ESTADÍSTICA

Tema 1

(DEI 031)

1.1. DEFINICIONES DE ESTADÍSTICA

¿Qué es la estadística?

Estadística es la ciencia de*:*

- Recolectar
- Describir
- Organizar
- Interpretar

Datos, para transformarlos en información, para la toma mas eficiente de decisiones.

¿Quienes usan la estadística?

Todos…

- Organismos oficiales.
- Diarios y revistas.
- Políticos.
- Deportes.
- Marketing.
- Control de calidad.
- Administradores.
- Investigadores científicos.
- Médicos
- etc.

Tipos de Estadística

LA, DESCRIPTIVA E INFERENCIAL

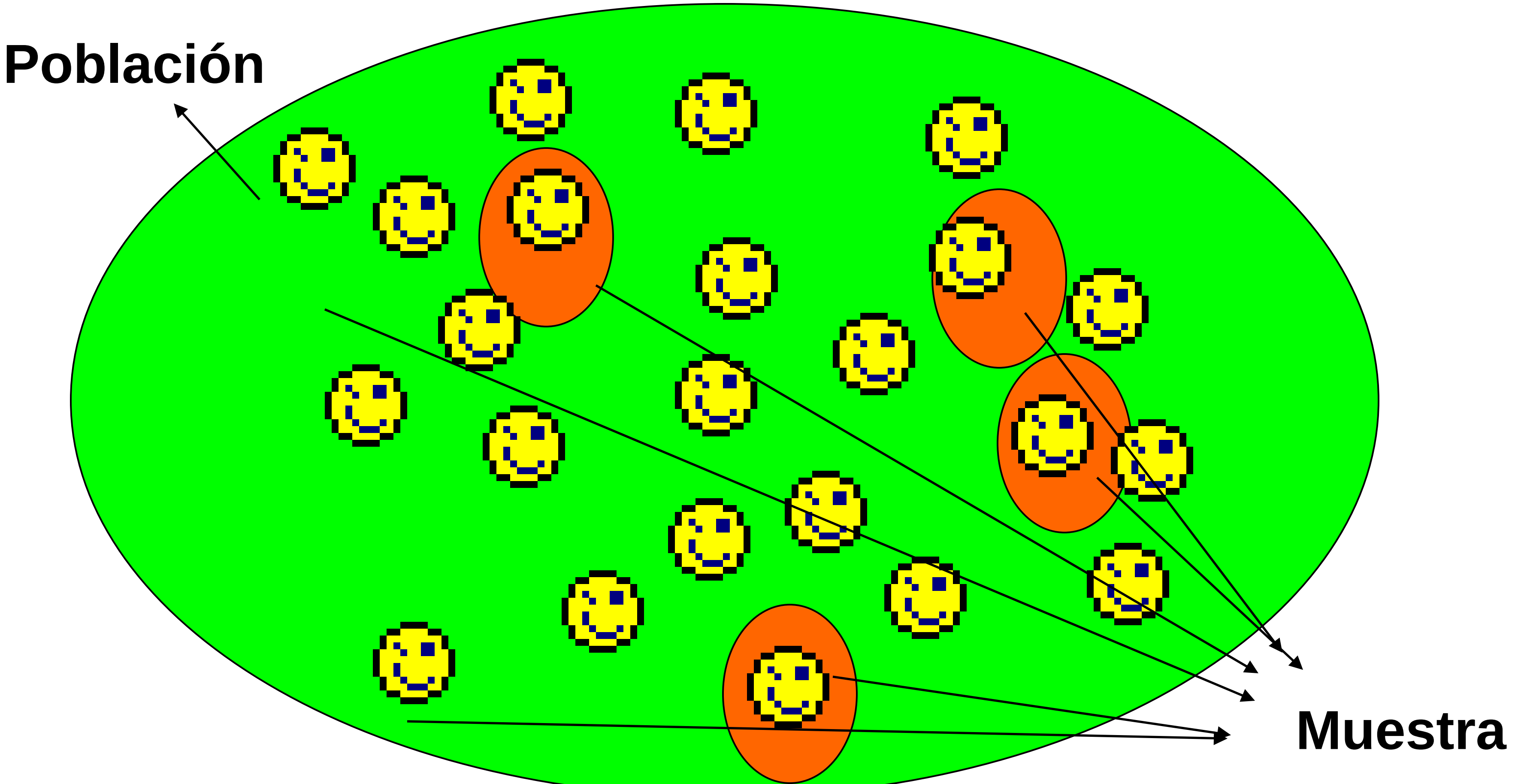
- *Estadística Descriptiva:* Método de recolectar, organizar, resumir y presentar los datos en forma informativa.

- Ejemplo 1: Los datos del Censo de población de 2001.
- Ejemplo 2: La cantidad de robos ocurridos el último mes en el municipio.
- Ejemplo 3: La cantidad de pacientes atendidos en el Hospital municipal el último año, etc.

- *Estadística inferencial:* Métodos usados para determinar algo acerca de la *población*, basado en una *muestra.*
- *Población*(1) es la colección, o conjunto, de individuos, objetos, características o eventos cuyas propiedades serán analizadas.
- *Muestra* es un subconjunto de la población de interés.

(1) Algunos autores utilizan Universo como sinónimo

Población y Muestra



1.1.2. VARIABLE Y TÍPOS DE VARIABLES

Variable

- *Variable:* Característica de interés sobre cada elemento individual de una *población* o *muestra*.
- *Dato:* Valor de la variable asociada a un elemento de la población o muestra. Este valor puede ser un número, una palabra o un símbolo.

Tipos de Variables

- *Cualitativa* o de *Atributos* Clasifica o describe un elemento de la población. Los valores que puede asumir no constituyen un espacio métrico, por lo tanto las operaciones aritméticas, como sumar y obtener promedios, no son significativas.

Tipos de Variables(cont.)

- *Cuantitativa* o *Numérica* Cuantifica un elemento de la población. Los valores que puede asumir constituyen un espacio métrico, por lo tanto las operaciones aritméticas, como sumar y obtener promedios, son significativas.

- Ejemplos: Cantidad de Habitaciones, Número de hijos, Kilómetros recorridos, Tiempo de reacción, Ingreso, etc..

- Las variables cuantitativas se pueden clasificar a su vez en *discretas* o *continuas*.

- *Cuantitativas Discretas:* solo pueden asumir ciertos valores y normalmente hay huecos entre ellos. Normalmente solo son conteos .

- Ejemplo1: cantidad de materias aprobadas.(1, 2,3 ......)
- Ejemplo2: cantidad de productos (1, 2, 3,4...)

- *Cuantitativas Continuas:* Pueden asumir valores continuos, que se obtienen o determinan en un experimento.

- Ejemplo1: La cantidad (kg) de un cierto producto embasado para la venta. (5,01 - 5,10 - 5,12 - 4,99....)
- Ejemplo2: La densidad de una muestra de un producto terminado (kg/m3) (0,86 – 0,88 – 0,85 – 0,87...)

- Las variables cualitativas se miden en escala *nominal* o *ordinal*.
- *Nominal:* los elementos solo pueden ser clasificados en categorías pero no se da un orden o jerarquía.
- Ordinal: Bueno, regular, primero, segundo….

- Ejemplo 1: Barrio de residencia de los alumnos .
- Ejemplo 2: Color de ojos
- Ejemplo 3: Aprovechamiento de los alumnos en un curso.

El fin último de cualquier estudio es aprender sobre las poblaciones. Pero es usualmente necesario, y más práctico, estudiar solo una muestra de cada una de las poblaciones. Definimos:

Usamos una muestra para conocer o estimar características de la población, denominamos:

PARÁMETRO ⇒ una medida resumen calculada sobre la población

ESTADÍSTICO ⇒ una medida resumen calculada sobre la muestra

1.1.2.1. TIPOS DE DATOS

En esta parte presentaremos los distintos tipos de datos o variables que podemos encontrar en una investigación e comentaremos algunas estrategias para el manejo de datos con una computadora.

1.1.2.2. CARACTERÍSTICAS DE LOS CONJUNTOS DE DATOS.

Denominaremos:

- UNIDAD DE ANÁLISIS O DE OBSERVACIÓN al objeto bajo estudio. El mismo puede ser una persona, una familia, un país, una región, una institución o en general, cualquier objeto.

- VARIABLE a cualquier característica de la unidad de observación que interese registrar, la que en el momento de ser registrada puede ser transformada en un número.

- VALOR de una variable, OBSERVACIÓN o MEDICIÓN, al número que describe a la característica de interés en una unidad de observación particular.

- CASO o REGISTRO al conjunto de mediciones realizadas sobre una unidad de observación.

Consideremos el siguiente ejemplo:

Caso	Sexo	Lugar nacimiento	Edad	PAS	
1	F	J1	35	110	
2	M	J2	28	120	⇐ REGISTRO
3	M	J2	59	136	

⇑ VARIABLE

OBSERVACIÓN

Cuando se diseña una investigación, se intenta estudiar de qué modo una o más variables (*variables independientes*) afectan a una o más variables de interés (*variables dependientes*).

Por ejemplo en un experimento, el investigador impone a los sujetos condiciones (variable independiente) y estudia el efecto de la misma sobre una característica del sujeto (aparición de una cierta característica, modificación de una condición, etc.).

Un paso importante al comenzar a manejar un conjunto de datos es identificar *cuántas variables* se han registrado y *cómo* fueron registradas esas variables, lo que permitirá definir la estrategia de análisis.

1.2. DISTRIBUCIÓN Y FRECUENCIA

Las distribuciones de frecuencia es la organización de datos crudos en forma de tablas, organizando dichos datos en clases y frecuencias.

- Las gráficas o diagramas proveen de inmediato un sentido de proporción.
- Se debe explicar por sí misma y cumplen con especificaciones de presentación.
- Un listado de todas las puntuaciones observadas (o de las categorías) de una variable y de la frecuencia, *f* , de cada puntuación o categoría
  - Presenta los valores de los datos y su frecuencia o las veces que se repite la observación
- La frecuencia de una puntuación o de una categoría no ofrece mucha información por sí mismo así que computamos proporciones y porcentajes

Frecuencia

- ***Frecuencia absoluta***: (n_i) El número de veces que se repite cada valor o dato de la variable.

- ***Frecuencia relativa***: (f_i) La frecuencia absoluta dividida por el número de datos.

$$f_i = \frac{n_i}{N}$$

donde N es el número de datos.

Frecuencia

- ***Frecuencia absoluta acumuladas***: (N_i). Es el número de datos que hay igual al considerado o inferiores a él.

- ***Frecuencia relativa acumuladas***: (F_i). Es cada frecuencia acumulada dividida por el número de datos.

Frecuencia

- Frecuencia absoluta:

$$n_1 + n_2 + \ldots + n_i + \ldots + n_n = n = \sum_{i=1}^{n} n_i$$

- Frecuencia absoluta acumulada:

$$\longrightarrow \left\{ \begin{array}{l} N_1 = n_1 \\ N_2 = n_1 + n_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ N_k = n_1 + n_2 + \ldots + n_k \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ N_n = n_1 + n_2 + \ldots + n_k + \ldots + n_n = N \end{array} \right.$$

- Frecuencia relativa:

$$f_i = \frac{n_i}{n} \longrightarrow \sum_{1}^{n} f_i = 1$$

- Frecuencia relativa acumulada.

$$F_i = f_1 + \ldots + f_i$$

Tabla de frecuencias

Valor	Frecuencia absoluta	Frecuencia absoluta accum.	Frecuencia relativa	Frecuencia relativa accum.
x_i	n_i	N_i	f_i	F_i
x_1	n_1	N_1	f_1	F_1
x_2	n_2	N_2	f_2	F_2
x_3	n_3	N	f_3	1
	$\Sigma n_i = N$		1	

1.2.1. HISTOGRAMAS POLÍGONOS Y OTROS GRÁFICOS

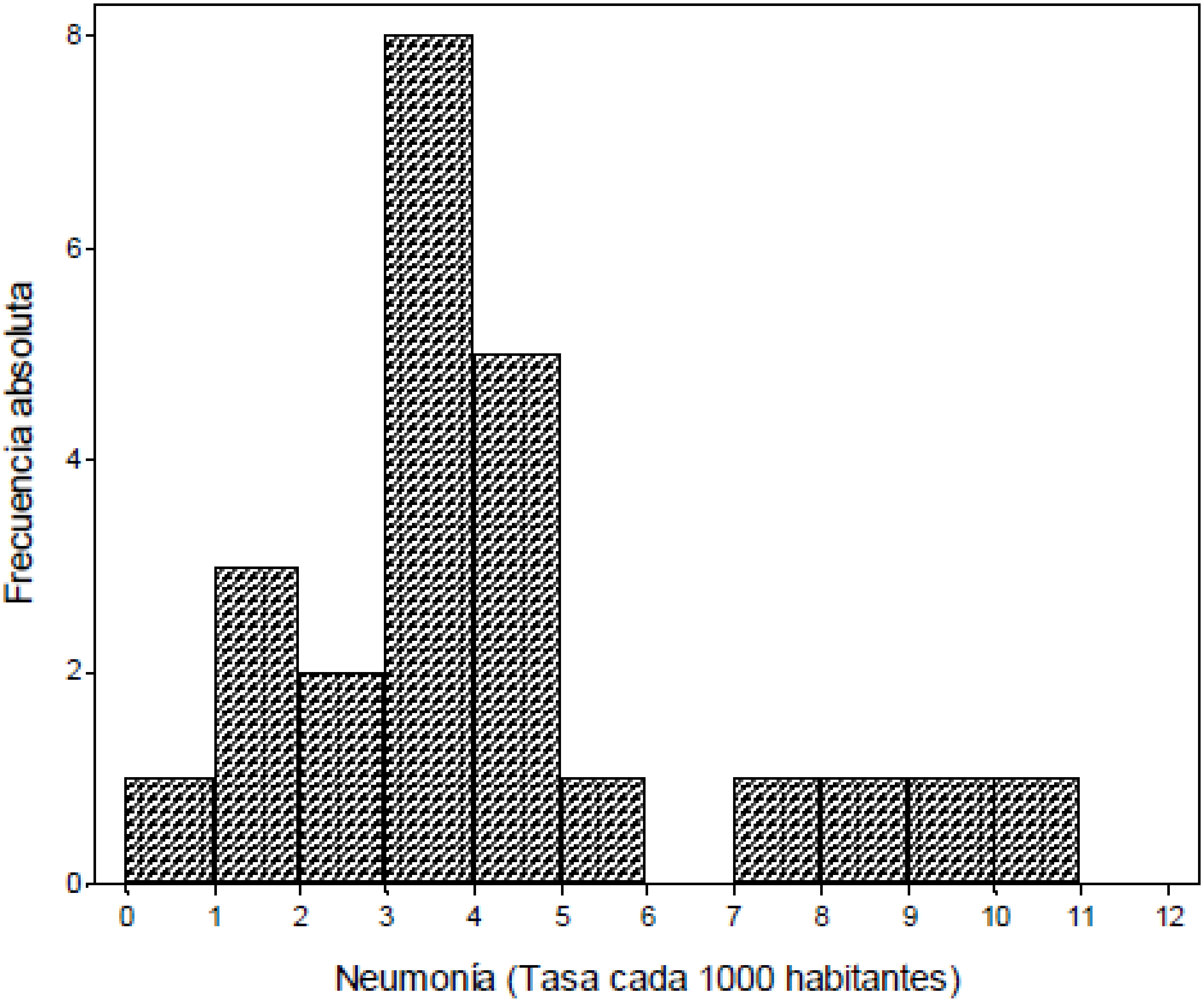
El modo más simple de presentar datos categóricos es por medio de una tabla de frecuencias.

1.2.2 HISTOGRAMA

El histograma es el más conocido de los gráficos para resumir un conjunto de datos numéricos y pretende responder a las mismas preguntas que un gráfico de barras.

Para construir un histograma es necesario previamente construir una *tabla de frecuencias.*

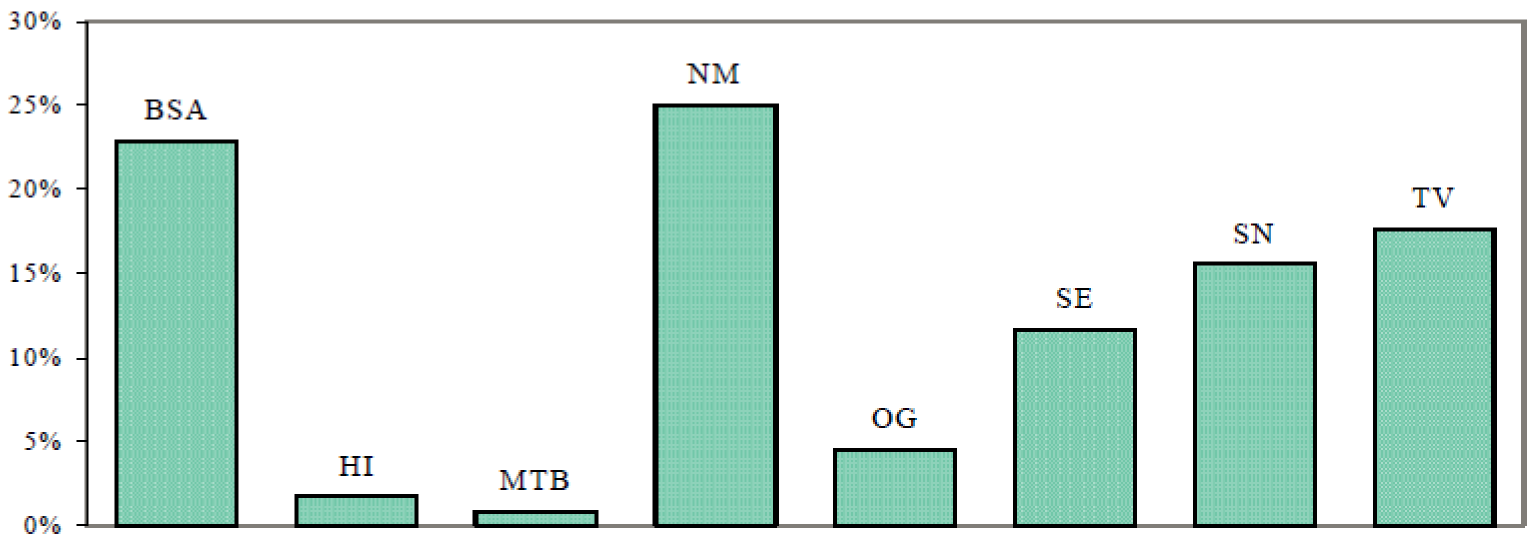
Histograma



Frecuencia absoluta
0
2
4
6
8
0
1
2
3
4
5
6
7
8
9
10
11
12
Neumonía (Tasa cada 1000 habitantes)

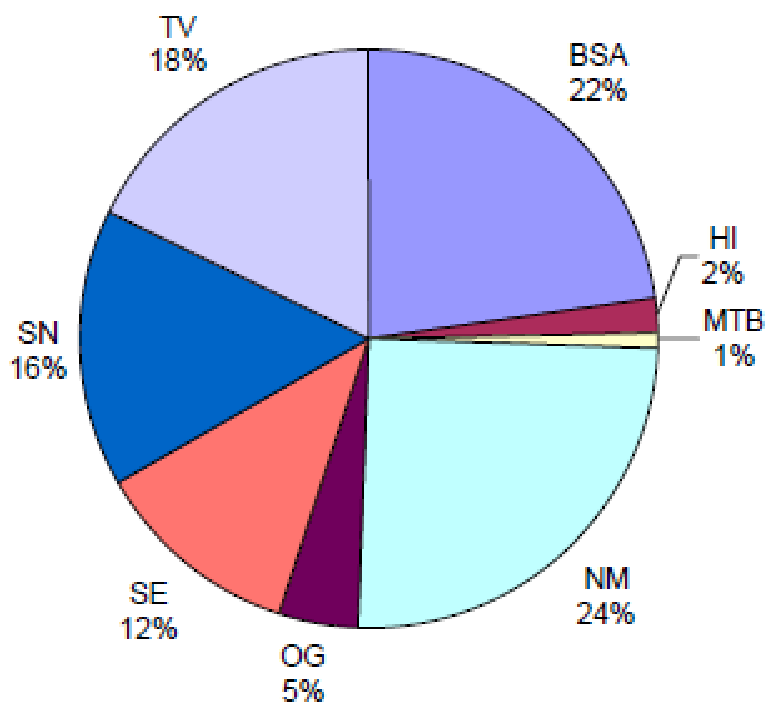
1.2.3. GRÁFICO DE BARRAS

Este gráfico es útil para representar datos categóricos nominales u ordinales. A cada categoría o clase de la variable se le asocia una barra cuya *altura* representa la frecuencia o la frecuencia relativa de esa clase. Las barras difieren sólo en altura, no en ancho.



1.2.4. GRÁFICO DE TORTAS

En este gráfico, ampliamente utilizado, se representa la frecuencia relativa de cada categoría como una porción de un círculo, en la que el ángulo se corresponde con la frecuencia relativa correspondiente. Como en todo gráfico es importante indicar el número total de sujetos.




Histogram
10
8
6
4
2
0
Frequency
12.5
13.5
14.5
15.5
16.5
17.5
18.5
19.5
millas por galón
Frequency

1.3. MEDIDAS RESÚMENES O DE TENDENCIA CENTRAL

En este punto, introduciremos distintas formas de resumir la distribución muestral o poblacional de una variable NUMÉRICA y finalmente presentaremos un tipo de gráfico que se construye a partir de medidas resúmenes.

Resumir un conjunto de datos es pasar de una visión detallada a una generalización simple e informativa tratando de preservar las características esenciales.

Las medidas resúmenes son útiles para comparar conjuntos de datos cuantitativos y para presentar los resultados de un estudio y se clasifican en dos grupos principales:

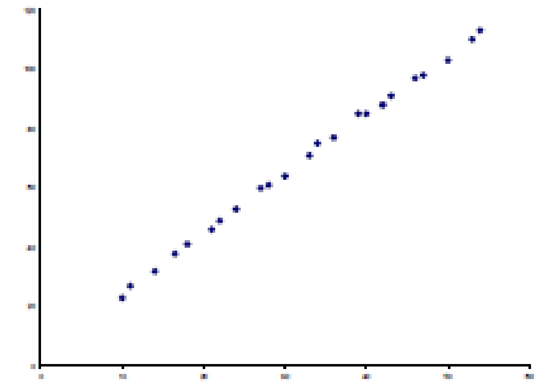
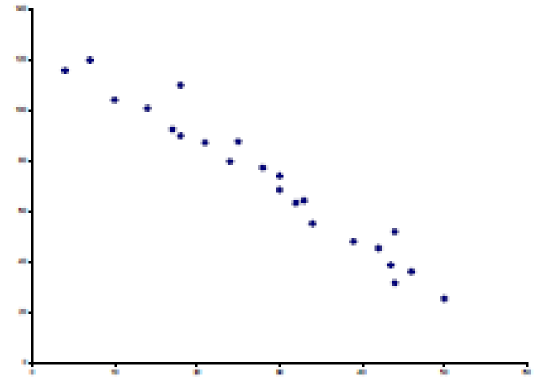
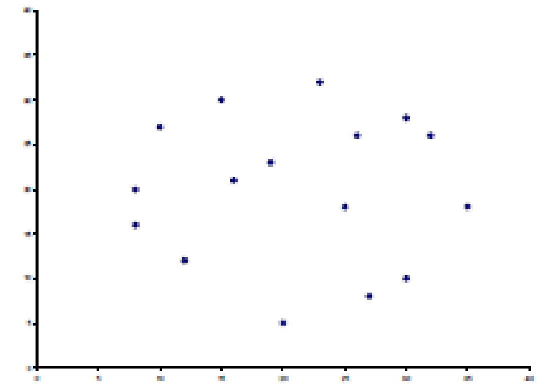
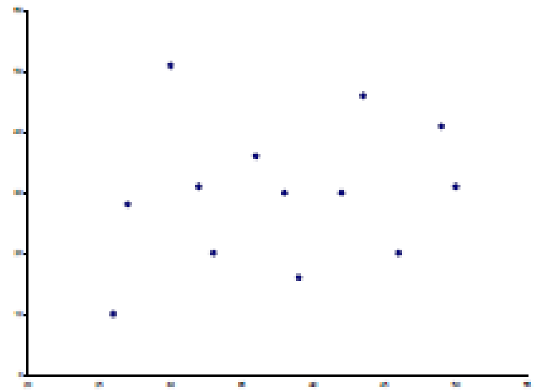
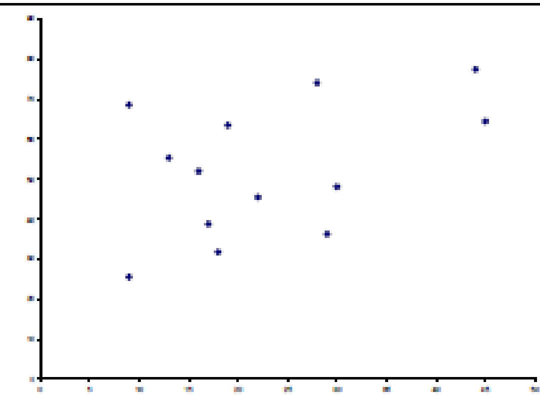
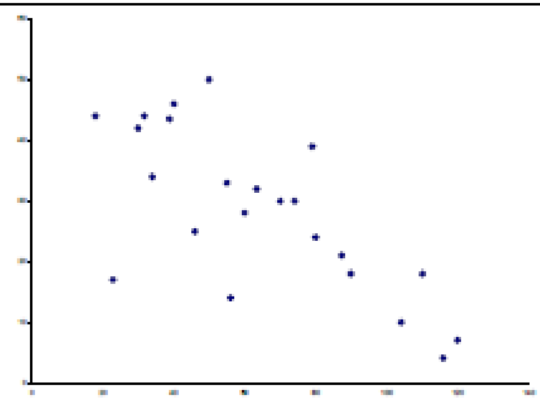
Medidas de posición o localización $\Rightarrow$ describen un valor alrededor del cual se encuentran las observaciones.

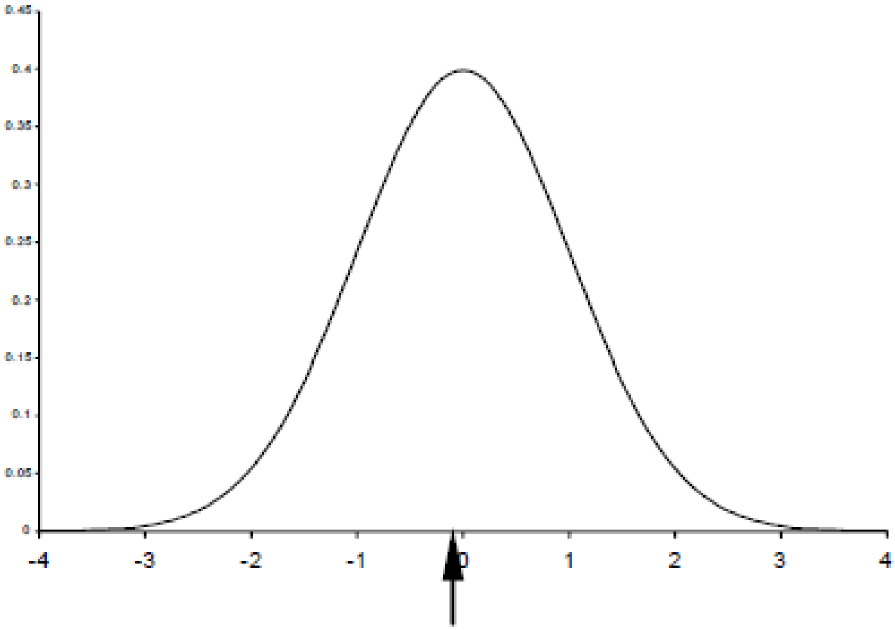
Medidas de dispersión o escala $\Rightarrow$ pretenden expresar cuan variable es un conjunto de datos.

1.3.1. MEDIDAS DE POSICIÓN O LOCALIZACIÓN

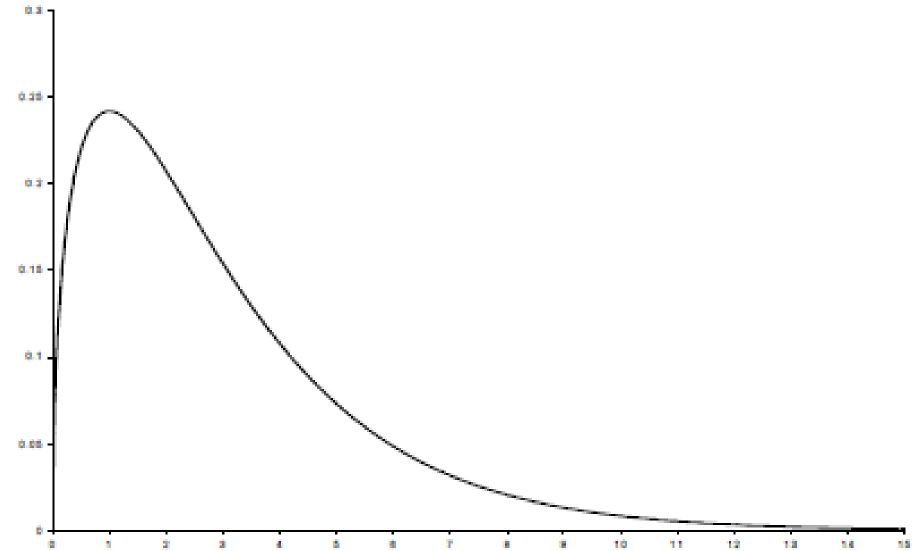
La **medidas de centralización** nos indican en torno a qué valor (centro) se distribuyen los datos.

Un modo de resumir un único conjunto de datos numéricos es a través de un número que debería ser *típico* para el grupo. No debería ser ni demasiado grande, ni demasiado pequeño y debería estar tan cerca del "centro" de la distribución como sea posible. Por lo tanto, una *medida de posición* es un número que pretende indicar dónde se encuentra el *centro* de la distribución de un conjunto de datos. Pero, ¿dónde se encuentra el "centro" de una distribución?

	
r = 0.9993	r = -0.9775
	
r = 0.0139	r = 0.418
	
r = 0.2241	r = - 0.718



CENTRO



¿CENTRO?

El centro es fácil de identificar si la distribución es simétrica, pero es difícil si la distribución es asimétrica. Por esta razón, no hay una única medida de posición para resumir una distribución.

a) PROMEDIO O LA MEDIA ARITMÉTICA

Es la medida de posición más frecuentemente usada. Para calcular la media aritmética o promedio de un conjunto de observaciones se suman todos los valores y se divide por el número total de observaciones.

Definición

Si tenemos una muestra de *n* observaciones y denotadas por X1, X2, ..., Xn, definimos la *media muestral* X del siguiente modo:

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

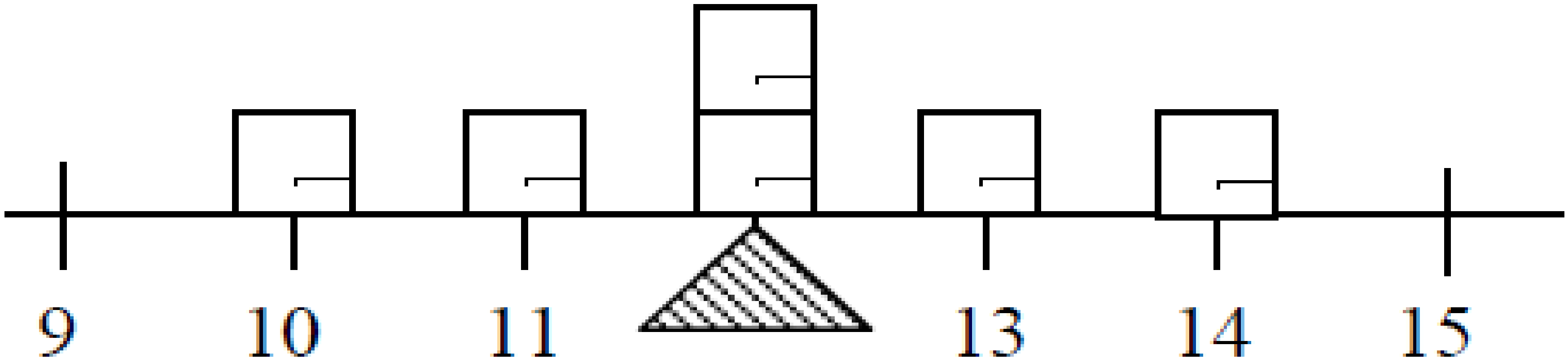
Ejemplo.

$X_1 = 10 \qquad X_2 = 14 \qquad X_3 = 12 \qquad X_4 = 11 \qquad X_5 = 12 \qquad X_6 = 13$

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_6}{n} = \frac{10+14+12+11+12+13}{6} = \frac{72}{6} = 12$$

Características y propiedades de la media.

a) Se usa para datos numéricos.
b) Representa el *centro de gravedad* o el punto de equilibrio de los datos.
Podemos imaginar a los datos como un sistema físico, en el que cada dato tiene una "masa" unitaria y lo ubicamos sobre una barra en la posición correspondiente a su valor. La media representa la posición en que deberíamos ubicar el punto de apoyo para que el sistema esté en equilibrio.



c) La suma de las distancias de los datos a la media es cero. Esta propiedad está relacionada con el hecho que la media es el centro de gravedad de los datos.

d) Es muy sensible a la presencia de datos atípicos

b) LA MEDIANA MUESTRAL

La mediana es el dato que ocupa la posición central en la muestra ordenada de menor a mayor.

¿Cómo calculamos la mediana de una muestra de n observaciones?

1. Ordenamos los datos de menor a mayor.
2. La mediana es el dato que ocupa la posición (n+1/2) en la lista ordenada.

Si el número de datos es *impar*, la mediana X es el dato que ocupa la posición central.

Si el número de datos es *par*, la mediana X es el promedio de los dos datos centrales.

Ejemplo

- *n impar*

$X_1 = 10 \qquad X_2 = 14 \qquad X_3 = 12 \qquad X_4 = 18 \qquad X_5 = 11$

Ordenamos los datos:

$10 \qquad 11 \qquad 12 \qquad 14 \qquad 18$

La posición de la mediana es $\frac{n+1}{2} = \frac{5+1}{2} = 3$ (tercer dato), es decir $\tilde{X} = 12$.

- *n par*

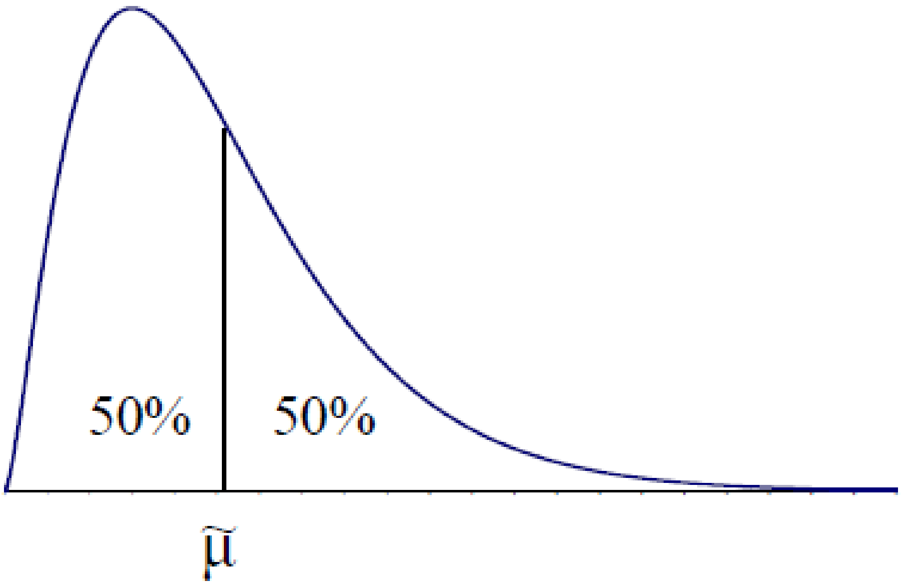
$X_1 = 10 \qquad X_2 = 14 \qquad X_3 = 12 \qquad X_4 = 18 \qquad X_5 = 11 \qquad X_6 = 23$

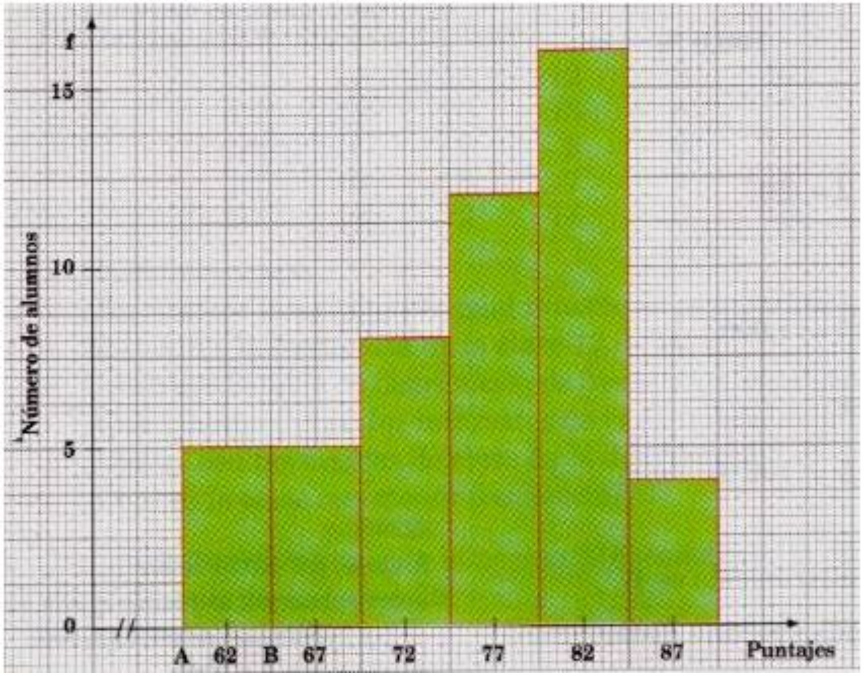
Ordenamos los datos:

$10 \qquad 11 \qquad 12 \qquad 14 \qquad 18 \qquad 23$

Posición de la mediana $\Rightarrow \frac{6+1}{2} = 3.5$

Obtenemos la mediana promediando el tercer y cuarto dato: $\tilde{X} = \frac{12+14}{2} = 13$.





Propiedades de la mediana

a) La mediana puede ser usada no sólo para *datos numéricos* sino además para *datos ordinales*, ya que para calcularla sólo es necesario establecer un orden en los datos.
b) Si la distribución de los datos es aproximadamente simétrica la media y la mediana serán aproximadamente iguales.

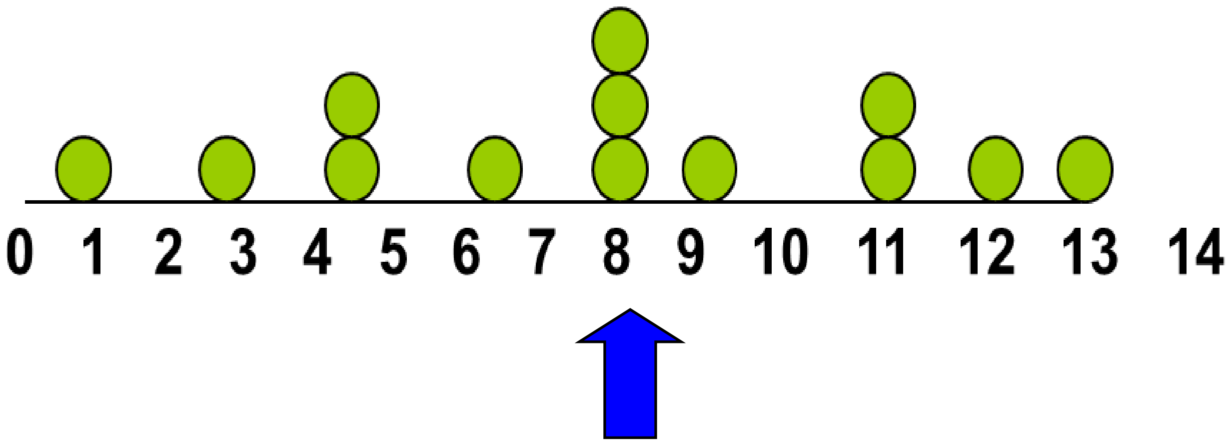
c) LA MODA

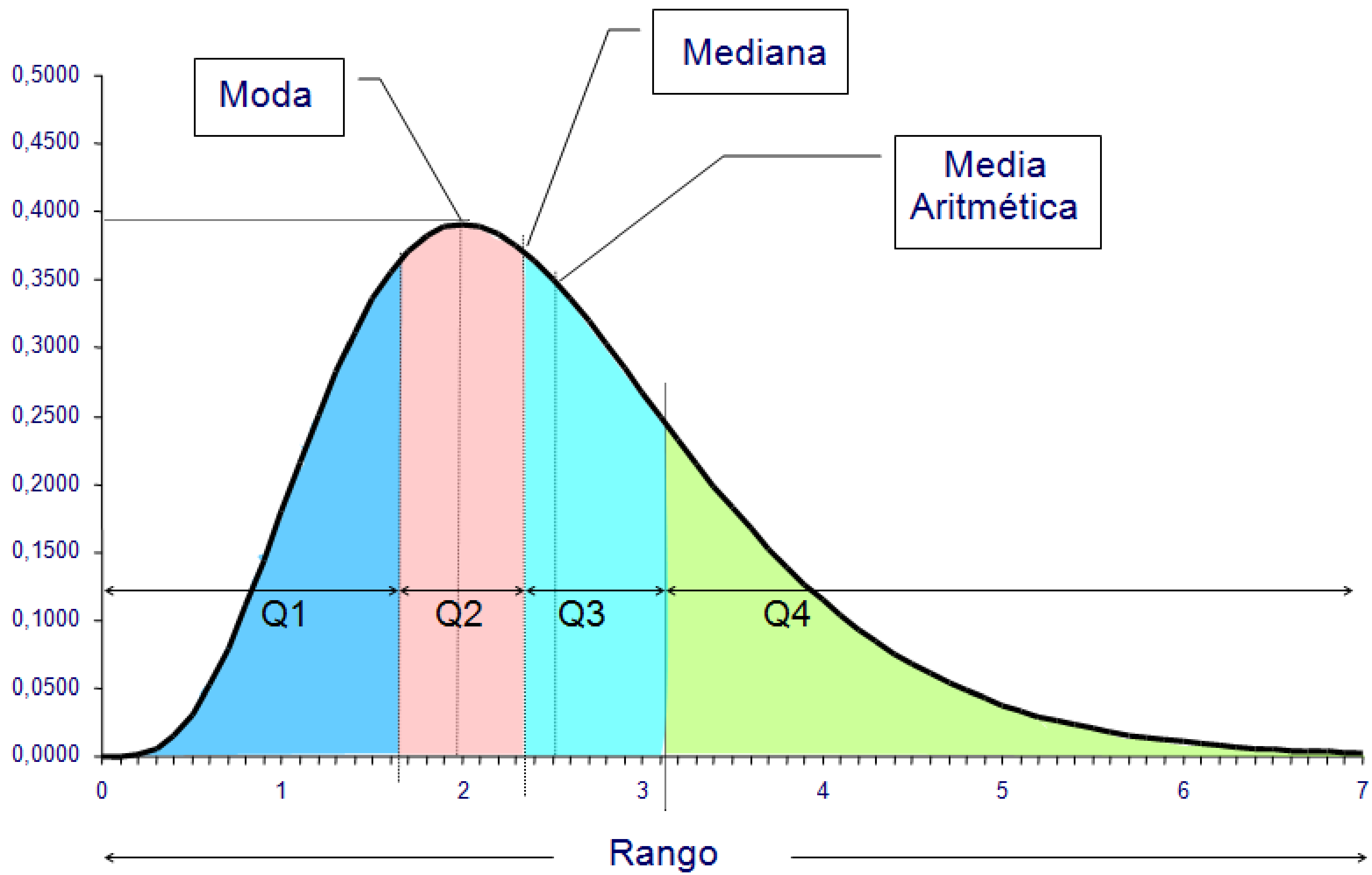
La moda es el dato que ocurre con mayor frecuencia en el conjunto.
Es una medida de poca utilidad salvo para datos categóricos en los que suele interesar identificar la categoría con mayor cantidad de datos.

ES EL DATO DE MAYOR FRECUENCIA

Cuando se considera la distribución poblacional de una variable continua, decimos que esta es UNIMODAL si presenta un pico y BIMODAL si aparecen dos picos claros.

- Valor que **ocurre más** a menudo.
- **No** es afectada por **valores extremos.**
- Puede **no existir** una moda.
- Pueden **haber varias** modas.

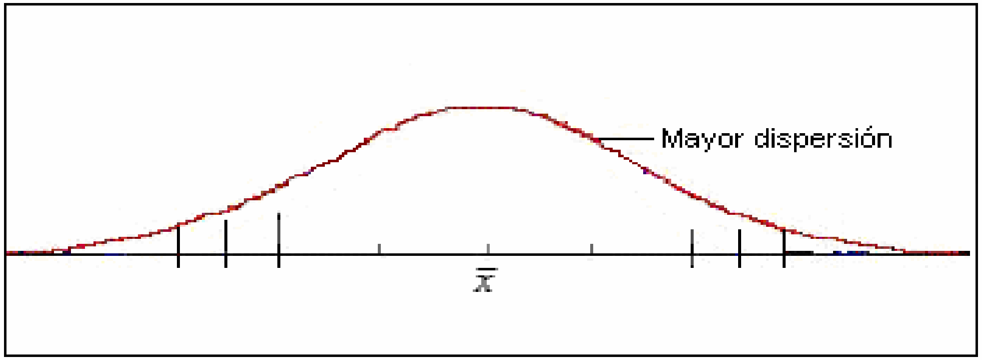


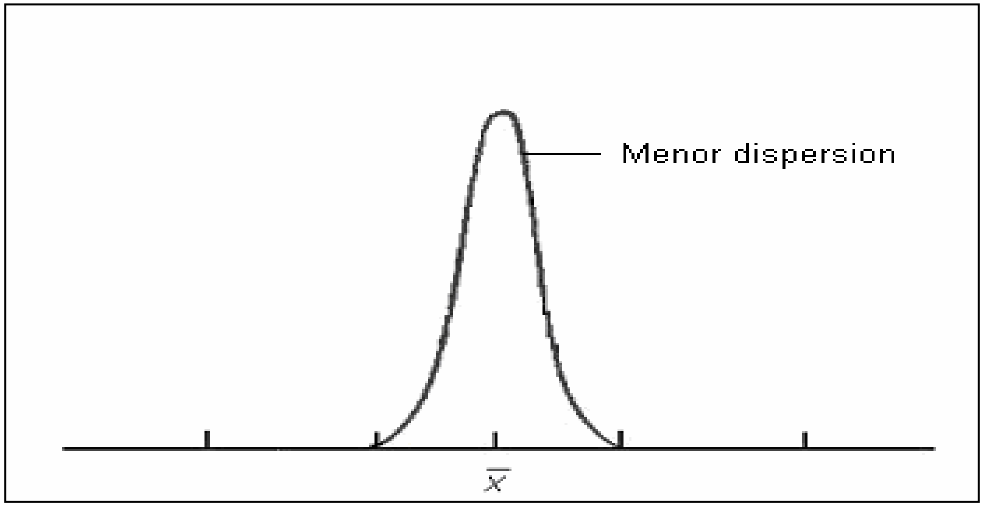

Moda
Mediana
Media
Aritmética
0,5000
0,4500
0,4000
0,3500
0,3000
0,2500
0,2000
0,1500
0,1000
0,0500
0,0000
Q1
Q2
Q3
Q4
0
1
2
3
4
5
6
7
Rango

Medidas de Dispersión

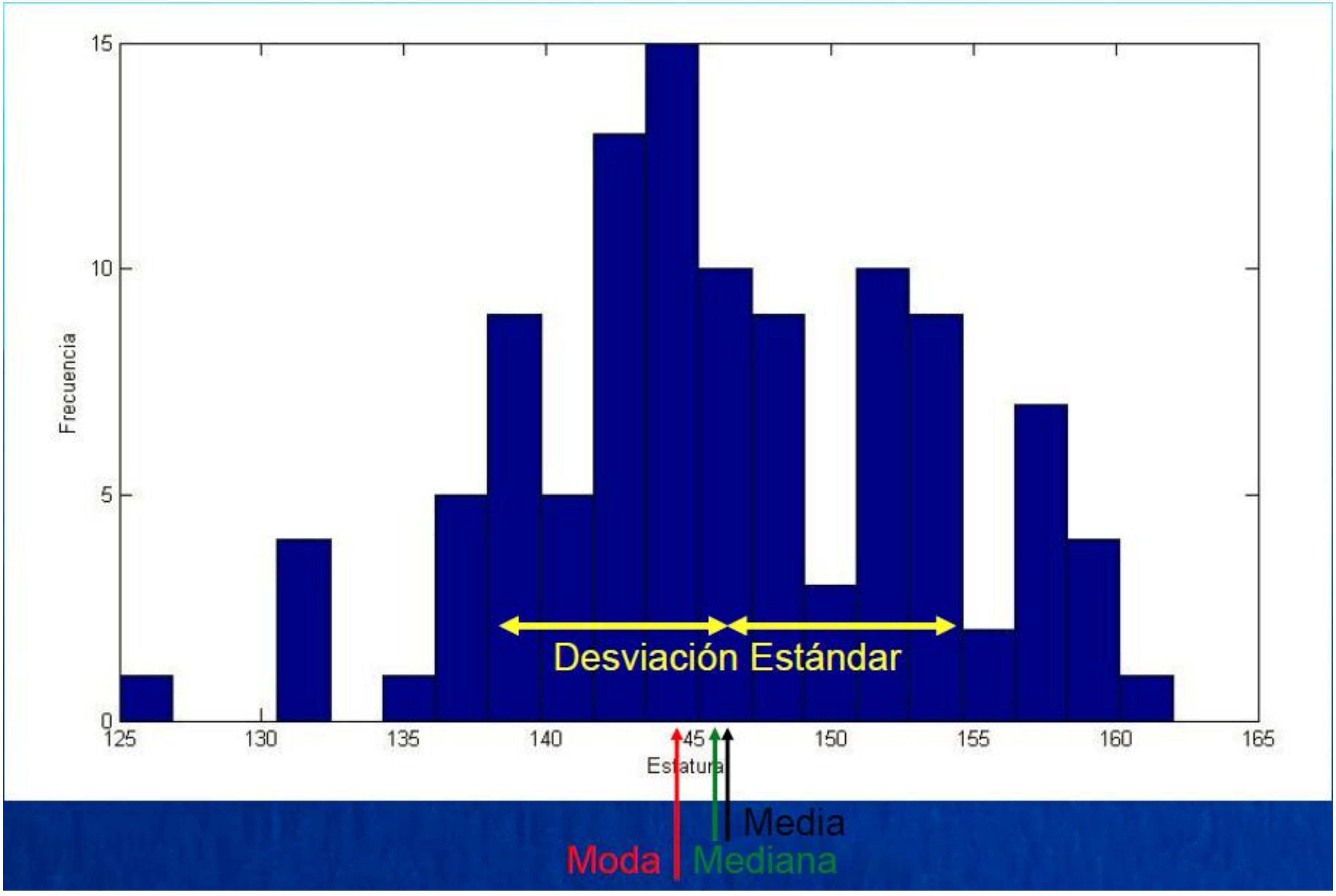
MEDIDAS DE DISPERSIÓN

Miden qué tanto se dispersan las observaciones alrededor de su media.





Medidas de tendencia central y dispersión:



Existen diversas medidas estadísticas de dispersión, pero las principales son:

- **Rango**
- **Varianza**
- **Desviación estándar**
- **Coeficiente de variación**

Alcance o Rango

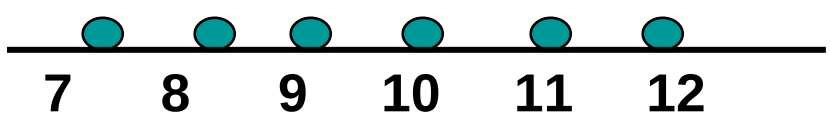
$$\boxed{R = \max(x_i) - \min(x_i)}$$

- Diferencia entre la mayor y la menor de las observaciones

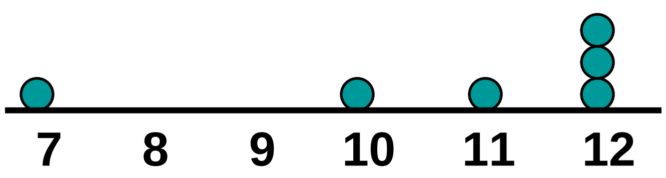
  Alcance = $x_{mayor} - x_{menor}$

- No toma en cuenta la forma en que están distribuidos los datos.

Alcance: 12 - 7 = 5



Alcance de 12-7 = 5



Ejemplo 2.

Hay dos conjuntos sobre la cantidad de lluvia (mm) en Villa Montes y Yacuiba en un año.

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Villa Montes	86	135	178	170	231	290	231	305	244	122	66	71
Yacuiba	40	77	83	89	147	168	184	252	209	101	32	13

Calcula el rango en cada una de las ciudades.

Solución.

Aplicando la fórmula correspondiente tenemos:

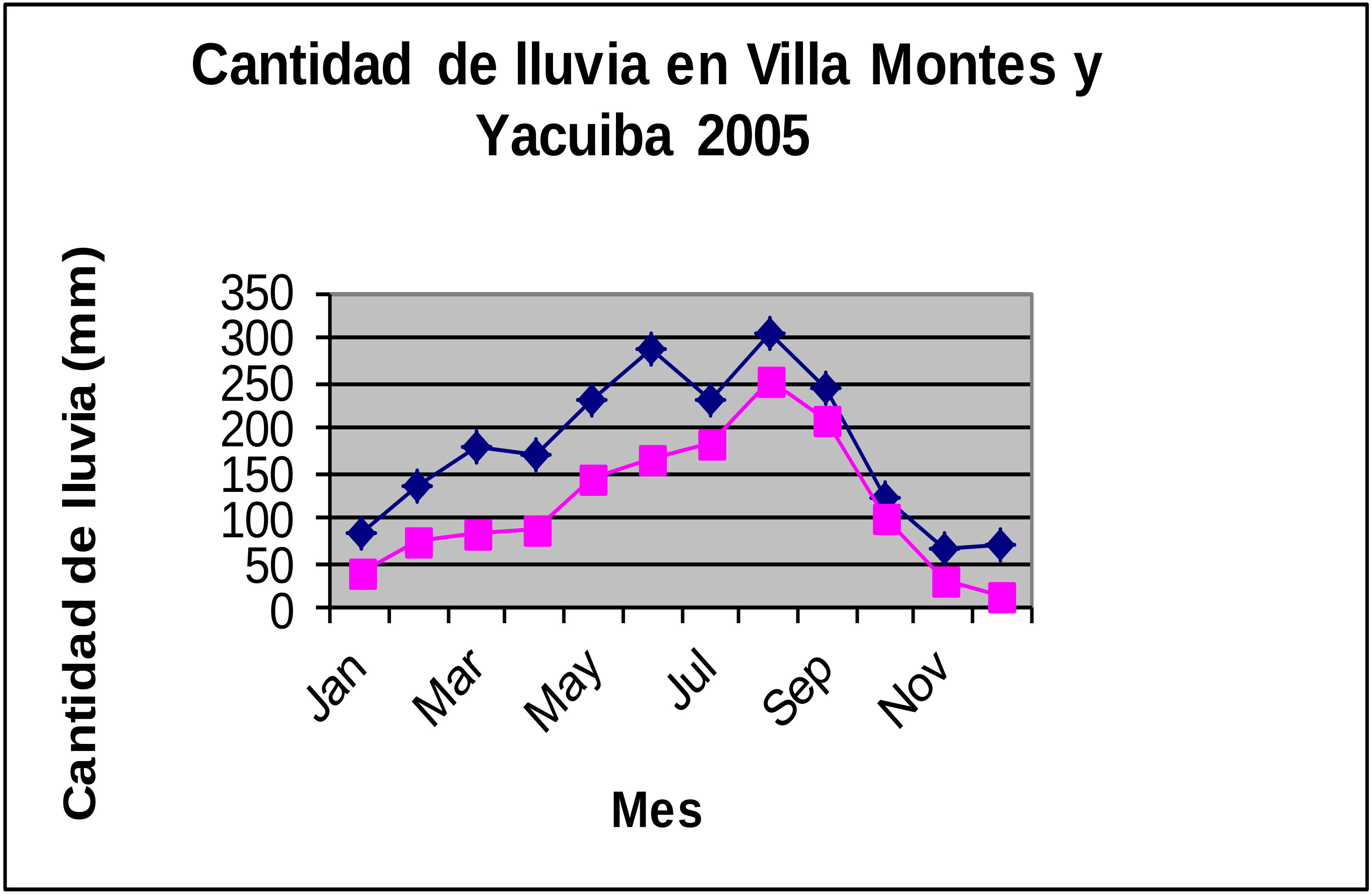
Villa Montes

$$Rango = 305mm - 66mm = 239mm$$

Yacuiba

$$Rango = 252mm - 13mm = 239mm$$

En este caso se puede observar que el rango es el mismo para ambos casos aunque las cantidades sean diferentes.


Cantidad de lluvia en Villa Montes y Yacuiba 2005
Cantidad de lluvia (mm)
350
300
250
200
150
100
50
0
Jan
Mar
May
Jul
Sep
Nov
Mes

VARIANZA (Datos no agrupados)

Mide la distancia existente entre los valores de la serie y la media. Se calcula como sumatoria de las diferencias al cuadrado entre cada valor y la media, multiplicadas por el número de veces que se ha repetido cada valor. La sumatoria obtenida se divide por el tamaño de la muestra.

FÓRMULA

Muestral $\longrightarrow$

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

Poblacional $\longrightarrow$

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

La varianza siempre será mayor que cero. Mientras más se aproxima a cero, más concentrados están los valores de la serie alrededor de la media. Por el contrario, mientras mayor sea la varianza, más dispersos están.

Ejemplo 1.

Calcula la varianza para los siguientes datos

2 1 2 4 1 3 2 3 2 0 5 1

Solución.

Primero es necesario obtener la media. En este caso $\bar{x} = 2.16$

Ahora aplicamos la fórmula correspondiente

$$s^2 = \frac{(2-2.16)^2+(1-2.16)^2+(2-2.16)^2+(4-2.16)^2+(1-2.16)^2+(3-2.16)^2+(2-2.16)^2+(3-2.16)^2+(2-2.16)^2+(0-2.16)^2+(5-2.16)^2+(1-2.16)^2}{12-1}$$

$$s^2 = \frac{21.6672}{11} = 1.9697$$

Ejemplo 2.

A continuación se muestran dos conjuntos de datos obtenidos a partir de un experimento químico que realizaron dos estudiantes distintos. Calcular la varianza.

			Volumen de ácido medido (cm^3)							
Estudiante A	8	12	7	9	3	10	12	11	12	14
Estudiante B	7	6	7	15	12	11	9	9	13	11

Solución.

Primero es necesario obtener la media de cada conjunto de datos. En este caso

Estudiante A

$$\bar{x} = \frac{8+12+7+9+3+10+12+11+12+14}{10} = 9.8$$

Estudiante B

$$\bar{x} = \frac{7+6+7+15+12+11+9+9+13+11}{10} = 10$$

Ahora aplicamos la fórmula correspondiente

Solución (Continuación).

Estudiante A

$$s^2 = \frac{(8-9.8)^2 + (12-9.8)^2 + (7-9.8)^2 + (9-9.8)^2 + (3-9.8)^2 + (10-9.8)^2 + (12-9.8)^2 + (11-9.8)^2 + (12-9.8)^2 + (14-9.8)^2}{10-1}$$

$$s^2 = \frac{91.6}{10} = 9.16$$

Estudiante B

$$s^2 = \frac{(7-10)^2 + (6-10)^2 + (7-10)^2 + (15-10)^2 + (12-10)^2 + (11-10)^2 + (9-10)^2 + (9-10)^2 + (13-10)^2 + (11-10)^2}{10-1}$$

$$s^2 = \frac{76}{10} = 7.6$$

DESVIACIÓN ESTÁNDAR (Datos no agrupados)

También llamada desviación típica, es una medida de dispersión usada en estadística que nos dice cuánto tienden a alejarse los valores puntuales del promedio en una distribución.

Específicamente, la desviación estándar es "el promedio de la distancia de cada punto respecto del promedio". Se suele representar por una S o con la letra sigma,σ, según se calcule en una muestra o en la población.

Una desviación estándar grande indica que los puntos están lejos de la media, y una desviación pequeña indica que los datos están agrupados cerca de la media.

Muestral

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

Poblacional

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

Ejemplo 1.

Si retomamos el ejemplo 1 que corresponde a la varianza:

Calcula la desviación estándar para los siguientes datos

2 1 2 4 1 3 2 3 2 0 5 1

Solución.

Una vez que hemos calculado la media y la varianza, sólo resta calcular la raíz cuadrada de la varianza.

$$\overline{x} = 2.16$$

$$s^2 = \frac{21.6672}{11} = 1.9697$$

$$S = \sqrt{1.9697} = 1.4034$$

Ejemplo 2.

Considerando nuevamente el segundo ejemplo que estudiaste para calcular la varianza, tenemos:

A continuación se muestran dos conjuntos de datos obtenidos a partir de un experimento químico que realizaron dos estudiantes distintos. Calcular la varianza.

			Volumen de ácido medido (cm^3)							
Estudiante A	8	12	7	9	3	10	12	11	12	14
Estudiante B	7	6	7	15	12	11	9	9	13	11

Solución.

Una vez que has calculado la media y la varianza, es necesario calcular la desviación estándar a partir de la obtención de la raíz cuadrada de la varianza.

Estudiante A

$$s^2 = \frac{91.6}{10} = 9.16$$

$$S = \sqrt{9.16} = 3.026$$

Estudiante B

$$s^2 = \frac{76}{10} = 7.6$$

$$S = \sqrt{7.6} = 2.756$$

COEFICIENTE DE VARIACIÓN

Es una medida de dispersión que se utiliza para poder comparar las desviaciones estándar de poblaciones con diferentes medias y se calcula como cociente entre la desviación típica y la media.

Muestral

$$CV = \frac{S}{\overline{x}} \cdot 100\%$$

Poblacional

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

Ejemplo 1.

En dos materias A y B los promedios que sacaron sus alumnos fueron 6.1 y 4.3 y las desviaciones estándar respectivas fueron 0.6 y 0.45 respectivamente. ¿En qué curso hay mayor dispersión?

Solución

Para responder esto, debemos obtener el coeficiente de variación aplicando la fórmula

$$\boxed{CV = \frac{S}{\overline{x}} \cdot 100\%}$$

$$\boxed{CV_A = \frac{0.6}{6.1}(100\%) = 9.8\%}$$

$$\boxed{CV_B = \frac{0.45}{4.3}(100\%) = 10.4\%}$$

Claramente, la materia A tiene una dispersión menor que la B, pese a presentar una mayor desviación estándar.

VARIANZA Y DESVIACIÓN ESTÁNDAR (Datos agrupados)

Cuando los datos están agrupados en tablas de frecuencias, el significado de las medidas de dispersión es el mismo, sin embargo la manera de calcularlas es diferente.

Enseguida se muestra la fórmula para la varianza, pero recuerda que la desviación estándar es igual a la raíz cuadrada de la primera.

Muestral

$$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{k} f_i x_i^2 - \frac{\left(\sum_{i=1}^{k} x_i f_i\right)^2}{n}}{n-1}$$

Poblacional

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

Ejemplo 1.

Se han registrado durante 20 días, el número de pasajes que hacen reservaciones a una agencia de viajes pero que no las hacen efectivas:

i	Número de viajeros (x_i)	Frecuencia (f_i)
1	12	3
2	13	3
3	14	6
4	15	3
5	16	5
Total	**70**	**20**

Calcula las medidas de dispersión de la variable en estudio.

Solución.

Tal como lo indica la fórmula, primero es necesario multiplicar la variable (x_i) por la frecuencia (f_i) y añadirlo como una columna a la tabla.

$$s^2 = \frac{\sum_{i=1}^{k} \cdots \frac{\left(\sum_{i=1}^{k} x_i f_i\right)^2}{\cdots}}{\cdots}$$

i	Número de viajeros (x_i)	Frecuencia (f_i)	$x_i f_i$
1	12	3	36
2	13	3	39
3	14	6	84
4	15	3	45
5	16	5	80
Total	**70**	**20**	**284**

Solución (Continuación).

Después se obtiene el cuadrado de la variable x, o sea, $(x_i)^2$.

$$s^2 = \frac{\sum_{i=1}^{k} \ldots x_i^2 \ldots}{\ldots}$$

i	Número de viajeros (x_i)	Frecuencia (f_i)	$x_i f_i$	x_i^2
1	12	3	36	144
2	13	3	39	169
3	14	6	84	196
4	15	3	45	225
5	16	5	80	256
Total	**70**	**20**	**284**	**990**

Solución (Continuación).

Ahora se multiplica el cuadrado de la variable por la frecuencia, es decir, $(f_i x_i^2)$.

$$s^2 = \frac{\sum_{i=1}^{k} f_i x_i^2 - \frac{\cdots}{\cdots}}{\cdots}$$

i	Número de viajeros (x_i)	Frecuencia (f_i)	$x_i f_i$	x_i^2	$f_i x_i^2$
1	12	3	36	144	432
2	13	3	39	169	507
3	14	6	84	196	1176
4	15	3	45	225	675
5	16	5	80	256	1280
Total	**70**	**20**	**284**	**990**	**4070**

Solución (Continuación).

Una vez obtenidos todos los datos anteriores, se procede a aplicar la fórmula

$$s^2 = \frac{\sum_{i=1}^{k} f_i x_i^2 - \frac{\left(\sum_{i=1}^{k} x_i f_i\right)^2}{n}}{n-1}$$

i	Número de viajeros (x_i)	Frecuencia (f_i)	$x_i f_i$	x_i^2	$f_i x_i^2$
1	12	3	36	144	432
2	13	3	39	169	507
3	14	6	84	196	1176
4	15	3	45	225	675
5	16	5	80	256	1280
Total	**70**	**20**	**284**	**990**	**4070**

Solución (Continuación).

i	Número de viajeros (x_i)	Frecuencia (f_i)	$x_i f_i$	x_i^2	$f_i x_i^2$
1	12	3	36	144	432
2	13	3	39	169	507
3	14	6	84	196	1176
4	15	3	45	225	675
5	16	5	80	256	1280
Total	**70**	**20**	**284**	**990**	**4070**

$$s^2 = \frac{4070 - \frac{284^2}{20}}{19} = 1.9579$$

$$s = \sqrt{1.9579} = 1.3992$$

Ejemplo 2.

De acuerdo a la siguiente tabla, calcula la varianza y la desviación estándar:

NOTA x	FREC. ABSOLUTA f	FREC. ABSOLUTA ACUMULADA	FREC. RELATIVA %	FREC RELATIVA ACUMULADA %
1.2	1	1	0.1	0.1
1.4	2	3	0.2	0.3
1.6	3	6	0.3	0.6
1.8	8	14	0.8	1.4
2.0	14	28	1.4	2.8
2.2	18	46	1.8	4.6
2.4	19	65	1.9	6.5
2.6	22	87	2.2	8.7
2.8	25	112	2.5	11.2
3.0	26	138	2.6	13.8
3.2	27	165	2.7	16.5
3.4	31	196	3.1	19.6
3.6	35	231	3.5	23.1
3.8	38	269	3.8	26.9
4.0	45	314	4.5	31.4
4.2	46	360	4.6	36.0
4.4	48	408	4.8	40.8
4.6	52	460	5.2	46.0
4.8	58	518	5.8	51.8
5.0	60	578	6.0	57.8
5.2	56	634	5.6	63.4
5.4	54	688	5.4	68.8
5.6	51	739	5.1	73.9
5.8	50	789	5.0	78.9
6.0	46	835	4.6	83.5
6.2	44	879	4.4	87.9
6.4	40	919	4.0	91.9
6.6	32	951	3.2	95.1
6.8	31	982	3.1	98.2
7.0	18	1000	1.8	100
TOTAL	**1000**	**4717**	**23970.12**	

Solución.

El primer paso es calcular $x_i\ f_i$:

NOTA x	FREC. ABSOLUTA f	FREC. ABSOLUTA ACUMULADA	FREC. RELATIVA %	FREC RELATIVA ACUMULADA %	$x_i\ f_i$
1.2	1	1	0.1	0.1	1.2
1.4	2	3	0.2	0.3	2.8
1.6	3	6	0.3	0.6	4.8
1.8	8	14	0.8	1.4	14.4
2.0	14	28	1.4	2.8	28
2.2	18	46	1.8	4.6	39.6
2.4	19	65	1.9	6.5	45.6
2.6	22	87	2.2	8.7	57.2
2.8	25	112	2.5	11.2	70
3.0	26	138	2.6	13.8	78
3.2	27	165	2.7	16.5	86.4
3.4	31	196	3.1	19.6	105.4
3.6	35	231	3.5	23.1	126
3.8	38	269	3.8	26.9	144.4
4.0	45	314	4.5	31.4	180
4.2	46	360	4.6	36.0	193.2
4.4	48	408	4.8	40.8	211.2
4.6	52	460	5.2	46.0	239.2
4.8	58	518	5.8	51.8	278.4
5.0	60	578	6.0	57.8	300
5.2	56	634	5.6	63.4	291.2
5.4	54	688	5.4	68.8	291.6
5.6	51	739	5.1	73.9	285.6
5.8	50	789	5.0	78.9	290
6.0	46	835	4.6	83.5	276
6.2	44	879	4.4	87.9	272.8
6.4	40	919	4.0	91.9	256
6.6	32	951	3.2	95.1	211.2
6.8	31	982	3.1	98.2	210.8
7.0	18	1000	1.8	100	126
TOTAL	**1000**	**4717**	**23970.12**		

Solución (Continuación).

Después se obtiene el cuadrado de la variable x, o sea, $(x_i)^2$.

NOTA x	FREC. ABSOLUTA f	FREC. ABSOLUTA ACUMULADA	FREC. RELATIVA %	FREC RELATIVA ACUMULADA %	$x_i f_i$	x_i^2
1.2	1	1	0.1	0.1	1.2	1.44
1.4	2	3	0.2	0.3	2.8	1.96
1.6	3	6	0.3	0.6	4.8	2.56
1.8	8	14	0.8	1.4	14.4	3.24
2.0	14	28	1.4	2.8	28	4
2.2	18	46	1.8	4.6	39.6	4.84
2.4	19	65	1.9	6.5	45.6	5.76
2.6	22	87	2.2	8.7	57.2	6.76
2.8	25	112	2.5	11.2	70	7.84
3.0	26	138	2.6	13.8	78	9
3.2	27	165	2.7	16.5	86.4	10.24
3.4	31	196	3.1	19.6	105.4	11.56
3.6	35	231	3.5	23.1	126	12.96
3.8	38	269	3.8	26.9	144.4	14.44
4.0	45	314	4.5	31.4	180	16
4.2	46	360	4.6	36.0	193.2	17.64
4.4	48	408	4.8	40.8	211.2	19.36
4.6	52	460	5.2	46.0	239.2	21.16
4.8	58	518	5.8	51.8	278.4	23.04
5.0	60	578	6.0	57.8	300	25
5.2	56	634	5.6	63.4	291.2	27.04
5.4	54	688	5.4	68.8	291.6	29.16
5.6	51	739	5.1	73.9	285.6	31.36
5.8	50	789	5.0	78.9	290	33.64
6.0	46	835	4.6	83.5	276	36
6.2	44	879	4.4	87.9	272.8	38.44
6.4	40	919	4.0	91.9	256	40.96
6.6	32	951	3.2	95.1	211.2	43.56
6.8	31	982	3.1	98.2	210.8	46.24
7.0	18	1000	1.8	100	126	49
TOTAL	**1000**	**4717**	**23970.12**			

Solución (Continuación).

Ahora se multiplica el cuadrado de la variable por la frecuencia, es decir, ($f_i x_i^2$).

NOTA x	FREC. ABSOLUTA f	FREC. ABSOLUTA ACUMULADA	FREC. RELATIVA %	FREC RELATIVA ACUMULADA %	$x_i f_i$	x_i^2	$f_i x_i^2$
1.2	1	1	0.1	0.1	1.2	1.44	1.44
1.4	2	3	0.2	0.3	2.8	1.96	3.92
1.6	3	6	0.3	0.6	4.8	2.56	7.68
1.8	8	14	0.8	1.4	14.4	3.24	25.92
2.0	14	28	1.4	2.8	28	4	56
2.2	18	46	1.8	4.6	39.6	4.84	87.12
2.4	19	65	1.9	6.5	45.6	5.76	109.44
2.6	22	87	2.2	8.7	57.2	6.76	148.72
2.8	25	112	2.5	11.2	70	7.84	196
3.0	26	138	2.6	13.8	78	9	234
3.2	27	165	2.7	16.5	86.4	10.24	276.48
3.4	31	196	3.1	19.6	105.4	11.56	358.36
3.6	35	231	3.5	23.1	126	12.96	453.6
3.8	38	269	3.8	26.9	144.4	14.44	548.72
4.0	45	314	4.5	31.4	180	16	720
4.2	46	360	4.6	36.0	193.2	17.64	811.44
4.4	48	408	4.8	40.8	211.2	19.36	929.28
4.6	52	460	5.2	46.0	239.2	21.16	1100.32
4.8	58	518	5.8	51.8	278.4	23.04	1336.32
5.0	60	578	6.0	57.8	300	25	1500
5.2	56	634	5.6	63.4	291.2	27.04	1514.24
5.4	54	688	5.4	68.8	291.6	29.16	1574.64
5.6	51	739	5.1	73.9	285.6	31.36	1599.36
5.8	50	789	5.0	78.9	290	33.64	1682
6.0	46	835	4.6	83.5	276	36	1656
6.2	44	879	4.4	87.9	272.8	38.44	1691.36
6.4	40	919	4.0	91.9	256	40.96	1638.4
6.6	32	951	3.2	95.1	211.2	43.56	1393.92
6.8	31	982	3.1	98.2	210.8	46.24	1433.44
7.0	18	1000	1.8	100	126	49	882
TOTAL	1000	4717	23970.12		4717		23970.12

Solución (Continuación).

Una vez obtenidos todos los datos anteriores, se procede a aplicar la fórmula

$$s^2 = \frac{\sum_{i=1}^{k} f_i x_i^2 - \frac{\left(\sum_{i=1}^{k} x_i f_i\right)^2}{n}}{n-1}$$

$$s^2 = \frac{23970.12 - \frac{4717^2}{1000}}{1000-1} = 1.7217$$ Varianza

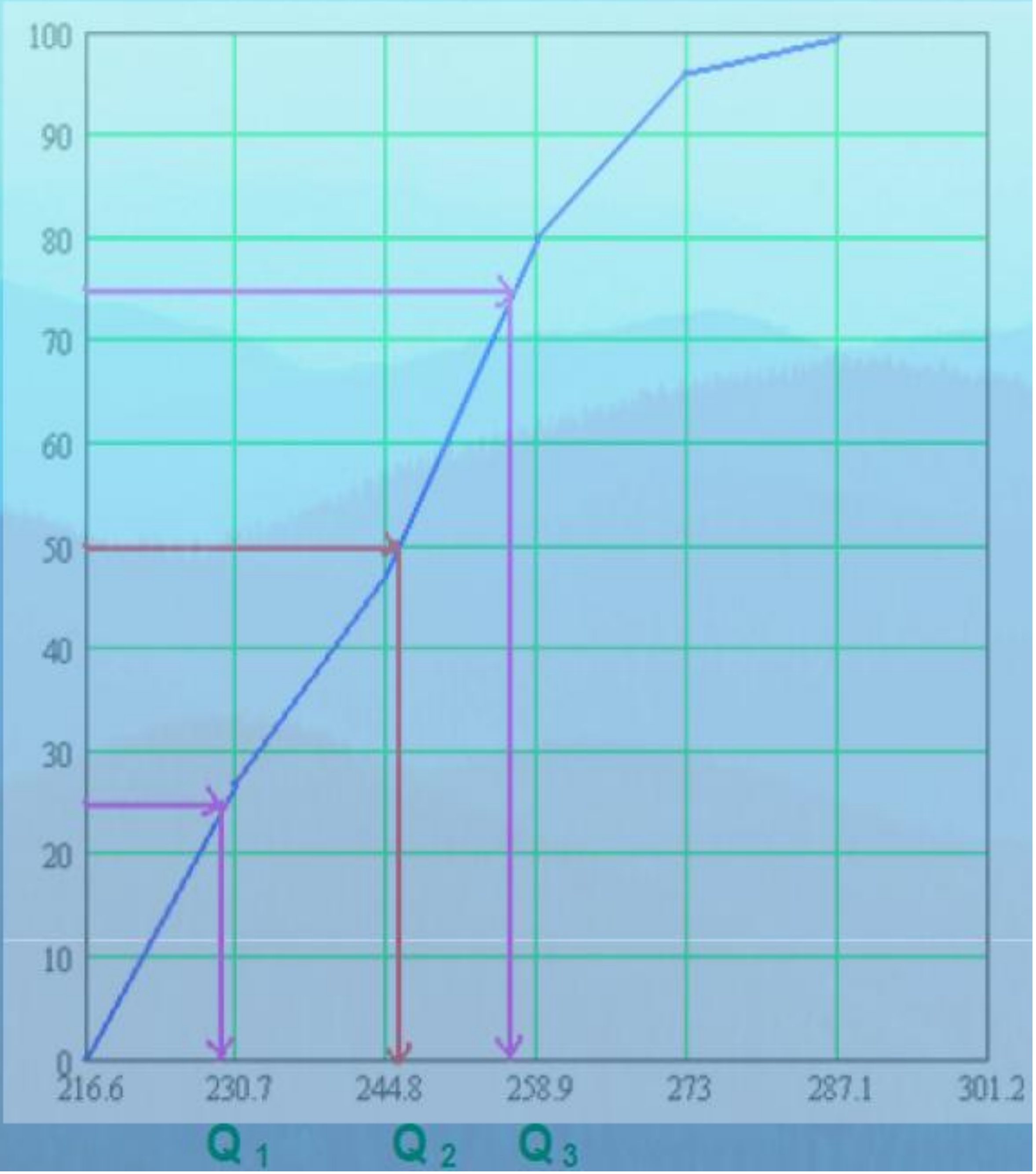
$$s = \sqrt{1.7217} = 1.3121$$ Desviación estándar

Cuartiles y Deciles

Estas medidas de dispersión se parecen mucho a la mediana en cuanto a que dividen a la distribución en partes iguales y se encuentra el valor que corresponde, los cuartiles la dividen en cuatro y los deciles en diez.

Cuartiles Al dividir a la distribución en cuatro partes iguales, los cuartiles contendrán entre uno y otro al 25% del total de datos. Al primer cuartil se le denota ***Q1*** y separa al primer 25% del total de datos; el segundo cuartil, ***Q2***, separa al primer 50% de los datos, (por lo que coincide con la mediana; el tercer cuartil, ***Q3***, separa al 75% de los datos.

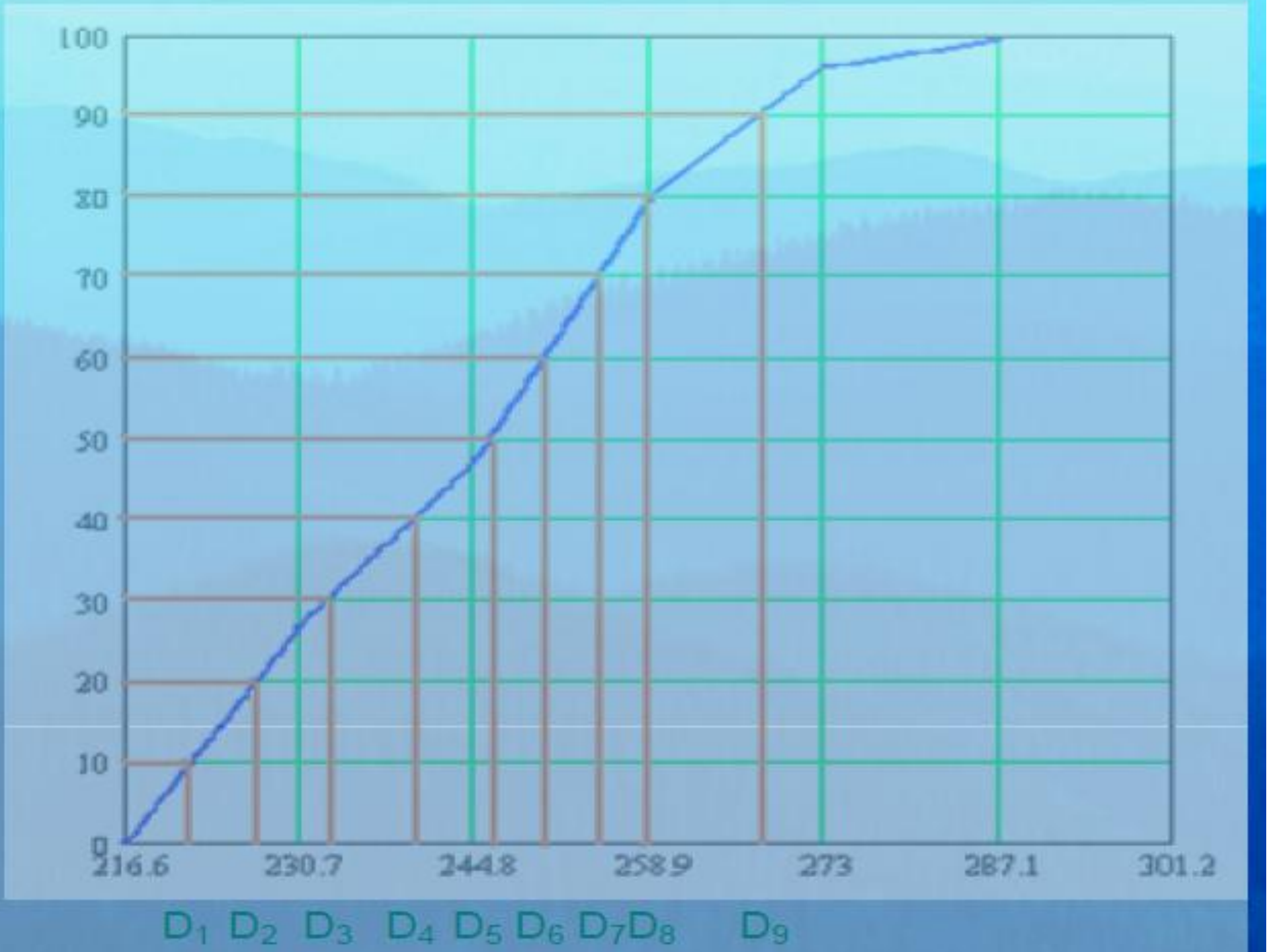
- el cuartil 1 (Q_1) divide a la población en 25% - 75%.
- el cuartil 2 (Q_2) divide a la población en 50% - 50%. Por lo que es igual a la Mediana
- el cuartil 3 (Q_3) divide a la población en 75% - 25%.



Los deciles

Son muy similares a los cuartiles pero dividen a la distribución en diez partes iguales:

- el decil 1 (D_1)
- el decil 2 (D_2)
- el decil 3 (D_3)
- el decil 4 (D_4)
- el decil 5 (D_5)
- el decil 6 (D_6)
- el decil 7 (D_7)
- el decil 8 (D_8)
- el decil 9 (D_9)



FIN

Defensa la próxima clase