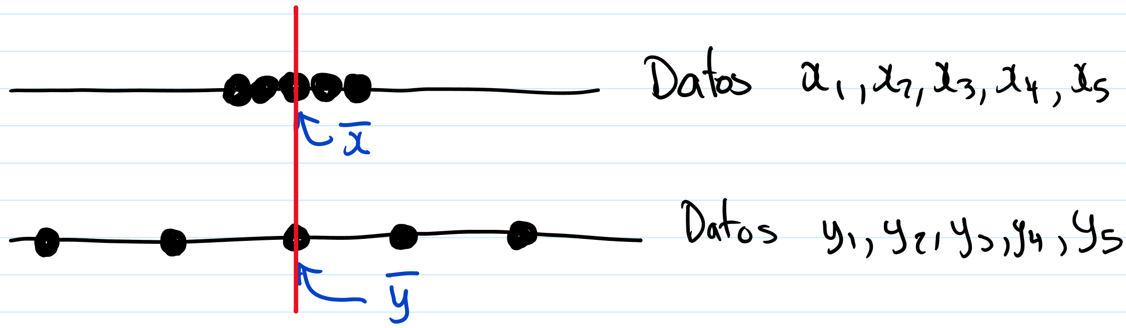
- Medidas de dispersión -

Recordemos que las medidas de tendencia central tratan de resumir o sintetizar un conjunto de valores numéricos en un sólo número (con excepción de la moda). Analicemos un ejemplo al respecto:

Es fácil mostrar que la media se puede ver como el "centro de gravedad" del conjunto de datos. Pensando en la media de este modo, considere los sig. dos conjuntos de datos:



Puede sorprenderle, pero ambos conjuntos de datos tienen la misma media, aunque es claro que son datos MUY distintos (recuerde que la media está "resumiendo" los datos, y en este ejemplo, los resume en lo mismo). En particular, los datos y's están más dispersos que los datos x's.

El ejemplo anterior muestra que también es importante medir qué tan dispersos están los datos. Las medidas de dispersión nos permiten hacerlo. Analizaremos solo 3 de estas medidas (aunque hay más):

- **Rango.-** Es simplemente la diferencia entre los valores mayor y menor:

$$\boxed{\text{Rango} := \text{Máx} - \text{Mín}}$$

- **Ejemplos:**

a) Para los datos: 1, 2, 3, 4, 5, 6, 7, 8, 9 rango = $9-1=8$

b) Para los datos: 1, 1, 1, 1, 1, 1, 1, 1, 9 rango = $9-1=8$

c) Para los datos: 1, 2, 3, 4, 5, 6, 7, 8, <u>1000</u> rango = $1000-1=999$

Una ventaja del rango es que es muy fácil de calcular. Sin embargo tiene desventajas manifiestas. Por ejemplo, se basa sólo en 2 datos e ignora todo lo demás. Así, los ejemplos a) y b) tienen el mismo rango, pero en b) los datos varían menos. Otra desventaja es que el rango se ve fuertemente afectado por los datos e ignora todo lo demás. Así, los ejemplos a) y b) tienen el mismo rango, pero en b) los datos varían menos. Otra desventaja es que el rango se ve fuertemente afectado por los outliers (inciso c)).

• Varianza.- Es el promedio de los cuadrados de las "desviaciones" de los datos con respecto a la media.

$$\boxed{\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

Población

x_i = datos

μ = "miu" = media pob

x_i = datos

$\bar{x}$ = media muestra

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

Muestra

Note que lo que estamos haciendo es medir "qué tan alejados están los datos de la media" y los elevamos al cuadrado. Note que entre más dispersos los datos, el rango y la varianza aumentan. Pero, ¿por qué elevar esas "distancias" al cuadrado?

• Ejemplo.- Muestre que si no ponemos al cuadrado en las fórmulas anteriores tendríamos:

$$\sigma^2 = 0, \quad s^2 = 0$$

sin importar los datos.

(en clase).

• Desviación Estándar.- Es simplemente la raíz cuadrada de la varianza

$$\boxed{\sigma = \left[\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}\right]^{1/2}}$$

$$\boxed{s = \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}\right]^{1/2}}$$

• Ejemplo Google.- Hallar el rango, varianza y desviación estandar de los datos de la clase pasada.

Surge la pregunta de la clase pasada. ¿Qué les pasa a las medidas de dispersión si a todos los datos se les suma o multiplica una misma constante? Piense bien su respuesta, recuerde que estas medidas tienen que ver con cómo varían los datos entre ellos.

• Ejercicio en Google (para subir):

→ media, mediana, moda
tamaño de la muestra

• Ejercicio en Google (para subir):

→ media, mediana, moda
tamaño de la muestra

a) Halle el rango, varianza y desviación estándar del conjunto de datos: Muestrales

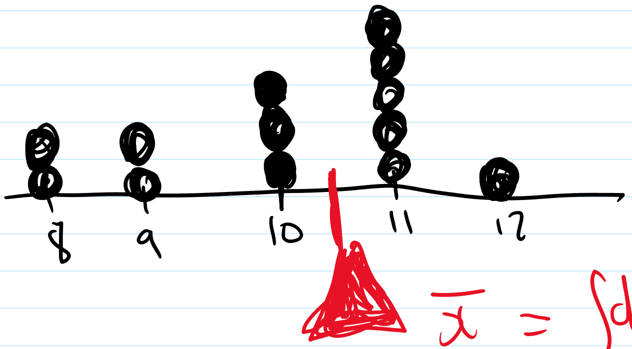
$$\boxed{5, 6, 8, 7, 7, 5, 7, 9, 8, 7, 10, 6}$$

10, 11, 13,

b) Repita el ejercicio anterior sumando a todos los datos 5.

c) Repita el inciso a) multiplicando todos los datos por 10.

Compare los resultados de los incisos a), b) y c).



$$\bar{x} = \frac{\int d\,peso}{\int Area} \rightsquigarrow n$$

μ

• Fórmulas en Google:

- Media: =promedio(celda_inicial:celda_final)
- Mediana: =mediana(celda_inicial:celda_final)
- Moda (múltiple): =mode.mult(celda_inicial:celda_final)

4

- Rango: =max(celda_inicial:celda_final) − min(celda_inicial:celda_fin)
- Varianza (muestral) =var(celda_inicial:celda_final)
- Desviación Estándar (muestral) =desvest(celda_inicial:celda_final)
- tamaño de mesta =count(celda_inicial:celda_final)