Medidas de Tendencia Central

Cuando se desea realizar la medición de una característica de calidad, el primer dato a investigar debe ser la *tendencia central* de los datos, en otras palabras, identificar un valor en torno al cual los datos tienden a aglomerarse o concentrarse. Esto nos permitirá conocer si el proceso está centrado; es decir, 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 1.20).

A continuación, veremos las siguientes medidas de tendencia central: la media, la mediana y la moda, utilizando el ejemplo 2.1.

EJEMPLO 2.1

Veamos el ejemplo que sucede durante un proceso de inyección de plástico una característica de calidad del producto es su grosor de disco, que debe ser de 1.20 mm con una tolerancia de ±0.10 mm.

Para poder considerar que el proceso de inyección cumple con los requerimientos, el grosor del disco debe estar entre la EI o especificación inferior, EI=1.10 y la ES o especificación superior, ES=1.30. Para realizar un estudio de capacidad para este proceso, es necesario contestar preguntas como las siguientes: ¿qué tipo de discos y en cuanto a grosor se están produciendo? ¿El grosor medio es adecuado? ¿La variabilidad del grosor es mucha o poca?

Para poder contestar estas preguntas, durante una semana se recabaron de una línea de producción los 125 datos de la tabla 2.1. El muestreo fue sistemático: cada cierto tiempo se tomaban cinco productos y se medían y al finalizar la semana se recolectaron los datos referidos.

A continuación, se analizarán los datos recabados a traves de diferentes *medidas de tendencia central*.

TABLA 2.1 Datos para el grosor de los discos, ejemplo 2.1.

1.15	1.20	1.17	1.16	1.16	1.15	1.17	1.20	1.16	1.19	1.17	1.13	1.15	1.20	1.18	1.17	1.16
1.20	1.17	1.17	1.20	1.14	1.19	1.13	1.19	1.16	1.18	1.16	1.17	1.15	1.21	1.15	1.20	1.18
1.17	1.17	1.13	1.16	1.16	1.17	1.20	1.18	1.15	1.13	1.20	1.17	1.19	1.23	1.20	1.24	1.17
1.17	1.17	1.17	1.18	1.24	1.16	1.18	1.16	1.22	1.23	1.22	1.19	1.13	1.15	1.15	1.22	1.19
1.18	1.19	1.17	1.16	1.17	1.18	1.19	1.23	1.19	1.16	1.19	1.20	1.17	1.13	1.22	1.19	1.21
1.20	1.19	1.17	1.19	1.22	1.19	1.18	1.11	1.19	1.19	1.17	1.19	1.17	1.20	1.16	1.19	1.20
1.20	1.17	1.25	1.16	1.16	1.20	1.20	1.16	1.18	1.21	1.20	1.22	1.19	1.14	1.19	1.17	1.20
1.16	1.15	1.20	1.12	1.11	1.18											

MEDIA MUESTRAL $\bar{X}$

Obtenemos la media maestral sumando todos los datos y el resultado de la suma se divide entre el número de datos (*n*).

Considerando que $x1$, $x2$, $x3$,..., xn son las observaciones numéricas de una muestra, entonces, la medida más usual de su tendencia central es proporcionada por la *media* (o promedio) muestral, que es igual a la media aritmética de todos los datos:

$$\overline{X} = \frac{x_1 + x_2 + \dots x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

En nuestro ejemplo 2.1, al obtener la media muestral, la media de los datos de la tabla 2.1 es $\bar{x}$=1.179 mm, esto significa que el grosor promedio de los discos de la muestra es de 1.179 mm. Sin embargo, esto no significa que todos o la mayoría de los discos tengan un grosor de 1.179 mm, en el ejemplo, ningún disco tiene ese grosor. En este caso, debido a que la media muestral procede de una muestra significativamente grande que abarca el periodo de una semana, entonces hay evidencia de que el proceso esta descentrado de forma moderada a la izquierda o hacia un valor inferior, ya que el valor objetivo para el grosor es de 1.20 mm.

MEDIA POBLACIONAL O DEL PROCESO $\boldsymbol{\mu}$

Como ya sabemos, para calcular la media se utilizan todos los elementos de la población (todos los individuos, especímenes, objetos o medidas de interés sobre los que se hace un estudio), por ejemplo, el grosor de todos los discos producidos en la última semana o mes; entonces, por otro lado, el promedio calculado es la media del proceso (o media poblacional) y se denota con la letra griega $\boldsymbol{\mu}$ (mu).

Es importante destacar que la *media del proceso* μ 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 las piezas que se seleccionan ($\bar{x}$ es una variable aleatoria). Por lo anterior, el valor que se observa de la media muestral, $\bar{x}$, por lo general es diferente a la media del proceso μ. Luego, es preciso tener cuidado con las afirmaciones basadas en $\bar{x}$ sobre la media del proceso o población.

MEDIANA O PERCENTIL 50

Veamos otra medida de tendencia central para un conjunto de datos llamada *mediana* $\tilde{x}$ la cual divide a la mitad a los datos cuando son ordenados de menor a mayor. Calcularla es muy sencillo, cuando tenemos un numero de datos impar, estos se ordenan de manera creciente y el que quede en medio de dicho ordenamiento será la mediana. Pero en el caso de que el numero de datos sea par, entonces la mediana se calcula dividiendo entre dos la suma de los dos números que están en el centro del ordenamiento.

Volviendo al ejemplo 2.1, la mediana es 1.18 mm, lo cual significa que 50% de los grosores de los discos de la muestra son menores o iguales a 1.18, y que el otro 50% son mayores o iguales a 1.18.

MODA

La moda es otra forma para medir la tendencia central de un conjunto de datos, la moda no es más que el dato que se repite más veces en el conjunto. Si varios datos se presentan con la frecuencia más grande, entonces cada uno de ellos es una moda y se dice que el conjunto de datos es *multimodal*.

En nuestro ejemplo del grosor de los discos hay una sola moda y es 1.17 debido a que esta medición fue la más frecuente, se repitió 23 veces. De esta manera, en nuestro ejemplo tenemos que la media es 1.179, la mediana 1.18 y la moda 1.17. Debido a que la media es la medida de tendencia central más empleada, en ocasiones se tiene confusión al creer que esta divide los datos a la mitad o que es el dato más frecuente; es decir, se puede confundir el concepto de media con el de mediana y moda, respectivamente.

Algo muy importante cuando se utiliza la media, es que esta resulta afectada por datos extremos o atípicos. Por ejemplo, la media y la mediana para los siguientes datos:

1 100, 1 300, 1 000, 1 500, 800, 1 600, 1 100.

Son $\bar{x}$ =1 200 $\tilde{x}$ =1= 100. Pero si a la lista anterior añadimos un dato atípico (el 7 600), entonces: $\bar{x}$ = 2 000 y $\tilde{x}$ =1 200 los resultados cambian, debido a que 7 600 ha modificado a la media, y ahora ya no es una buena medida de tendencia central porque solo un dato está por arriba de la media. En este tipo de casos, la mediana no es afectada por el dato atípico, lo cual tampoco ocurre cuando la distribución de los datos es sesgada. Por lo tanto, podemos decir que la mediana es una de las mejores medidas de tendencia central.

Sin embargo, para describir la tendencia central de cualquier serie de datos, es necesario apoyarse tanto en la media como en la mediana y la moda. En los casos en los que la media es muy diferente a la mediana, es señal de que existen datos atípicos o existe un sesgo importante, por lo que se recomienda reportar como medida de tendencia central a la mediana e indagar la causa de los datos atípicos, ya que en ocasiones reflejan un aspecto importante del proceso.

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