

MEDIDAS DE DISPERSIÓN (VARIABILIDAD)

En el análisis estadístico no basta el cálculo e interpretación de las medidas de tendencia central o de posición, ya que, por ejemplo, cuando pretendemos representar toda una información con la media aritmética, no estamos siendo absolutamente fieles a la realidad, pues suelen existir datos extremos inferiores y superiores a la media aritmética, los cuales, en honor a la verdad, no están siendo bien representados por este parámetro. En dos informaciones con igual media aritmética, no significa este hecho, que las distribuciones sean exactamente iguales, por lo tanto, debemos analizar el grado de homogeneidad entre sus datos. Por ejemplo, los valores 5, 50, 95 tiene igual media aritmética, y mediana que los valores 49, 50,51; sin embargo, para la primera información la media aritmética, se encuentra muy alejada de los valores extremos 5 y 95, cosa que no ocurre con la segunda información que posee igual media aritmética y mediana, vemos entonces que la primera información es más heterogénea o dispersa que la segunda.

Para medir el grado de dispersión de una variable, se utilizan principalmente los siguientes indicadores:

- 1 Rango o recorrido
- 2 Desviación media
- 3 Varianza y desviación típica o estándar
- 4 Coeficiente de variabilidad.

1.- Rango o Recorrido

Es la medida de dispersión más sencilla ya que solo considera los dos valores extremos de una colección de datos, sin embargo, su mayor utilización está en el campo de la estadística no paramétrica.

$$R = V_{\max} - V_{\min}$$

V_{\max} y V_{\min} son los valores máximos y mínimos de la variable.

2.- DESVIACIÓN MEDIA

La desviación media, mide la distancia absoluta promedio entre cada uno de los datos, y el parámetro que caracteriza la información. Usualmente se considera la desviación media con respecto a la media aritmética

$$DM = \frac{\sum_{i=1}^m |x_i - \bar{x}| \cdot f_i}{n}$$

donde

DM = desviación media

x_i = valores de la variable

f_i = frecuencia absoluta de x_i

\bar{x} = media aritmética de la muestra

n = tamaño de la muestra

m = número de intervalos

Varianza y desviación típica

El problema de los signos en la desviación media, es eludido tomando los valores absolutos de las diferencias de los datos con respecto a la media aritmética. Ahora bien, la varianza obvia los signos elevando las diferencias al cuadrado, lo cual resulta ser más elegante, aparte de que es supremamente útil en el ajuste de modelos estadísticos que generalmente conllevan formas cuadráticas.

La varianza es uno de los parámetros más importantes en estadística paramétrica, se puede decir que, teniendo conocimiento de la varianza de una población, se ha avanzado mucho en el conocimiento de la población misma.

Numéricamente definimos la varianza, como desviación cuadrática media de los datos con respecto a la media aritmética:

$$s^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 f_i}{n}, \quad \text{donde}$$

$s^2 = \text{varianza}$

$x_i = \text{valor de la variable } x$

$f_i = \text{frecuencia absoluta}$

$n = \text{tamaño de la muestra}$

$m = \text{número de intervalos}$

Ejemplo de las baldosas

kg / cm² x_i f_i $x_i f_i$ $x_i - \bar{x}$ $(x_i - \bar{x})^2 f_i$

100-200	150	4	600	-298	355216
200-300	250	10	2500	-198	392040
300-400	350	21	7350	-98	201684
400-500	450	33	14850	2	132
500-600	550	18	9900	102	187272
600-700	650	9	5850	202	367236
700-800	750	5	3750	302	456020
Suma			44800		1959600

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{n}} = \sqrt{19596} = 140 \text{ kg/cm}^2$$

4. Coeficiente de variabilidad

Generalmente interesa establecer comparaciones de la dispersión, entre diferentes muestras que posean distintas magnitudes o unidades de medida.

El coeficiente de variabilidad tiene en cuenta el valor de la media aritmética, para establecer un número relativo, que hace comparable el grado de dispersión entre dos o más variables, y se define como:

$$CV = \frac{s}{x} \cdot 100$$

Ejemplo: en el caso de las baldosas, tenemos:

$$CV = \frac{140}{448} = 0,3125 \rightarrow C = 31,25\%$$