MEDIDAS DE VARIABILIDAD

Una desviación estándar baja no siempre significa que los datos están "bien" o "normales".

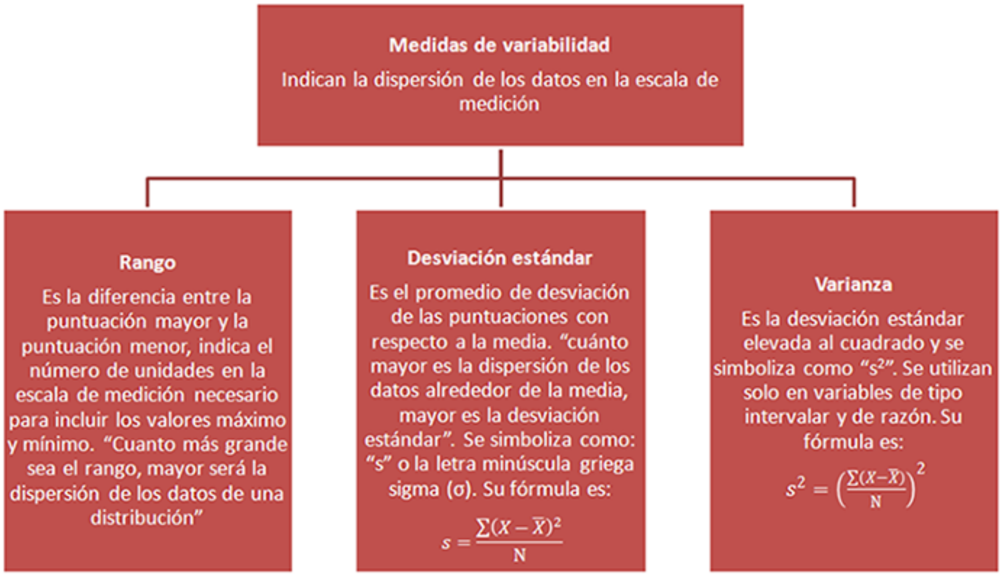
La desviación estándar es **0**, porque no hay variabilidad. Pero... en muchos contextos eso es **sospechoso** o **poco realista** y podría significar que:

- Los datos fueron manipulados
- Hubo errores en la recolección
- Las respuestas fueron copiadas
- O incluso fraude en encuestas

Medidas de variabilidad

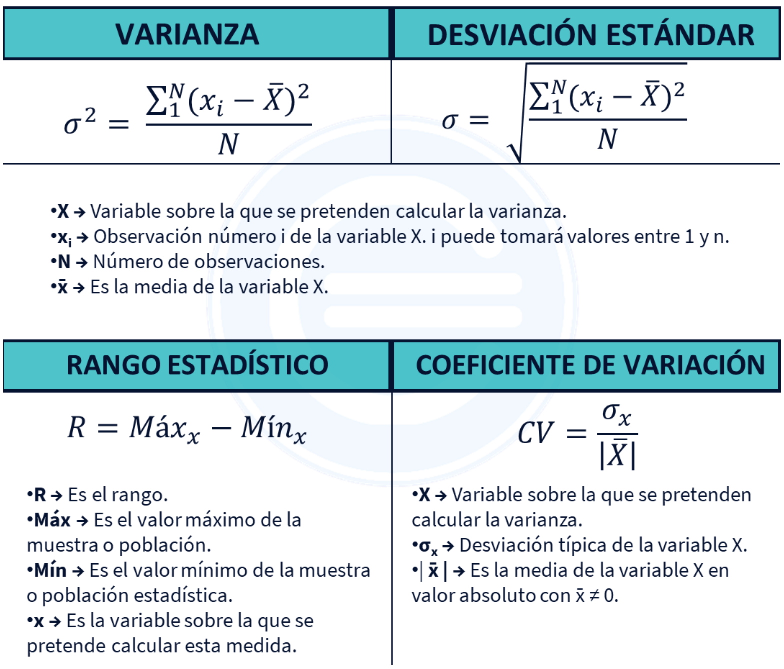
Las medidas de variabilidad o dispersión cuantifican el grado en que los datos de un conjunto se alejan entre sí o respecto a una medida central. Mientras que las medidas de tendencia central nos dicen *dónde está el centro*, las de variabilidad nos dicen *qué tan dispersos están alrededor de ese centro*. Es decir, muestran qué tan alejados o cercanos están los valores individuales con relación a la media, mediana u otra medida de posición.

Una variabilidad alta significa que los datos tienen valores muy diferentes entre sí; una baja variabilidad indica que los datos están más concentrados.



https://share.google/images/683TLi47WRsjZ9S4G

Las medidas de dispersión son fundamentales para entender la variabilidad de un conjunto de datos, ya que nos indican qué tan dispersos o concentrados están los valores respecto a la media o cualquier otra medida de tendencia central.



https://share.google/images/AIjRWPkjRnConhZbQ

Principales medidas de dispersión:

1. **Rango (R)**
   - Es la diferencia entre el valor máximo y el mínimo.
   - Fórmula:

$R = X_{máx} - X_{min}$

   - Mide la extensión total de los datos.

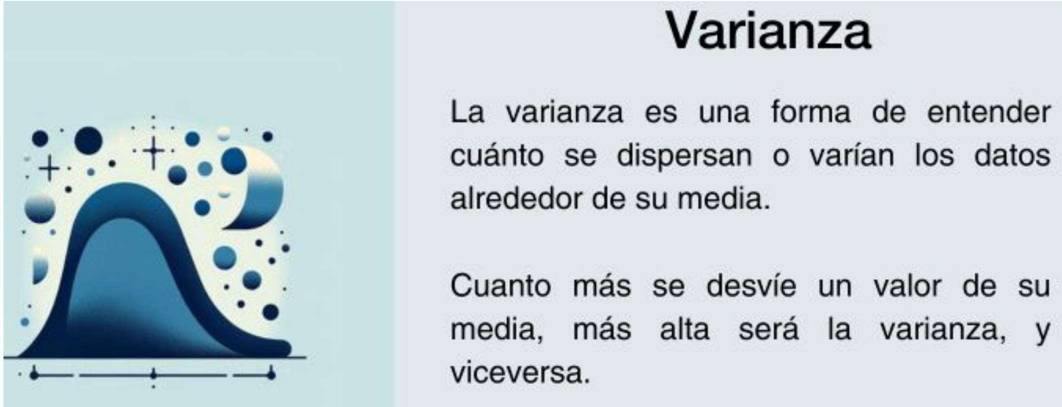
2. **Varianza (σ^2 o s^2)**
   - Mide el promedio de las diferencias al cuadrado respecto a la media.
   - Se calcula distinto si se trata de población o muestra.

3. **Desviación estándar (σ o s)**
   - Es la raíz cuadrada de la varianza.
   - Expresa, en las mismas unidades de los datos, cuánto se alejan los valores de la media.
   - Es la medida más usada de dispersión.

4. **Rango intercuartílico (RIC o IQR)**
   - Diferencia entre el tercer cuartil (Q3) y el primero (Q1).
   - Se usa para detectar valores atípicos.

Dos de las medidas de dispersión más comunes son la varianza y la desviación estándar. Ambas nos permiten analizar cuánto se alejan los datos de su media y son fundamentales en análisis estadísticos, investigación científica, procesos de control de calidad, educación y más.

VARIANZA



https://share.google/images/bE2g7vUGh31Zyfamd

📊 **¿Qué es?**

La **varianza** mide el promedio de las **diferencias al cuadrado** entre cada valor del conjunto de datos y la **media**. Es decir, indica **cuánto varían los datos respecto al valor promedio**.

🔷 **Fórmulas:**

- **Para una población completa**:

$\sigma^2 = (1 / N) \times \Sigma (x_i - \mu)^2$

Donde:

  - σ^2 = varianza poblacional
  - N = tamaño de la población
  - x_i = cada valor
  - μ = media poblacional

- **Para una muestra**:

$s^2 = (1 / (n - 1)) \times \Sigma (x_i - \bar{x})^2$

Donde:

  - s^2 = varianza muestral
  - n = número de datos
  - x_i = cada valor del conjunto
  - $\bar{x}$ = media aritmética
  - Σ = suma de todos los valores

Interpretación de la varianza

Una **varianza baja** indica que los datos están **muy cercanos entre sí y a la media**.

Una **varianza alta** sugiere que los datos están **muy dispersos o alejados de la media**.

Ejemplo:

Dos grupos de estudiantes tienen la misma media (85), pero en el Grupo A todos sacaron entre 83 y 87, mientras que en el Grupo B hubo calificaciones de 70 y 100.

Ambos tienen la misma media, pero el **Grupo B tendrá una varianza mucho mayor**.

📊 Varianza para datos no agrupados (σ^2)

La **varianza** mide el promedio de las distancias al cuadrado de cada dato con respecto a la media del conjunto de datos. Es decir, la varianza mide qué tan dispersos están los datos alrededor de la media. La varianza es igual a la desviación estándar elevada al cuadrado.

La fórmula para calcular la varianza de n datos es:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Donde:

- Xi es cada valor individual de los datos.
- μ es la **media** de los datos.
- n es el número total de observaciones.

Si los datos representan una muestra en lugar de toda la población, la fórmula se ajusta dividiendo entre n−1 en lugar de n, para corregir el sesgo en la estimación de la varianza de la población (s^2):

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Donde:

▯ es la media muestral.

Ejemplo:

Las calificaciones de 5 estudiantes son: 70, 75, 80, 85, 9

Paso 1: Calcular la media aritmética ($\bar{x}$)

Media ($\bar{x}$) = (70 + 75 + 80 + 85 + 90) ÷ 5 = 400 ÷ 5 = 80

Paso 2: Calcular la diferencia de cada dato con respecto a la media y elevar al cuadrado

Dato (x_i)	**Diferencia ($x_i - \bar{x}$)**	**Cuadrado de la diferencia ($(x_i - \bar{x})^2$)**
70	70 - 80 = -10	$(-10)^2 = 100$
75	75 - 80 = -5	$(-5)^2 = 25$
80	80 - 80 = 0	$0^2 = 0$
85	85 - 80 = 5	$5^2 = 25$
90	90 - 80 = 10	$10^2 = 100$

Paso 3: Sumar los cuadrados de las diferencias

100 + 25 + 0 + 25 + 100 = 250

Paso 4: Calcular la varianza muestral (s^2)

Varianza muestral (s^2) = Suma de los cuadrados ÷ (n - 1)

$s^2 = 250 \div (5 - 1) = 250 \div 4 = 62.5$

La varianza de las calificaciones es: $s^2 = 62.5$.

Interpretación

La varianza indica la dispersión de los datos respecto a la media. Para expresar esta dispersión en las mismas unidades que los datos originales, calculamos la desviación estándar, que es la raíz cuadrada de la varianza.

Desviación estándar (s) = $\sqrt{62.5} \approx 7.91$

Esto significa que, en promedio, las calificaciones se alejan aproximadamente 7.91 puntos de la media.

MEDIDAS DE DISPERSIÓN PARA DATOS AGRUPADOS

Cuando los datos están agrupados en clases o intervalos, se utilizan las frecuencias de las clases y los puntos medios de las mismas para calcular las medidas de dispersión.

Varianza para datos agrupados

La fórmula de la varianza en datos agrupados se calcula de manera similar a la de los datos no agrupados, pero utilizando los puntos medios de las clases (x_i) y las frecuencias de las clases (f_i):

$$\sigma^2 = \sum_{i=1}^{k} \frac{f_i (x_i - \mu)^2}{n}$$

Donde:

- *x_i es el punto medio de la i-ésima clase.*
- *f_i es la frecuencia de la i-ésima clase.*
- *n es el número total de observaciones ($n=\sum f_i$).*
- *μ es la **media** del conjunto de datos.*

Si se trata de una **muestra**, la fórmula se ajusta dividiendo por n−1 en lugar de n.

$$\sigma^2 = \sum_{i=1}^{k} \frac{f_i (x_i - \mu)^2}{n-1}$$

Ejemplo:

La altura en cm de los jugadores de un equipo de baloncesto está en la siguiente tabla. Calcular la varianza.

	x_i	f_i
$[160, 170)$	165	1
$[170, 180)$	175	2
$[180, 190)$	185	4
$[190, 200)$	195	3
$[200, 210)$	205	2

En primer lugar, rellenar la siguiente tabla:

	x_i	f_i	$x_i f_i$	$x_i^2 f_i$
$[160, 170)$	165	1	165	27225
$[170, 180)$	175	2	350	61250
$[180, 190)$	185	4	740	136900
$[190, 200)$	195	3	585	114075
$[200, 210)$	205	2	410	84050
		12	2250	423500

Se debe calcular la media

$$\bar{x} = \frac{2250}{12} = 187.5$$

para poder aplicar la fórmula.

Se calcula entonces la varianza

$$\omega^2 = \frac{423500}{12} - 187.5^2 = 135.42$$

https://www.sangakoo.com/es/temas/varianza-y-desviacion-tipica

DESVIACIÓN ESTÁNDAR

https://share.google/images/mPkgwYXvBcKwzd0F1

La desviación estándar es una medida que nos dice, en promedio, cuánto se alejan los datos con respecto a la media. Es una de las medidas de dispersión más utilizadas, ya que se expresa en las mismas unidades que los datos originales, lo que facilita su interpretación.

Fórmula (muestral):

$s = \sqrt{(s^2)}$

Donde:

- s = desviación estándar de una muestra
- s^2 = varianza muestral

¿Cómo se interpreta?

- Si **s es pequeña**, significa que los datos están **muy cerca de la media**.
- Si **s es grande**, indica que los datos están **muy dispersos**.

Ejemplo:

Usando el siguiente conjunto de calificaciones, calcula la desviación estándar.

70, 75, 80, 85, 90

media = 80

varianza muestral (s^2) = 62.5

Entonces:

Desviación estándar (s) = √62.5 ≈ 7.91

Interpretación del resultado: Esto significa que, en promedio, las calificaciones se desvían **7.91 puntos** respecto a la media. Este valor nos ayuda a entender **la consistencia o variabilidad** del rendimiento de los estudiantes.

Desviación Estándar para datos agrupados

La **desviación estándar** para datos agrupados es simplemente la raíz cuadrada de la **varianza**.

$$\sigma = \sqrt{\sigma^2}$$

Nota: Toma en cuenta que, si tienes una serie de valores de una muestra y necesitas calcular su varianza y su desviación estándar, deberás calcular primero la media poblacional x̄ con la siguiente fórmula:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Ejemplo: Calcula la varianza y desviación estándar para los siguientes datos agrupados:

Intervalo de clase	Frecuencia (fi)
0-10	5
10-20	8
20-30	12
30-40	10
40-50	5

Paso 1: Calcular los puntos medios:

Intervalo de clase	Frecuencia (fi)	Punto medio (x_i)
0-10	5	5
10-20	8	15
20-30	12	25
30-40	10	35
40-50	5	45

Paso 2: Calcular la media:

La media para los datos agrupados se calcula como:

$$\mu = \frac{\sum_{i=1}^{k} f_i * x_i}{n}$$

Donde n = 5+8+12+10+5=40

$$\mu = \frac{5(5) + 8(15) + 12(25) + 10(35) + 5(45)}{40} = 25.5$$

Paso 3: Calcular la varianza:

$$\sigma^2 = \frac{1}{40}[5(420.5) + 8(110.25) + 12(0.25) + 10(90.25) + 5(380.25)] = 144.725$$

Paso 4: Calcular la desviación estándar:

$$\sigma = \sqrt{\sigma^2} = \sqrt{144.725} = 12.02$$

Por lo tanto, la varianza es aproximadamente 144.73 y la desviación estándar es aproximadamente 12.02

📊 **Desviación Estándar para datos no agrupados**

La desviación estándar es simplemente la raíz cuadrada de la varianza. Al tomar la raíz cuadrada, convertimos las unidades de la varianza (que están al cuadrado) a las mismas unidades de los datos originales, lo que la hace más interpretable.

La fórmula para la desviación estándar de una población es:

$$\sigma = \sqrt{\sigma^2}$$

Mientras, que la fórmula para la desviación estándar de una muestra es:

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

Ejemplo: Supongamos que un grupo de 6 estudiantes obtuvo las siguientes calificaciones en un examen:

Datos (x_i): 72, 75, 80, 85, 90, 95

Paso 1: Calcular la media (promedio)

Se suman todos los datos y se dividen entre la cantidad total de datos:

Media (□) = (72 + 75 + 80 + 85 + 90 + 95) ÷ 6

Media (□) = 497 ÷ 6 = 82.83

Paso 2: Calcular la diferencia de cada dato con respecto a la media y elevarla al cuadrado

Dato (x_i)	**$x_i - \bar{x}$**	**$(x_i - \bar{x})^2$**
72	-10.83	117.32
75	-7.83	61.30
80	-2.83	8.00
85	2.17	4.70
90	7.17	51.41
95	12.17	148.18
Total	–	**390.91**

Paso 3: Calcular la varianza muestral

Varianza (s^2) = Suma de los cuadrados ÷ (n - 1)

$s^2 = 390.91 \div (6 - 1) = 390.91 \div 5 = 78.18$

Paso 4: Calcular la desviación estándar

Desviación estándar (s) = Raíz cuadrada de la varianza

$s = \sqrt{78.18} \approx 8.84$

Resultado

La desviación estándar de las calificaciones es 8.84 puntos aproximadamente.

Interpretación

Esto significa que, en promedio, las calificaciones se desvían 8.84 puntos respecto a la media, que fue de 82.83. Una desviación estándar moderada como esta indica que las calificaciones no están ni demasiado agrupadas ni demasiado dispersas, lo que sugiere cierta variabilidad entre los resultados de los estudiantes.

Ejemplo de cálculo de varianza y desviación estándar

Del siguiente conjunto de datos: 2, 4, 6 y 8 determina la varianza y la desviación estándar:

Paso 1: **Calcular la media**:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + x_3 + x_4}{N} = \frac{2 + 4 + 6 + 8}{4} = \frac{20}{4} = 5$$

Paso 2: **Calcular la varianza**:

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N} = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + (x_4 - \mu)^2}{N}$$

$$\sigma^2 = \frac{(2-5)^2 + (4-5)^2 + (6-5)^2 + (8-5)^2}{4} = \frac{(-3)^2 + (-1)^2 + (1)^2 + (3)^2}{4}$$

$$\sigma^2 = \frac{9 + 1 + 1 + 9}{4} = \frac{20}{4} = 5$$

Paso 3: **Calcular la desviación estándar**:

$$\sigma = \sqrt{\sigma^2}$$

$$\sigma = \sqrt{5} = 2{,}236$$

Ejemplo 2:

Tenemos dos conjuntos de datos:

Conjunto de Datos A	**Conjunto de Datos B**
X=[10,12,14,16,18]	Y= [30,40,50,60,70]

Paso 1: Calcula la media y desviación estándar de ambos conjuntos

Conjunto A:

Media:

μ = (10+12+14+16+18)/5 =14

Desviación estándar: Para calcular la desviación estándar, primero calculamos la **varianza**

$$\sigma^2 = \frac{(10-14)^2 + (12-14)^2 + (14-14)^2 + (16-14)^2 + (18-14)^2}{5} = 8$$

$$\sigma = \sqrt{8} = 2.83$$

Conjunto B

Media: σ =50

Desviación estándar: σ^2 =200

σ =14.14

Las medidas de dispersión como la varianza y la desviación estándar son esenciales para comprender la distribución de los datos, ya que nos indican cuán dispersos o concentrados están los datos alrededor de la media. Ambas medidas aportan información clave para:

- Evaluar la homogeneidad o dispersión de los datos.
- Determinar qué tan representativa es la media como medida central.
- Comparar la variabilidad entre distintos conjuntos de datos.

Aunque los cálculos para datos agrupados requieren el uso de los puntos medios y las frecuencias, los principios básicos para calcular la varianza y la desviación estándar son los mismos.

.Referencia:

Redacción Mate móvil (2022) Varianza y desviación estándar, ejemplo y ejercicios. Recuperado de: https://matemovil.com/varianza-y-desviacion-estandar-ejemplos-y-ejercicios

Redacción El mundo de los datos. (2023) Entendiendo la Varianza: Claves para Interpretar la Dispersión de Datos. El mundo de los datos. Recuperado de: https://elmundodelosdatos.com/entendiendo-la-varianza-claves-para-interpretar-la-dispersion-de-datos/

Walpone, Ronald. (2012) Probabilidad y estadística para ingenierías y ciencias. México. Pearson.

Redacción Universidad Abierta y a Distancia de México. (2019). Unidad 2: Medidas de tendencia central y dispersión. Estadística básica. UnADM. Recuperado de: https://dmd.unadmexico.mx/contenidos/TRONCO_COMUN/EBA/U2/descargables/EBA_U2_Contenidos.pdf

Redacción LibreTexts Español. (s.f.). 2.3: Medidas de variabilidad. Estadísticas Introductorias. LibreTexts Español. Recuperado de: https://espanol.libretexts.org/Estadisticas/Estadisticas_Introductorias/Libro%3A_Estad%C3%ADsticas_Introductorias_%28Shafer_y_Zhang%29/02%3A_Estad%C3%ADstica_Descriptiva/2.03%3A_Medidas_de_Variabilidad

Montgomery, D. C., & Runger, G. C. (2018). Estadística aplicada y probabilidad para ingenieros (6ª ed.). México: Limusa.

Triola, M. F. (2019). Estadística (13ª ed., adaptada al español). México: Pearson Educación.

Navidi, W. (2016). Estadística para ingenieros y científicos (4ª ed.). McGraw-Hill Education.