



ESTADÍSTICA DESCRIPTIVA

Bioestadística

CLAVE:	PERIODO LECTIVO: primer semestre	HT: 1	HP: 3	TH: 4	CRÉDITOS: 5
--------	-------------------------------------	-------	-------	-------	-------------

LICENCIATURA EN MEDICINA VETERINARIA Y
ZOOTECNIA
UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO

Septiembre de 2018

Dr. en BCA. Jorge Osorio Avalos

Contenido: Unidad 1

Estadística Descriptiva

1.1 Introducción

1.2 Datos y variables

1.3 Población y muestra

1.4 Medidas de tendencia central

1.5 Medidas de dispersión

1.6 Presentación de la información

1.7 Ejercicios y problemas

1.1 Introducción

- ▶ **Definición:** La ESTADÍSTICA es la rama de la matemáticas que se encarga de recopilar, organizar y analizar datos numéricos.
- ▶ La ***BIOESTADÍSTICA*** estudiará los fenómenos biológicos, biomédicos y sociales; para solucionar problemas relacionados con la salud y la producción de las poblaciones animales; e influir posteriormente en las funciones del sector agropecuario.

ESTADÍSTICA	DESCRIPTIVA	Su finalidad es describir, ordenar, representar gráficamente y analizar los resultados de dicha muestra de manera sencilla y clara para un conjunto de datos.
	INFERENCIAL	La finalidad es inferir el comportamiento de los parámetros de la población a partir de la información contenida en una muestra.



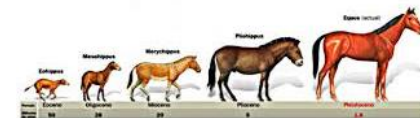
1.2 Datos y variables

Variable: Es la propiedad de interés que se presenta en diferentes modalidades en la población bajo estudio, pero ya asociados los símbolos a las modalidades. Estas se clasifican por un lado en:

a) **Cuantitativas:** cuando expresan magnitud de las modalidades de una propiedad, ejemplo: peso, producción láctea, número de huevos, etc.



b) **Cualitativas:** Cuando expresan tan solo diferentes manifestaciones de la propiedad, no magnitudes, ejemplo: sexo, raza, condición corporal, etc.



Fuente: fotografía <https://igc.fampedia.wordpress.com/tag/inquis-bonaverse/>

Las variables también pueden ser clasificadas como:

c) **Discretas:** Aquellas que identifican propiedades cuyas modalidades cambian de una unidad en unidad, y no existen puntos intermedios, ejemplo: número de parásitos, etc.



d) **Continuas:** Aquellas que identifican propiedades que cambian de un objeto a otro de forma gradual, ejemplo: alzada a la cruz, diámetro torácico, etc.



Medir

- ▶ Es tipificar, etiquetar, clasificar las diferentes modalidades con las que se presenta cierta propiedad de interés de un fenómeno.
- ▶ Asigna número o nombres a las modalidades con las que se presentan las propiedades particulares de los fenómenos naturales o sociales, para representar las relaciones existentes entre dichas propiedades; es por ello que es importante conocer e identificar las escalas de medición.

los tipos de **escalas** de medición son:

Nominal: Asigna nombres o números a las categorías.

- Puede cambiarse sin alterar la información y no se pueden realizar operaciones entre los números asignados.
- Clasifica a los objetos en una y solo una de las categorías.
- Ejemplo: sexo, raza, tipo de sangre, color, etc.

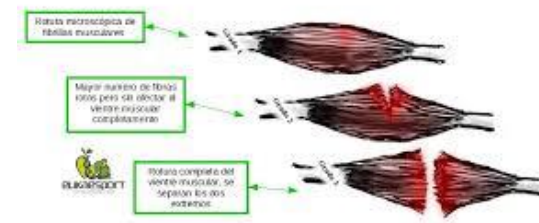


los tipos de **escalas** de medición son:

Ordinal: Las modalidades guardan orden en cuanto a la intensidad de la propiedad.

- ✓ Se asignan nombres o números a las categorías pero guardando la relación de orden.
- ✓ Se pueden cambiar los identificadores siempre y cuando se preserve el orden.
- ✓ No se pueden realizar operaciones aritméticas con los números asignados a las categorías.
- ✓ Ejemplo: Condición corporal, gravedad de una lesión, calidad de la carne, etc.

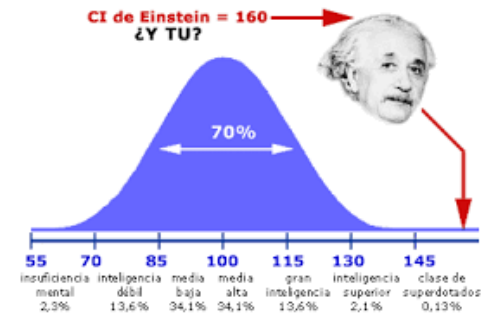
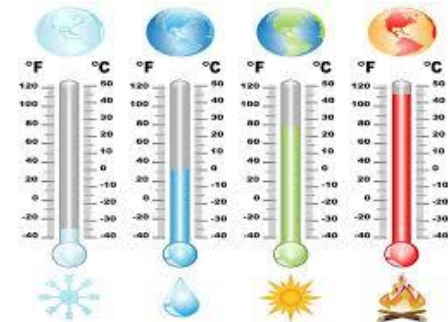
GRADOS DE ROTURA DE FIBRAS



los tipos de escalas de

Intervalo: Se asignan números a las modalidades, pero estos no representan la magnitud, si no solo sirven para evaluar cambios en la intensidad de la propiedad.

- ❖ Como no hay un estado donde la propiedad no existe, el cero es arbitrario.
- ❖ Se conoce el orden entre modalidades y la distancia entre ellas. La unidad de medida es arbitraria.
- ❖ Es la primera escala verdaderamente cuantitativa (todos los métodos numéricos descriptivos).
- ❖ Ejemplo: Temperatura, Inteligencia, movimientos telúricos, etc. Se puede obtener la media, moda, mediana, desviación estándar, etc.



los tipos de escalas de

Razón: Las modalidades presentan cambios graduales en la intensidad de la propiedad. Existe un estado donde la propiedad no se manifiesta.

- ❑ Se asignan números a las modalidades, se asigna “cero” al estado donde la propiedad no existe.
- ❑ Hay un orden entre modalidades y entre los números asignados.
- ❑ Las diferencias entre números corresponde a diferencias entre magnitudes de las modalidades.
- ❑ Se puede realizar cualquier operación aritmética con los números asociados a las modalidades.
- ❑ Ejemplo: Peso de los animales, longitud de una pata, producción láctea, ácido úrico en orina, etc. Se puede obtener la media, mediana, moda, desviación estándar, etc.



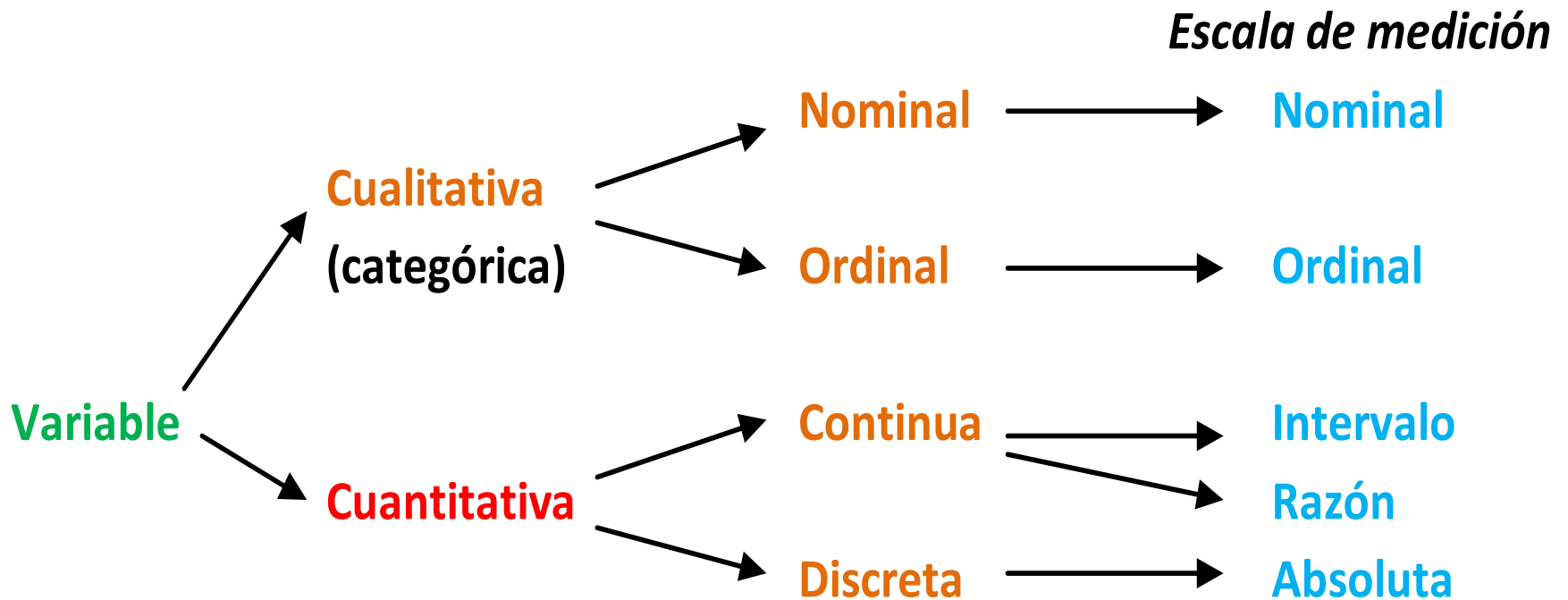
los tipos de **escalas** de

Absoluta: Cuando la propiedad a medir corresponde a un conteo. Se asignan números enteros positivos incluyendo al cero a las modalidades de la propiedad.

- El cero está fijo y la unidad de medida está fija.
- Se pueden realizar todas las operaciones aritméticas con los números asociados a las modalidades.
- Ejemplo: número de garrapatas, número de lechones, número de huevos, etc. Se puede obtener la media, mediana, moda, desviación estándar, etc.



Resumen de variables



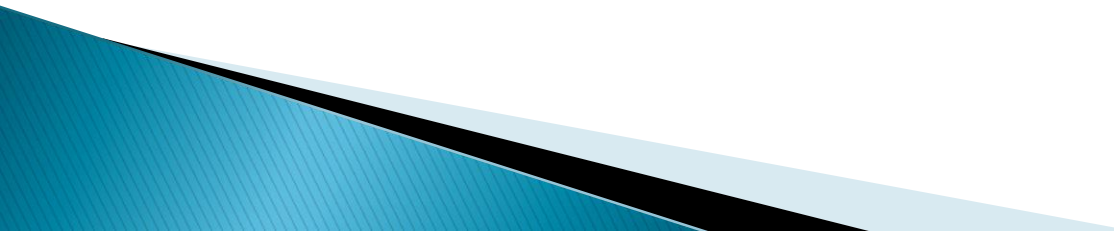
▶ Las variables pueden ser:

a) **Independientes** (explicativas), son la causa: es x , ejemplo: % de proteína en la dieta.

b) **Dependientes** (respuesta), son el efecto; es y , ejemplo: G.D.P.

Proceso aleatorio

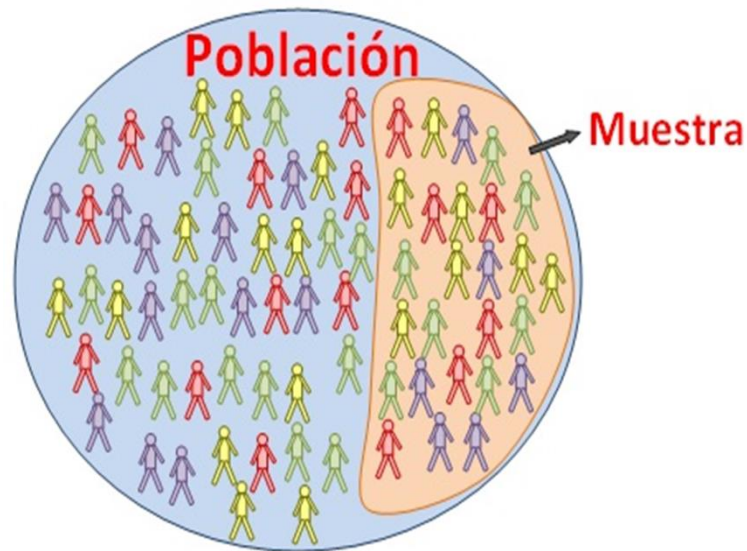
Al lanzar un dado, desconocemos en general cuál va a ser la puntuación obtenida, y por lo tanto sucede cuando por ejemplo extraemos una carta de un mazo o arrojamos una moneda en el aire. Los juegos de azar (cuyo estudio surgió la teoría matemática de la probabilidad) se caracterizan precisamente por el hecho de desconocerse cuál es, de entre toda una serie de resultados posibles el que va a obtenerse tras una determinada jugada, por esta razón los llamamos aleatorios.



Ejemplo de variables y su clasificación:

Tipos de cepas	Explicativa	<u>Cuanti/cualitativa</u>	Nominal	Virus
Lesiones	Respuesta	Categórica	Ordinal	
Niveles de anticuerpos	Respuesta	Categórica	Razón	Suero
Genotipos/familias	Explicativa	Categórica	Nominal	Camaron
% sobrevivencia fam.	Respuesta	Cuantitativa	Razón	Familia
Virus (presencia o no)	Respuesta	Categórica	Nominal	Virus
<u>Defensividad</u>	Respuesta	Cuantitativa	Absoluta	Colonia
Marcadores genéticos	Explicativa	Categórica	Nominal	Colonia
Genotipo (estirpe)	Explicativa	Categórica	Nominal	Colonia
Producción de leche	Respuesta	Cuantitativa	Razón	Cabra
<u>Prod. de</u> <u>proteína</u>	Respuesta	Cuantitativa	Razón	Cabra
Producción de grasa	Respuesta	Cuantitativa	Razón	Cabra
Modelo de <u>evaluac.</u>	Explicativa	Categórica	Nominal	Modelo

1.3 Muestra y población



Población

- ▶ **Definición 1.** Habitualmente se considera como una colección de entidades, por lo general personas, o bien animales, máquinas, plantas o células. En sí es una colección más grande de entidades de interés en un momento particular. Si una población de valores consiste en un **número fijo** de esos valores, se dice que la población es *finita*, o bien si una población consiste en una **sucesión interminable** de valores entonces es una población *infinita*.
- ▶ **Definición 2.** Es el conjunto sobre el que estamos interesados en obtener conclusiones (hacer inferencia). Normalmente es demasiado grande para poder abarcarlo.

Muestra

- ▶ **Definición 1.** Es un subconjunto o parte de la población; es decir, si se tiene únicamente una parte de la población se tiene entonces una *muestra*. Se obtienen por tener una probabilidad estadística de la población. Es la que vamos a medir a través de una escala de medición.
- ▶ **Definición 2.** Es un subconjunto cuyo al que tenemos acceso y sobre el que realmente hacemos las observaciones (mediciones). Debe ser representativo. Está formado por miembros “seleccionados” de la población (individuos, unidades experimentales).
- ▶ **Tamaño de muestra:** es el número de observaciones en una muestra, ejemplo: número de colonias de bacterias en una caja de petri.

Conceptos importantes: Población y Muestra

- **Población** (*population*) es el conjunto sobre el que estamos interesados en obtener conclusiones (hacer inferencia).
 - Normalmente es demasiado grande para poder abarcarlo.



- **Muestra** (*sample*) es un subconjunto suyo al que tenemos acceso y sobre el que realmente hacemos las observaciones (mediciones)
 - Debería ser “representativo”
 - Esta formado por miembros “seleccionados” de la población (individuos, unidades experimentales).

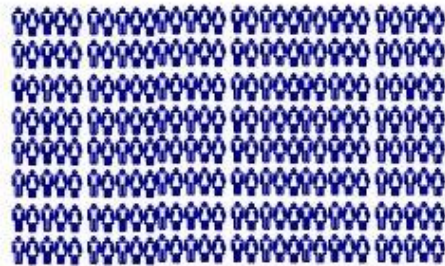


Proceso Estadístico



Población Total vs Muestra

Población



Muestra



Parámetros

Estadísticos

Media

μ

\bar{X}

Desviación típica

σ

S

Varianza

σ^2

S^2

Correlación

ρ_{xy}

r_{xy}

A partir de los estadísticos de la muestra (n) inferimos los valores o parámetros de la población (N)  Estadística Inferencial

1.4 Medidas de tendencia central

- ▶ **Moda:** Es el valor de las observaciones con mayor frecuencia en el conjunto de datos, unimodal, bimodal, trimodal, etc.
- ▶ **Mediana:** Es el valor de la observación que se encuentra al centro de un conjunto de datos ordenados.
- ▶ **Media (\bar{x}):** Promedio o media aritmética. Suma de los valores de las observaciones de un conjunto de datos dividida entre el número de observaciones.
- ▶ $\mu = \frac{\sum xi}{N}$, $\bar{x} = \frac{\sum xi}{n}$ o $\bar{y} = \frac{\sum yi}{n}$

▶ **Media:** Es una medida aritmética. Es la suma de todos los datos dividida entre el total de ellos.

▶ Media de una población:

$$\mu = \frac{\sum_{i=1}^N X_i}{N} = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N}$$

▶ X_i = dato i -ésimo

▶ N = Número total de unidades en la población.

- ▶ Media de una muestra:
- ▶ Utiliza toda la información disponible y es muy útil para hacer inferencias.

$$\bar{x} = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

<u>Escala</u>		<u>Medida de tendencia central</u>
Nominal	→	Moda
Ordinal	→	Moda y mediana
Intervalo	→	Moda, mediana y media
Razón	→	Moda, mediana y media
Absoluta	→	Moda, mediana y media

1.5 Medidas de dispersión



Rango

Es la diferencia del valor máximo al mínimo, valor máximo - valor mínimo. Nos indica que tan dispersos se encuentran los datos y sensible a valores extremos.

- ▶ Es muy sensible a los valores extremos.
- ▶ Ignora la información contenida en medio de los valores extremos.
- ▶ La escala de medición mínima es ordinal.

- ▶ Ejemplo de los pollos..... Rango = $2.20 - 1.35 = 0.85$

- ▶ El **rango intercuartílico (R.I)** es la diferencia entre el valor del tercer y primer cuartil:
$$Vq3 - Vq1 = 1.92 - 1.42 = 0.50$$

- ▶ No se ve afectado por los valores extremos
- ▶ Representa el recorrido cubierto por la mitad central de las observaciones.

Varianza

- ▶ **Varianza:** Promedio de las desviaciones cuadradas con respecto a la media:

$$\sigma^2 = \frac{\sum (x_i - 4)^2}{N} \quad s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

cuando lo tenemos aplicamos los grados de libertad

Una medida de dispersión puede construirse evaluando cuanto se alejan las observaciones de su centro. La escala mínima de medición es de intervalo.

$$\underbrace{(x_i - \bar{x})}$$

Distancia del dato con respecto a su media

Varianza

- ▶ Si un dato es menor que la media la diferencia es negativa, si el dato es mayor que la media la diferencia es positiva. Entonces al sumar todas las diferencias o desviaciones el resultado siempre será igual a “cero”.

- ▶
$$\sum_{i=1}^n (x_i - \bar{x}) = (x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \dots + (x_n - \bar{x}) = 0$$

- ▶ Esta medida no es útil, lo que se hace para evitar los signos es elevar al cuadrado las desviaciones:

- ▶
$$\sum_{i=1}^n (x_i - \bar{x})^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2$$

Varianza

- ▶ La magnitud de esta suma no depende únicamente de la magnitud de las desviaciones, si no del número de datos.
- ▶ Cuando la varianza se obtiene de una población finita:

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

- ▶ Cuando la varianza se obtiene con datos de una muestra:

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

Desviación estándar

Como la varianza se obtiene con la suma de cuadrados, su unidad es el cuadrado de la unidad en la que los datos originales fueron expresados; por ejemplo: si los datos fueron metros, la varianza estaría en metros cuadrados, o bien si son kilos, la varianza estaría en kilos cuadrados.

Para obtener una medida de variabilidad en la misma unidad que los datos, se obtiene la raíz cuadrada y se llama desviación estándar. De la misma forma que la varianza, existe:

- ▶ Una desviación estándar poblacional: $\sigma = \sqrt{\sigma^2}$
- ▶ Una desviación estándar muestral: $s = \sqrt{s^2}$
- ▶ Ejemplo del peso de los pollos: $s = 0.2938$, $s^2 = 0.0863$

Coeficiente de variación (C.V)

- ▶ Representa la desviación estándar, expresada en términos de % con respecto a la media.

- ▶ $C.V. = (S/\bar{x}) * 100$

- * Permite comparar la variación de dos variables sin importar las unidades.

1.6 Presentación de la información

1.7 Ejercicios y problemas

Ejemplo de los datos de los pesos de los pollos (kg):

$$\bar{x} = \frac{\sum x_i}{n} = \frac{27.57}{16} = 1.723 \text{ kg}$$

Mediana = 1.70 kg

Moda = 1.35 kg

Varianza = $s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{1.2947}{15} = 0.863$ D. estándar = $s = .2938$

$$s^2 = \frac{\sum x_i^2 - 16 (1.7223)^2}{15} = \frac{48.80 - 16 (1.723)^2}{15} = \frac{1.3020}{15} = 0.0868 \quad s = 0.2938$$

$$CV = \frac{s}{\bar{x}} * 100 = \frac{0.2938}{1.723} * 100 = 17.05\%$$

Xi	desviación	desviación ²	xi ²
	(xi - 1.72)	(xi - 1.72) ²	
1.35	-0.37	0.1369	1.8225
1.35	-0.37	0.1369	1.8225
1.38	-0.34	0.1156	1.9044
1.40	-0.32	0.1024	1.96
1.43	-0.29	0.0841	2.0449
1.57	-0.15	0.0225	2.4649
1.60	-0.12	0.0144	2.56
1.70	-0.02	0.0004	2.89
1.75	0.03	0.0009	3.0625
1.85	0.13	0.0169	3.4225
1.90	0.18	0.0324	3.61
1.91	0.19	0.0361	3.6481
1.95	0.23	0.0529	3.8025
2.10	0.38	0.1444	4.41
2.13	0.41	0.1681	4.5369
2.20	0.48	0.2304	4.84
Total	0.00	1.2953	48.8017

Ejemplo de los datos del peso (kg) en cabritos a 75 días de edad:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{111.0}{5} = 22.2 \text{ kg}$$

Mediana = 22.0 kg

Moda = ninguna

$$\text{Varianza} = s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{14.80}{4} = 3.71$$
$$\text{D. estándar} = s = .1.924$$

$$\text{CV} = \frac{s}{\bar{x}} * 100 = \frac{1.924}{22.2} * 100 = 8.67\%$$

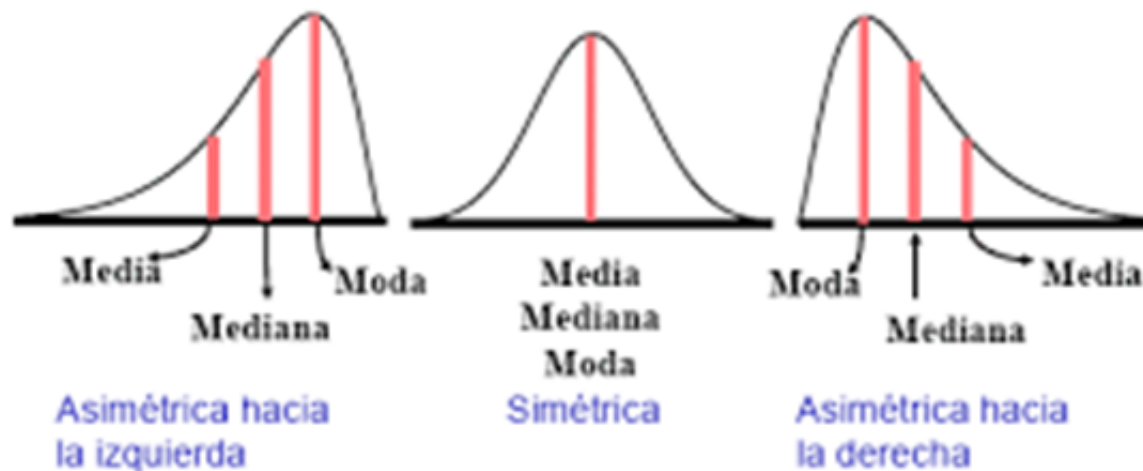
Xi	desviación	desviación²	xi²
	(xi - 22.2)	(xi - 22.2) ²	
20.0	-2.20	4.84	400.0
21.0	-1.20	1.44	441.0
22.0	-0.20	0.04	484.0
23.0	0.80	0.64	529.0
25.0	2.80	7.84	625.0
Total	0.00	14.80	2479.0

Marca de clase

----- Marca de clase: punto medio de cada intervalo; graficando estos puntos se obtiene el polígono de frecuencias.

Moda = 95, 110; Media (\bar{X}) = 98.27; Mediana = 94.5; Rango = 178.

Forma de distribución de frecuencias basadas en polígonos de frecuencias:



Referencias bibliográficas

- Aviva P, Watson P (2006). Statistics for veterinary and animal science. Oxford, Ames, Iowa. Blackwell Publishing. (ISBN: 9781405127813)
- Daniel, Wayne W (1997). Bioestadística: Base para el Análisis de las Ciencias de la Salud. 3a ed. Ed UTEHA. México. (ISBN: 968-18-596-X).
- Navarro Fierro, Ricardo (1988). Introducción a la Bioestadística. Ed. Mcgraw-Hill. México, DF. (ISBN: 9789684223875)
- Steel, Robert George Douglas Torrie, James Hiram, Martínez B, Ricardo, TR (1987). Bioestadística: Principios y Procedimiento. Ed McGraw-Hill México, DF. (ISBN: 9789684514959)