## CAPITULO II

## MEDIDAS DE CENTRALIZACION Y DISPERSION

Del mismo modo que las gráficas pueden mejorar la presentación de los datos, las descripciones numéricas también tienen gran valor. Estas descripciones son números que sintetizan la información contenida en una serie de datos y le dan plena validez objetiva a una evaluación.

El valor central o típico del conjunto de datos, en el sentido que es el más representativo del grupo de datos, se le denomina promedio. Este valor pertenece a la serie de *Medidas de Tendencia Central*. Los promedios basados en sus propiedades matemáticas pueden clasificarse en dos categorías: *promedios computados* (media aritmética, media geométrica y media armónica), y *promedios de posición* (mediana y moda).

Además de la tendencia central de los valores a agruparse en las cercanías de un valor promedio, es necesario saber cuanto se dispersan o varían; es decir, si están cerca uno del otro o alejados. Las medidas de este acercamiento o alejamiento se conocen como *Medidas de Variabilidad o de Dispersión* y las más usadas son el rango, la varianza, desviación estándar y coeficiente de variación.

### 2.1. Medidas de centralización

Una característica importante de un conjunto de números es su localización o tendencia central. Esto se evalúa a través del promedio, mediana, moda, media geométrica, cuartiles, etc.

a) Promedio ó Media Aritmética:

La media aritmética, por su facilidad de computación y largo uso, es el promedio más conocido y más comúnmente usado. A veces se le llama simplemente *la media* o *promedio*; pero cuando se hace referencia a otros tipos de medias deben emplearse los adjetivos apropiados.

Para un grupo de datos, $x_1, x_2, x_3,...,x_n$, la medida más conocida de centralización es el promedio aritmético de *n* observaciones. Ya que los datos usualmente corresponden a una muestra de una gran población (en el ejemplo de la Tabla 1.2, 90 muestras de bloques de concreto tomadas de la cantidad total o población), este promedio se refiere al *promedio de muestra* y se define como:

$$\bar{x} = \frac{1}{n}( x_1 + x_2 + x_3 +...+ x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Nótese que la suma de las desviaciones de cada muestra con respecto al promedio, $x_i - \bar{x}$ debe dar cero; esto es:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - (n\bar{x}) = 0$$

Se puede imaginar al promedio como el punto de balance que mantiene una barra completamente horizontal. Es decir, los valores de las observaciones de un lado de la media igualan a los valores de las observaciones del otro lado de ella.

b. Media Ponderada

Teóricamente, todas las medias aritméticas son promedios ponderados. Si no se asignan pesos específicos a todos y cada uno de los valores de una serie, a cada una se le asigna un peso igual a 1. Al computar la media aritmética con datos agrupados, las frecuencias de clases pueden considerarse como una serie de pesos para los puntos medios. Cuando se usan pesos diferentes en el cómputo, puede decirse apropiadamente que la media aritmética es *ponderada.*

En algunas circunstancias, no todas las observaciones tienen el mismo peso. Sean los valores $X_1, X_2 ,..., X_n$ asociados a pesos $w_1, w_2,..., w_n$ que dependen de la significación e importancia de esos números; en este caso:

$$\bar{x}_w = \frac{w_1X_1 + w_2X_2 +...+ w_nX_n}{w_1 + w_2 + ... + w_n} = \frac{\sum wX}{\sum w}$$

c. Media a partir de datos agrupados

Debido a que los valores individuales de la muestra se pierden al agruparse, para calcular la media aritmética se usa los puntos medios o centros de clase como representantes de clase. En consecuencia, si $m_i$ es el centro de clase y $f_i$ la frecuencia de clase, la media aritmética se define como:

$$\bar{x} = \frac{f_1 m_1 + f_2 m_2 + f_3 m_3 + \ldots + f_n m_n}{n} = \frac{\sum_{i=1}^{K} f_i \; m_i}{n}$$

donde *k* es el número de clases.

d. Media geométrica, *G*

La media geométrica de una serie de '*n*' números, $x_1, x_2, \ldots, x_n$, es la raíz *N-ésima* del producto de los números:

$$G = \sqrt[n]{x_1 x_2 \ldots x_n} = \sqrt[n]{\prod_{i=1}^{n} X_i}$$

La computación de la media geométrica se facilita reduciendo la fórmula a su forma logarítmica:

$$\log G = \frac{\log x_1 + \log x_2 + \log x_3 + \ldots + \log x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} \log X_i$$

Por ejemplo, la media geométrica de 2, 4 y 8 libras es:

$$G = \sqrt[3]{(2)(4)(8)} = 4 \;\; lb$$

Para la misma serie el promedio aritmético es 4.7 lb. Es cierto siempre que la media aritmética es mayor que la media geométrica para cualquier serie de valores positivos, a menos que las partidas que se promedian sean del mismo valor, en cuyo caso los dos promedios son iguales.

A causa de esta relación, para computar la media geométrica a partir de datos agrupados se puede utilizar la siguiente expresión:

$$\log G = \frac{f_1 \log m_1 + f_2 \log m_2 + f_3 \log m_3 + \ldots + f_n \log m_n}{n} = \frac{1}{n} \sum_{i=1}^{n} f \log m_i$$

Puede observarse que la media geométrica sólo es significativa para un conjunto de observaciones que son todos positivos. La media geométrica se emplea por ejemplo en microbiología para calcular títulos de disolución promedio y para promediar cantidades en forma de proporciones y tasas de crecimiento o de cambios y en general cuando convenga hacer una transformación logarítmica. La media geométrica se adapta especialmente a las razones de promedios, índices de cambio y series distribuidas logarítmicamente. En ciertos casos especiales de razones de promedios o porcentajes, como la computación del índice de precios, la media geométrica puede dar resultados significativos y lógicos que la media aritmética no da.

La media geométrica da igual ponderación a las tasas iguales. Es decir, al promediar tasas de cambio geométricamente, la tasa que muestra el doble de su base es compensada por otra muestra la mitad de su base; la tasa que muestra cinco veces su base, es compensada por otra que muestra un quinto su base; y así sucesivamente. Las tasas de cambio son ordinariamente expresadas en porcentajes. Puesto que la base para cada proporción expresada en porciento es siempre igual a 100%, el promedio de dos proporciones las cuales se compensan, deberá ser 100 % también. El siguiente cuadro da una ilustración de que la media geométrica proporciona una mejor respuesta que la que proporciona la media aritmética:

| Elemento | Unidades vendidas | | Tasa de cambio | |
|---|---|---|---|---|
| | 1984 | 1985 | 1984 (base) | 1985 |
| A | 5 yd | 25 yd | 100 % | 500 % |
| B | 50 lb | 10 lb | 100 % | 20 % |
| | | | | |
| Media Aritmética | | | 100 % | 260% = $\frac{20\% + 500\%}{2}$ |
| Media Geométrica | | | 100 % | 100% = $\sqrt{20 \; x \; 500}$ |

e. La Media Harmónica.

Si se toma el recíproco del valor de cada partida, se calcula la media aritmética de los recíprocos y se toma el recíproco de esta media, el resultado se conoce como media harmónica. O más sintéticamente, la media harmónica es el recíproco de la media aritmética de los recíprocos de las observaciones. La formula es:

$$H = \frac{1}{\frac{1/x_1 + 1/x_2 + ... + 1/x_n}{n}} = \frac{1}{\frac{\sum(1/x)}{n}} = \frac{n}{\sum(1/x)}$$

Para facilitar los cálculos se puede utilizar:

$$\frac{1}{H} = \frac{1/x_1 + 1/x_2 + ... + 1/x_n}{n} = \frac{\sum(1/x)}{n}$$

La media harmónica para los siguientes valores: 2, 4 y 8 es:

$$\frac{1}{H} = \frac{1/2 + 1/4 + 1/8}{3} = \frac{7}{24}$$

$$H = 3.43$$

Para los mismos datos, la media aritmética es 4.7 y la media geométrica es 4. Para cualquier serie cuyos valores no sean iguales y que no tengan ningún valor de cero, la media harmónica es siempre menor que la media aritmética y la media geométrica. Así, la media harmónica se considera que tiene una tendencia hacia abajo, en tanto que la media aritmética la tiene hacia arriba.

La media harmónica se utiliza en el cálculo de tasas medias de tiempo en ciertas condiciones y ciertos tipos de precios. También se adapta bien a una situación en que las observaciones se expresan inversamente a lo que se requiere en el promedio; es decir cuando, por ejemplo, se desea el costo medio por unidad, pero los datos muestran el número de unidades producidas por cantidad de costo. Obsérvese la siguiente ilustración que describe lo señalado.

Supóngase que se ha gastado lo siguiente:

- Un dólar por 3 docenas de naranjas;
- Un dólar por 4 docenas de naranjas;
- Un dólar por 5 docenas de naranjas.

De esto resulta que los precios por docena son respectivamente: 33 1/3, 25 y 20 centavos, cuyo promedio es:

(33 1/3 + 25 + 20)/3 = 26.1 centavos.

Siendo que se compró 12 docenas de naranjas, resultaría que 26 *12 = 3.13 dólares sería el cálculo de pago efectuado, lo cual contradice la realidad que correspondió a 3.0 dólares. Se puede observar que los datos expresan 'tantas docenas por dólar'. Estas son expresiones inversas; en consecuencia, el promedio puede obtenerse por la media harmónica, según:

$$\frac{1}{H} = \frac{1/33.33 + 1/25 + 1/20}{3} = 0.04$$

$$H = 25 \text{ centavos}$$

Se puede comprobar que la media aritmética ponderada también da resultados precisos.

f. Mediana

Es una medida de centralización mucho menos sensible a variaciones en una observación. La mediana es la observación del medio cuando las observaciones se ordenan en forma creciente en un grupo impar de *n* observaciones. Si *n* es par, la mediana es el promedio de las dos observaciones intermedias.

Sean $X_1, X_2 ... X_n$ una muestra acomodada en orden creciente de magnitud, entonces la mediana se define como la parte media , ó:

$$Med = \begin{cases} X_{([n+1]/2)} & impar \\ \frac{X_{(n/2)} + X_{(([n/2]+1))}}{2} & par \end{cases}$$

Si se toma por ejemplo las 10 primeras observaciones de la Tabla 1.2 (Cap. I) y se les ordena crecientemente, se tiene:

31.3 32.3 42.2 42.3 44.5 47.5 49.2 50.0 53.9 60.9

↑ ↑

46.0

Para los 90 datos de la Tabla 1.2, la mediana es el promedio de los valores de las posiciones 45 y 46 y corresponde a 45.5.

La característica típica de la mediana es que divide la distribución en dos partes iguales. En este sentido, la mediana es también un punto de equilibrio. Así, la mediana es especialmente significativa para describir observaciones que se anotan o puntúan, como tasas, calificaciones y clasificaciones, en vez de contarse o medirse. Sin embargo, no tiene sentido para datos completamente cualitativos.

g. Moda

Representa el valor más frecuente en el conjunto de datos. En un histograma, el *pico* representa la clase de mayor frecuencia; el punto medio de esta clase, también se llama moda. En el ejemplo de los bloques de concreto, la mayor clase es la que corresponde al intervalo [44 – 48] y el punto medio de esta clase ea 46 que corresponde a la moda.

Puesto que la moda es el punto de mayor concentración, la moda es el promedio más común para una distribución. A causa de esta propiedad, la moda carece de significado si la distribución no tiene un gran número de observaciones. La moda es un promedio muy inestable y su verdadero valor es difícil de determinar.

Para distribuciones simétricas, el promedio, la mediana y la moda tienen aproximadamente el mismo valor. Para el caso particular de las 90 observaciones de la resistencia a la compresión de los bloques de concreto se tiene:

$\bar{x} = 45.54$; Med = 45.5; Moda = 46.0

Para las distribuciones sesgadas, estos valores no son iguales. Para una distribución positivamente asimétrica, la media tiene el valor más grande, la moda el más pequeño y la mediana aproximadamente un tercio de la distancia de la media a la moda. Para una distribución negativamente asimétrica, la media es menor, la moda es mayor y la mediana se encuentra a una distancia de un tercio desde la media hacia la moda.

La moda es con frecuencia el concepto que la mayoría de las personas tienen en mente cuando hablan de promedios. El 'consumidor medio' suele significar el consumidor que aparece con mayor frecuencia en relación a su cuadro de consumo u otra cualidad; el tamaño modal de zapatos para hombre es el tamaño típico comprado, porque más personas compraran ese tamaño que cualquier otro. Así, se usa la moda con preferencia a otros promedios si se desea indicar el valor más típico de la serie.

h. Cuartiles.

La mediana (ya sea de una población ó de una muestra) divide a los datos en dos partes iguales. También es posible dividir los datos en más de dos partes. Cuando se divide un conjunto ordenado de datos en cuatro partes iguales, los puntos de la división se conocen como cuartiles. *El primer cuartil o cuartil inferior*, $q_1$, es un valor que tiene aproximadamente la cuarta parte (25%) de las observaciones por debajo de el, y el 75% restante por encima de el. El *segundo cuartil*, $q_2$, tiene aproximadamente la mitad (50%) de las observaciones por debajo de él, a su vez equivale a la mediana. El tercer cuartil o cuartil superior, $q_3$, tiene aproximadamente tres cuartas partes (75%) de las observaciones por debajo de él. Al igual que en el caso de la mediana, es posible que los cuartiles no sean únicos, y en tales casos, si dos observaciones satisfacen la definición, se utiliza el promedio de ellas.

El rango intercuartil ($q_3 - q_1$) contiene el 50% central, con un 25% por debajo y otro 25% por arriba. A menudo se emplea como medida de dispersión.

h) Percentiles

Cuando un conjunto ordenado de datos se divide en cien partes iguales, los puntos de división reciben el nombre de percentiles. El 100 *k*-iésimo percentil $p_k$, es un valor tal que al menos 100 *k*% de las observaciones están en el valor o por debajo de él, y al menos el 100(1-*k*)% están en el valor o por encima de él.

Ejemplo:

Una serie de análisis químico de un gran lote de material tiene un contenido de 80 g de elemento útil por tonelada del material. Para comprobar esta afirmación se toma 25 muestras del material con los siguientes resultados:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 77 | 81 | 76 | 86 | 79 | 79 | 80 | 77 | 89 | 77 | 78 | 85 | 80 |
| 75 | 79 | 88 | 81 | 78 | 82 | 80 | 76 | 83 | 81 | 85 | 79 | |

Comprobar la ley promedio del concentrado y definir los cuartiles de la muestra tomada.

Los datos tomados se ordenan según:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 75 | 76 | 76 | 77 | 77 | 77 | 78 | 78 | 79 | 79 | 79 | 79 | 80 |
| 80 | 80 | 81 | 81 | 81 | 82 | 83 | 85 | 85 | 86 | 88 | 89 | |

Promedio = 80.44 g/T

Mediana, $X_{13}$ = 80

$$posicion\ de\ \ q_1 = n\left(\frac{1}{4}\right)+0.5 = \frac{25}{4}+0.5 = 6.75 \qquad posicion\ de\ \ q_3 = n\left(\frac{3}{4}\right)+0.5 = \left(25 * \frac{3}{4}\right)+0.5 = 19.25$$

$$q_1 = \left(\frac{X_6+X_7}{2}\right) = \frac{77+78}{2} = 77.5 = p_{0.25} \qquad q_3 = \left(\frac{X_{19}+X_{20}}{2}\right) = \frac{82+83}{2} = 82.5 = p_{0.75}$$

---

$$rango\ \text{inter}cuartil = q_3 - q_1 = 82.5 - 77.5 = 5.0$$

$$100\left[\frac{(i-0.5)}{n}\right]\ es\ el\ percentil\ de\ X_i$$

$$X_4 = 77\ es\ el\ cuarto\ valor \Rightarrow 100\left[\frac{(4-0.5)}{25}\right] = 14 \Rightarrow p_{014} = 77$$

$$X_{19} = 82 \Rightarrow 100\left[\frac{(19-0.5)}{25}\right] = 74 \Rightarrow p_{0.74} = 82$$

## 2.2. Variabilidad

Los datos cuantitativos, materia prima para el análisis estadístico, se caracterizan siempre por diferencias de valor entre las observaciones individuales. Estas diferencias cuantitativas son tan importantes como la tendencia de las cifras a agruparse alrededor de un valor central en una serie. De igual modo que decimos que la Estadística es la ciencia de los promedios, podemos decir igualmente que todos los métodos estadísticos son técnicas para estudiar la variación.

La estadística es una disciplina a la que le concierne el proceso de obtener datos y comprender los problemas en presencia de la variabilidad. Los métodos estadísticos tienen como propósito fundamental entender la variabilidad. Virtualmente todo proceso varía. Si los métodos de medición son suficientemente precisos, por ejemplo se encontrará que el diámetro de pernos varia de uno a otro, el rendimiento de un proceso químico varía con el tiempo, o que el porcentaje de artículos defectuosos varía de un lote a otro, etc.

Hay múltiples razones para la presencia de esta variabilidad. Proviene de ligeras diferencias en las condiciones en las cuales se realiza la producción. La variabilidad puede reflejar diferencias en las materias primas, diferencias entre maquinas u operadores, diferencias de las condiciones de operación por cambios en variables tales como temperatura, humedad, presión, etc. También el muestreo puede ser una causa de la variabilidad, puesto que esta operación aportará muestras representativas con diferentes características cada vez que se realiza la operación. Ya que la variabilidad proviene de los procesos, es común referirse a la variabilidad de procesos.

Parte de la variabilidad proviene de las operaciones de medición. Por ejemplo, para determinar la humedad de cierto mineral, esta no tendrá exactamente el mismo valor luego de varias mediciones de la misma muestra. Semejante fenómeno se verifica en todo proceso productivo. La variabilidad de medición es el ruido que distorsiona la verdadera señal del proceso.

Por otro lado, mucha variación o variabilidad degrada la calidad de la producción y ocasiona pérdidas a una empresa. Si un gran número de productos son producidos fuera de las especificaciones, y si los productos no son inspeccionados antes de salir de una fábrica, entonces la empresa puede enfrentar los siguientes problemas:

a) La cantidad de quejas se incrementa;

b) Recursos adicionales se deberán emplear para reparar los artículos que estén bajo garantía;

c) Los clientes buscarán un producto de mayor confiabilidad.

El problema de variabilidad se cuantifica de acuerdo a los límites de especificación que son la referencia para la inspección de los productos y asegurar ausencia de defectos en el embarque de productos terminados. Un producto es aceptable si se encuentra dentro de los límites de especificación o será inaceptable si se encuentra fuera de los límites de especificación.

Las pérdidas de la compañía se deben a producción fuera de los límites de especificación. Cuando el problema aparece, la solución no es incrementar la inspección. Esto requeriría mayor fuerza laboral y aumento de costos por retratamiento y pérdidas. Consecuencia de ellos los precios finales se incrementarán con el lógico malestar en los consumidores. Eventualmente, el consumidor encontrará un mejor proveedor, redundando en pérdidas de mercado.

Para obtener un producto de calidad a costo reducido y de buena calidad, se debe diseñar el proceso para producir según el valor del objetivo, y no para producir dentro de las especificaciones. Para minimizar las pérdidas debidas a variación con respecto al objetivo, la variabilidad del producto debe ser reducida, como se muestra en la Fig. 2.1. El producto *A* tiene mayor proporción de artículos cerca del objetivo, por lo tanto es una producción con menores pérdidas

A fin de encontrar efectiva y económicamente los valores de las variables de entrada que optimicen el proceso, se requiere de apropiados diseños experimentales. Las pruebas de diseños experimentales indicarán las variables que deben ser controladas precisamente y cuales de ellas afectarán significativamente en las condiciones finales del producto.

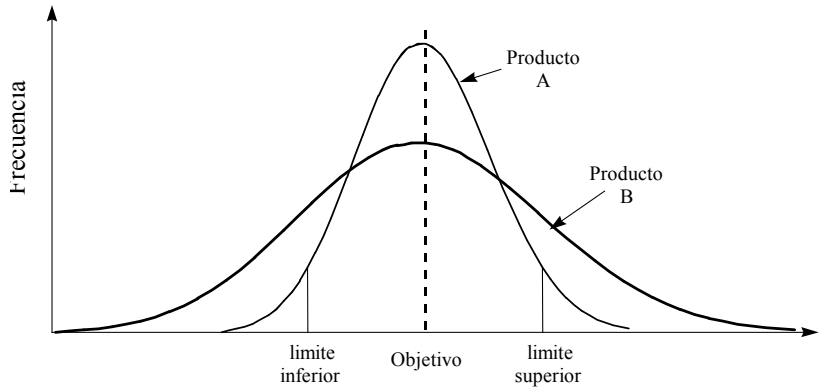



Figura 2.1: Alta calidad implica reducción de la varianza alrededor del objetivo.

## 2.3. Medidas de variación

La localización o tendencia central no necesariamente proporciona información suficiente para describir datos de manera adecuada. Las medidas de la variación de los datos de uso más común son el ***rango, varianza*** y la ***desviación estándar.*** Por ejemplo, considérese los datos de resistencia a la tensión (en psi) de dos muestras, *X, Y*, de una aleación metálica:

| | | | | | | |
|---|---|---|---|---|---|---|
| Muestra *X* | 130 | 150 | 145 | 158 | 165 | 140 |
| Muestra *Y* | 90 | 128 | 205 | 140 | 165 | 160 |

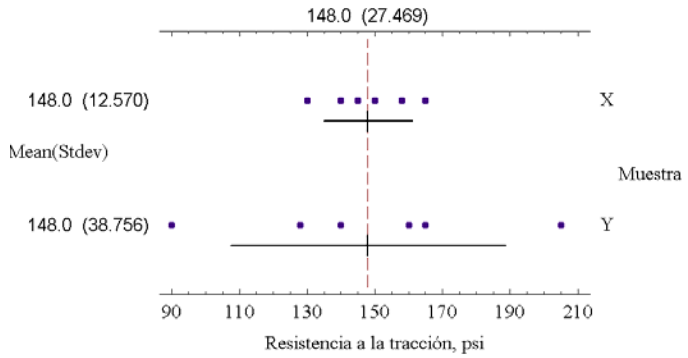



Figura 2.2: Diagrama de puntos de la resistencia a la tracción.

La media de ambas muestras es 148 psi, sin embargo, en la Figura 2.2 se observa mayor *variabilidad* para la muestra *Y.*

La más simple medida de variación es el ***rango,*** definido como la diferencia entre la mayor y la menor de las observaciones. La medida más común de la variabilidad es la ***varianza*** de muestra, $\mathbf{s^2}$, y la desviación estándar de muestra, **s**, que es la raíz cuadrada de la varianza. El cuadrado de la distancia $|x_i - \bar{x}|$ que es $(x_i - \bar{x})^2$ para las $x_i$ observaciones con respecto al promedio general $\bar{x}$ proporciona alguna información sobre la variabilidad. La varianza de muestra es un promedio especial del cuadrado de estas distancias. Para una muestra de *n* determinaciones, $x_1, x_2,...,x_n$ se define como:

$$s^2 = \frac{1}{n-1}\left[\left(x_1 - \bar{x}\right)^2 + \left(x_2 - \bar{x}\right)^2 + ... + \left(x_n - \bar{x}\right)^2\right] = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$$

y para una población con *n* determinaciones, se define como:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

A pesar de haber definido la varianza como un promedio de *n* distancias, la sumatoria correspondiente se divide entre *(n-1)*. Para justificar este procedimiento tengamos en cuenta que las *n* desviaciones $(x_i - \bar{x})$ sumadas entre si dan cero. Por ello se necesitan solo *(n-1*) de estas desviaciones para calcular $\boldsymbol{s^2}$. Además, el utilizar el divisor *(n-1*) en lugar de *n* no representa una importante diferencia numérica si la cantidad de observaciones es lo suficientemente grande.

Una varianza $s^2 \geq 0$ mide la separación de las observaciones alrededor del promedio. Si $\boldsymbol{s^2} = 0$ se dice que no hay ninguna variación porque las *n* observaciones deben ser las mismas. Ya que $\boldsymbol{s^2}$ es un promedio de cuadrados, sus unidades son el cuadrado de las unidades de medición de *x*. Si por ejemplo *x* es medida en Kg., $\boldsymbol{s^2}$ tiene unidades de $Kg^2$.

La raíz cuadrada de la varianza de muestra proporciona la medida conocida como la desviación estándar de muestra:

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

Del ejemplo de los bloques de concreto tenemos que para las diez primeras observaciones: $\bar{x}$ = 45.41; $s^2$ = 82.62 y *s* = 9.1 y para las 90 observaciones: $\bar{x}$ = 45.51; $s^2$ = 58.74 y *s* = 7.66

En el ejemplo de los bloques de concreto, Tabla 1.2, para *n* = 10 observaciones, se encuentra que 7 de las 10 observaciones caen dentro a una desviación estándar del promedio; las otras tres (31.3, 32.3 y 60.9) caen fuera de estos límites. En general, se observa que *algo más de la mitad* de todas las observaciones que se

efectúen se encontraran dentro del intervalo ( $\bar{x} \pm s$ ). Más de tres cuartos de las observaciones caerán en el intervalo ( $\bar{x} \pm 2s$ ) y ***todas*** las observaciones caerán en el intervalo **(** $\bar{x} \pm 3s$ **).**

2.3.1 Coeficiente de variación

Es válido comparar las desviaciones estándar si deseamos comparar las dispersiones de dos ó más series que tienen la misma o casi la misma media y que se expresan con la misma unidad. Sin embargo, hay casos en que diferentes distribuciones pueden tener diferentes medias o se expresan en diferentes unidades. Así resulta difícil comparar las desviaciones estándar. Por ejemplo, la desviación estándar de la aleación *X* y la aleación *Y* de la Figura 2.2. son 12.57 y 38.76 respectivamente. Los valores absolutos de estas variaciones alrededor de los promedios (300 % mayor una que la otra) no proporcionan la base de una comparación adecuada. Para estos propósitos, la medida más adecuada es el cálculo del ***coeficiente de variación CV*** que expresa la desviación estándar como porcentaje del promedio, según:

$$CV = \frac{s}{\bar{x}} \quad \acute{o} \quad \%\, CV = \frac{s}{\bar{x}} * 100$$

Esta medida indica que las observaciones caen, en promedio, a aproximadamente *CV* % del promedio. Se usa para comparar distribuciones con diferentes unidades o para comparar las dispersiones de dos distribuciones diferentes. En el ejemplo de las 10 primeras observaciones de los bloques de concreto se tiene que:

$$\%\, CV = \frac{100\ s}{\bar{x}} = \frac{100 * 9.1}{45.41} = 20.04$$

así, las observaciones que están a una desviación estándar del promedio están a 20 % del valor de la media aritmética. En el caso de las aleaciones *X* , *Y*, Fig. 2.2, se tiene:

$$CV_X = \frac{100 * 12.57}{148.0} = 8.49 \qquad CV_Y = \frac{100 * 38.76}{148.0} = 26.19$$

Ejemplo: En la determinación del análisis químico de una solución de lixiviación, se realizó por dos métodos; uno por espectrofotometría de absorción atómica y otro por volumetría. Se obtuvieron los siguientes resultados: a) Por la A. A. un promedio de 3,92 g/L de cobre con una desviación estándar de 0,015; y b) por volumetría, un promedio de 1,54 g/L y una desviación estándar de 0,008. ¿Cuál de los métodos es relativamente más preciso?.

$$CV_{AA} = \frac{0{,}015 * 100}{3{,}92} = 0,38\ \% \quad CV_{Vol} = \frac{0{,}008 * 100}{1{,}54} = 0{,}52\ \%$$

Por lo tanto, las mediciones realizadas con el método de la absorción atómica son relativamente más precisas.

2.3.2 ¿Cómo mide la varianza muestral la variabilidad?

Considérese la Figura 2.3, que ilustra las variaciones de los valores medidos de resistencia y las desviaciones $x_i - \bar{x}$. Entre más grande sea la variabilidad en los datos, mayor será la magnitud de las desviaciones $x_i - \bar{x}$. Puesto que la suma de las desviaciones $x_i - \bar{x}$ siempre es cero, se debe utilizar una medida de la variabilidad que cambie las desviaciones negativas en cantidades positivas. El elevar al cuadrado las desviaciones es el método que emplea la varianza muestral. En consecuencia, si $s^2$ es pequeña, entonces existe una variabilidad pequeña en los datos, pero si $s^2$ es grande, entonces la variabilidad también lo es.

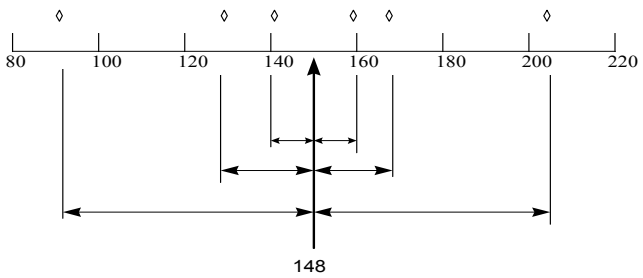



Figura 2.3: Manera en que la varianza mide la variabilidad mediante las desviaciones $x_i - \bar{x}$

De los datos de la Figura 2.2, se tiene:

$$\frac{1}{n-1}\sum_{i=1}^{6}\left(X_i-\overline{X}\right)^2=158=S_X^2 \qquad \frac{1}{n-1}\sum_{i=1}^{6}\left(Y_i-\overline{Y}\right)^2=1502=S_Y^2$$

Naturalmente, la muestra *Y* presenta mayor dispersión o variabilidad de los datos, lo cual se observa en el diagrama de puntos y en los valores calculados de la varianza.

2.3.3. Ejemplo de aplicación

El ingeniero jefe de una planta concentradora de hierro recibió quejas respecto al producto de la planta de peletización referidas a la falta de uniformidad en la distribución del tamaño de los pellets. Es importante mantener lo más uniforme posible el tamaño de los pellets para asegurar un adecuado funcionamiento del posterior proceso de tratamiento de tales pellets.

Se desea evaluar el problema y sugerir modificaciones con el uso de técnicas estadísticas. Primeramente, se recolecta información de los reportes de producción de planta durante los pasados meses. Se obtienen valores de la producción pellets cuyo 80% pasa por determinada malla, Tabla 2.1 (a).

Si la especificación del producto es 80 % malla 45 μm ± 5 μm, que modificaciones se propondrían al ingeniero jefe en base al análisis estadístico de la información obtenida?

Tabla 2.1 (a) Datos de los meses anteriores para la producción de Pellets; malla por la que pasa el 80 % de la producción diaria. Datos tal como se obtuvieron.

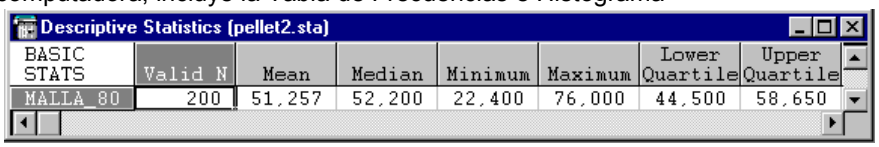
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 49.4 | 46.2 | 69.5 | 61.1 | 45.5 | 34.1 | 50.9 | 56.8 | 54.1 | 45.0 |
| 32.9 | 43.7 | 50.4 | 54.8 | 32.8 | 50.7 | 49.8 | 71.5 | 60.5 | 41.8 |
| 53.2 | 30.2 | 61.3 | 35.0 | 65.2 | 50.7 | 76.0 | 68.5 | 30.5 | 49.5 |
| 54.7 | 31.8 | 58.1 | 30.2 | 47.0 | 45.0 | 35.8 | 54.7 | 52.7 | 66.3 |
| 41.3 | 31.8 | 64.4 | 59.3 | 49.9 | 39.3 | 48.5 | 66.5 | 50.0 | 54.3 |
| 73.3 | 60.8 | 43.4 | 65.0 | 51.3 | 59.9 | 59.5 | 48.0 | 54.2 | 59.3 |
| 53.5 | 51.6 | 49.3 | 53.1 | 60.8 | 70.2 | 39.8 | 56.1 | 54.1 | 54.1 |
| 65.5 | 59.9 | 72.1 | 46.8 | 53.0 | 39.2 | 51.1 | 38.4 | 54.6 | 45.2 |
| 50.8 | 34.5 | 54.6 | 41.2 | 45.6 | 52.1 | 56.7 | 60.8 | 55.9 | 43.8 |
| 36.8 | 62.8 | 50.1 | 59.9 | 53.9 | 62.5 | 67.2 | 33.2 | 60.0 | 37.2 |
| 64.8 | 58.7 | 73.0 | 45.5 | 40.1 | 44.7 | 51.0 | 45.8 | 55.9 | 55.2 |
| 54.1 | 50.1 | 47.3 | 58.3 | 54.7 | 54.2 | 44.0 | 62.3 | 61.5 | 58.8 |
| 35.9 | 49.4 | 35.5 | 55.8 | 37.8 | 52.2 | 54.2 | 62.2 | 32.6 | 55.0 |
| 34.0 | 38.6 | 59.9 | 46.3 | 43.6 | 47.9 | 60.2 | 57.4 | 54.4 | 53.5 |
| 58.8 | 47.1 | 56.1 | 62.0 | 52.2 | 38.0 | 56.5 | 36.6 | 58.0 | 45.9 |
| 54.8 | 37.2 | 48.1 | 29.4 | 52.9 | 58.8 | 50.5 | 53.5 | 43.9 | 58.0 |
| 41.9 | 58.6 | 57.8 | 58.6 | 38.8 | 43.8 | 47.0 | 49.7 | 39.4 | 54.8 |
| 55.1 | 42.1 | 45.6 | 50.9 | 74.0 | 48.4 | 51.3 | 65.8 | 45.7 | 28.8 |
| 69.7 | 58.8 | 50.7 | 37.4 | 43.1 | 41.9 | 46.4 | 53.6 | 60.0 | 62.9 |
| 50.2 | 58.1 | 43.6 | 58.0 | 44.3 | 62.6 | 61.7 | 57.1 | 60.5 | 22.4 |

Es muy útil para el análisis que sigue, ordenar los datos obtenidos, Tabla 2.1 (b). Esta operación permitirá identificar el valor máximo y el valor mínimo y definir el rango como una medida de la variabilidad. La mediana y los cuartiles también son fácilmente identificables con este ordenamiento.

Tabla 2.1 (b) Datos de los meses anteriores para la producción de Pellets; malla por la que pasa el 80 % de la producción diaria. Datos ordenados.

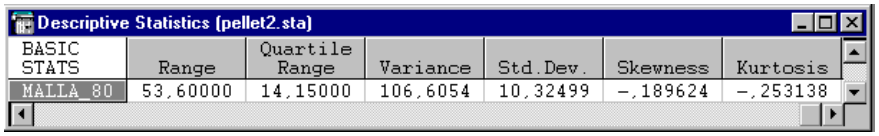
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 22.4 | 36.8 | 43.1 | 45.9 | 49.9 | 52.2 | 54.6 | 57.4 | 59.9 | 62.9 |
| 28.8 | 37.2 | 43.4 | 46.2 | 50.0 | 52.7 | 54.6 | 57.8 | 59.9 | 64.4 |
| 29.4 | 37.2 | 43.6 | 46.3 | 50.1 | 52.9 | 54.7 | 58.0 | 60.0 | 64.8 |
| 30.2 | 37.4 | 43.6 | 46.4 | 50.1 | 53.0 | 54.7 | 58.0 | 60.0 | 65.0 |
| 30.2 | 37.8 | 43.7 | 46.8 | 50.2 | 53.1 | 54.7 | 58.0 | 60.2 | 65.2 |
| 30.5 | 38.0 | 43.8 | 47.0 | 50.4 | 53.2 | 54.8 | 58.1 | 60.5 | 65.5 |
| 31.8 | 38.4 | 43.8 | 47.0 | 50.5 | 53.3 | 54.8 | 58.1 | 60.5 | 65.8 |
| 31.8 | 38.6 | 43.9 | 47.1 | 50.7 | 53.5 | 54.8 | 58.3 | 60.8 | 66.3 |
| 32.6 | 38.8 | 44.0 | 47.3 | 50.7 | 53.5 | 55.0 | 58.6 | 60.8 | 66.5 |
| 32.8 | 39.2 | 44.3 | 47.9 | 50.7 | 53.6 | 55.1 | 58.6 | 60.8 | 67.2 |
| 32.9 | 39.3 | 44.7 | 48.0 | 50.8 | 53.9 | 55.2 | 58.7 | 61.1 | 68.5 |
| 33.2 | 39.4 | 45.0 | 48.1 | 50.9 | 54.1 | 55.8 | 58.8 | 61.3 | 69.5 |
| 34.0 | 39.8 | 45.0 | 48.4 | 50.9 | 54.1 | 55.9 | 58.8 | 61.5 | 69.7 |
| 34.1 | 40.1 | 45.2 | 48.5 | 51.0 | 54.1 | 55.9 | 58.8 | 61.7 | 70.2 |
| 34.5 | 41.2 | 45.5 | 49.3 | 51.1 | 54.1 | 56.1 | 58.8 | 62.0 | 71.5 |
| 35.0 | 41.3 | 45.5 | 49.4 | 51.3 | 54.2 | 56.1 | 59.3 | 62.2 | 72.1 |
| 35.3 | 41.8 | 45.6 | 49.4 | 51.3 | 54.2 | 56.5 | 59.3 | 62.3 | 73.0 |
| 35.8 | 41.9 | 45.6 | 49.5 | 51.6 | 54.2 | 56.7 | 59.5 | 62.5 | 73.3 |
| 35.9 | 41.9 | 45.7 | 49.7 | 52.1 | 54.3 | 56.8 | 59.9 | 62.6 | 74.0 |
| 36.6 | 42.1 | 45.8 | 49.8 | 52.2 | 54.4 | 57.1 | 59.9 | 62.8 | 76.0 |

Con estos datos, se obtiene los siguientes cuadros con información obtenida mediante el uso de un programa estadístico de computadora; incluye la Tabla de Frecuencias e Histograma



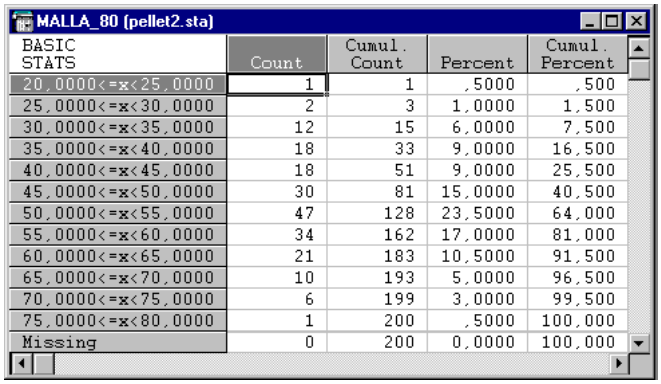
Descriptive Statistics (pellet2.sta)

| BASIC STATS | Valid N | Mean | Median | Minimum | Maximum | Lower Quartile | Upper Quartile |
|---|---|---|---|---|---|---|---|
| MALLA_80 | 200 | 51,257 | 52,200 | 22,400 | 76,000 | 44,500 | 58,650 |



Descriptive Statistics (pellet2.sta)

| BASIC STATS | Range | Quartile Range | Variance | Std.Dev. | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| MALLA_80 | 53,60000 | 14,15000 | 106,6054 | 10,32499 | -,189624 | -,253138 |



MALLA_80 (pellet2.sta)

| BASIC STATS | Count | Cumul. Count | Percent | Cumul. Percent |
|---|---|---|---|---|
| 20,0000<=x<25,0000 | 1 | 1 | ,5000 | ,500 |
| 25,0000<=x<30,0000 | 2 | 3 | 1,0000 | 1,500 |
| 30,0000<=x<35,0000 | 12 | 15 | 6,0000 | 7,500 |
| 35,0000<=x<40,0000 | 18 | 33 | 9,0000 | 16,500 |
| 40,0000<=x<45,0000 | 18 | 51 | 9,0000 | 25,500 |
| 45,0000<=x<50,0000 | 30 | 81 | 15,0000 | 40,500 |
| 50,0000<=x<55,0000 | 47 | 128 | 23,5000 | 64,000 |
| 55,0000<=x<60,0000 | 34 | 162 | 17,0000 | 81,000 |
| 60,0000<=x<65,0000 | 21 | 183 | 10,5000 | 91,500 |
| 65,0000<=x<70,0000 | 10 | 193 | 5,0000 | 96,500 |
| 70,0000<=x<75,0000 | 6 | 199 | 3,0000 | 99,500 |
| 75,0000<=x<80,0000 | 1 | 200 | ,5000 | 100,000 |
| Missing | 0 | 200 | 0,0000 | 100,000 |

Figura 2.4: Resultados obtenidos por computadora de los datos de pellets

El histograma de la Figura 2.5 muestra la distribución de los datos bajo la figura de una campana, lo que demuestra una distribución normal de la producción de los pellets. Es evidente también la simetría mostrada alrededor del intervalo [50, 55]. Este intervalo a su vez incluye al promedio de esta muestra.

El intervalo de especificación de producto pelletizado es [40, 50] μm. De la Tabla 2.1(b) se puede observar que $X_{33}$ = 39.8 y $X_{83}$ = 50.1 cuyos percentiles son 16.25% y 41.25% respectivamente. De esto último se deduce que el 16.25% de la producción se encuentra por debajo del límite inferior de especificación y el (100 - 41.25) = 58.75% de la producción se encuentra por encima del límite superior de especificación. Es decir, que el (16.25 + 58.75) = 75 % de los pellets son producidos fuera de los límites de especificación.

Si bien, la simple determinación del promedio de tamaños nos da una indicación de la calidad de producción, ese valor no es suficiente para el análisis de la producción. El dato más importante tal vez, es el que nos indica que 75 de cada 100 pellets son producidos fuera de las especificaciones requeridas.

Como conclusiones del análisis realizado, se recomienda reducir el tamaño de descarga de los pellets a tamaños alrededor de 45 μm y reducir la variabilidad en el tamaño del producto, de modo que se produzcan pellets más ajustados a los límites de especificación.

*Histograma: Distribución de la Producción de Pellets*

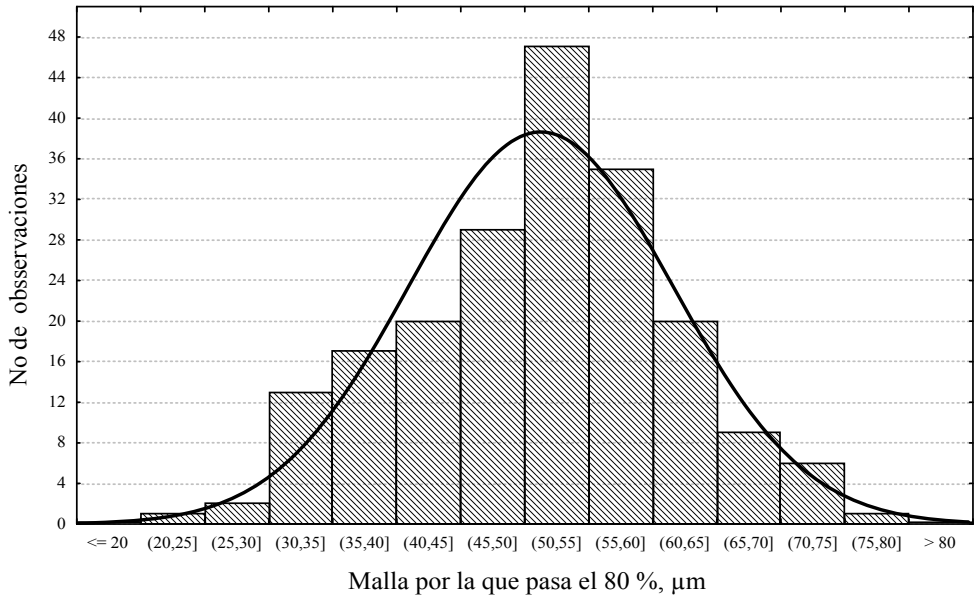



Figura 2.5: Histograma de la producción de la planta pelletizadora de hierro.

## 2.4. Diagramas de caja

El diagrama de caja (ó diagrama de caja y bigote) es una representación visual que describe al mismo tiempo varias características importantes de un conjunto de datos, tales como el centro, la dispersión, la desviación de la simetría y la identificación de observaciones que se alejan de manera poco usual del resto de los datos.

El diagrama de caja presenta en un rectángulo los tres cuartiles, y los valores mínimo y máximo de los datos. El rectángulo delimita el rango intercuartil con la arista izquierda (ó inferior) ubicada en el primer cuartil, $q_1$, y la arista derecha (ó superior) en el tercer cuartil, $q_3$. Se dibuja una línea a través del rectángulo en la posición que corresponde al segundo cuartil, $q_2$ ó mediana. De cualquiera de las aristas del rectángulo se extiende una línea ó bigote que va hacia los valores extremos. Estas son observaciones que se encuentran entre 0 y 1.5 veces el rango intercuartil a partir de las aristas del rectángulo. Las observaciones que están 1.5 a 3 veces el rango intercuartil a partir de las aristas del rectángulo reciben el nombre de valores atípicos. Las observaciones que están más allá de tres veces el rango intercuartil se conocen como valores atípicos extremos

Por ejemplo, para los datos de la Tabla 2.1, el diagrama de caja se muestra en la Figura 2.6 obtenido mediante el uso de un programa estadístico de computadora. Se observa bastante simetría con respecto al valor central, lo cual da una indicación adicional de la distribución normal de los datos obtenidos. En este caso no se presentan valores atípicos.

Los diagramas de caja son muy útiles para hacer comparaciones gráficas entre conjuntos de datos, ya que tienen un gran impacto visual y son fáciles de comprender. Por ejemplo en la Figura 2.7 se presentan los

diagramas comparativos para un índice de calidad de fabricación de dispositivos electrónicos de tres plantas distintas. El examen de este diagrama revela que existe mucha variabilidad en la planta 2 y que es necesario que las plantas 2 y 3 incrementen sus índices de calidad.

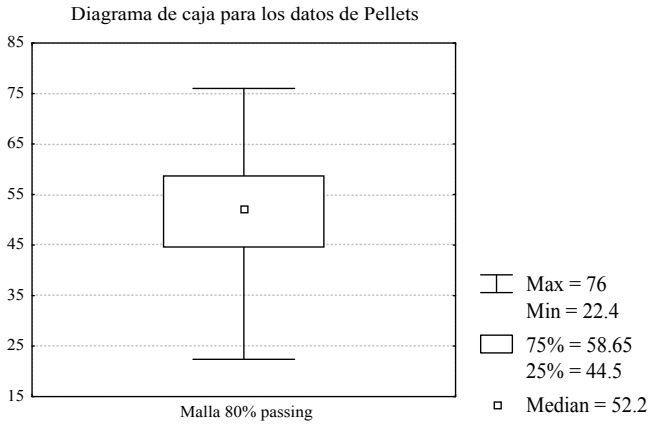



Figura 2.6: Diagrama de caja para los datos de la pelletizadora de mineral de hierro.

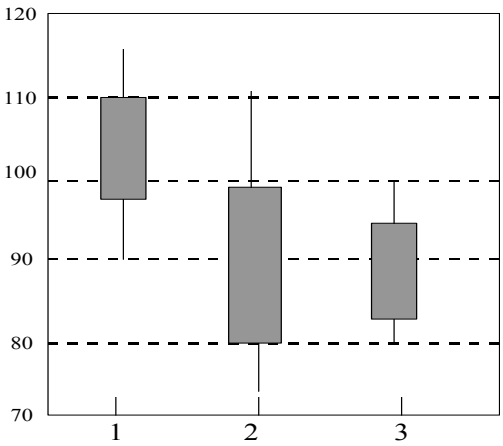



Figura 2.7: Diagramas de caja comparativos de un índice de calidad de tres plantas distintas.

2.5. Otros tipos de medición.

Existen dos tipos de mediciones que se utilizan mucho para analizar un grupo de datos. Se trata de la Asimetría o Sesgo (Skewness en Inglés) y la Curtosis (Kurtosis en Inglés).

a. Sesgo ó Asimetría.

El sesgo o grado de asimetría, es la falta de simetría de una distribución. Si la curva de frecuencias de una distribución tiene una cola más larga a la derecha del máximo central que a la izquierda, se dice que la distribución esta sesgada a la derecha o que tiene sesgo positivo. Si es al contrario se dice que esta sesgada a la izquierda o tiene sesgo negativo.

En distribuciones sesgadas, la media tiende a situarse con respecto a la moda al mismo lado que la cola más larga. Este parámetro proporciona un valor que indica falta de simetría en los datos. Su formula general es:

$$a_3 = \frac{m^3}{S^3} = \frac{\dfrac{\sum_{i=1}^{h}\left(X_i - \overline{X}\right)^3}{n}}{S^3}$$

La asimetría es un número que indica el grado de desviación de la simetría. Si el valor de $a_3$ es cero, los datos son simétricos; si es mayor de cero (positivo), los datos se inclinan hacia la derecha, lo cual significa

que la base larga está a la derecha; si es menor de cero (negativo), los datos se inclinan hacia la izquierda, es decir, que la base larga de la curva está a la izquierda.

b. Curtosis.

Es la medida de agudeza de los datos; normalmente se toma en relación a la distribución normal. Una distribución que presenta un apuntamiento relativo alto, tal como la de la curva de la Figura 2.8(a), se llama *leptocurtica,* mientras que la curva de la Figura 2.8(b), que es más achatada, se llama *platicúrtica.* La distribución normal de la Figura 2.8(c), que ni es muy apuntada ni achatada, se llama *mesocúrtica*.

(a) Leptocúrtica (b) Platicúrtica (c) Mesocúrtica

Figura 2.8: Curtosis en las distribuciones.

Una medida de curtosis emplea el momento de cuarto orden con respecto a la media, expresado en forma adimensional y dado por :

$$a_4 = \frac{m^4}{S^4} = \frac{\dfrac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^4}{n}}{S^4}$$

En el caso de una distribución normal, mesocúrtica, el valor de curtosis es $a_4 = 3.0$. Si $a_4 > 3.0$, la altura de la distribución es mayor de lo normal, es decir, es leptocúrtica. Si $a_4 < 3.0$ la altura de la distribución es inferior que la normal o platicúrtica. Existen algunos paquetes de software para normalizar los datos a cero, restándole a la respuesta 3.

## ***EJERCICIOS GRUPO 2***

1. La fuerza de tensión (en lb/pulg$^2$) de unas muestras de fibra sintética son:

   12 15 18 16 15 14 16 17.

   a) Construir el diagrama de puntos.

   b) Calcular el promedio de muestra, varianza y desviación estándar.

2. A continuación se presentan los datos de consumo de $O_2$ (oxígeno) por el salmón (mm$^3$/hr):

| | | | |
|---|---|---|---|
| 105 | 95 | 94 | 112 |
| 83 | 80 | 96 | 93 |
| 69 | 71 | 108 | 75 |
| 94 | 84 | 102 | 94 |

   Calcular la media, mediana, moda, varianza y desviación estándar.

3. Los siguientes son los números de imperfecciones observadas en 50 muestras, tomadas de rollos de telas:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 0 | 4 | 4 | 1 | 4 | 0 | 3 | 2 | 0 |
| 0 | 1 | 1 | 1 | 0 | 1 | 2 | 4 | 1 | 1 |
| 1 | 5 | 2 | 2 | 5 | 3 | 4 | 0 | 4 | 0 |
| 0 | 0 | 3 | 0 | 1 | 4 | 2 | 1 | 2 | 0 |
| 3 | 1 | 3 | 4 | 2 | 0 | 5 | 6 | 3 | 2 |

   a) Calcular la media, varianza, desviación estándar, mediana y moda.

   b) Agrupar los datos en una tabla de frecuencias, mostrando clases, límites, frecuencias, frecuencias acumuladas y % acumulado.

   c) Construir una ojiva para los datos agrupados y calcular $q_1$ *y* $q_3$

4. Los siguientes datos son mediciones de viscosidad de un producto químico tomadas cada hora:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 47.9 | 47.9 | 48.6 | 48.0 | 48.4 | 48.1 | 48.0 | 48.6 |
| 48.8 | 48.1 | 48.3 | 47.2 | 48.9 | 48.6 | 48.0 | 47.5 |
| 48.6 | 48.0 | 47.9 | 48.3 | 48.5 | 48.1 | 48.0 | 48.3 |
| 43.2 | 43.0 | 43.2 | 43.1 | 43.0 | 42.9 | 43.6 | 43.3 |
| 43.0 | 42.8 | 43.1 | 43.2 | 43.6 | 43.2 | 43.5 | 43.0 |

   a) Construya un histograma y un diagrama de caja para esta serie de datos.

   b) ¿Cuáles son los percentiles 90 y 10 de estos datos?

   c) Las especificaciones sobre la viscosidad del producto son 48 $\pm$ 2, ¿qué conclusiones puede obtener sobre el desempeño del proceso?

5. Un fabricante de aleaciones metálicas esta preocupado por las quejas de sus clientes acerca de la falta de uniformidad en el punto de fusión de filamentos metálicos producidos. 50 filamentos se seleccionaron y sus puntos de fusión determinados. Los siguientes son los resultados obtenidos:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 320 | 325 | 314 | 314 | 313 | 329 | 320 | 329 | 317 | 316 |
| 331 | 326 | 328 | 312 | 308 | 327 | 316 | 308 | 321 | 319 |
| 322 | 320 | 325 | 319 | 318 | 305 | 314 | 329 | 323 | 327 |
| 323 | 335 | 320 | 318 | 310 | 313 | 328 | 330 | 322 | 310 |
| 324 | 324 | 318 | 317 | 322 | 324 | 320 | 324 | 311 | 317 |

   a) Construir la distribución de frecuencias y mostrar el histograma.

   b) Calcular el promedio, mediana, varianza de muestra y desviación estándar.

   c) Cuantas observaciones caen dentro del limite de un desviación estándar alrededor del promedio?. Dentro de dos desviaciones estándar?

6. Se ha medido el espesor de orejas de recipientes de pintura. Se toma 5 muestras a intervalos periódicos desde el recipiente en donde se colecta la producción de dos máquinas. Se mide el espesor de las orejas producidas. Los resultados (en pulgadas multiplicados por 100) de 30 muestras son los que se anotan en la **Tabla 6a**. (ver abajo)

   a) Construir un histograma e interpretar la forma del histograma. Utilizar 15 clases.

   b) Experimentar con diferentes valores de, k, (clases) en dos histogramas diferentes.

      k= 6; k = 10. Interpretar las diferencias en los tres histogramas construidos.

7. Los números $x_1,x_2,...,x_k$, se presentan con frecuencias $f_1,f_2,...,f_n$, donde $f_1 + f_2, +...+,f_n = n$ es la frecuencia total. Hallar la media geométrica de los números.

8. La siguiente Tabla muestra la distribución de la carga máxima en toneladas que soportan ciertos cables producidos por una compañía. Determinar la media de la carga máxima.

| Máximo de Carga (Ton.) | Nro. de cables |
|---|---|
| 9.3 - 9.7 | 2 |
| 9.8 - 10.2 | 5 |
| 10.3 - 10.7 | 12 |
| 10.8 - 11.2 | 17 |
| 11.3 - 11.7 | 14 |
| 11.8 - 12.2 | 6 |
| 12.3 - 12.7 | 3 |
| 12.8 - 13.2 | 1 |

9. En algunos conjuntos de datos se aplica una transformación matemática a los datos originales, tal como raíz cuadrada de y ó log(y), que pueden dar como resultado datos con los que es más fácil trabajar desde el punto de vista estadístico. Para ilustrar el efecto de una transformación, considere los siguientes datos, los cuales representan el número de vueltas transcurridas antes de una falla para un producto de hilo:

   675 3650 175 1150 290 2000

   a. Construya un diagrama de caja y comente la forma que tiene la distribución.

   b. Transforme los datos utilizando logaritmos; o sea: $y^* = log(y)$ ($y$ es el valor original). Construya un diagrama de caja para los datos transformados y comente el efecto de la transformación.

10. Remítase a la bibliografía y describa, con ejemplos, las aplicaciones de la media geométrica, media harmónica, mediana y moda.

11. Describa, con ejemplos, la forma de calcular la media geométrica, media harmónica, mediana y moda a partir de datos agrupados.

12. Calcule el sesgo y curtosis para los datos del problema Nro. 5

**Tabla 6a: Espesor de orejas de recipientes de pintura**

| Muestra | M e d i c i o n e s | | | | |
|---|---|---|---|---|---|
| 1 | 29 | 36 | 39 | 34 | 34 |
| 2 | 29 | 29 | 28 | 32 | 31 |
| 3 | 34 | 34 | 39 | 38 | 37 |
| 4 | 35 | 37 | 33 | 38 | 41 |
| 5 | 30 | 29 | 31 | 38 | 29 |
| 6 | 34 | 31 | 37 | 39 | 36 |
| 7 | 30 | 35 | 33 | 40 | 36 |
| 8 | 28 | 28 | 31 | 34 | 30 |
| 9 | 32 | 36 | 38 | 38 | 35 |
| 10 | 35 | 30 | 37 | 35 | 31 |
| 11 | 35 | 30 | 35 | 38 | 35 |
| 12 | 38 | 34 | 35 | 35 | 31 |
| 13 | 34 | 35 | 33 | 30 | 34 |
| 14 | 40 | 35 | 34 | 33 | 35 |
| 15 | 34 | 35 | 38 | 35 | 30 |
| 16 | 35 | 30 | 35 | 29 | 37 |
| 17 | 40 | 31 | 38 | 35 | 31 |
| 18 | 35 | 36 | 30 | 33 | 32 |
| 19 | 35 | 34 | 35 | 30 | 36 |
| 20 | 35 | 35 | 31 | 38 | 36 |
| 21 | 32 | 36 | 36 | 32 | 36 |
| 22 | 36 | 37 | 32 | 34 | 34 |
| 23 | 29 | 34 | 33 | 37 | 35 |
| 24 | 36 | 36 | 35 | 37 | 37 |
| 25 | 36 | 30 | 35 | 33 | 31 |
| 26 | 35 | 30 | 29 | 38 | 35 |
| 27 | 35 | 36 | 30 | 34 | 36 |
| 28 | 35 | 30 | 36 | 29 | 35 |
| 29 | 38 | 36 | 35 | 31 | 31 |
| 30 | 30 | 34 | 40 | 28 | 30 |