

Cálculo práctico de la varianza y la desviación típica

Para calcular la variación media, la varianza o la desviación típica usando la definición es necesario utilizar todas las desviaciones de los datos, que se basan en el cálculo previo de la media. Esto presenta algunos problemas prácticos:

- * La media fácilmente puede ser un número con muchas cifras decimales; en ese caso, las operaciones se complican y, además, se van acumulando errores.
- * Si cambia alguno de los datos, cambia la media, y por tanto hay que recalcular todas las desviaciones.

Por estos motivos, se utiliza para calcular la varianza y la desviación típica otro método más sencillo desde el punto de vista práctico, que se basa en una propiedad de la varianza.

Propiedad de la varianza

La varianza de un conjunto de datos es igual a la media de los cuadrados de los datos menos el cuadrado de la media.

Expresión simbólica

Consideramos los valores x_1, x_2, \dots, x_n ; llamamos \bar{x} a su media y σ^2 a su varianza. Entonces, se verifica:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum x_i^2}{n} - \bar{x}^2$$

Demostración

Para facilitarte la comprensión de la demostración general, comenzamos por la demostración en un caso más sencillo y la usamos de guía para la generalización.

Demostración para tres datos

Consideramos los datos a, b y c . Llamamos m a su media: $m = (a+b+c) : 3$.

Calculamos la varianza como la media de los cuadrados de las tres desviaciones:

$$\begin{aligned} \frac{(a-m)^2 + (b-m)^2 + (c-m)^2}{3} &= \frac{a^2 - 2am + m^2 + b^2 - 2bm + m^2 + c^2 - 2cm + m^2}{3} = \\ &= \frac{a^2 + b^2 + c^2 - 2m(a+b+c) + 3m^2}{3} = \frac{a^2 + b^2 + c^2 - 2m \cdot 3m + 3m^2}{3} = \\ &= \frac{a^2 + b^2 + c^2 - 6m^2 + 3m^2}{3} = \frac{a^2 + b^2 + c^2 - 3m^2}{3} = \frac{a^2 + b^2 + c^2}{3} - m^2 \end{aligned}$$

Vemos que hemos obtenido la media de los cuadrados de los datos menos el cuadrado de la media.

Demostración general

Llamamos \bar{x} a la media de los valores x_1, x_2, \dots, x_n . Sabemos que $\bar{x} = (\sum x_i) : n$

Calculamos la varianza como la media de los cuadrados de todas las desviaciones:

$$\begin{aligned} \frac{\sum (x_i - \bar{x})^2}{n} &= \frac{\sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n} = \frac{\sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2}{n} = \frac{\sum x_i^2 - 2\bar{x}n\bar{x} + n \cdot \bar{x}^2}{n} = \\ &= \frac{\sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2}{n} = \frac{\sum x_i^2 - n\bar{x}^2}{n} = \frac{\sum x_i^2}{n} - \bar{x}^2 \end{aligned}$$

Vemos que hemos obtenido la media de los cuadrados de los datos menos el cuadrado de la media.