

# Medidas de dispersión

- También llamadas medidas de variación, muestran la variabilidad de una muestra, indicando mediante un número, si las diferentes puntuaciones de una variable están muy alejadas de la media o no. Cuanto mayor sea el valor de cualquiera de las medidas de dispersión, mayor será la variabilidad de los datos y por tanto la muestra se denomina heterogénea. En el caso contrario, en que estas medidas sean “pequeñas”, la muestra será homogénea y las observaciones estarán cercanas al valor del promedio.
- Al calcular el promedio por lo general es común que éste vaya acompañado por alguna medida de dispersión, ya que la media en sí no nos brinda información, por ejemplo, de la existencia de valores extremos, mientras que las medidas de dispersión sí lo hacen. De aquí la importancia de estos parámetros ya que de alguna manera evalúan la confiabilidad o no de las medidas de tendencia central.
- Resumiendo, si las medidas de dispersión son “pequeñas”, las medidas de tendencia central son válidas y representativas de la muestra. Caso contrario no son tan confiables.

# Medidas de dispersión

- **Rango (R):** Es la diferencia entre los dos valores extremos de la variable, esto es:

$$R = \text{Max} - \text{Min}$$

- Es la medida de dispersión más sencilla de calcular y también la que menos información proporciona, pues solo se calcula teniendo en cuenta el valor de dos observaciones de la muestra, lo cual puede provocar un análisis inadecuado.
- Informa sobre el recorrido de la variable al registrar de alguna manera la distancia entre el mínimo y el máximo de la variable.

- **Ejemplo:** Retomemos nuevamente el ejemplo de las calificaciones de los 50 estudiantes. La muestra es:

6- 12- 20 -17- 3- 6- 8- 10- 10- 10- 9- 14- 15- 16- 18- 17- 13- 11- 11- 9- 9- 4- 7- 14- 13- 12-  
10- 9- 13- 15- 19- 5- 19- 20 -3 - 5- 7- 11- 18- 17- 6 -7- 8- 14- 12- 18- 10- 9- 14- 16.

- Claramente, el mínimo es 3 y el máximo es 20. Luego  **$R = 20 - 3 = 17$** .

# Medidas de dispersión

- **Desviación estándar (s):** También llamada desviación típica, siempre será una medida positiva, aunque en algunos casos ya veremos que puede valer 0.
- Esta medida nos permite averiguar el promedio aritmético de fluctuación de los datos respecto a la media aritmética.
- Si bien existe una fórmula matemática para calcularla, es una de las medidas que se puede obtener usando el modo estadístico de cualquier calculadora científica.
- A continuación, les dejo un enlace que les mostrará cómo calcular la desviación estándar y la media teniendo en cuenta los dos modelos más comunes de calculadora:

<https://www.youtube.com/watch?v=4eXrcn6QLgs>

- En el video se habla de desviación poblacional y desviación muestral. Recordar que en este curso siempre trabajaremos con la desviación muestral.

# Medidas de dispersión

- Luego de ver el video, les pido que intenten calcular la desviación estándar para esta pequeña muestra, donde se muestran las edades en años de los niños en la sala de espera de un consultorio pediátrico:

5- 10- 12- 4- 3- 1- 1- 8

Deberías obtener en este caso,  $s = 3,8405$ .

- La unidad de la desviación estándar coincide con la unidad de la variable estadística que se esté estudiando.
- Esto es, por ejemplo, en el caso anterior, la desviación estándar es de 3,8405 años.



# Medidas de dispersión

Para pensar y discutir en el foro:

¿En qué situación, de acuerdo a la definición que vimos de la desviación estándar, consideras que este parámetro puede valer cero? ¿Por qué?



# Medidas de dispersión

- **Varianza ( $s^2$ ):** Es otra medida de dispersión, que calcula el promedio de las desviaciones de la media, elevadas al cuadrado.
- De manera más sencilla, podemos obtenerla elevando al cuadrado la desviación estándar  $s$ .
- Por ejemplo, si  $s = 3,8405$  años en el ejemplo de la variable que representaba las edades de los niños en un consultorio pediátrico, la varianza será en este caso:

$$s^2 = (3,8405 \text{ años})^2 = 14,75 \text{ años}^2$$

- Naturalmente, la unidad de medición para la varianza es la unidad de la variable al cuadrado, algo que pueden observar fácilmente a partir del cálculo anterior.
- La información que proporcionan la desviación estándar y la varianza es la misma desde un punto de vista práctico, aunque es más sencillo analizar la desviación estándar ya que está expresada en la misma unidad que la variable.

# Medidas de dispersión

- **Coeficiente de variación (CV):** Es una medida que proporciona información sobre la dispersión relativa del conjunto de datos y se calcula como el cociente entre la desviación estándar y su media aritmética. En la mayoría de los casos este resultado se multiplica por 100 para expresarlo como un porcentaje.

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

- Puesto que tanto la desviación estándar como la media “heredan” las unidades de medición de la variable, el CV es una medida **independiente** de las unidades originales de los datos.
- Debido a la propiedad anterior, el CV es el parámetro adecuado para comparar la variabilidad de dos muestras que podrían estar medidas en diferentes unidades.

# Medidas de dispersión

- **Ejemplo:** Supongamos que conocemos la media y la desviación estándar de dos muestras distintas. En ambas se estudia el peso de un grupo de personas que siguen un plan de alimentación determinado.

- **Muestra 1:** Los datos se miden en kg, y se sabe que  $\bar{x} = 75,5 \text{ kg}$  y  $s = 3,5 \text{ kg}$

Luego:

$$CV_1 = \frac{3,5 \cancel{\text{kg}}}{75,5 \cancel{\text{kg}}} \cdot 100\% = 0,046 \cdot 100\% = 4,6\%$$

- **Muestra 2:** Los datos se miden en g, y se sabe que  $\bar{x} = 54300 \text{ g}$  y  $s = 4800 \text{ g}$

$$CV_2 = \frac{4800 \cancel{\text{g}}}{54300 \cancel{\text{g}}} \cdot 100\% = 0,088 \cdot 100\% = 8,8\%$$

- Aquí debemos concluir que la muestra más heterogénea o con mayor variabilidad es la segunda ya que posee un coeficiente de variación más elevado.
- En la mayoría de los casos, esta medida será menor a 100 aunque los valores por encima de 100 indicarán que estamos en presencia de una muestra con una muy elevada dispersión.



# Medidas de dispersión

## Ejercicio para pensar:

- Dos laboratorios de medicamentos tienen maneras diferentes de pagar a sus visitadores médicos. La compañía A lo hace mediante un sueldo fijo mensual y la compañía B mediante un porcentaje sobre las ventas efectuadas. La distribución de los salarios (en pesos) es la siguiente:
- **Laboratorio A:** 42000- 35000- 54000- 74300- 29200- 43780- 95000-120320- 48500
- **Laboratorio B:** 10000- 8500- 2500- 25300- 7500- 4800- 9700- 2900- 6500

Usando las medidas que creas conveniente, determinar ¿en qué laboratorio el sueldo medio es más equitativo? ¿Por qué?