Manejo de la información para sistemas de gestión

Variación

Servicio Nacional de Aprendizaje SENA

Variación

La variación es el grado en que los datos numéricos tienden a extenderse alrededor de un valor, que en su mayoría es el valor medio.

¿Por qué es importante la variación?

1. A menudo, una medida de posición de un conjunto de datos se vincula con la indicación de representatividad para la población. Se requiere contar con la información que proporcionan las medidas de variación. Además, el conocimiento de un estadístico de tendencia central no aclara o define toda la distribución, por lo que es significativo tener una idea de la dispersión de los valores y determinar cómo se ubican alrededor de la media.

2. La medida de tendencia central no indica relación entre datos, por esta razón es necesario conocer las medidas de variabilidad o dispersión.

3. Al tratar problemas con datos dispersos se requiere conocer sus implicaciones y el punto en que la dispersión tiene un riesgo aceptable o inaceptable en la toma de decisiones.

4. Al comparar dos distribuciones se enfatiza en la posición central y en la dispersión.

Medidas de dispersión utilizadas con mayor frecuencia

Rango o recorrido (amplitud): mide la dispersión de la totalidad de los datos, es la distancia entre los valores máximo y mínimo. El rango o recorrido permite comprender el grado de variación en la población, con frecuencia los resultados pueden ser engañosos, pues depende de los valores extremos e ignora la variación de las demás observaciones. Está afectado por ocurrencias raras o extraordinarias.

Entre las propiedades del rango se encuentran: la facilidad de calcular, sus unidades son las mismas que las de la variable, no utiliza todas las observaciones (sólo dos de ellas), se puede ver muy afectado por alguna observación extrema, y aumenta con el número de observaciones, o bien se queda igual, pero en ningún caso nunca disminuye.

La varianza (S): la varianza es la media de los cuadrados de las diferencias entre cada valor de la variable y la media aritmética de la distribución.

$$S_x^2 = \frac{\sum_{j=1}^{n}\left(x_j - Media\left(x\right)\right)^2}{n}$$

Se obtiene al sumar las diferencias de cuadrados, y tiene como unidades de medida el cuadrado de las unidades de medida de la variable.

Desviación típica: cuando se utiliza la varianza como medida de dispersión, surge el problema de tener que trabajar con distintas dimensiones, por ello se define la desviación típica como la raíz cuadrada de la varianza.

$$S_x = \sqrt{\frac{\sum_{j=1}^{n}\left(x_j - Media\left(x\right)\right)^2}{n}}$$

¿Por qué es importante la variación?

1. Es afectada por el valor de cada observación
2. Como consecuencia de considerar desviaciones cuadráticas pone mayor énfasis en las desviaciones extremas.
3. Si en el eje X de la distribución de frecuencias normal se mide a ambos lados de la media una distancia igual a:

- Una desviación típica: se forma un intervalo en el cual se encuentra el 68,27 % de los valores centrales de la variable.
- Dos desviaciones típicas: se forma un intervalo donde se encuentra el 95,43 % de los valores centrales.
- Tres desviaciones típicas: se forma un intervalo que contiene el 99,73 % de los valores centrales.

4. Al construir la tabla de frecuencias de una variable discreta, y a partir de ella calcular la desviación típica; no hay pérdida de información porque la desviación para los datos observados es igual que para los datos tabulados.

Como medidas de variabilidad más importantes conviene destacar algunas características de la varianza y la desviación típica:

- Son índices que describen la variabilidad o dispersión; cuando los datos están muy alejados de la media el numerador de sus fórmulas será grande, por consiguiente la varianza y la desviación típica lo serán.
- Al aumentar el tamaño de la muestra disminuye la varianza y la desviación típica. Para reducir a la mitad la desviación típica la muestra se tiene que multiplicar por 4.
- Cuando todos los datos de la distribución son iguales, la varianza y la desviación típica son iguales a 0.
- Para su cálculo se utilizan todos los datos de la distribución; por tanto, cualquier cambio de valor será detectado.

Coeficiente de Variación: Cuando se necesita comparar dos o más series de datos, a veces no es posible hacerlo con las medidas absolutas, en esos casos deben utilizarse cantidades relativas. El coeficiente de variación se define como el cociente de la desviación típica entre la media aritmética y es adimensional. Frecuentemente se expresa en porcentaje.

$$CV = \frac{S}{\bar{x}}$$

¿Por qué es importante la variación?

1. Se debe calcular solo para ——variables con todos los valores positivos, y todo índice de variabilidad es esencialmente no negativo. Las observaciones pueden ser positivas o nulas, pero su variabilidad debe ser siempre positiva. De ahí que sólo se debe trabajar con variables positivas, para las que se tiene con seguridad que x > 0.

2. No es invariante ante cambios de origen. Es decir, si a los resultados de una medida se le suma una cantidad positiva, b > 0, para tener Y = X + b, entonces CV < CV .

3. Es invariante a cambios de escala. Así por ejemplo el coeficiente de variación de una variable medida en metros es una cantidad adimensional que no cambia si la medición se realiza en centímetros.