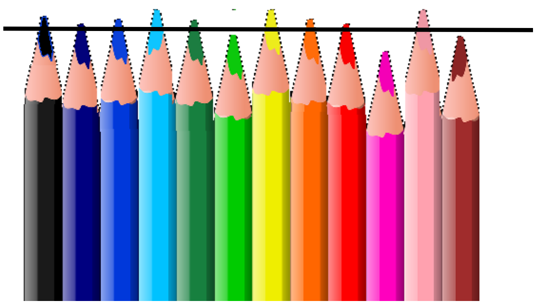
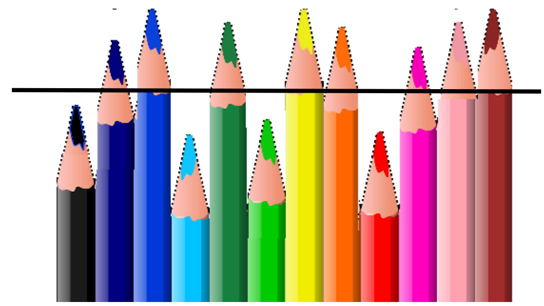
Varianza
& Desviación estándar

Dr. Jesús Alberto Mellado Bosque

La varianza mide la dispersión de los datos respecto a la media, es decir, cuando los datos están compactos respecto a la media, se tiene menos varianza, pero cuando los datos están más separados de la media se tiene mayor varianza.

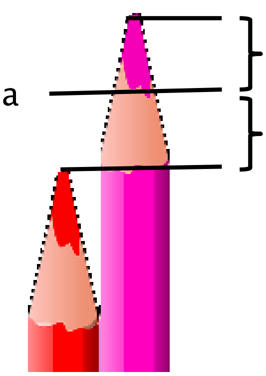


Menor varianza en la longitud de los colores



Mayor varianza en la longitud de los colores

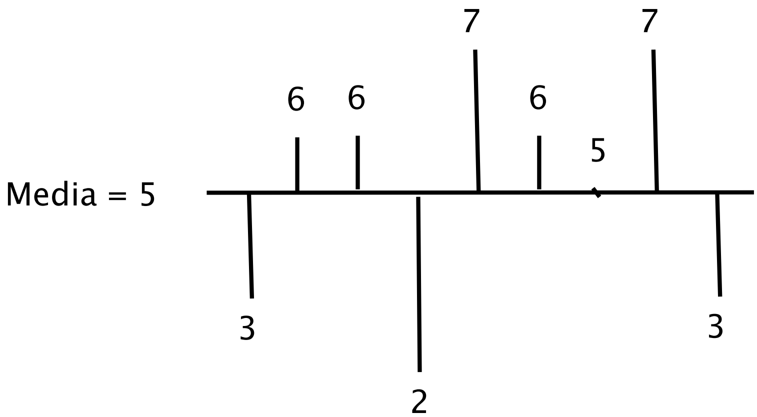
La varianza se va a medir obteniendo la distancia que existe entre cada dato y la media del conjunto de los datos.



Suponiendo que se tiene el siguiente conjunto de datos:
3, 6, 6, 2, 7, 6, 5, 7, 3

La media es 5

La distancia de cada dato respecto a la media es como se muestra



La distancia de cada dato respecto a la media es
2, 1, 1, 3, 2, 1, 0, 2, 2

Una forma de calcular la distancia es restar de cada dato la media, de la siguiente manera:

3 – 5 = -2
6 – 5 = 1
6 – 5 = 1
2 – 5 = -3
7 – 5 = 2
6 – 5 = 1
5 – 5 = 0
7 – 5 = 2
3 – 5 = -2

CUIDADO

Hay algunos resultados negativos, y como no hay distancias negativas, hay que pasarlas a positivo. Es por eso que se tomó la decisión de elevar al cuadrado cada resta para que el resultado sea positivo

Cada resta se eleva al cuadrado para hacer los resultados positivos:

$(3 - 5)^2 = 4$
$(6 - 5)^2 = 1$
$(6 - 5)^2 = 1$
$(2 - 5)^2 = 9$
$(7 - 5)^2 = 4$
$(6 - 5)^2 = 1$
$(5 - 5)^2 = 0$
$(7 - 5)^2 = 4$
$(3 - 5)^2 = 4$

Para calcular la varianza poblacional se obtiene el promedio de las distancias calculadas

$$\sigma^2 = \frac{4 + 1 + 1 + 9 + 4 + 1 + 0 + 4 + 4}{9} = \frac{30}{9} = 3.33$$

Para indicar que es una varianza poblacional se usa la letra sigma (σ) y está elevada al cuadrada para indicar que las distancias se elevaron al cuadrado previamente

La ecuación de la varianza poblacional es:

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

Medidas de variabilidad o dispersión

- Rango
- Varianza poblacional
- Varianza muestral
- Desviación estándar poblacional
- Desviación estándar muestral
- Coeficiente de variación

Varianza poblacional

La varianza poblacional se obtiene cuando se toman datos de toda una población. Esta varianza casi no se usa porque lo práctico es obtener una muestra de la población y estimar la varianza poblacional. Para calcularla se realizan los siguientes pasos:

	Acción	Objetivo
1	Se obtiene la media de la población: μ	Para poder calcular la distancia de cada valor respecto a la media.
2	Para cada valor x_i se le resta la media. $(x_i - \mu)$	La distancia de cada valor respecto a la media es la diferencia entre sus valores.
3	Cada diferencia se eleva al cuadrado. $(x_i - \mu)^2$	Algunos valores de distancia quedarán negativos, pero como físicamente no existen distancias negativas, es necesario elevarlas al cuadrado para que todas sean positivas.
4	Se obtiene el promedio de las distancias. $\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$	Al sumar todas las distancias y dividirlas entre n se obtiene el promedio de la disperción de los datos.

Varianza muestral

Cuando se desea obtener la varianza de una muestra, se siguen los mismos pasos que para una varianza poblacional, solamente que cambia el paso 4.

4	Se obtiene el promedio de las distancias. $s^2 = \frac{\sum (x_i - x)^2}{n-1}$	Lo mas seguro es que en la muestre no entren los valores extremos de la población, así que seguramente la varianza debe ser un poco mayor, para ajustarla se divide entre ***n - 1***.

En la varianza muestral la letra sigma se cambia por la letra “s”

Rango El rango es la diferencia entre el mayor valor de las observaciones y el menor

Desviación estándar poblacional

La varianza poblacional no es el promedio de las distancias de cada valor respecto a la media, es el promedio de los cuadrados de las distancias. Para que el valor sea más significativo, se puede expresar como la desviación estándar poblacional, cuya fórmula es:

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

Desviación estándar muestral

Al igual que la varianza poblacional, la desviación estándar muestral se obtiene:

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

Coeficiente de variación

El coeficiente de variación es el porciento que representa la desviación estándar respecto a la media. Se puede expresar en decimales o en porciento (se multiplica por 100 y se agrega el signo %)

$$C.V = \frac{s}{\bar{x}}$$

$$C.V = \frac{s}{\bar{x}} \times 100\%$$

Ejemplo:

Encontrar la media, la mediana, la moda, la varianza muestral, la desviación estándar muestral, el rango y coeficiente de variación de los siguientes datos:

5, 8, 8, 7, 6, 8, 4, 9, 8

Media: $\bar{x} = \frac{63}{9} = 7$

Mediana: $4, 5, 6, 7, \boxed{8}, 8, 8, 8, 9$

Moda: 8

Desviación estándar muestral: $s = \sqrt{2.75} = 1.65$

Coeficiente de variación: $C.V = \frac{1.65}{7} = 0.23$

23%

Varianza muestral:

$$s^2 = \frac{(5-7)^2 + (8-7)^2 + (8-7)^2 + (7-7)^2 + (6-7)^2 + (8-7)^2 + (4-7)^2 + (9-7)^2 + (8-7)^2}{9-1} = 2.75$$