Física experimental

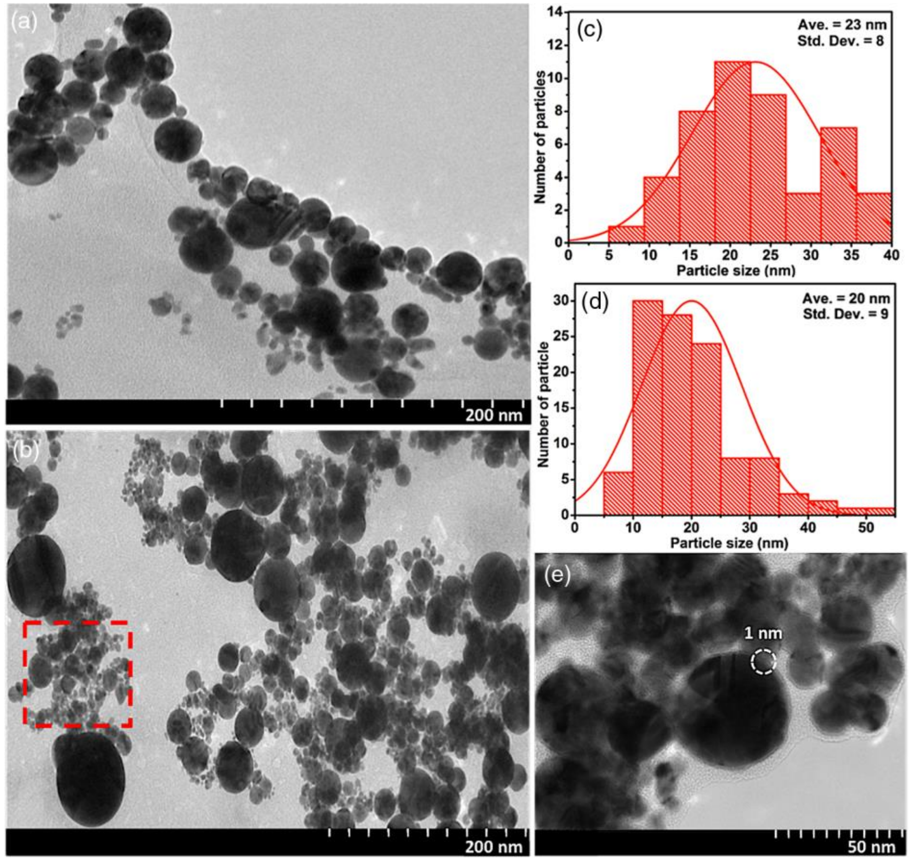
- Histograma, distribución estadística.
- Medidas de tendencia central y medidas de dispersión

Docente: Álvaro Herrera Carillo

UNIVERSIDAD
DE ANTIOQUIA
1 8 0 3

Histogramas y distribución estadística

Veamos algunas ***elementos útiles*** para el <u>**tratamiento estadístico**</u> del resultado de una serie de medidas***. El problema central de la estadística es inferir propiedades de una población a partir de las propiedades de una muestra de algunos individuos de la misma***. Por población entendemos no sólo un grupo de personas, sino que podemos referirnos a objetos más generales como la población de todos los resultados posibles de una medida.



Concentrémonos en una propiedad x de una población. Si tomamos una muestra de tamaño *N* y observamos el valor de esta propiedad para cada individuo, encontramos *N* resultados:, x_1, x_2, x_3,…….., x_n comprendidos en un intervalo de valores $[x_{min}, x_{max}]$ entre el menor y mayor valor registrado.

Una manera útil de visualizar las características de este conjunto de datos consiste en dividir el intervalo $[x_{min}, x_{max}]$ en *m* sub-intervalos o clases delimitados por los puntos y_1, y_1, y_1,…….., y_m . Luego, contamos el número n_1 de individuos de la muestra cuyas edades estén comprendidas en el primer intervalo $[y_j, y_{j+1}]$, etc., el número de los individuos de la muestra que están en el *j*-ésimo intervalo , etc., hasta llegar al *m*-ésimo sub-intervalo. Estos valores son la ***frecuencia absoluta de cada clase***.

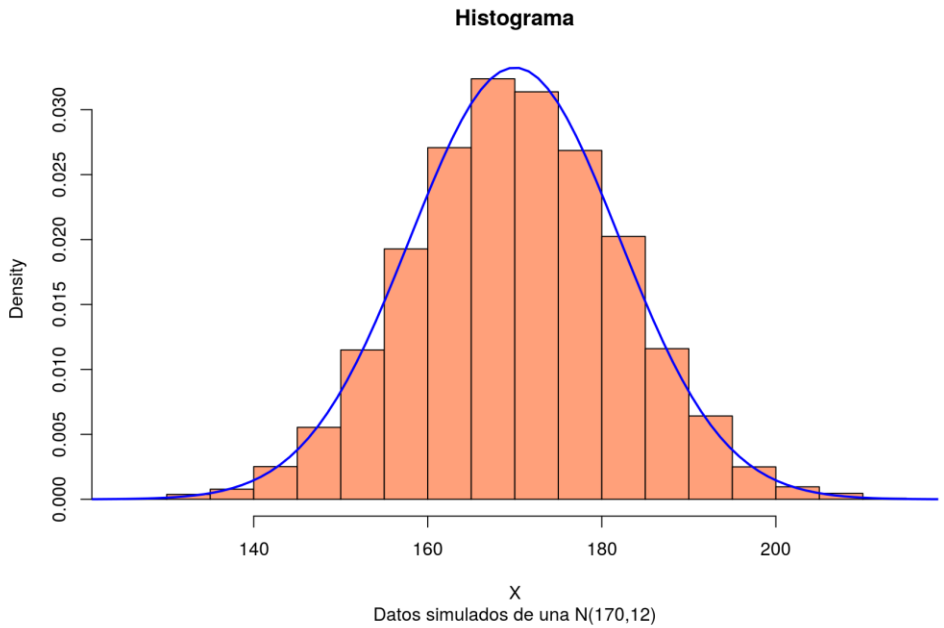
A partir de la frecuencia absoluta definimos la función f_j de distribución como:

$$f_j = \frac{n_j}{\sum_j n_j}$$

La función de distribución está normalizada, es decir:

$$\sum_{j=1}^{m} f_j = 1$$

Si construimos un gráfico de f_j en función de x_j para $j = 1,2,3, \dots, m$ donde $x_j = (y_{j-1} + y_j)/2$ son los centros de los intervalos, nos dará una idea clara de cómo se distribuyen la variable de una población de individuos de la muestra en estudio. Este tipo de gráfico se llama ***histograma***. Usualmente estos gráficos se representan con barras.

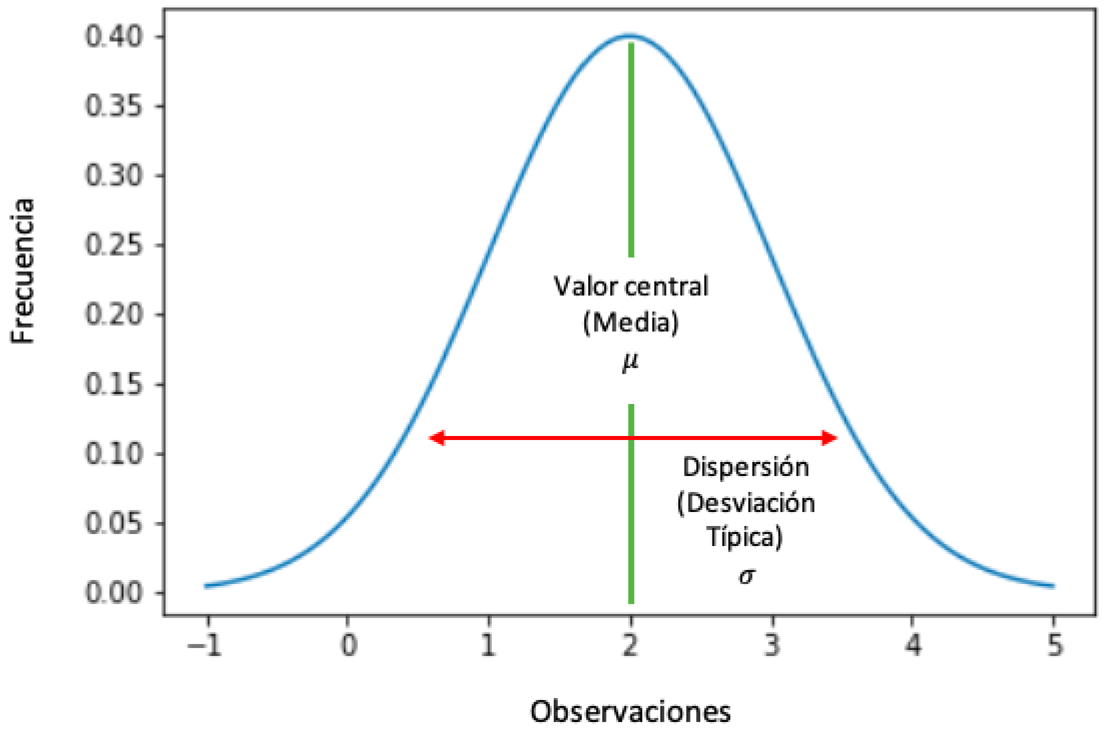


Es una herramienta usada para representar una distribución por medio de barras. La altura de la barra está en función de la frecuencia (eje y) y el rango (eje x) de una variable continua.

Nos ofrece un vistazo general del comportamiento de las variables, donde logramos analizar aspectos como ***distribución, dispersión, aleatoriedad y tendencia***.

Distribución Normal o Gaussiana

La distribución de probabilidad más importante es la distribución gaussiana o normal, que tiene una forma de campana . Corresponde a la función



La media, moda, y mediana coinciden

$$y = f(x) = N(\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty$$

- Toma en cuenta μ y σ
- Área bajo la curva es igual a uno
- Es simétrica respecto al centro
- 50% de valores a cada lado de μ
- El factor que multiplica la exponencial garantiza que el area total bajo la curva sea igual a uno

Distribución normal estándar: $N(0,1)$, recordando que $N(\mu,\sigma)$

La distribución normal con media $\mu = 0$ y desviación típica $\sigma = 1$ se llama distribución normal estándar, o tipificada, o reducida.

$$y = f(x) = N(0,1) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \quad -\infty < x < \infty$$

http://www.est.uc3m.es/esp/nueva_docencia/comp_col_leg/ing_tec_inf_gestion/estadistica/Documentacion/Tablas/tablas2caras.pdf

Para poder utilizar la tabla tenemos que transformar la variable x que sigue una distribución $N(\mu,\sigma)$ en otra variable Z que siga una distribución $N(0,1)$. Por lo que la operación necesaria es la siguiente:

$$Z = \frac{x-\mu}{\sigma}$$

Construcción de un histograma

Tiempo en minutos por usuario				
11,50	10,26	10,08	13,00	11,14
13,73	13,41	10,44	11,36	14,40
11,64	12,39	12,82	14,25	15,41
14,35	9,35	12,40	9,04	15,30
14,79	15,27	10,63	14,30	15,48
14,80	8,78	14,00	13,09	10,00
12,20	11,70	15,37	11,81	10,06
12,49	8,58	11,32	12,20	12,45
11,28	12,60	14,36	13,08	13,50
12,68	9,19	14,32	12,17	9,10

Paso 1: Determinar el rango: El valor mínimo y el valor máximo de los datos

Valor Max=15,48   valor min= 8,58

R=15,48-8,58=6,9

Paso 2. Se calcula el número de intervalos de clase(k).

$$K = \sqrt{N} = \sqrt{50} \approx 7{,}07 \quad \Rightarrow \quad K=7$$

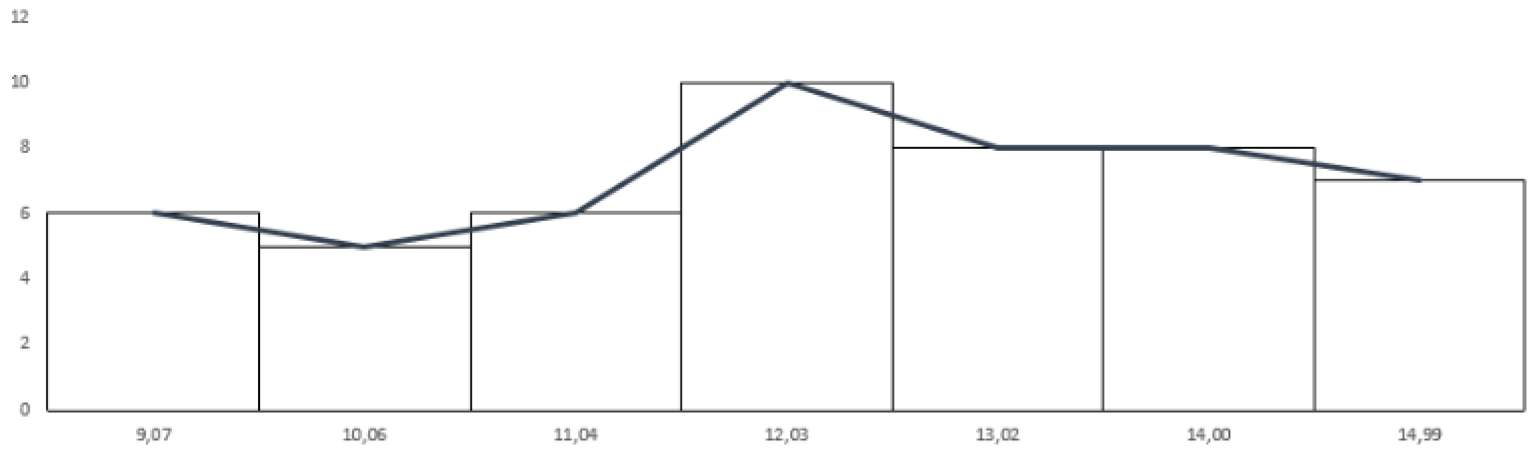
Paso 3: Se calcula la amplitud o ancho del intervalo

$$A = \frac{R}{K} = \frac{6{,}9}{7} = 0{,}99$$

Paso 4 : Se define las clases sumándole al valor más pequeño , el ancho del intervalo hasta que obtenga K intervalos

Paso 5: Se agrupa cada valor dentro del intervalo de clase, o dicho de otra forma, **determinamos la frecuencia**

Intervalo de clase		Frecuencia
Desde	Hasta	
8,58	9,57	6
9,57	10,55	5
10,55	11,54	6
11,54	12,52	10
12,52	13,51	8
13,51	14,49	8
14,49	15,48	7



Ejercicio: Considere la siguiente tabla, donde se detalla la lita de N=30 medida de la masa m de una muestra de cierto material

1,09	1,01	1,10	1,14	1,16
1,11	1,04	1,16	1,13	1,17
1,14	1,03	1,17	1,09	1,09
1,15	1,06	1,12	1,08	1,20
1,08	1,07	1,14	1,11	1,05
1,06	1,12	1,00	1,10	1,07

A). Realice el histograma correspondiente a estos valores.

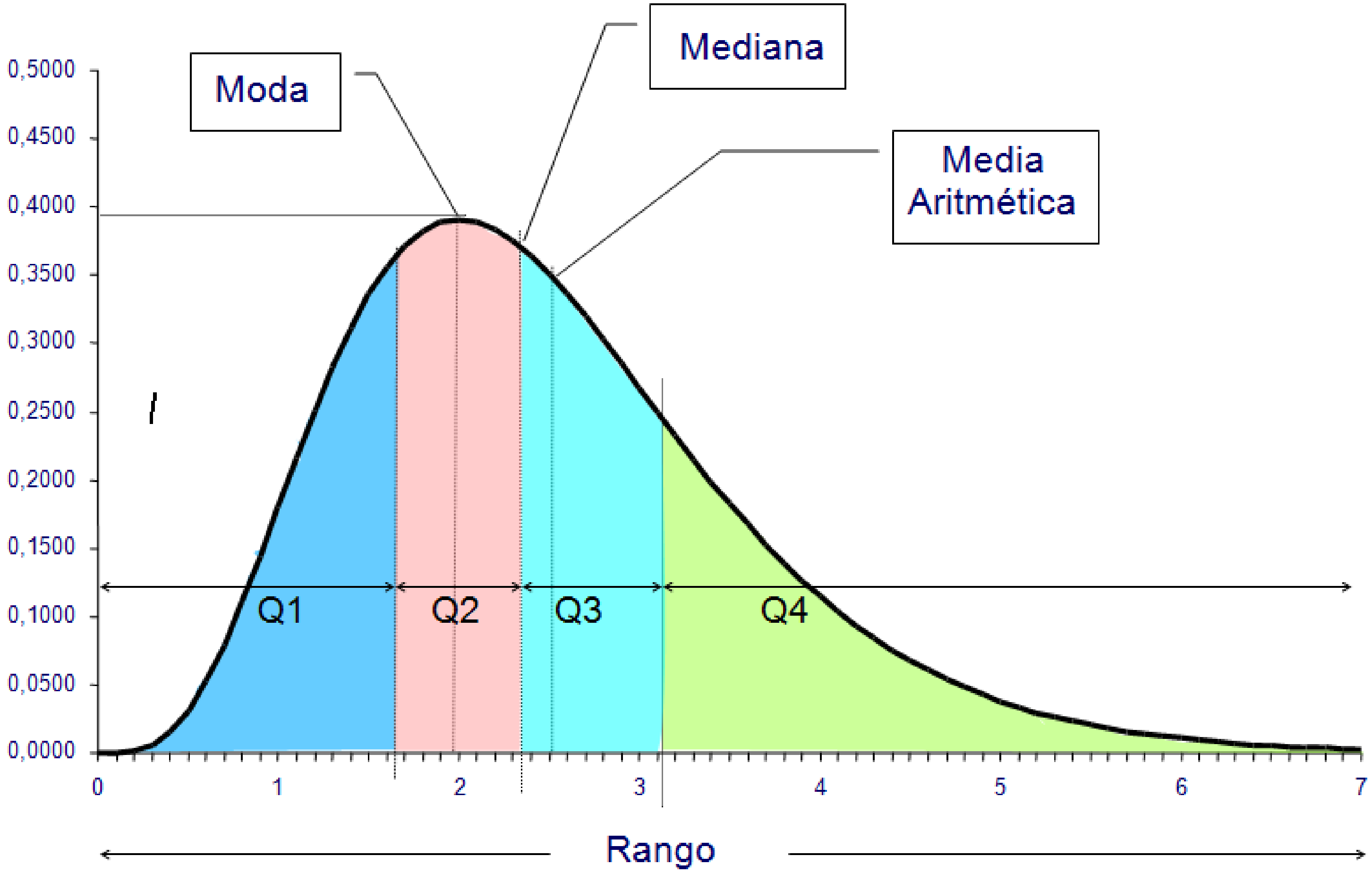
B).Calcule la desviación estándar de las medidas

MEDIDAS DE TENDENCIA CENTRAL
Y MEDIDAS DE DISPERSIÓN

Objetivo: Conocer y calcular las medidas de tendencia central y medidas de dispersión

- Medidas de tendencia central
  - Media
  - Mediana
  - Moda
- Medidas de dispersión
  - Rango
  - Varianza
  - Desviación estándar

La mayor parte de las serie de datos muestran una clara tendencia a agruparse alrededor de un cierto punto central. Así pues, dada cualquier serie de datos particular, por lo general es posible seleccionar algún valor o promedio típico para describir toda la serie de datos. Este valor descriptivo típico es una **medición de tendencia central** o de **ubicación (medidas de posición)**



Fuente: Allende H y Ahumada S, ILI-280

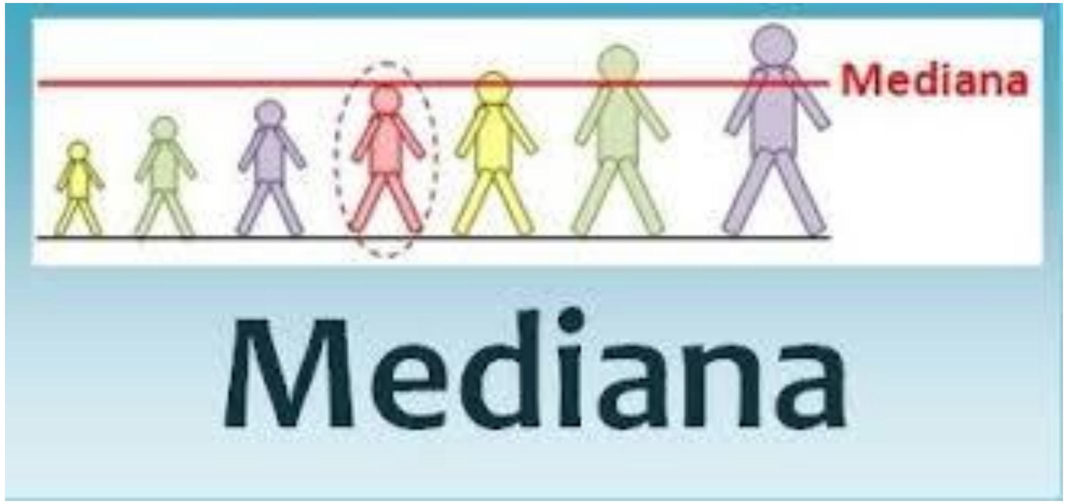
Media

Se representa por $\bar{X}$ (se lee como "*x* barra" o " media de la muestra"). Es la suma de todos los valores de la variable *x* (la suma de valores *x* se simboliza como $\sum X$) y dividiendo entre el número total de datos, *n*. Lo podemos expresar como:

$$\bar{X} = \frac{\sum X}{n}$$

Mediana

Valor de los datos que ocupa la posición central cuando los datos se ordenan según su tamaño. Se representa por $\tilde{X}$ (se lee como "*x* tilde" o "mediana de la muestra").



La **media** (**promedio**) de un conjunto de datos se encuentra al sumar todos los números en el conjunto de datos y luego al dividir entre el número de valores en el conjunto. La **mediana** es el valor medio cuando un conjunto de datos se ordena de menor a mayor

Procedimiento para calcular la mediana

1. Ordene los datos
2. Utiliza la expresión:

$$\tilde{X} = \begin{cases} X_{(n+1)/2} \; si \; n \; es \; impar \\ \dfrac{1}{2}\left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1}\right) si \; n \; es \; par \end{cases}$$

Ejemplo 1

Suponga que un conjuntos de datos está dado por: 2.0, 2.5, 5.3, 4.2, 3.8. Calcular la mediana

Solución:

- **Ordenamos los datos:** 2.0, 2.5, 3.8, 4.2, 5.3
- N es impar; 5
- **Utilizamos la expresión:**
- $X_{(n+1)/2}$= 3, la mediana es el número ubicado en la tercera posición: 3.8

Ejemplo 2

Una toma de datos de un experimenta arroja los siguientes resultados

X_i	$\boldsymbol{t(s) \pm 0.1}$
X_1	0.48
X_2	0.50
X_3	0.52
X_4	0.55
X_5	0.56
X_6	0.58
X_7	0.60
X_8	0.63
X_9	0.66
X_{10}	0.69

Calcular la mediana

Solución:

- Los datos están ordenados
- N es par, utilizamos la siguiente expresión:

$$\frac{1}{2}\left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1}\right) = \frac{1}{2}\left(X_{\frac{10}{2}} + X_{\frac{10}{2}+1}\right) = \frac{1}{2}(X_5 + X_6) = 0.57$$

Moda

La moda de un conjunto de datos es el dato que **más veces se repite**, es decir, aquel que tiene **mayor frecuencia absoluta**. Se denota por **Mo**. En caso de existir dos valores de la variable que tengan la mayor frecuencia absoluta, habría dos modas. Si no se repite ningún valor, no existe moda.

Anexos de ejemplos en Excel

Medidas de dispersión

- **Rango**
- **Varianza**
- **Desviación estándar**

La dispersión se refiere a la extensión de los datos, es decir al grado en que las observaciones se distribuyen

Nos proporcionan informacional adicional que nos permite juzgar la confiabilidad de nuestra medida de tendencia central, si los datos están muy dispersos la posición central es menos representativas de los datos, como un todo, que cuando estos se agrupan mas estrechamente alrededor de la media

Rango: Es la diferencia entre el mayor y el meno de los valores observados

$$R = X_{Mayor} - X_{Menor}$$

Como medida de dispersión es limitada

Varianza y desviación estándar: Las descripciones más comprensibles de la dispersión son aquellas que tratan con la desviación promedio con respecto a laguna medida de tendencia central. Veremos dos medidas que nos dan una tendencia promedio con respecto a la media de la distribución: Varianza y desviación estándar.

DATOS SIN AGRUPAR

Varianza Muestral:

$$s^2 = \frac{\sum(x-\overline{x})^2}{n-1} = \frac{\sum x^2}{n-1} - \frac{n\overline{x}^2}{n-1}$$

$\boldsymbol{s^2}$ Varianza de la muestra

x Elemento u observación

$\overline{x}$ Media de la muestra

n Número de elementos de la muestra

Desviación estándar muestral

$$s = \sqrt{\frac{\sum(x-\overline{x})^2}{n-1}} = \sqrt{\frac{\sum x^2}{n-1} - \frac{n\overline{x}^2}{n-1}}$$

DATOS AGRUPADOS

Varianza Muestral:

$$s^2 = \frac{\sum(m_i-\overline{x}).f_i}{n-1}$$

$\boldsymbol{s^2}$ Varianza de la muestra

s Desviación estándar de la muestra

f_i frecuencia absoluta de la clase i

$\mathrm{m_i}$ Marca de clase de la clase i

n Número de elementos de la muestra

$\overline{x}$ Tamaño de la muestra

Desviación estándar muestral

$$s = \sqrt{\frac{\sum(m_i-\overline{x}).f_i}{n-1}}$$

Actividad de retroalimentación del aprendizaje

Estudiantes de física experimental 1 midieron con un micrómetro el diámetro de un conjunto de frijoles obteniendo los siguientes resultados:

DIÁMETRO DE FRIJOLES (mm ±0.01)								
10.45	10.2	10.41	9.6	10.19	7.76	9.25	10.6	9.7
10.0	9.2	10.2	9.5	9.6	10.2	10.5	10.9	8.9
10.22	9.0	8.4	8.83	9.7	10.8	9.7	9.65	9.2
9.3	9.9	9.8	10.0	8.9	9.1	8.9	10.1	9.8
10.1	9.89	9.5	8.8	9.18	9.0	9.0	8.7	10.03

Hacer un histograma, la distribución normal y calcular la desviación estándar, utilizar cualquier software.

PRÁCTICA INSTRUMENTAL

Practicar la utilización de instrumentos como el micrómetro y el pie de rey en casa calculando diámetros de diferentes elementos que tengas en casa.