra;
- entre $\bar{x} - 2S$ y $\bar{x} + 2S$ está 95%, y
- entre $\bar{x} - 3S$ y $\bar{x} + 3S$ está 99.7%.

Todos los intervalos anteriores son válidos solo para los datos muestrales y no necesariamente para toda la población o proceso. Sin embargo, si los intervalos se calculan con la media y la desviación estándar del proceso o población, entonces serán válidos para toda la población. Por lo que en la medida en la que se tengan muestras aleatorias grandes y representativas, los intervalos anteriores podrán dar una idea aproximada de lo que ocurre en el proceso.

Lo que afirma el teorema de Chebyshev se aplica para cualquier tipo de datos, independientemente de su comportamiento o distribución.[3] Mientras que la regla empírica, como su nombre lo dice, se ha obtenido por medio de la observación empírica y es válida para muchos de los casos que se dan en la práctica, sobre todo si los datos tienen un comportamiento con cierto grado de similitud a una campana o a la distribución normal (vea el apéndice). De cualquier manera, ambos casos ilustran muy bien la manera en la que la desviación estándar mide la variabilidad en torno a la media.

Si se aplica la regla empírica a los datos del ejemplo 8.1 del peso de las bolsas de azúcar, se tiene que el intervalo $\bar{x} \pm 3S$ está dado por:

[2] En general, la desigualdad de Chebyshev afirma que al menos $(1 - 1/k^2) \times 100$ de los datos están entre $\bar{x} - kS$ y $\bar{x} + kS$; es decir, ese porcentaje de datos estará dentro de $k$ desviaciones estándar a partir de la media, donde $k$ es cualquier número mayor que 1.

[3] Apoyando la regla empírica existe una extensión a la desigualdad de Chebyshev, hecha por Camp y Meidel (vea Duncan, 1989), que aumenta el porcentaje que cubren los intervalos. Concretamente, esta extensión afirma que si la distribución de $X$ es unimodal, la probabilidad de que $X$ se desvíe de su media en más de $k$ veces su desviación estándar, es igual o menor que $1/2.25k^2$. Con lo que bajo estas circunstancias entre $\bar{x} \pm 2S$ se encontraría al menos 89% de los datos muestrales y entre $\bar{x} \pm 3S$ estaría al menos 95%.

Marca A: $502.3 - 3(7.23) = 480.61$  $502.3 + 3(7.23) = 523.99$
Marca B: $498.1 - 3(3.68) = 487.06$  $498.1 + 3(3.68) = 509.14$

De esto se deduce que alrededor de 99% de los pesos de las bolsas de muestra de la marca A varía entre 480.61 y 523.99 gramos. Mientras que en el caso de la marca B, esta variación está entre 487.06 y 509.14 gramos. Con esto se confirma la evidencia a favor de la marca B.

### Ejemplo 8.2 Capacidad de proceso

En una fábrica de piezas de asbesto, una característica importante de calidad es el grosor de las láminas, que, para cierto tipo de lámina, el óptimo es de 5 mm, y se establece como discrepancia tolerable ±0.8 mm, ya que si la lámina tiene un grosor menor que 4.2 mm se considera demasiado delgada y no reunirá las condiciones de resistencia exigidas por el cliente. Pero si la lámina tiene un grosor mayor que 5.8 mm, entonces se gastará demasiado material para su elaboración y se elevarán los costos del fabricante. Por lo tanto, es necesario que el proceso de fabricación de este tipo de láminas garantice que su grosor cumple con especificaciones: $EI = 4.2$ y $ES = 5.8$ mm.

Se plantean las siguientes interrogantes: ¿qué tipo de láminas en cuanto a grosor se están produciendo? ¿El grosor medio es adecuado? ¿La variabilidad del grosor es mucha o poca? Para responder a estas interrogantes de la producción de una semana, mediante muestreo sistemático, se mide el grosor de 60 láminas y se obtiene:

$$\bar{X} = 4.73 \quad \tilde{X} = 4.7 \quad S = 0.48.$$

Tomando en cuenta el tamaño de muestra y la forma en la que se obtuvo, se puede asegurar con un buen nivel de confianza que el grosor promedio no fue satisfactorio, ya que es algo menor que 5 mm; con el índice $K$ que se detalla en el siguiente capítulo, se ve que la media está desfasada 34% a la izquierda del óptimo de la especificación. De la mediana se ve que 50% de las 60 láminas medidas tuvo un grosor menor o igual que 4.7 mm.

Para investigar la variabilidad y saber si al menos el espesor de las 60 láminas cayó dentro de las especificaciones, se aplica la regla empírica, con lo que se ve que entre:

$$4.73 - 3 \times 0.48 \text{ y } 4.73 + 3 \times 0.48 \ (3.29 \text{ y } 6.17 \text{ mm})$$

estuvieron prácticamente todas las láminas de la muestra en cuanto a grosor, por lo que hay serios problemas pues deberían haber estado, a lo más, entre 4.2 y 5.8.

De acuerdo con lo anterior, el proceso no es capaz de cumplir los requerimientos de calidad, porque se tiene demasiada variación y el proceso está descentrado, donde se da una tendencia marcada a producir láminas más delgadas (4.73). Para atender estos problemas, la experiencia indica que se debe trabajar primeramente en centrar el proceso y luego en reducir la variabilidad.

Como se vio en los ejemplos 8.1 y 8.2, con base en la relación entre la media y la desviación estándar expresada por el teorema de Chebyshev y la regla empírica, es posible determinar si la variabilidad es mucha. Se debe tener cuidado en la aplicación de la regla empírica cuando se tienen muestras pequeñas y/o poco representativas, ya que las conclusiones serán poco confiables respecto a toda la población.

## Límites reales o naturales

Debido a las propiedades de la distribución normal (vea el apéndice), expresada por la regla empírica, lo que se ha estudiado da origen a lo que se conoce como **límites naturales** o **reales** de un proceso. Sea $\mu$ la media y $\sigma$ la desviación estándar del proceso, entonces los límites reales o naturales del proceso están dados por:

$$\text{Límite real inferior (LRI)} = \mu - 3\sigma$$
$$\text{Límite real superior (LRS)} = \mu + 3\sigma$$

Dentro de estos límites se ubicarán los valores de la variable de salida correspondiente.

**límites naturales o reales**
Se obtienen con $\mu - 3\sigma$ y $\mu + 3\sigma$, e indican de dónde a dónde varía la salida de un proceso.