

MEDIDAS DE DISPERSIÓN

Un promedio no dice nada acerca de la diseminación de los datos. El promedio no es representativo cuando se tiene una amplia dispersión. Se puede comparar cuán dispersas están dos o más distribuciones.

Una medida de dispersión puede utilizarse para evaluar la confiabilidad de dos o más promedios.

- **Medidas de dispersión:**

Varias medidas de dispersión:

- **Amplitud de Variación:** Tal intervalo especial se utiliza ampliamente en las aplicaciones del control estadístico de procesos.

AMPLITUD DE VARIACIÓN = VALOR MÁS GRANDE – VALOR MÁS PEQUEÑO

- **Desviación media:** Es la media aritmética de los valores absolutos de las desviaciones con respecto a la media aritmética.

$\sum (X - \bar{X})$

DESVIACIÓN MEDIA DM =

n

Donde:

X valor de cada observación

$\bar{X}$ media aritmética de los valores

n número de observaciones de la muestra

// valor absoluto

No se consideran los signos de las desviaciones respecto de la media dado que las desviaciones positivas y negativas se compensarían exactamente y la desviación media siempre sería cero. Cero es un dato estadístico inútil.

Dos ventajas:

- Utiliza en su cálculo todos los valores de la muestra.
- Fácil de comprender pues es el promedio en que los valores se desvían con respecto a la media.

Desventaja:

- El uso de valores absolutos.
- **Variancia y Desviación Estándar:** se basan en las desviaciones al cuadrado con respecto a la media.

Variancia: la media aritmética de las desviaciones cuadráticas con respecto al a media

La **Variancia Poblacional** para datos no agrupados o para datos no tabulados en una distribución de frecuencias se calcula como:

$$\sum (X - \bar{X})^2$$

=

N

Donde:

 símbolo de variancia poblacional

X valor de una observación de población

N número total de observaciones en la citada población

 media aritmética de la población

La **variancia** es difícil de interpretar a causa de las unidades.

La **Desviación Estándar Poblacional** es el promedio de las desviaciones respecto de la media, se presenta en las mismas unidades que los datos. Fórmula:

$$\sqrt{\sum (X - \bar{X})^2}$$

=

N

$$\sqrt{\sum (X - \bar{X})^2}$$

Variancia Muestral s² =

$$\frac{\sum (X - \bar{X})^2}{n - 1}$$

Donde:

X valor de las observaciones en la muestra

X media de la muestra

N número total de observaciones de la muestra

Puede demostrarse que:

$$\sum (X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n}$$

Fórmula operativa de la variancia muestral:

$$\sum (X - \bar{X})^2 / n$$

$$s^2 =$$

$$n - 1$$

La **Desviación Estándar Muestral** se utiliza como un estimador, es la raíz cuadrada de la variancia muestral:

$$\sum (X - \bar{X})^2 / n$$

$$s =$$

$$n - 1$$

• **Medidas de dispersión para datos agrupados en distribución de frecuencias:**

- **Amplitud de Variación**: se resta el límite inferior de la clase más pequeña del límite superior de la clase mayor.
- **Desviación estándar**: para datos no agrupados. Fórmula:

$$\sum fX^2 - (\sum fX)^2 / n$$

$$s =$$

$$n - 1$$

Donde:

$\sum$ desviación estándar muestral

$\bar{X}$ punto medio de la clase

f frecuencia de clase

n número total de observaciones en la muestra

Para encontrar la desviación de los datos agrupados en una distribución de frecuencias se produce:

Paso 1. Cada frecuencia de clase se multiplica por su punto medio.

Paso 2. Se calcula $\sum fX^2$, esto podría explicarse como $\sum fX \cdot X$.

Paso 3. Se suman las columnas $\sum fX$ y $\sum fX^2$.

• **Interpretación y usos de la desviación estándar:**

La desviación estándar se emplea como una medida para comparar la dispersión en dos o más conjuntos de observaciones.

- ♦ **Teorema de Chebyshev**: Para un conjunto cualquiera de observaciones (muestra o población), la proporción mínima de los valores que se encuentran dentro de k desviaciones estándares desde la media es al menos $1 - 1/k^2$, donde k es una constante mayor que 1.

- ♦ **Regla empírica:** Para una distribución de frecuencias simétrica de campana, aproximadamente 68% de las observaciones estará a más y menos una desviación estándar desde la media, aproximadamente 95% de tales observaciones se encontrará a más y menos dos desviaciones estándares de la misma; y prácticamente todas las observaciones (99,7%) se hallarán a más y menos tres desviaciones con respecto a la media.

Forma gráfica:

Curva simétrica de campana que muestra las relaciones entre la desviación estándar y la media

-3s -2s -1s X 1s 2s 3s

70 80 90 100 110 120 130

68%

95%

99,7%

Si una distribución es simétrica con forma de campana, prácticamente todas las observaciones se encuentran entre la media más o menos tres desviaciones estándares.

- **Dispersión Relativa:**

Karl Pearson (1857–1936) desarrolló una medida relativa denominada coeficiente de variación(CV). Es una medida útil cuando:

- Los datos están en unidades diferentes(como U\$S y días de asistencia).
- Los datos están en la mismas unidades, pero las medias muy distantes (ingresos de superiores e ingresos de empleados).

- ♦ **Coeficiente de variación:** es la razón (cociente) de la desviación estándar a la media aritmética, expresada como un porcentaje:

s

CV = (100)

X

Karl Pearson desarrolló tb una medida para evaluar el grado de orientación al sesgo, denominada **coeficiente de asimetría (CA):**

3 (media – mediana)

CA =

Desviación Estándar

- **Otras medidas de dispersión:**

Un método es determinar la ubicación de los valores que dividen un conjunto de observaciones en partes iguales. Estas medidas son:

- ♦ Los **cuartiles**, que dividen un conjunto de observaciones en 4 partes iguales (conjuntos ordenados de menor a mayor). El **primer cuartil (Q1)** es el valor abajo del cual se encuentra el 25% de las observaciones, y, el **tercer cuartil (Q3)** es el valor por abajo del cual se encuentra el 75% de las observaciones. **Q2** es la mediana.
- ♦ Los **deciles** dividen un conjunto de observaciones en 10 partes iguales.
- ♦ Los **centiles** se utilizan para reportar resultados acerca de ciertas pruebas nacionales estandarizadas, empleado para calificar la admisión a programas.
- **Cuartiles, Deciles y Centiles (o Porcentiles):**

Para formalizar el procedimiento, sea **L_p** la ubicación del centil deseado.

Ej: percentil 33 L₃₃ . El número de observaciones es **n**. Entonces se aplica: **(n +1) /2**

Ubicación de un centil $L_p = (n +1) P/100$

- ♦ **Diagramas de caja:** representación gráfica basada en cuartiles, que ayuda a ilustrar un conjunto de dato. Se necesitan 5 valores estadísticos: el valor mínimo; Q1 ; la mediana; Q3 ; y el valor máximo.

Mediana

Q1 Q3

Valor mínimo Valor Máximo

La distancia entre los extremos de la caja se denomina **amplitud cuartílica (o intercuartílica)**. Dicho intervalo es la distancia entre el primero y el tercer cuartiles.

Se indican dos asteriscos (**) . Uno indica n dato impropio. Un **dato incongruente** es un valor inconsciente con el resto de los datos. Es como aquel valor que más de 1,5 veces el valor de la amplitud intercuartílica, mayor que Q3 o bien, menor que Q1.

Dato incongruente = $Q1 - 1,5 (Q3 - Q1)$