Taller Formativo 5

Mario Castillo Sanhueza

Docente: Dr. Julio Rojas Mora

Departamento de Ingeniería Informática
Facultad de Ingeniería
Universidad Católica de Temuco

Agosto 11, 2022

1. Problemática.

1. ¿Qué sentido tiene elevar al cuadrado las diferencias?.
2. Demuestre que para un conjunto de datos $X = x_i, \forall_i = 1, ..., n$:

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

   es decir, si no se elevan al cuadrado las diferencias el valor promedio será siempre cero.
3. ¿Qué sentido tiene sacar la raíz cuadrada a la varianza?.
4. El rango intercuartílico es la versión robusta del rango. ¿Se imagina por qué?.
5. ¿Cuándo no debe usarse el coeficiente de variación?.
6. ¿Por qué se llama al rango intercuartílico relativo una versión robusta del coeficiente de variación?.

2. Desarrollo

1. El objetivo de elevar al cuadrado las diferencias es que si esto no se hace, la suma resulta en 0,dado que al calcular la varianza sin que el término interior posea un elemento al cuadrado, su suma resultante siempre es 0, a su vez al elevar estos elementos al cuadrado restringimos que el resultante sea un valor negativo.
2.

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \tag{2}$$

   Separamos la sumatoria de la forma:

$$\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}\right) \tag{3}$$

$$\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i - n \cdot \bar{x}\right) \tag{4}$$

$$\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i - n \cdot \frac{\sum_{i=1}^{n} \cdot x_i}{n}\right) \tag{5}$$

$$\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i\right) \tag{6}$$

Dado que restamos los mismos elementos obtenemos:

$$\frac{1}{n-1} \cdot 0 \tag{7}$$

De esta forma, si los elementos no son elevados al cuadrado, se llega a la situación en donde se restan mismos elementos y obtenemos como resultado 0.

3. Al calcular la varianza se debe elevar elementos al cuadrado como hemos explicado anteriormente, por esta razón para una homogeneidad absoluta al momento de observar la dispersión de los datos se requiere obtener la raíz y así evaluar en las unidades de medida iniciales de los datos.

4. Dado que esta trabaja con la distancia entre el primer y tercer cuartil, estos serian los extremos de los datos encontrados en la mediana, por ende podríamos especular que contendrá una mayor concentración de los datos totales de la muestra, dado que a diferencia del rango que solo trabaja con 2 datos y estos pueden ser extremos el intercuartílico trabaja con intervalos de datos asegurando así una mayor fidelidad a estas distancias.

5. Dado que es un cociente sin unidades, este no debe usarse cuando tenemos una necesidad explicita de obtener un resultado con una unidad, por ejemplo al calcular la fuerza con la que se repelen n cargas podríamos calcular el coeficiente de variación de 2 sistemas, pero si una está en μC (micro Coulomb) y la otra en C(Coulomb) la diferencia entre estas dos medidas está demasiado dispersa y no se establecería una comparación entre unidades homogéneas.

6. La diferencia en el uso de la mediana por el rango intercuartílico relativo, a diferencia de la media aritmética que posee una alta sensibilidad a valores extremos para el cálculo de esta misma, al usar

la mediana nos aseguramos de ubicarnos en la mitad de los datos independiente de los valores de estos.