Medidas de Dispersión o Variabilidad

Aparte de conocer la tendencia central de un conjunto de datos, es indispensable conocer cuáles son sus diferencias entre sí, para esto es necesario determinar su variabilidad o dispersión, la cual es un elemento muy importante en cualquier estudio de capacidad de un proceso. A continuación, veremos cuatro formas en las que podemos medir la variabilidad.

La primera forma es utilizando la ***desviación estándar muestral***, la cual es una medida muy usada para determinar la variabilidad, ya que nos muestra qué tan dispersos están los datos con respecto a la media; la desviación estándar muestral se representa con la letra ***S*** y se calcula con la siguiente expresión:

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

Donde *x*1, *x*2, *xn* son las observaciones numéricas de la muestra, *n* su tamaño y es la media muestral. Como podemos ver, ***S*** mide la distancia que hay en "promedio" entre los datos y la media; es por eso que entre más grande sea el valor de ***S***, mayor será la variabilidad en los datos.

La desviación estándar se representa empleando las mismas unidades de medición que los datos estudiados (ya sea que se trate de gramos, milímetros, etc.). Es importante recordar que ***S*** no muestra la magnitud de los datos, sino que refleja lo retirado que están los datos de la media y, al igual que esta, es afectada por datos atípicos.

Desviación estándar poblacional o del proceso, σ

Cuando para calcular la desviación estándar se emplean todos los elementos de la población o proceso, lo que se obtiene es la desviación estándar poblacional y se representa con la letra griega sigma (σ). Como se comentó antes, es posible considerar a la población como las mediciones de toda la producción de las últimas semanas, o si las mediciones se toman por muestras, entonces una buena idea es obtener los parámetros poblacionales (μ y σ) con todas las mediciones realizadas en las últimas semanas, siempre y cuando estas no sean pocas; de 120 a 150 mediciones en adelante es una buena cantidad.

Por otra parte, el cuadrado de la desviación estándar, $S2$, conocido como *varianza muestral*, es muy importante para propósitos de inferencia estadística. Y en forma equivalente $\sigma 2$ es la varianza (o variancia) poblacional.

Otra de las medidas importantes de dispersión es lo que se conoce como ***rango*** o ***recorrido***, ***R***, que es igual a la diferencia entre el dato mayor y el dato menor de un conjunto de datos. El rango mide la amplitud de la variación de un grupo de datos y también es independiente de la magnitud de los datos; por ejemplo, sean los dos conjuntos de datos:

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

Podemos observar que la magnitud de los datos es diferente, y eso es reflejado por la media, que es de 12 y 161, respectivamente. Pero en cuanto a la variabilidad, los datos de ambos conjuntos de datos están dispersos de la misma manera, 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* o ***CV***, es una medida de variación que es relativa a la magnitud de los datos, ya que es igual a la magnitud relativa de la desviación estándar en comparación con la media de los datos, es decir:

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

El *CV* nos sirve para comparar la variación de dos o más variables que están medidas o expresadas en diferentes escalas o unidades de medición (por ejemplo, metro frente a centímetro o metro frente a kilogramo). Este coeficiente de variación suele entenderse 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 acaban de ver en la definición anterior de rango, se tiene que sus correspondientes *CV* son:

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

Por lo tanto, 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%.

Volviendo al primer ejemplo (2.1), en el caso del grosor de los discos, tenemos que S =0.027, S^2=0.0007, R=1.25 -1.11 =0.14, y CV=2.29%.

Podemos interpretar el rango de manera muy directa, ya que nos indica la amplitud máxima de la dispersión; de esta forma 0.14 mm es la discrepancia máxima que existió entre los grosores de los discos en la muestra.

De manera general, la desviación estándar se puede interpretar en combinación con la media, como lo veremos en seguida, ya que su interpretación en forma individual se realiza en forma comparativa con respecto a la desviación estándar de otras líneas de producción o lotes. Es indispensable considerar, en caso de hacer estas comparaciones, que lo que se observa en una muestra es variable, y que cuando existen pequeñas diferencias muestrales no implican diferencias entre procesos o lotes.

Ahora, respecto al coeficiente de variación o *CV*=2.29%, nos muestra que la variación del grosor es de 2.29%, lo cual se puede interpretar como un porcentaje relativamente bajo.

Límites reales o naturales

Los *límites reales* o *naturales* de un proceso nos muestran los puntos entre los cuales varía la salida de un proceso y se obtienen de la siguiente forma:

Límite real inferior (*LRI*)=$\mu - 3\sigma$ y Límite real superior (*LRS*)=$\mu + 3\sigma$

Cuando se realiza un estudio de capacidad, los límites reales se comparan con las especificaciones deseadas de una característica de calidad. Por ejemplo, si las especificaciones para una característica de calidad son que debe tener dimensiones de 800±5; luego, la especificación inferior es *EI* =795 (800-5), y la superior es *ES*=805 (800+5). Y si conocemos la media y la desviación estándar de tal característica de calidad son: μ=800.6 y σ=1.2, respectivamente, entonces los limites reales son:

LRI=800.6 −3(1.2)=**797.0** y ***LRS***=800.6 + 3(1.2)= **804.2**

Como resultado, se espera que esta característica de calidad varíe de 797.0 a 804.2, con una media de 800.6. Al comparar el resultado con las especificaciones, se aprecia que los límites reales están dentro de las mismas, entonces es posible deducir que el proceso es capaz de cumplir con tales especificaciones.

REFERENCIA:

Gutiérrez, H. y De la Vara, R. (2009). Control estadístico de calidad y seis sigma. Recuperado de: https://www.uv.mx/personal/ermeneses/files/2018/05/6-control-estadistico-de-la-calidad-y-seis-sigma-gutierrez-2da.pdf