

# BIOESTADISTICA



Ing. Joel Herrera Ordoñez

# DIAGNOSTICO DE LA MATERIA

1. Para ti, ¿Qué es la Bioestadística?

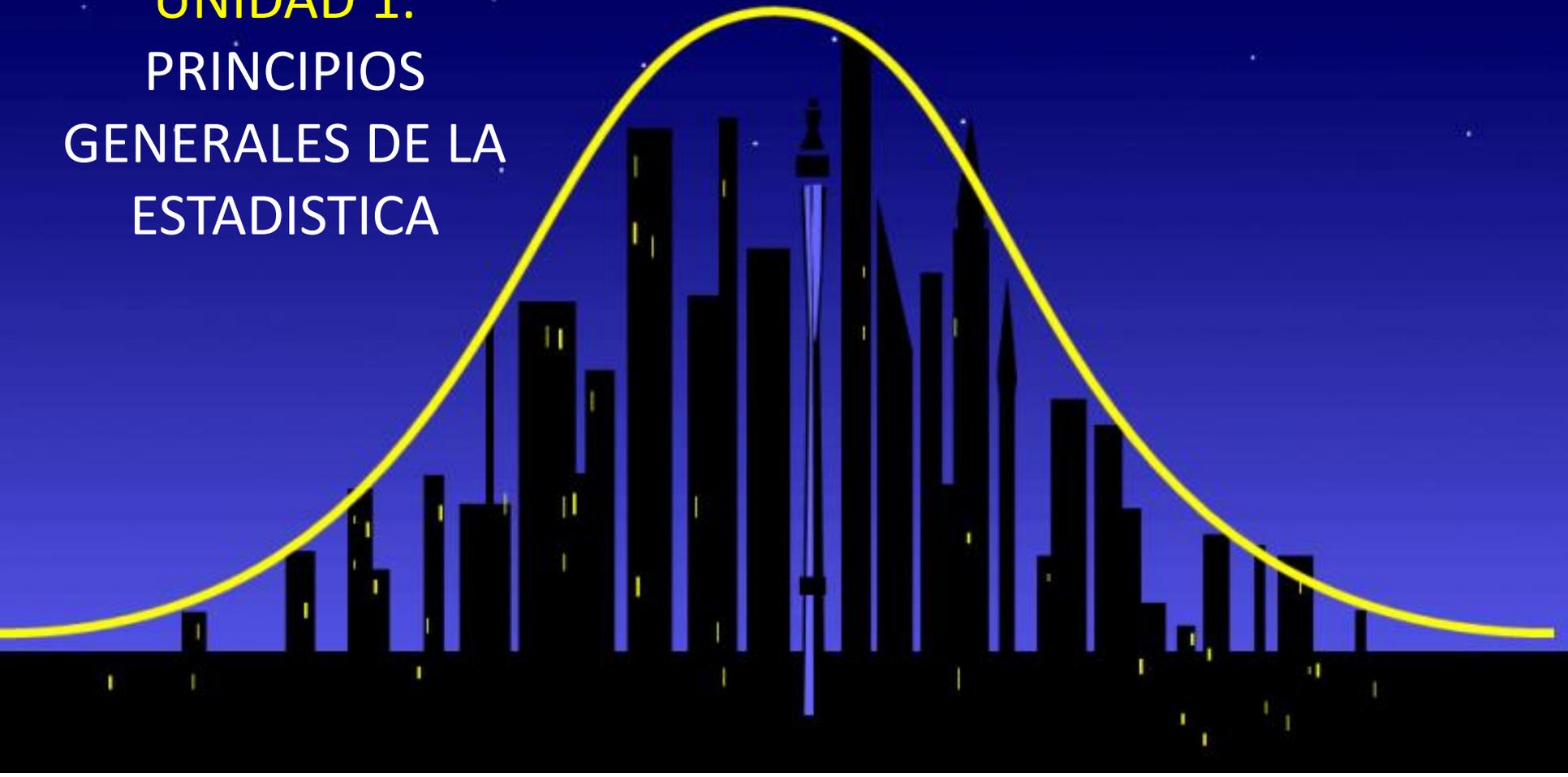
2. ¿Por qué consideras que es importante la Bioestadística en las ciencias de la salud?



**UDS**

# ESTADÍSTICA

UNIDAD 1.  
PRINCIPIOS  
GENERALES DE LA  
ESTADISTICA



# BIOESTADISTICA

Es una ciencia que  
abarca métodos y  
procedimientos



Ciencias de la vida como  
la biología y la medicina.



**¿CUAL CREEES QUE ES  
LA APLICACIÓN DE LA  
ESTADISTICA EN  
ENFERMERIA?**

<b>Área de aplicación</b>	<b>Ejemplo</b>
<b>Estudios de variación</b>	Edad, peso, estatura, presión sanguínea, niveles de colesterol, albúmina sérica, recuento de plaquetas.
<b>Diagnóstico de enfermedades</b>	Valoración de los síntomas declarados o recabados en los individuos para realizar un diagnóstico de salud
<b>Predicción probable de un programa de intervención</b>	Programa de intervención nutricional para determinar el impacto de la aplicación de un suplemento alimenticio
<b>Elección apropiada de intervención en paciente</b>	Evaluación de la eficacia de un fármaco y/u otros métodos de tratamiento
<b>Administración sanitaria y planificación</b>	Determinar el perfil sanitario de la población en términos de distribución de la enfermedad
<b>Realización y análisis en la investigación en salud pública</b>	Probabilidad de cáncer de próstata en individuos con edad mayor a 60 años

# INTRODUCCION HISTORICA



**Pierre Charles-Alexandre  
Louis (1787-1872)**

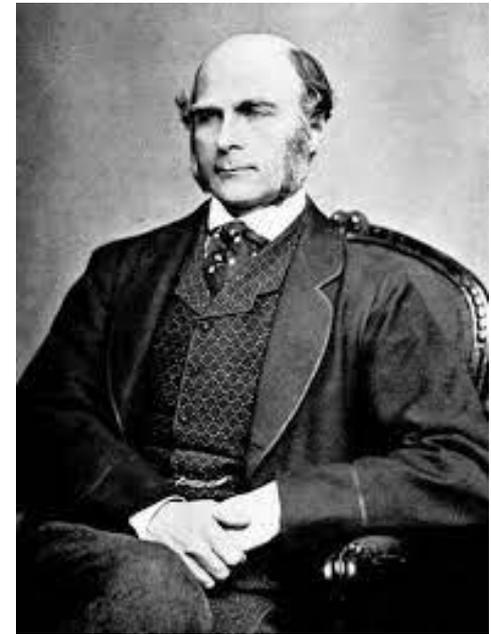
Fue el primer médico que utilizó métodos matemáticos para cuantificar variables de pacientes y sus enfermedades. La primera aplicación del método numérico en su clásico estudio de la tuberculosis.



Hicieron los primeros mapas epidemiológicos usando métodos cuantitativos y análisis epidemiológicos.

**René Villermé (1782-1863) y  
William Farr (1807-1883)**

Fundó la biometría estadística.



**Francis Galton (1822-1911)**



Propuso un modelo temporal discreto en un intento de explicar la ocurrencia regular de la epidermis de sarampión.

**William Heaton Hamer**  
**(1862-1936)**

Exploró la aplicación matemática de la teoría de las probabilidades con la finalidad de determinar la relación entre el número de mosquitos y la incidencia de malaria en situaciones endémicas y epidérmicas.



**Ronald Ross**  
**(1857-1932)**



Su Mayor contribución fue el épico trabajo que correlacionó el tabaco y el cáncer de pulmón.

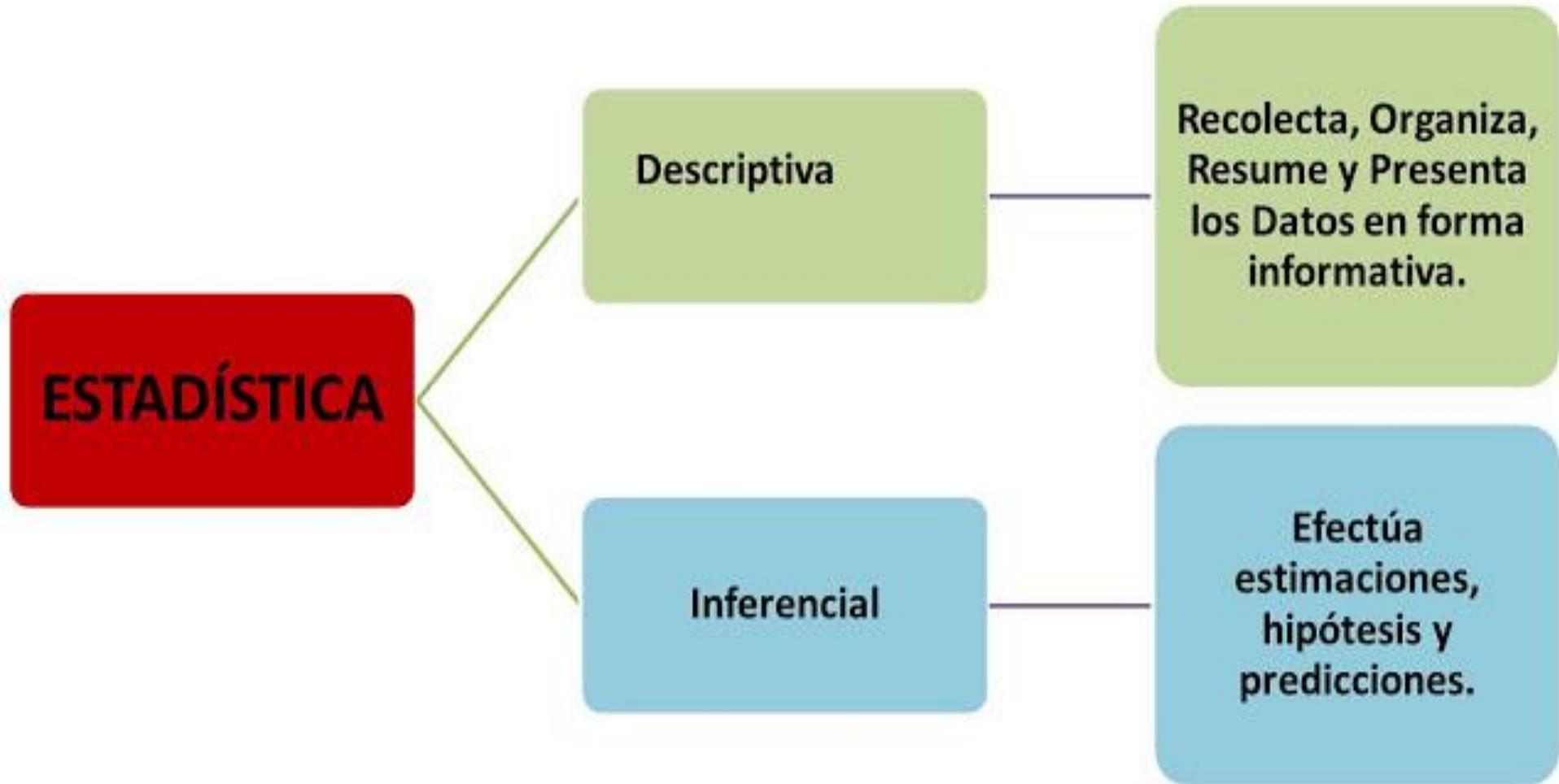
**Bradford Hill**  
**(1897-1991)**

Realizo los primeros trabajos bioestadísticas en enfermería a mediados del siglo XIX.



**Florence Nightingale**  
**(1820- 1910)**

# CONCEPTOS BASICOS



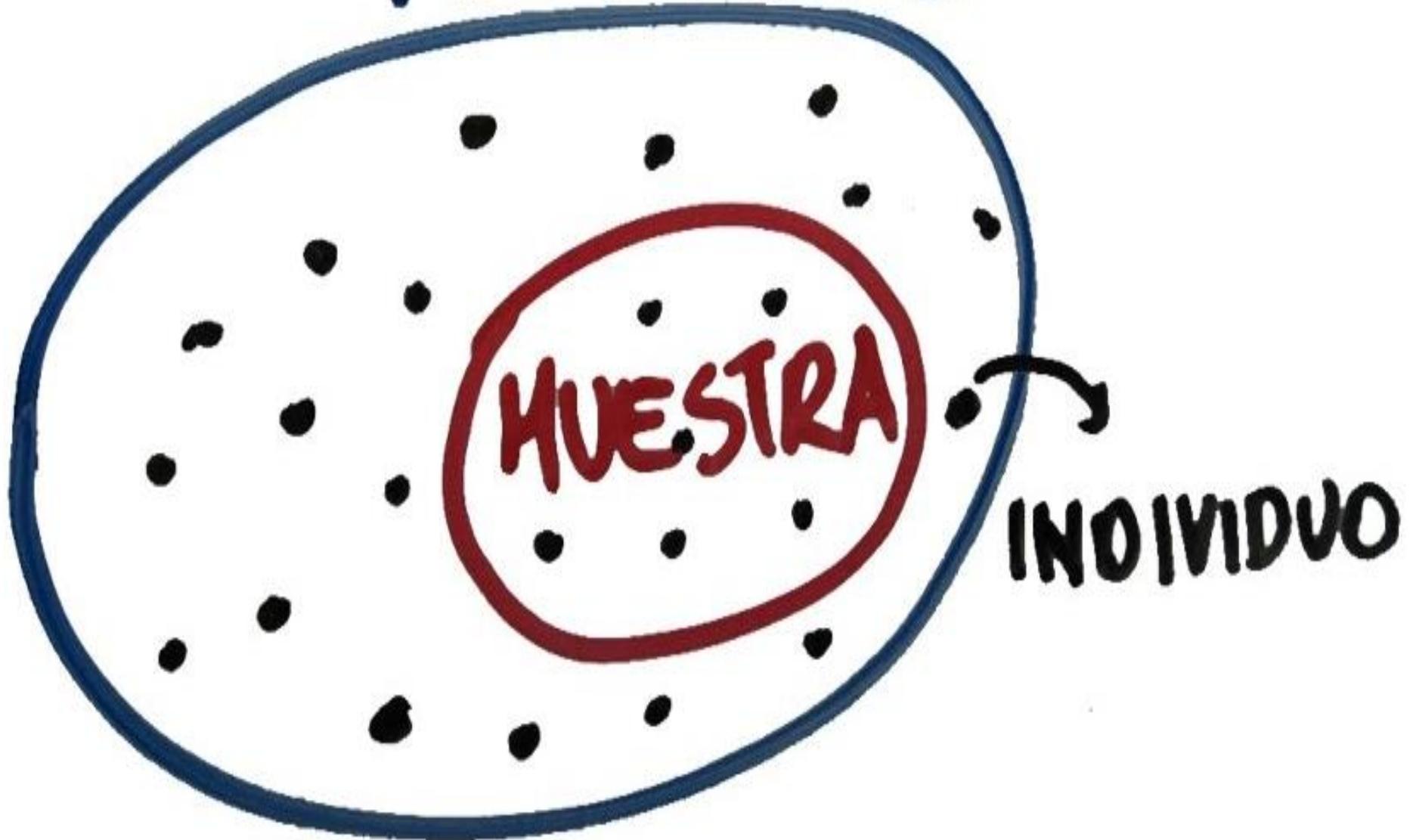
**UDS**

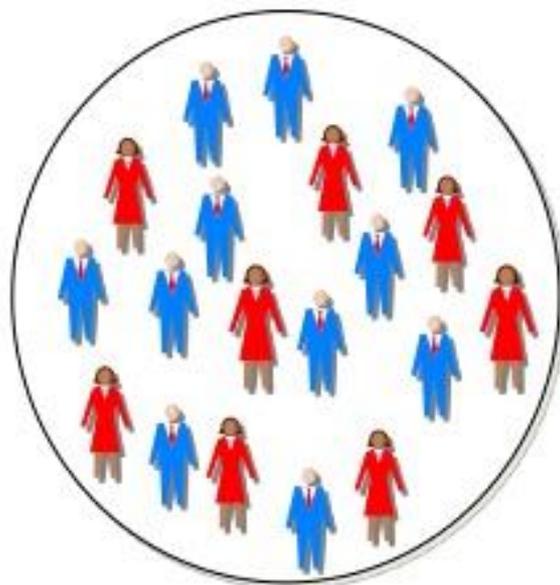
# ***Parámetros y estadísticos***

Los ***parámetros*** son las características de la población que no conocemos pero queremos estimar: por ejemplo la media y la desviación estándar.

Por su parte, los ***estadísticos*** son las características de la muestra, por ejemplo su media y desviación estándar.

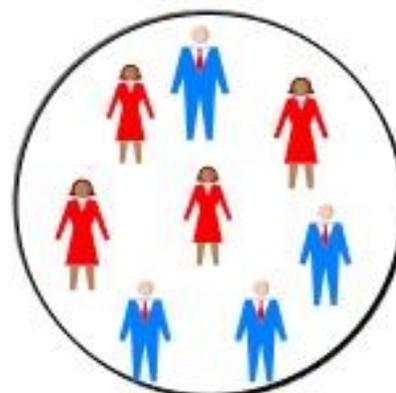
POBLACIÓN





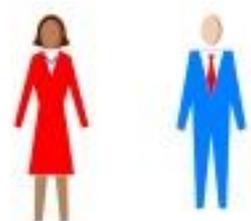
## Población

Es el conjunto de todos los individuos que poseen información sobre el fenómeno que se estudia.



## Muestra

Es un subconjunto de elementos pertenecientes a una población.



## Unidad Estadística

Cada individuo, animal o cosa al que se le mide u observa una o más características

## Variables:

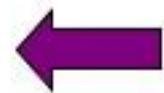
Características que se observan en las unidades estadísticas.



**Cualitativas**

**Cuantitativas discretas**

**Cuantitativas continuas**



# VARIABLE ESTADISTICA

Es la característica de los elementos de la población que se investiga.



Aquella variable que no es medible.

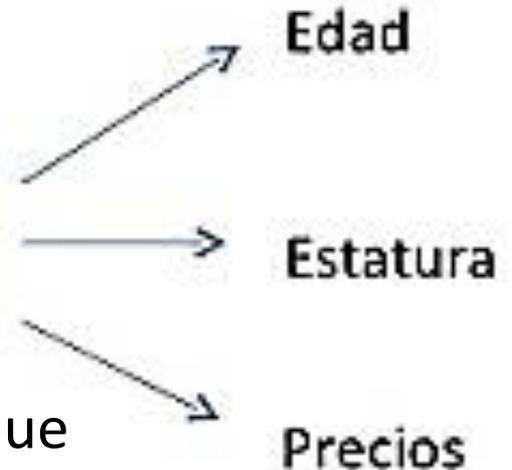


Cualitativas



Tipos de Variable

Cuantitativas



Aquella variable que se puede contar o medir.

# Tipos de Variables

## ■ Cualitativas:

Si sus valores (*modalidades*) no se pueden asociar naturalmente a un número (*no se pueden hacer operaciones algebraicas con ellos*)

❖ **Nominales:** Si sus valores no se pueden ordenar

- Género, Grupo Sanguíneo, Religión, Lugar de nacimiento, Fumar (Si/No).

❖ **Ordinales:** Si sus valores se pueden ordenar

- Escolaridad, Grado de satisfacción, Intensidad del dolor.

## ■ Cuantitativas o Numéricas:

Si sus valores son numéricos (*tiene sentido hacer operaciones algebraicas con ellos*)

❖ **Discretas:** Si toma valores enteros

- Número de alumnos, Número de grupos.

❖ **Continuas:** Si entre dos valores, son posibles infinitos valores intermedios.

- Altura, edad.

## ACTIVIDAD: CUESTIONARIO

1. ¿Qué es la estadística descriptiva?
2. ¿Qué es la estadística inferencial?
3. ¿Qué es una muestra?
4. ¿Qué es un parámetro?
5. ¿Qué es una clase o intervalo de clase?
6. ¿Qué es la marca de clase?
7. ¿Qué es la frecuencia de clase?
8. ¿Se representa como  $S^2$ ?
9. ¿Qué representa la desviación típica y con qué letra se representa?
10. ¿Cuál es la diferencia entre mediana y moda?
11. ¿Qué es la media aritmética?
12. ¿Qué es una variable cuantitativa y menciona ejemplos?
13. ¿Qué es una variable cualitativa y menciona ejemplos?

**Para responder estas preguntas, descarga del apartado de recursos el archivo denominado **conceptos básicos**.**

## DATOS NO AGRUPADOS

**Los datos no agrupados** son el conjunto de observaciones que se presentan en su forma original tal y como fueron recolectados para obtener información directamente de ellos.

**EJEMPLO:** Se pregunto la edad de 10 estudiantes de licenciatura en enfermería y los datos obtenidos son los siguiente:

22, 25, 21, 22, 23, 22, 23, 23, 24, 22

## DATOS AGRUPADOS

Los datos agrupados son aquellos que se encuentran ordenados y clasificados.

### AGRUPACION PUNTUAL

**EJEMPLO:** Se pregunto la edad a 30 pacientes de un hospital y los datos obtenidos son los siguientes:

22	23	21	24	20	EDAD	F
25	21	24	21	24	20	4
25	20	21	22	24	21	8
23	25	20	22	21	22	7
23	21	20	21	22	23	3
22	25	21	22	22	24	4
					25	4

# DATOS AGRUPADOS

## TABLA EN INTERVALOS

### Ejercicio:

Agrupar los siguientes datos

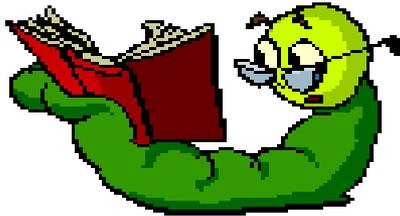
Edad (años)	Nº pacientes $f_i$	
[20, 30)	20	30 55 72 45 32 48 75 46 51 22 56 49 35 22 40 42 33 45 66 54 20 30 46 51 69 50 34 57 71 53 41 45 26 42 32 60 38 51 59 29 55 32 54 38 64 53 37 73 53 89 60 56 76 61 31 70 39 44 65 73 56 82 43 20 54 31 47 33
[30, 40)	35	63 24 54 45 51 25 50 65 51 30 51 30 58 47 78 59 37 75 53 46 62 67 51 49 43 71 59 63 69 70 50 41 47 60
[40, 50)	50	46 39 33 58 81 41 21 51 68 50 40 50 42 25 53 79 35 21 40 71 31 59 27 61 67 39 25 56 47 81 34 41 34 47
[50, 60)	49	39 47 50 25 40 43 59 70 50 40 21 39 48 41 80 41 61 51 44 32 54 62 52 44 68 56 62 29 48 61 58 30 28 47
[60, 70)	25	34 80 56 40 28 50 61 56 20 40 49 49 47 73 30 48 36 20 54 31 43 41 32 58 60 53 32 61 40 49
[70, 80)	15	
[80, 90)	6	

# Distribución de frecuencias

Edad (x)	Marca de Clase ( $X_i$ )	Frecuencia absoluta ( $f_i$ )	Frecuencia absoluta acumulada ( $F_i$ )	Frecuencia relativa ( $f_r$ )		Frecuencia relativa acumulada ( $F_r$ )	
[10 - 19)	14.5	5	5	0.1	10%	0.1	10%
[19 - 28)	23.5	11	16	0.22	22%	0.32	32%
[28 - 37)	32.5	8	24	0.16	16%	0.48	48%
[37 - 46)	41.5	5	29	0.1	10%	0.58	58%
[46 - 55)	50.5	8	37	0.16	16%	0.74	74%
[55 - 64)	59.5	6	43	0.12	12%	0.86	86%
[64 - 73]	68.5	7	50	0.14	14%	1	100%
	<b>Total</b>	<b>50</b>	<b>Total</b>	<b>1</b>	<b>100%</b>		



# La Moda, Mediana y Media



# Definición De Moda

❖ La moda es el o los números que ocurren con mayor frecuencia en un conjunto de datos.



## Ejemplos De Moda

❖ 3, 5, 3, 8, 5, 3, 6, 8, 1, 3, 9, 4, 3, 6, 8

La moda es 3

❖ 1, 6, 1, 5, 3, 8, 1, 5, 7, 5, 1, 2, 3, 5

La moda es 1 y 5, en este caso es moda compuesta.

# Definición De Mediana

- ❖ Mediana es el numero medio o el promedio de dos números medios cuando los datos están sin organizar en orden numérico en un conjunto de datos.

# Ejemplos De Mediana

❖ 38,40,43,50,56,61,63,68,74

La mediana es 56

❖ 40,46,48,53,56,59,63,68

La mediana es  $53+56=109$

# Definición De Media o promedio

- ❖ La Media es la suma de los datos en un conjunto de datos dividido por la cantidad de números



# Ejemplos de Media

❖ 8,10,7,6,8,15,4,3,8,21

La media es

$$\frac{8+10+7+6+8+15+4+3+8+21}{10} = \frac{90}{10} = 9$$

# Ejemplo de Media

❖ 12,5,7,4,6,5,8,3,5,2,5

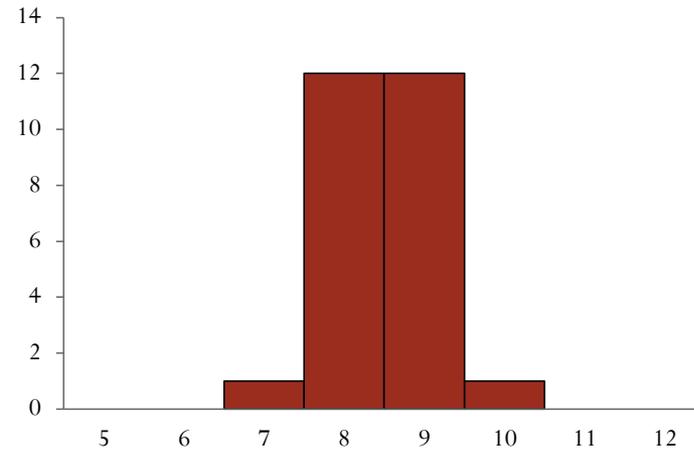
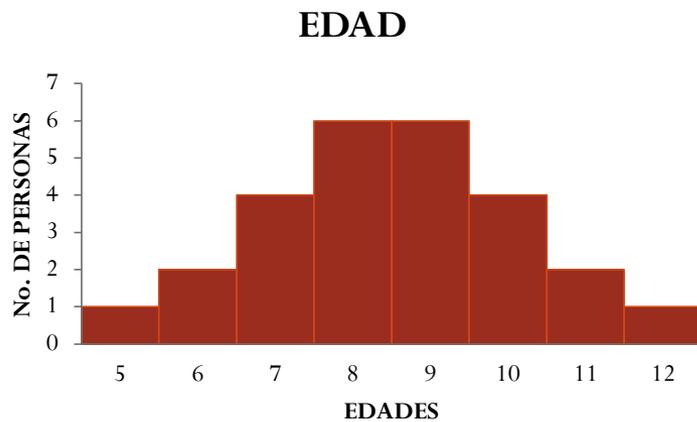
La media es

$$\frac{12+5+7+4+6+5+8+3+5+2+5}{11} = \frac{62}{11} = 5.64$$

# MEDIDAS DE DISPERSIÓN

Las **Medidas de Dispersión**, también llamadas medidas de variabilidad, muestran la variabilidad de una distribución, indicando por medio de un número si las diferentes puntuaciones de una variable están muy alejadas de la media. Cuanto mayor sea ese valor, mayor será la variabilidad, y cuanto menor sea, más homogénea será a la media. Así se sabe si todos los casos son parecidos o varían mucho entre ellos. Las medidas de dispersión nos informan sobre cuánto se alejan del centro los valores de la distribución.

- La *Dispersión* hace referencia a la forma en que se dispersan o alejan las puntuaciones de una distribución o lista de puntajes



MEDIA: 8.5 años

# TIPOS DE MEDIDAS DE DISPERSION

## Medidas de Dispersión Absolutas

- Rango o Recorrido
- Desviación media
- Desviación estándar o típica
- Varianza

## Medidas de Dispersión Relativas

- Coeficiente de variación



# RANGO O RECORRIDO

- El **Rango** es la diferencia entre el mayor y el menor de los datos de una distribución estadística.

$$Rango = X_{m\acute{a}x.} - X_{m\acute{i}n.}$$

## Características de Rango o Recorrido

- Solo suministra información de los extremos de la variable.
- Informa sobre la distancia entre el mínimo y máximo valor observado.
- Se limita su uso a una información inicial.

## Desviación Media

La Desviación Media es la diferencia entre cada valor de la variable estadística y la media aritmética.

La desviación media se representa por:

### Datos no agrupados:

$$D_{\bar{X}} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots \dots \dots * |x_n - \bar{x}|}{N}$$

o

$$D_{\bar{X}} = \frac{\sum_1^n |x_1 - \bar{x}|}{N}$$

### Datos agrupados:

$$D_{\bar{X}} = \frac{|x_{mc1} - \bar{x}| * f_1 + |x_{mc2} - \bar{x}| * f_2 + \dots \dots \dots * |x_{mcn} - \bar{x}| * f_n}{N}$$

o

$$D_{\bar{X}} = \frac{\sum_1^n |x_{mci} - \bar{x}| * f_n}{N}$$

## Características de la Desviación Media

- Todas las observaciones se usan en el cálculo.
- No tiene influencia debido a los valores altos y bajos.
- Es un poco difícil trabajar con los valores absolutos.

## Varianza

La Varianza es la media aritmética del cuadrado de las desviaciones respecto a la media de una distribución estadística.

**Datos no Agrupados:**

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

**Datos Agrupados:**

$$(\sigma^2) = \frac{\sum_{i=1}^n (x_{mci} - \bar{x})^2 \cdot f_i}{n}$$

### Características de la Varianza

- Es siempre un valor no negativo, que puede ser igual o distinta de 0.
- La Varianza es la medida de dispersión cuadrática óptima por ser la menor de todas.
- Si a todos los valores de la variable se le suma una constante la varianza no se modifica.
- Si todos los valores de la variable se multiplican por una constante la varianza queda multiplicada por el cuadrado de dicha constante.

## Desviación Estándar o Típica

La Desviación Típica es la raíz cuadrada de la varianza; es decir, la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación. La desviación típica se representa por  $S$  o  $\sigma$ .

**Datos no Agrupados:**

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

**Datos Agrupados:**

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_{mci} - \bar{x})^2 \cdot f_i}{n}}$$

### Características de la Desviación Estándar o Típica

- Es afectada por el valor de cada observación
- Como consecuencia de considerar desviaciones cuadráticas pone mayor énfasis en la desviaciones extremas que en las demás desviaciones.

# EJEMPLO Datos no Agrupados

- Calcular rango, Desviación media, varianza y desviación estándar para los siguientes puntajes:

10 – 12 – 15 – 18 – 20

$$\text{Rango} = 20 - 10 = 10 \quad \bar{x} = \frac{10 + 12 + 15 + 18 + 20}{5} = \frac{75}{5} = 15$$

$$D_{\bar{x}} = \frac{|10 - 15| + |12 - 15| + |15 - 15| + |18 - 15| + |20 - 15|}{5} = \frac{5 + 3 + 0 + 3 + 5}{5} = \frac{16}{5} = 3,2$$

$$\sigma^2 = \frac{(10 - 15)^2 + (12 - 15)^2 + (15 - 15)^2 + (18 - 15)^2 + (20 - 15)^2}{5} = \frac{25 + 9 + 0 + 9 + 25}{5} = \frac{68}{5} = 13,6$$

$$\sigma = \sqrt{13,6} \approx 3,7$$

# EJEMPLO Datos Agrupados

La siguiente tabla corresponde a la distribución del tiempo, en minutos, que una muestra de operarios de una fábrica demora en armar cierto producto. Calcular las medidas de dispersión.

Tiempo	Número de Operarios
10 - 20	7
20 - 30	15
30 - 40	19
40 - 50	9

$$\bar{x} = \frac{15*7+25*15+35*19+45*9}{50} = \frac{105+375+665+405}{50} = \frac{1550}{50} = 31$$

$$D_{\bar{x}} = \frac{|10-15|+|12-15|+|15-15|+|18-15|+|20-15|}{5} = \frac{5+3+0+3+5}{5} = \frac{16}{5} = 3,2$$

$$\sigma^2 = \frac{(10-15)^2+(12-15)^2+(15-15)^2+(18-15)^2+(20-15)^2}{5} = \frac{25+9+0+9+25}{5} = \frac{68}{5} = 13,6$$

$$\sigma = \sqrt{13,6} \approx 3,7$$

# MEDIDAS DE DISPERSION RELATIVAS

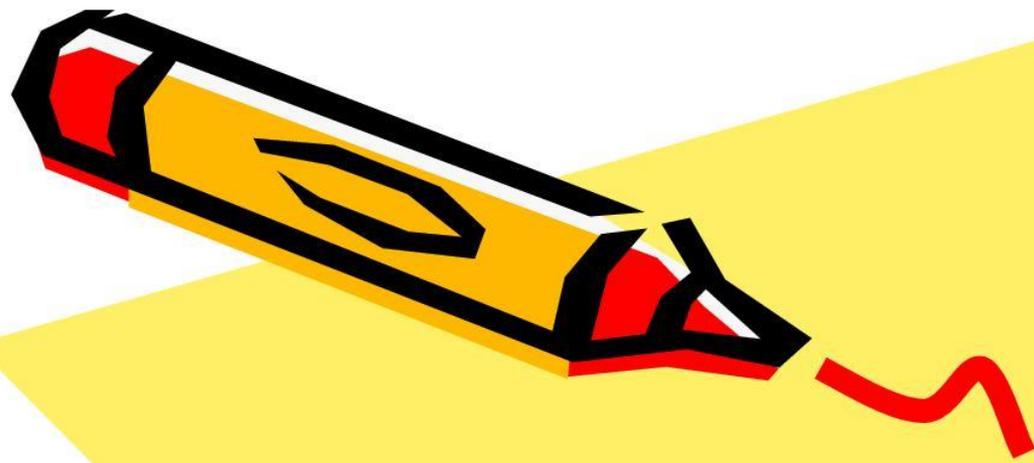
El coeficiente de variación (CV) permite realizar comparaciones entre conjuntos con respecto a la dispersión de sus datos, e incluso entre variables que se miden con diferentes unidades de medida. Matemáticamente, corresponde al cociente entre la desviación estándar y la media aritmética. Esto es:

$$CV = \frac{\sigma}{|\bar{x}|}$$

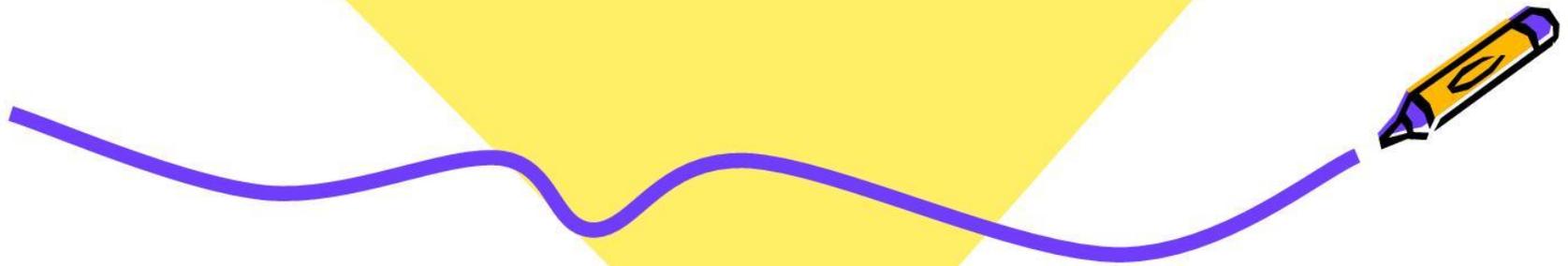
Para expresar el CV en porcentaje, basta con multiplicar el cociente obtenido por 100.

- Mientras menor sea el coeficiente de variación, el conjunto es más homogéneo (los datos son más parecidos entre sí).
- Mientras mayor sea el coeficiente de variación, el conjunto es más heterogéneo (los datos se diferencian más entre sí).

Para el ejemplo anterior :  $\sigma = 3,7$        $CV = \frac{3,7}{31} \approx 0,12$



# MEDIDAS DE POSICIÓN



# Definiciones de Medidas descriptivas

<u>Tendencia Central</u>	<u>Dispersión</u>	<u>De forma</u>	<u>De posición</u>
Permiten analizar los datos en torno a un valor central	Muestran la variabilidad de una distribución.	Evalúan la forma que toman la distribución de frecuencia respecto al grado de distorsión que registra respecto al valor promedio.	Son indicadores usados para señalar que porcentaje de datos dentro de una distribución de frecuencias superan estas expresiones

# Medidas de posición

- Permiten conocer otros puntos característicos de la distribución que no son los valores centrales.
- Se suelen utilizar una serie de valores que dividen la muestra en tramos iguales:
- Informan de como se distribuye el resto de los valores de la serie

# Medidas de posición

	Definición	Fórmula
<b>Percentil</b>	Dividen la población en 100 partes	$P_1, P_2, \dots, P_{99}$
<b>Decil</b>	Se divide la población en 10 partes iguales	$D_1, D_2, \dots, D_9$
<b>Cuartil</b>	se divide la población en 4 partes	$Q_1, Q_2, Q_3$



# Cuartiles

- Los **cuartiles** son los **tres valores** de la variable que **dividen** a un **conjunto de datos ordenados** en **cuatro partes iguales**.
- $Q_1$ ,  $Q_2$  y  $Q_3$  determinan los valores correspondientes al **25%**, al **50%** y al **75%** de los **datos**.
- $Q_2$  coincide con la **mediana**

# Deciles

- Los **deciles** son los **nueve valores** que **dividen** la serie de **datos** en **diez partes iguales**, ordenada la serie en forma creciente o decreciente
- Los **deciles** dan los valores correspondientes al 10%, al 20%... y al 90% de los datos.
- **D<sub>5</sub>** coincide con la **mediana**.

# Percentiles

- Los **percentiles** son los **99 valores** que **dividen** la serie de **datos** en **100 partes iguales**.
- Los **percentiles** dan los valores correspondientes al 1%, al 2%... y al 99% de los datos.
- **P<sub>50</sub>** coincide con la **mediana**.

# DATOS NO AGRUPADOS

No. par

$$Q_K = \frac{kn}{4}$$

$$D_K = \frac{kn}{10}$$

$$P_K = \frac{kn}{100}$$

No. Impar

$$Q_K = \frac{k(n+1)}{4}$$

$$D_K = \frac{k(n+1)}{10}$$

$$P_K = \frac{k(n+1)}{100}$$

DONDE:

k = Es número de Q, D, o P a encontrar

n = Es el total de datos del ejercicio

## DATOS AGRUPADOS PUNTUALMENTE

$$Q_K = \frac{kn}{4}$$

$$D_K = \frac{kn}{10}$$

$$P_K = \frac{kn}{100}$$

DONDE:

k = Es número de Q, D, o P a encontrar

n = Es el total de datos del ejercicio