



**Universidad Mariano Gálvez**  
**Estadística y probabilidad para Ingeniería**  
**Sección B.**


# Conceptos básicos. Unidad 1.

**Presenta**

**Dra. En Ing. Rita Victoria de León Ardón**



# Objetivo de la presentación

- ▶ Introducir al estudiante a los antecedentes teóricos para el manejo de datos.
- 



# Recomendaciones básicas para esta unidad

- ▶ Realice práctica adicional a los ejercicios y tareas brindados en clase.
- ▶ Comprenda los conceptos de manera precisa para que pueda aplicarlos adecuadamente.
- ▶ Como requisito de esta unidad es necesario que el estudiante realice un formulario. El cual será revisado y calificado por el catedrático.
- ▶ Se harán exámenes cortos sorpresa.

# Función general de medición

- ▶ La función general para el procedimiento de medición viene dada por  $f(x,y)$  en donde  $x$  es cualquier objeto, “ $y$ ”  $y$ =un valor numérico.

# Tipos de medición

- ▶ Dentro de los distintos niveles de medición se pueden encontrar los siguientes (Kerlinger, 1994):
- ▶ Medición **nominal-escala nominal**: que es el nivel más bajo de medición, los números asignados a los objetos son valores numéricos sin un significado numeral; no pueden ser ordenados o añadidos. **[1 99 0 3 34]**
- ▶ Medición **ordinal-escala ordinal**: requiere que los objetos de un conjunto puedan tener un rango y ser ordenados con base en alguna característica o propiedad definida de manera operacional. **[1 2 3 4] 1 el más aceptable 4 el menos aceptable**
- ▶ Medición **de intervalo-escala de intervalo**: cumple características ordinales y nominales, las distancias numéricamente iguales en la escalas de intervalo representan distancias iguales en la propiedad que se está midiendo. **[0-1]**
- ▶ **Medición de razón**: existe el cero absoluto o natural que tiene significado empírico, hay una base que permite determinar si el objeto no tiene ninguna de las propiedades que se están midiendo. **0 1 2 3 4 5**

# Clasificaciones de la estadística

univariable y multivariable

Descriptiva

Paramétrica (se conoce el comportamiento de la población) y no paramétrica (se desconoce el comportamiento de la población)

# Estadística descriptiva

- Distribución de Frecuencias (¿Qué tanto se repite un dato?)
- Gráficas

# Características de la medición

Misma  
medición en  
varias  
repeticiones

Confiabilidad

Validez

Resultados  
corresponden  
a un campo de  
conocimiento.  
Ejemplo:  
Ingeniería,  
medicina,  
entre otros.



Ejemplo: precio de una libra de café de una marca **CaféSol** en distintas tiendas.

Q32.30	Q33.30	Q1.34	Q32.60	Q34.80
Q38.00	Q38.90	Q32.10	Q34.60	Q36.80
Q33.00	Q34.60	Q34.60	Q23.50	Q33.80
Q33.00	Q23.40	Q134.90	Q1.45	Q32.90
Q33.00	Q25.40	Q34.80	Q34.60	Q31.90
Q39.00	Q31.40	Q35.90	Q36.80	Q32.90
Q35.00	Q32.60	Q32.90	Q190.90	Q34.60
Q38.00	Q33.80	Q26.90	Q193.45	Q35.80
Q34.00	Q45.30	Q23.70	Q32.50	Q36.90

**Observar los datos**

**Responda:**

**1. ¿Todos los datos son iguales?**

**2. ¿Hay confiabilidad?**

**3. ¿Qué haría si le dan estos datos?**

# Paso 1

➔ **Eliminar los datos que no son confiables.**

Q32.30	Q33.30	Q1.34	Q32.60	Q34.80
Q38.00	Q38.90	Q32.10	Q34.60	Q36.80
Q33.00	Q34.60	Q34.60	Q23.50	Q33.80
Q33.00	Q23.40	Q134.90	Q1.45	Q32.90
Q33.00	Q25.40	Q34.80	Q34.60	Q31.90
Q39.00	Q31.40	Q35.90	Q36.80	Q32.90
Q35.00	Q32.60	Q32.90	Q190.90	Q34.60
Q38.00	Q33.80	Q26.90	Q193.45	Q35.80
Q34.00	Q45.30	Q23.70	Q32.50	Q36.90

## Paso 2

➔ Ordenar los datos de menor a mayor

Q23.40	Q34.00
Q23.50	Q34.60
Q23.70	Q34.60
Q25.40	Q34.60
Q26.90	Q34.60
Q31.40	Q34.60
Q31.90	Q34.80
Q32.10	Q34.80
Q32.30	Q35.00
Q32.50	Q35.80
Q32.60	Q35.90
Q32.60	Q36.00
Q32.90	Q36.80
Q32.90	Q36.80
Q32.90	Q36.90
Q33.30	Q38.00
Q33.80	Q38.90
Q33.80	Q39.00
Q34.00	Q39.00
Q34.00	Q45.30

### **Paso 3**

## **Obtener rango**

**Rango= Dato mayor-Dato menor**

**Rango= 45.30-23.40**

**Rango= 21.90**

## Paso 4

Obtener número de intervalos

Varias formas

No. de Intervalos=  $\sqrt{n}$

Donde n es igual al número de datos trabajados en este caso 40.

No. de intervalos=  $\sqrt{40}=6.32$

Conviene establecer 6 intervalos

## Paso 5

➔ **Establecer ancho de intervalo**

**Ancho de intervalo= Rango/# de intervalos**

**Ancho de intervalo= 21.9/6**

**Ancho de intervalo=3.65**

# Paso 6 construir los intervalos

[23.4-27.05)

[27.05-30.7)

[30.7-34.35)

[34.35-38)

[38.00-41.65)

[41.65-45.3]

# Paso 7 Obtener la distribución de frecuencias

[23.4-27.05)	5
[27.05-30.7)	0
[30.7-34.35)	16
[34.35-38)	14
[38.00-41.65)	4
[41.65-45.3]	1

← Frecuencia

Q23.40	Q34.00
Q23.50	Q34.60
Q23.70	Q34.60
Q25.40	Q34.60
Q26.90	Q34.60
Q31.40	Q34.60
Q31.90	Q34.80
Q32.10	Q34.80
Q32.30	Q35.00
Q32.50	Q35.80
Q32.60	Q35.90
Q32.60	Q36.00
Q32.90	Q36.80
Q32.90	Q36.80
Q32.90	Q36.90
Q33.30	Q38.00
Q33.80	Q38.90
Q33.80	Q39.00
Q34.00	Q39.00
Q34.00	Q45.30



## PASO 8 Graficar

- Para graficar se necesitan tres medidas adicionales

Punto medio de los intervalos  
 $= (LS + LI) / 2$

**Marca de clase ( $x_i$ )**

$$F_i = \sum_{j=1}^i f_j$$

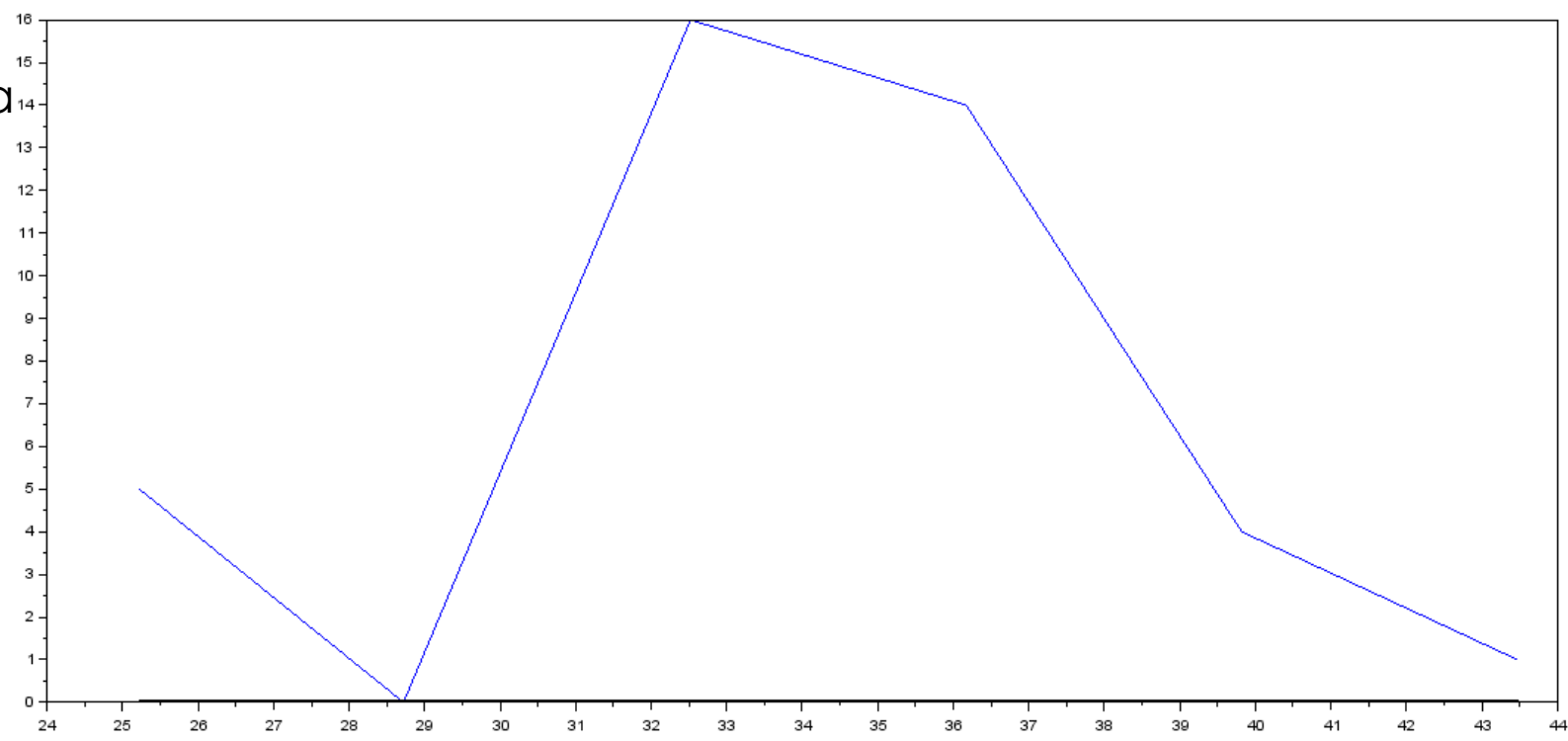
**Frecuencia acumulada**

$$f_i^* = \frac{f_i}{n} = \frac{f_i}{\sum_{j=1}^m f_j}$$

**Frecuencia relativa**

Límites	Frecuencia (fi)	Marca de clase	Frecuencia acumulada	Frecuencia relativa
[23.4-27.05)	5	$(23.4+27.05)/2$ 25.22	5	$5/40=0.125$
[27.05-30.7)	0	28.72	5	$0/40=0$
[30.7-34.35)	16	32.52	21	$21/40=0.525$
[34.35-38)	14	36.17	35	$35/40=0.875$
[38.00-41.65)	4	39.82	39	$39/40=0.975$
[41.65-45.3]	1	43.47	40	$40/40 = 1$
Frecuencia acumulada	40			

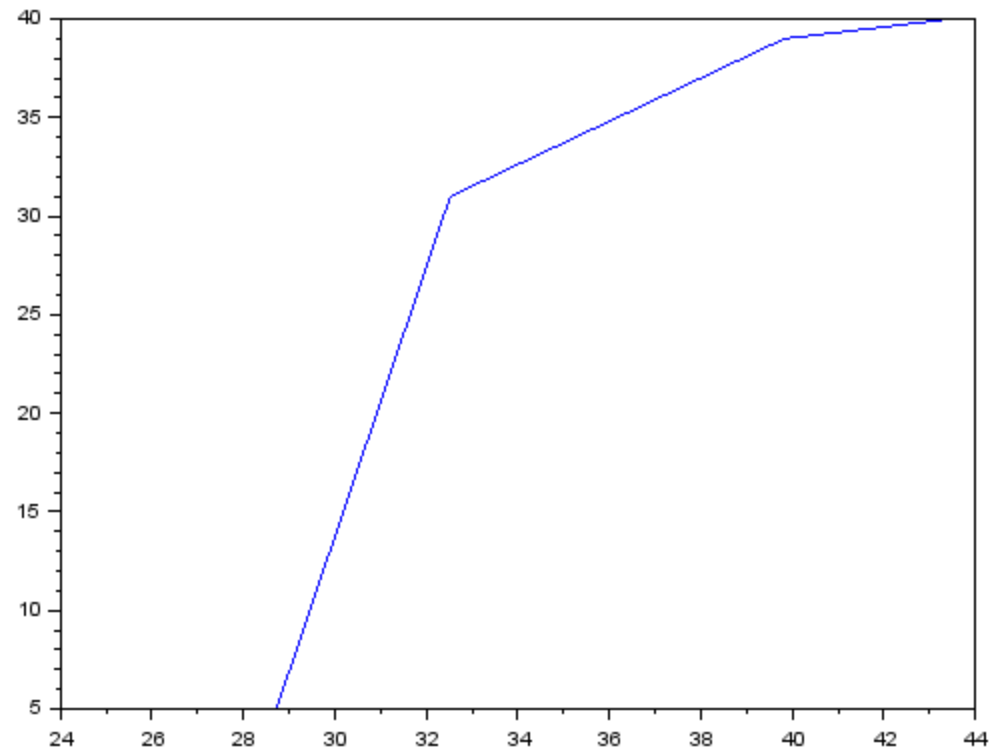
frecuencia



Marca de clase (xi)

Polígono de frecuencias

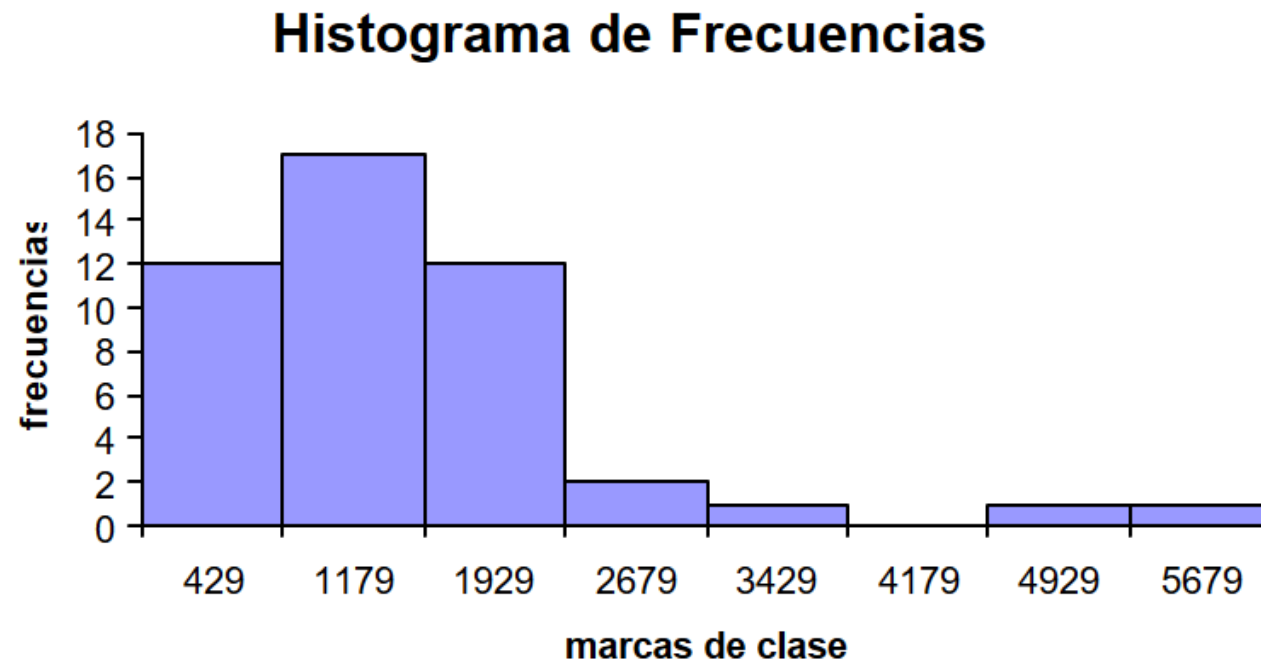
Frecuencia  
acumulada



Marca de clase ( $x_i$ )



# Histograma de frecuencias



Actividad complementaria 1.

1) Realice el histograma de frecuencias del ejercicio

# Otro concepto útil: fronteras de clase (límites reales)

- ▶ Se construyen a partir de los límites considerando que:
- ▶ Si los límites son números enteros, entonces, restar 0.5 al límite inferior y sumar 0.5 al límite superior.
- ▶ Si los límites no son números enteros, se debe restar y sumar a los intervalos de clase:
  - 0.05 si tienen un solo decimal,
  - 0.005 si tienen dos decimales
  - 0.0005 si tienen tres decimales, entre otros.

<b>[23.40-27.05)</b>
<b>[27.05-30.7)</b>
<b>[30.7-34.35)</b>
<b>[34.35-38)</b>
<b>[38.00-41.65)</b>
<b>[41.65-45.3]</b>

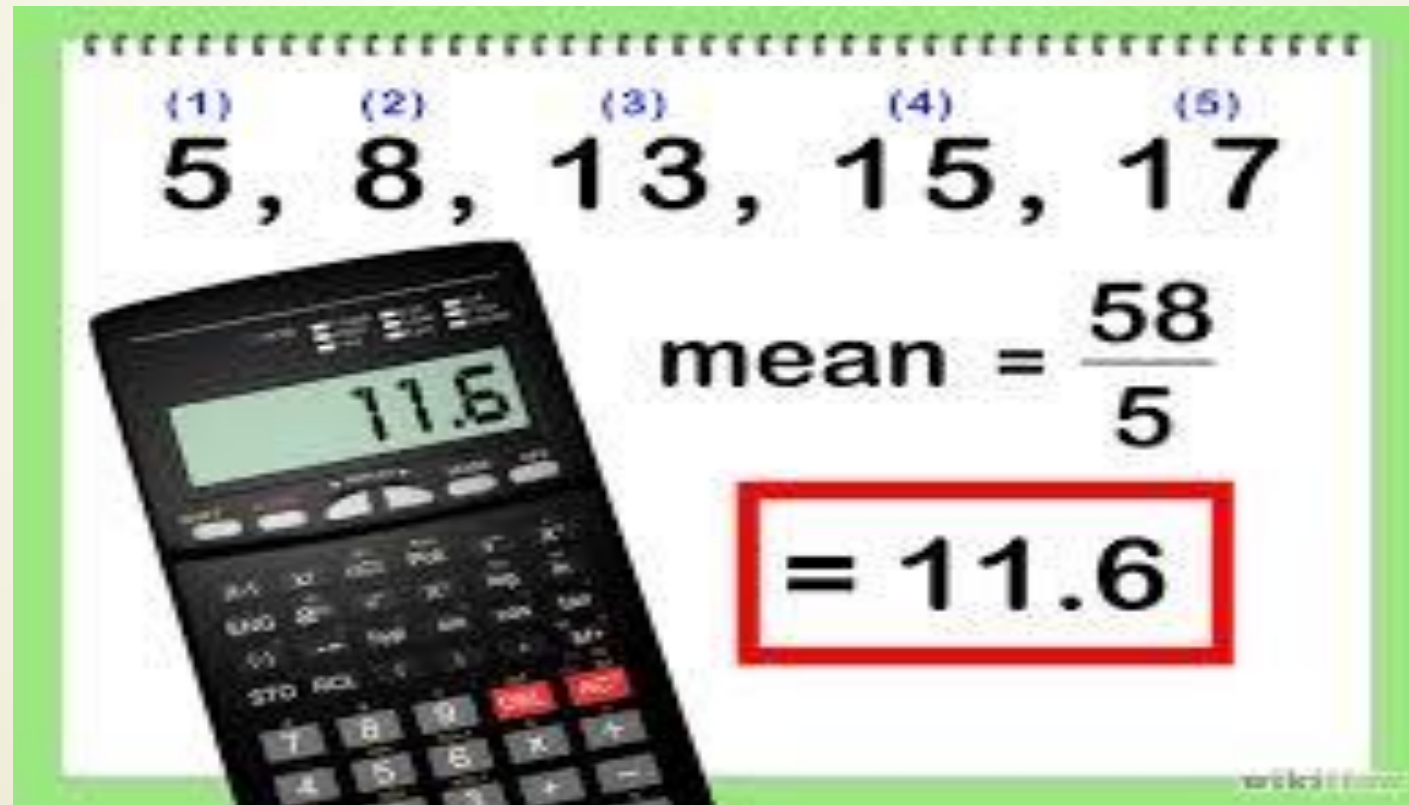
Actividad complementaria 2.

1) Calcular las fronteras de clase

## Actividad 3. Realice el análisis por intervalos para los siguientes datos.

31.5	34.4	36	35.5	36.1
18.8	37.1	34.2	33.2	31
33.9	24.6	23.3	35.3	29.6
35	32.8	30.8	31.8	34.7
35.2	36.8	33.7	33.4	34
27.6	33.6	30.8	25.4	16.4

# Media, mediana y moda



The image shows a black calculator with a green display showing the number 11.6. To the right of the calculator is a white worksheet with a green border. The worksheet contains a data set: (1) 5, (2) 8, (3) 13, (4) 15, (5) 17. Below the data set, the mean is calculated as  $\text{mean} = \frac{58}{5}$ . The result, 11.6, is enclosed in a red rectangular box.

(1)	(2)	(3)	(4)	(5)
5	8	13	15	17

mean =  $\frac{58}{5}$

**= 11.6**





The diagram features a central orange circle containing the text "Medidas de tendencia central". Three smaller orange circles are arranged around it: one at the top, one at the bottom left, and one at the bottom right. Each of these three circles overlaps with the central circle. To the right of the top circle is the text "Promedio de un conjunto de valores". To the left of the bottom left circle is the text "Dato que más se repite". To the right of the bottom right circle is the text "Divide al conjunto de datos en exactamente dos subconjuntos". On the far left, there are several thin, dark green lines that resemble blades of grass or reeds, extending from the bottom towards the top of the page.

# Medidas de tendencia central

Media

**Promedio de un conjunto de valores**

Moda

**Dato que más se repite**

Mediana

**Divide al conjunto de datos en exactamente dos subconjuntos**

# Clasificación a tomar en cuenta



**Datos  
agrupados**



**No. de datos mayor o  
igual a 25**



**Datos no  
agrupados**



**No. de datos menor a  
25**

# Media

Si  $n \leq 25$  usar la corrección  $n-1$

$$\bar{x} = \begin{cases} \frac{1}{n} \sum_{i=1}^n x_i & ; \text{ Datos no agrupados} \\ \frac{1}{n} \sum_{i=1}^m x_i f_i & ; \text{ Datos agrupados} \end{cases}$$

Debe observarse que para datos agrupados la suma va desde  $1$  hasta  $m$ , donde  $m$  es el número de intervalos, y  $x_i$  y  $f_i$  son la marca de clase y la frecuencia del intervalo, respectivamente.

# Moda

**Datos no agrupados: dato que más se repite.**

$$\text{MODA} \quad Mo = L_{Mo\text{inf}} + \left[ \frac{a}{a+b} \right] c_{Mo}$$

$$\text{donde:} \quad a = f_{Mo} - f_{Mo-1}, \quad b = f_{Mo} - f_{Mo+1}$$

$f_{Mo}$  = frecuencia absoluta de la clase que contiene a la moda

$c_{Mo}$  = longitud de la clase que contiene a la moda

$L_{Mo\text{inf}}$  = límite inferior de la clase que contiene a la moda

# Mediana

$$\text{MEDIANA } \hat{X} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & ; \text{ si } n \text{ es impar} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} & ; \text{ si } n \text{ es par} \end{cases}$$

**Datos no agrupados**

# Mediana

$$\text{MEDIANA } \tilde{X} = L_{i \text{ inf}} + \left[ \frac{\frac{n}{2} - F_{i-1}}{f_i} \right] c_i$$

donde:  $c_i$  = longitud de la clase que contiene a la mediana  
 $L_{i \text{ inf}}$  = límite inferior de la clase que contiene a la mediana

## Datos agrupados



Datos sin  
agrupar

Ejemplo:

➔ Calcule la media,  
mediana y moda de:

60, 60, **60**, 60, 100

5 datos  
(impar)


Media =  $(60+60+60+60+100)/5 = 68$

Moda = 60

Mediana = 60

3ª. Posición  
Posición =  $(5+1)/2$

► Calcule la media, mediana y moda de:



Datos  
agrupados

Intervalo	No. de muestras	Marca clase
28-32	1	30
33-37	2	35
38-42	5	40
43-47	12	45
48-52	24	50
53-57	7	55
58-62	3	60
63-67	3	65



# Medidas de dispersión

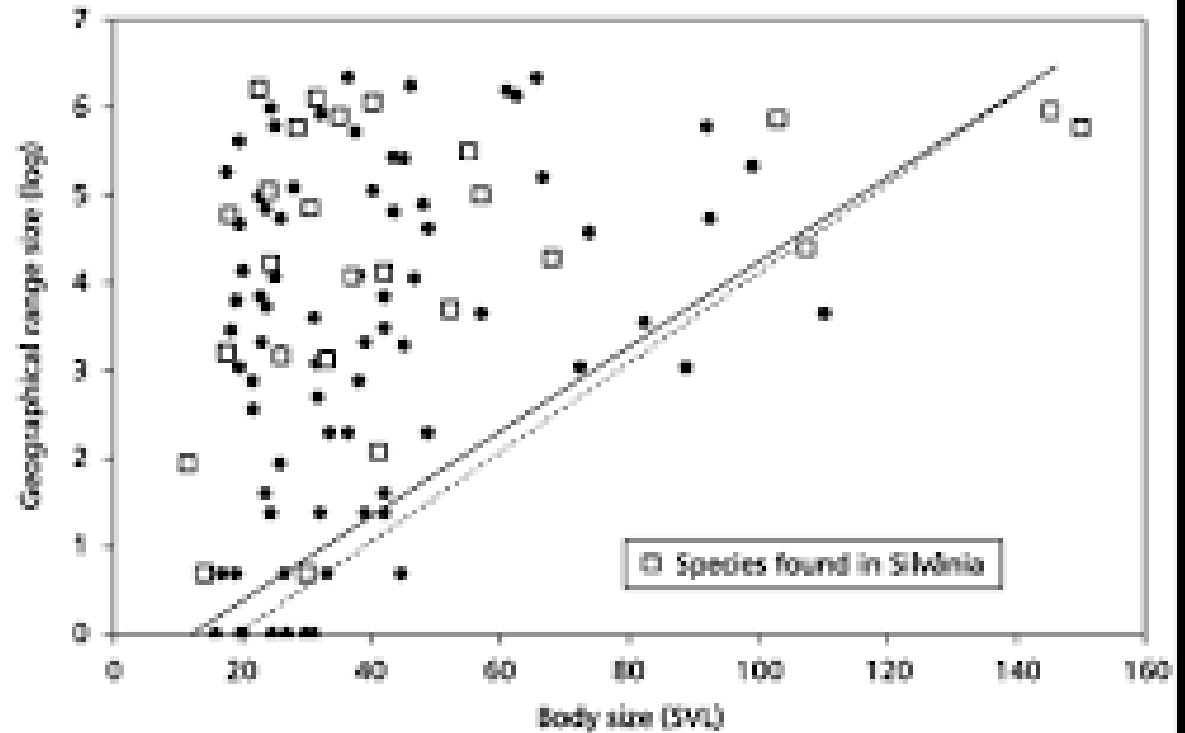
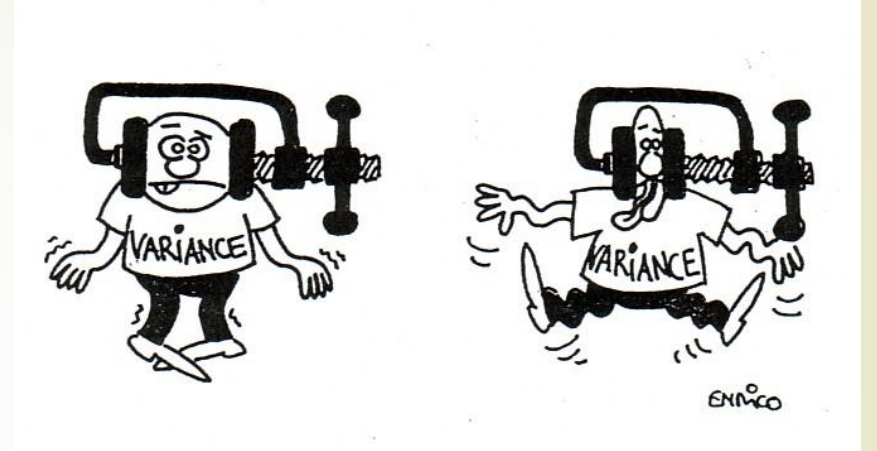


Fig. 3 — Relationship between GRS and SVL. The boundaries were fitted according to Gotelli & Enteniger (2000). Dashed line = overall data; continuous line = Silesia data.

# Medidas de dispersión



# Varianza

	población	muestra
DATOS SIN AGRUPAR	$\sigma^2 = \frac{\sum (x - \mu)^2}{n}$	$\hat{s}^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$
DATOS AGRUPADOS	$\sigma^2 = \frac{\sum f(\dot{x} - \mu)^2}{n}$	$\hat{s}^2 = \frac{\sum f(\dot{x} - \bar{x})^2}{n - 1}$

# Desviación estándar

Población

Muestra

Agrupar

$$\sigma = \sqrt{\frac{\sum f(X - \bar{X})^2}{N}}$$

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N f(x_i - \bar{x})^2}$$

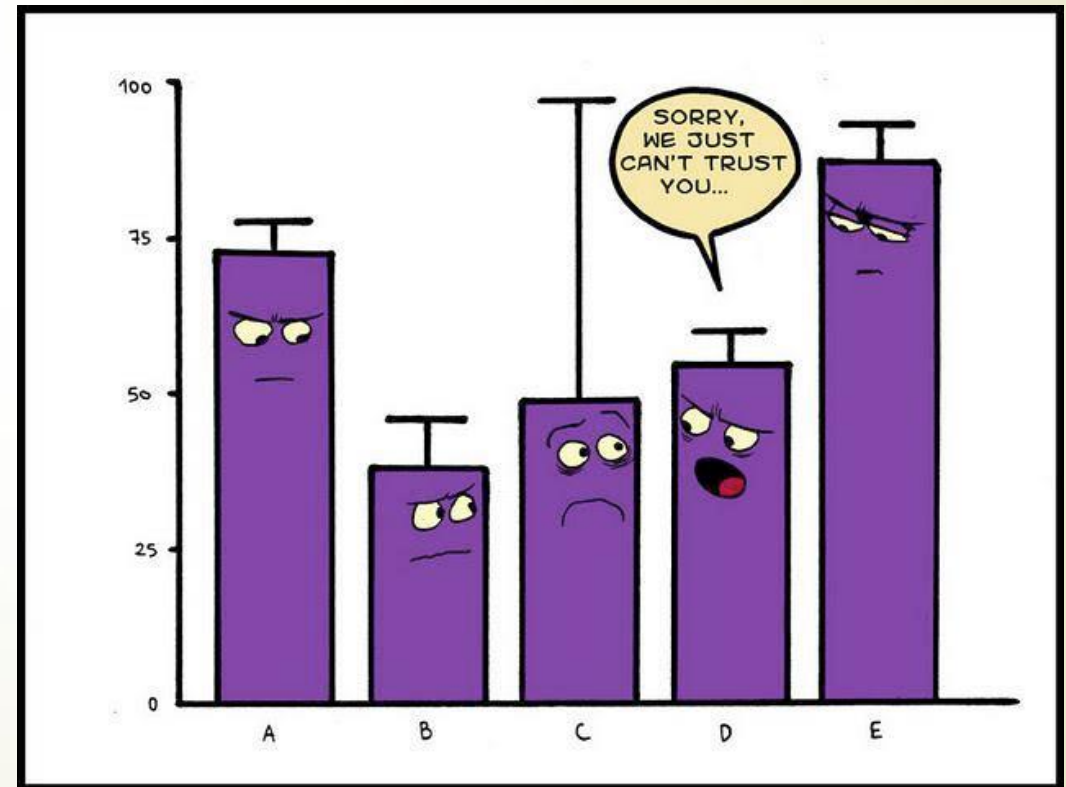
Sin  
agrupar

$$\sigma = \sqrt{\frac{\sum_i (Xi - \bar{X})^2}{n}}$$

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

# Coeficiente de variación de Pearson

$$V = \frac{s}{\bar{x}} * 100$$



# Momentos estadísticos

- Son formulaciones matemáticas que tienen amplia connotación dentro del estudio de curvas de distribución de frecuencias.

$$m'_r = \sum_{i=1}^n \frac{(x_i - A)^r}{n}$$

A es un punto cualquiera.  
A puede ser la media aritmética.  
Si A tiene un valor de 0 se le denomina momento al origen.

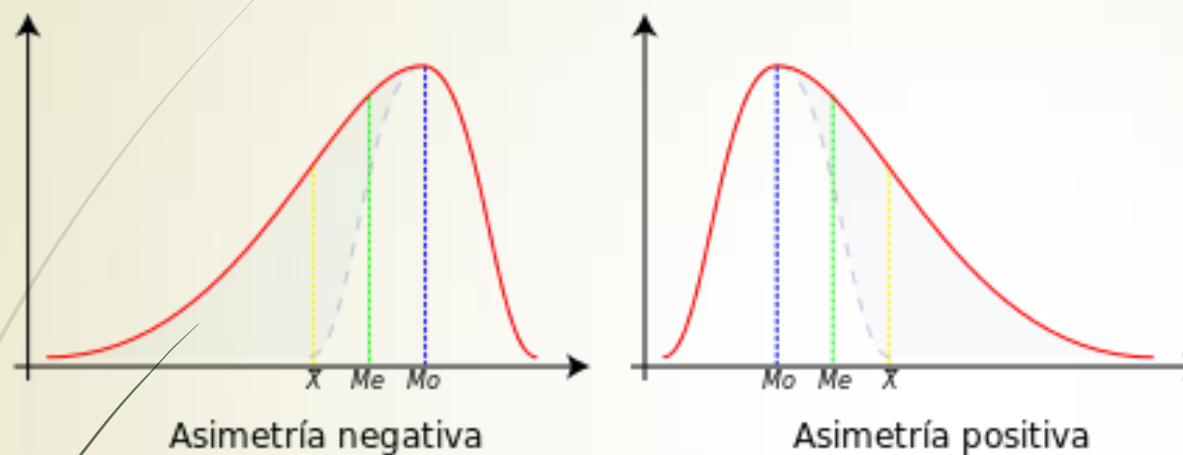
r es el número de momento



Ejemplo: calcule el segundo momento respecto al origen de: 2,0,1,4,0,3,1,1,0,2

$$m_2' = \frac{2^3 + 0^3 + 1^3 + 4^3 + 0^3 + 3^3 + 1^3 + 1^3 + 0^3 + 2^3}{10} = 11$$

# Sesgo en función de momentos



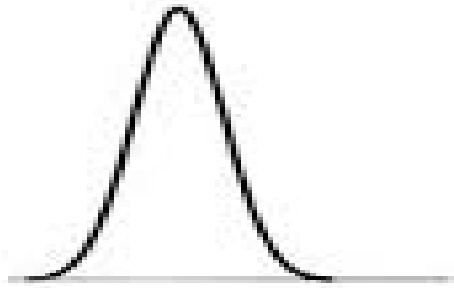
$$a_3 = \frac{m_3}{S^3}$$

$m_3$  = Momento 3 respecto a la media  
 $S$  = desviación estándar

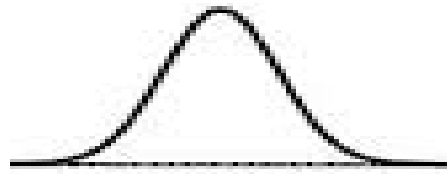
- $a_3 > 0$  distribución asimétrica positiva**
- $a_3 = 0$  distribución asimétrica positiva**
- $a_3 < 0$  distribución asimétrica negativa**



# Apuntamiento



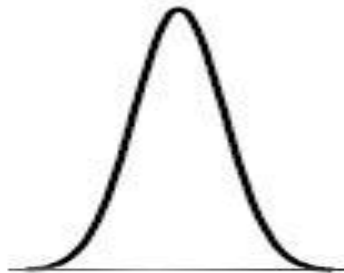
**Leptocúrtica**



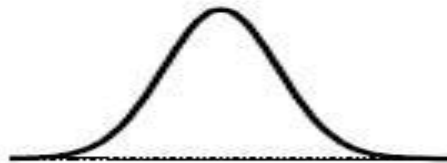
**Mesocúrtica**



**Platicúrtica**



**Datos agrupados  
Poca variabilidad**



**Datos normales**



**Datos dispersos  
Mucha variabilidad**



Para conocer que tipo de  
apuntamiento tiene la curva



**Coeficientes de curtosis**



Nuevos conceptos  
denominados fractilos

# Fractiles

- Facilitan la localización de la posición de un dato respecto a su grupo.

Cuartiles

Marcan el valor que está en el 25%, 75%, 50% y 100%

Marcan el valor que está en el 1%, 2%, 3%...100%

Centiles

Deciles

Marcan el valor que está en el 10%, 20%, 30%...100%

## Cuartiles

$$Q_j = \frac{j * N}{4} + 0.5$$

Qj: posición  
J: 1,2,3,4

## Deciles

$$D_j = \frac{j * N}{10} + 0.5$$

Dj: posición  
J: 1,2,3...9

## Centiles

$$C_j = \frac{j * N}{100} + 0.5$$

Cj: posición  
J: 1,2,3,4...99

- 1) Ordenar los datos de menor a mayor.
- 2) Calcular los fractiles
- 3) Determinar la posición correspondiente

# Datos no agrupados

$$Q_j = L_{Rinf} + \frac{\left(\frac{jN}{4} - F_{antQj}\right)}{f_{Qj}} * i$$

$$D_j = L_{Rinf} + \frac{\left(\frac{jN}{10} - F_{antDj}\right)}{f_{Dj}} * i$$

$$D_j = L_{Rinf} + \frac{\left(\frac{jN}{100} - F_{antCj}\right)}{f_{Cj}} * i$$

# Curtoosis

En función de fractilos

Curtoosis de Fisher

$$K_1 = \frac{0.5(Q_3 - Q_1)}{P_{90} - P_{10}}$$

$k_1 > 0.263$  es leptocúrtica  
 $k_1 = 0.263$  mesocúrtica  
 $k_1 < 0.263$  platicúrtica

$$K_{fisher} = \frac{m_4}{s^4}$$

$m_4$  es el momento de cuarto orden con respecto a la media aritmética

$k_{fisher} > 0$  es leptocúrtica  
 $k_{fisher} = 0$  mesocúrtica  
 $k_{fisher} < 0$  platicúrtica