

## *Estadística Aplicada a la Educación*

# Tema 5

Tutor.

**UNED Madrid-Sur (A.U. Parla)**

*Miguel Ángel Daza*

[migdaza@madridsur.uned.es](mailto:migdaza@madridsur.uned.es)

1

- La Estadística en el proceso de investigación pedagógica empírica.

2

- Problema, hipótesis / objetivos, variables y datos. Niveles de medida

4

- Organización de los datos. análisis exploratorio de datos.

5

- Reducción de datos. Medidas descriptivas básicas y representaciones gráficas.

6

- Medidas individuales.

7

- Relación entre variables. Las correlaciones. La regresión.

8

- Aplicaciones de la correlación: fiabilidad y validez de las medida.

9

- Modelos estadísticos y probabilidad. La curva normal de probabilidades.

10

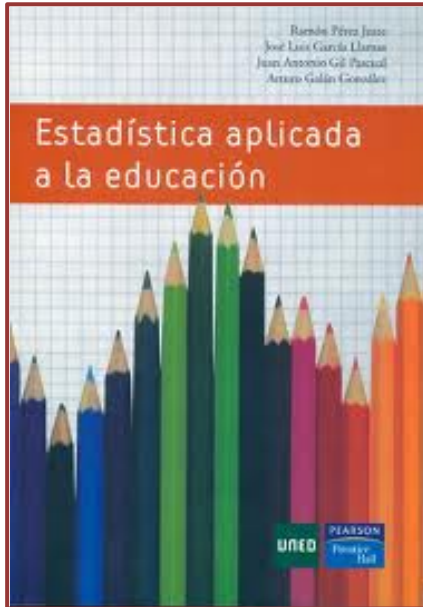
- Los baremos o normas. Muestreo. Aplicaciones.

11

- Estimación de parámetros. Errores de estimación.

12

- Introducción al contraste de hipótesis: la prueba t para el contraste de medias en los diseños de dos grupos.



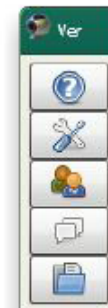
- 1) Introducción. De la organización a la descripción de datos.
- 2) Medidas de tendencia central: media, median y moda. Usos e interpretación.
- 3) Medidas de variabilidad.
  - 1) La desviación media.
  - 2) La desviación típica.
  - 3) La amplitud o recorrido.
  - 4) La desviación semi-intercuartílica.
  - 5) El cociente de variación.
- 4) Media y desviación típica para variables dicotómicas.
- 5) Asimetría y apuntamiento: relación con la curva normal.
- 6) Representaciones gráficas.
  - 1) Gráfico de sectores.
  - 2) Gráfico de barras.
  - 3) Histograma.
  - 4) Gráfico de Caja.
  - 5) Gráfico de tallo y hojas.

## 5. REDUCCIÓN DE DATOS. MEDIDAS DESCRIPTIVAS BÁSICAS Y REPRESENTACIONES GRÁFICAS.

+ Info +



## MEDIDAS DE TENDENCIA CENTRAL Y VARIABILIDAD. ASIMETRÍA Y CURTOSIS



### [http://www.intecca.uned.es/portalavip/grabacion.php?ID\\_Grabacion=56816&ID\\_Sala=60365&hashData=8841fa8f81551549edce84bdcf1a4005&paramsToCheck=SURfR3JhYmFjaW9uLEIEX1NhbGEGs](http://www.intecca.uned.es/portalavip/grabacion.php?ID_Grabacion=56816&ID_Sala=60365&hashData=8841fa8f81551549edce84bdcf1a4005&paramsToCheck=SURfR3JhYmFjaW9uLEIEX1NhbGEGs)

## 5.1 Introducción. De la organización a la descripción de datos.

Una vez tabulados los datos, comenzamos a trabajar con la matriz de datos y a realizar los análisis pertinentes...

...normalmente se parte de la Estadística Descriptiva para hacerse una primera idea del comportamiento de cada una de las variables contenidas en nuestra matriz de datos.

La **Estadística Descriptiva** consiste en una serie de procedimientos para organizar, clasificar y resumir conjuntos de datos a través de índices numéricos y por medio de representaciones gráficas.

- **Índices numéricos:** medidas de tendencia central (media, mediana y moda) y medidas de variabilidad (amplitud, varianza, desviación típica)

- **Representaciones gráficas:** gráfico de sectores, gráfico de barras, histograma, gráfico de caja...

Para caracterizar un grupo, necesitamos tanto medidas de **tendencia central** como **medidas de variabilidad**

(0,0,0,10,10,10) vs. (5,5,5,5,5,5)

## 5.1 Introducción. De la organización a la descripción de datos.

Para caracterizar un grupo, necesitamos tanto medidas de **tendencia central** como **medidas de variabilidad**

(0,0,0,10,10,10) vs. (5,5,5,5,5,5)

n	Grupo1	Grupo2
1	0	5
2	0	5
3	0	5
4	10	5
5	10	5
6	10	5

<b>media</b>	5	5
<b>desviación</b>	5	0

## 5.2 Medidas de tendencia central: media, mediana y moda. Usos e interpretación.

Una medida de tendencia central indica hacia que valor tiende el grupo, cuál es el valor que mejor lo representa. La más conocida es la **media aritmética**

$\bar{x} = \frac{\text{Suma de todas las puntuaciones}}{\text{Número total de puntuaciones}}$

$$\bar{x} = \frac{\sum x_i}{N}$$

Para interpretar correctamente una media aritmética, es conveniente conocer la puntuación **mínima** y **máxima** de la escala de medida de la variable y situar la media aritmética dentro de ese recorrido.

En sentido estricto, la **media aritmética** sólo puede utilizarse en variables cuyo nivel de medida sea de *intervalo* o de *razón*

## 5.2 Medidas de tendencia central: media, mediana y moda. Usos e interpretación.

La **Moda (Mo)** es el valor con frecuencia absoluta más alta, la puntuación que más se repite.

Cuando observemos la distribución de frecuencias, la puntuación directa a la que corresponda la frecuencia mayor será la moda.

*Atención: la moda no es la frecuencia (error frecuente en los estudiantes), es la puntuación.*

También hay distribuciones **bimodales y plurimodales**.

La Moda (Mo) es especialmente indicada para variables de tipo **categorístico** (nivel de **medida nominal**)

$X_i$	$f$	$f_a$	
22	2	50	■
21	4	48	■■■
20	1	44	■
19	5	43	■■■■■
18	1	38	■
17	2	37	■■
16	8	35	■■■■■■■■
15	2	27	■■
14	3	25	■■■
13	6	22	■■■■■■
12	5	16	■■■■■
11	2	11	■■
10	3	9	■■■
9	2	6	■■
8	3	4	■■■
5	1	1	■
	50		

**MODA**

La **MODA** es **16**,  
no **8**,  
**8** es **f**, la  
frecuencia absoluta



## 5.2 Medidas de tendencia central: media, mediana y moda. Usos e interpretación.

### Moda

### Datos sin agrupar

*La moda es aquel valor de la variable al que corresponde máxima frecuencia (absoluta o relativa)*

$$M_d = x_j \text{ tal que } n_j = \max \{n_1, \dots, n_k\}$$

## 5.2 Medidas de tendencia central: media, mediana y moda. Usos e interpretación.

La mediana ( $M_d$ ) es aquel valor que deja por encima y por debajo de sí al 50% de los sujetos de la muestra.

Si ordenamos las puntuaciones de menor a mayor (o al contrario), aquélla que ocupa la posición central, ésa es la mediana.

Mediana  
14,5

$$\text{si } N_{j-1} = \frac{n}{2} < N_j$$

$$N_8 = \frac{50}{2} = 25$$

$$N_9 = 27$$

$$N_8 = 50 < N_9 = 27$$

$$M_e = \frac{C_{j-1} + C_j}{2}$$

$$M_e = \frac{C_8 + C_9}{2} = \frac{14 + 15}{2} = 14,5$$

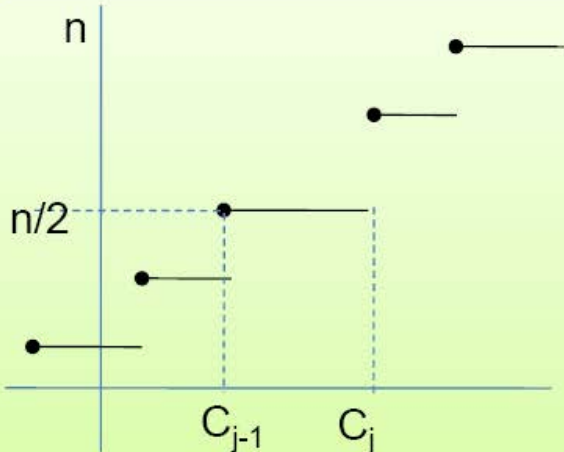
Puntuación Directa	Frecuencia	Frecuencia Relativa	Porcentaje	Frecuencia Acumulada	Frecuencia Relativa Acumulada	Porcentaje Acumulada
Xi	f	fi/N	p	fa	fa/N	pa
5	1	0,02	2,0%	1	0,02	2,0%
8	3	0,06	6,0%	4	0,08	8,0%
9	2	0,04	4,0%	6	0,12	12,0%
10	3	0,06	6,0%	9	0,18	18,0%
11	2	0,04	4,0%	11	0,22	22,0%
12	5	0,1	10,0%	16	0,32	32,0%
13	6	0,12	12,0%	22	0,44	44,0%
14	3	0,06	6,0%	25	0,5	50,0%
15	2	0,04	4,0%	27	0,54	54,0%
16	8	0,16	16,0%	35	0,7	70,0%
17	2	0,04	4,0%	37	0,74	74,0%
18	1	0,02	2,0%	38	0,76	76,0%
19	5	0,1	10,0%	43	0,86	86,0%
20	1	0,02	2,0%	44	0,88	88,0%
21	4	0,08	8,0%	48	0,96	96,0%
22	2	0,04	4,0%	50	1	100,0%
	50	1	100,0%			100,0%

## 5.2 Medidas de tendencia central: media, mediana y moda. Usos e interpretación.

### Mediana

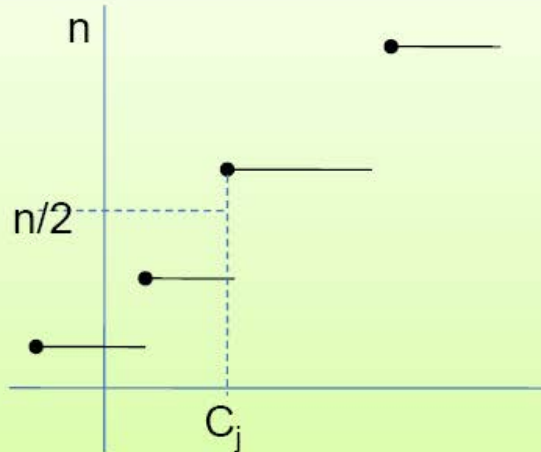
### Datos sin agrupar

La mediana es aquel valor de la variable, tal que supuestos ordenados los valores de ésta en orden creciente, la mitad son menores o iguales y la otra mitad son mayores o iguales.



$$\text{si } N_{j-1} = \frac{n}{2} < N_j$$

$$M_e = \frac{C_{j-1} + C_j}{2}$$



$$\text{si } N_{j-1} < \frac{n}{2} < N_j$$

$$M_e = C_j$$

## 5.2 Medidas de tendencia central: media, mediana y moda. Usos e interpretación.

La medida más precisa de tendencia central es la media aritmética, ya que tiene en cuenta todas las puntuaciones para su cálculo.

Podremos utilizar:

- **En variables de intervalo o razón:** media, mediana y moda
- **En variables de tipo ordinal:** mediana y moda
- **En variables de tipo nominal:** moda

Además, como la media se ve arrastrada por las puntuaciones extremas; la mediana puede ser una medida preferible a la media cuando existen puntuaciones extremas que distorsionen la verdadera tendencia central del grupo

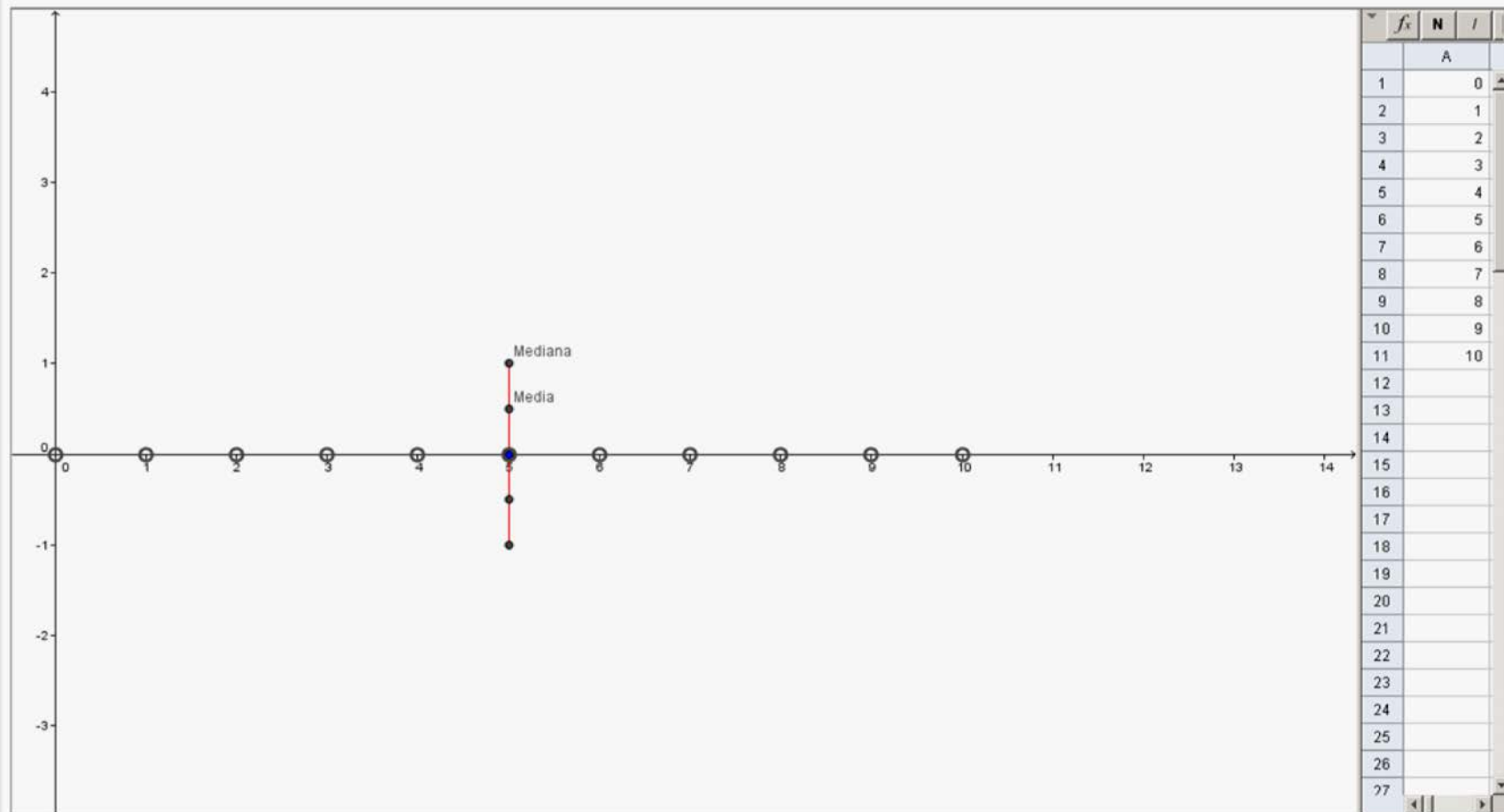
*(3,3,4,4,4,5,5) vs. (3,3,4,4,4,5,10)*

n	caso 1	caso 2
1	3	3
2	3	3
3	4	4
4	4	4
5	4	4
6	5	5
7	5	10
Suma	28	33
Media	4	4,71
Moda	4	4
Mediana	4	4

## Media y Mediana

11 Notas de Una Clase.

Ver el comportamiento de la MEDIA y la MEDIANA



MAD, 3 Marzo 2014, Creado con [GeoGebra](#)

## 5.3 Medidas de variabilidad.

Las medidas de tendencia central por sí solas no proporcionan suficiente información como para hacernos una idea completa de las características de un grupo...

...el índice de tendencia central debe ir acompañado por un **índice de dispersión o variabilidad** que indique en qué medida las puntuaciones de los sujetos se dispersan o varían en torno a la media aritmética

n	Grupo 1	Grupo 2
1	104	80
2	104	80
3	104	128
4	104	128
Suma	416	416
Media	<b>104</b>	<b>104</b>
Moda	<b>104</b>	<b>80 y 128</b>
Mediana	<b>104</b>	<b>104</b>

## 5.3.1 La desviación media.

El índice de dispersión más intuitivo y fácil de comprender es la **desviación media**.

La **desviación media (D.M.)** es una medida que indica la media aritmética de las desviaciones de las puntuaciones directa respecto de la media aritmética.

Es decir, si la D.M. es alta significa que las puntuaciones, en su conjunto, se desvían, están alejadas de la media aritmética del grupo (grupo heterogéneo)...

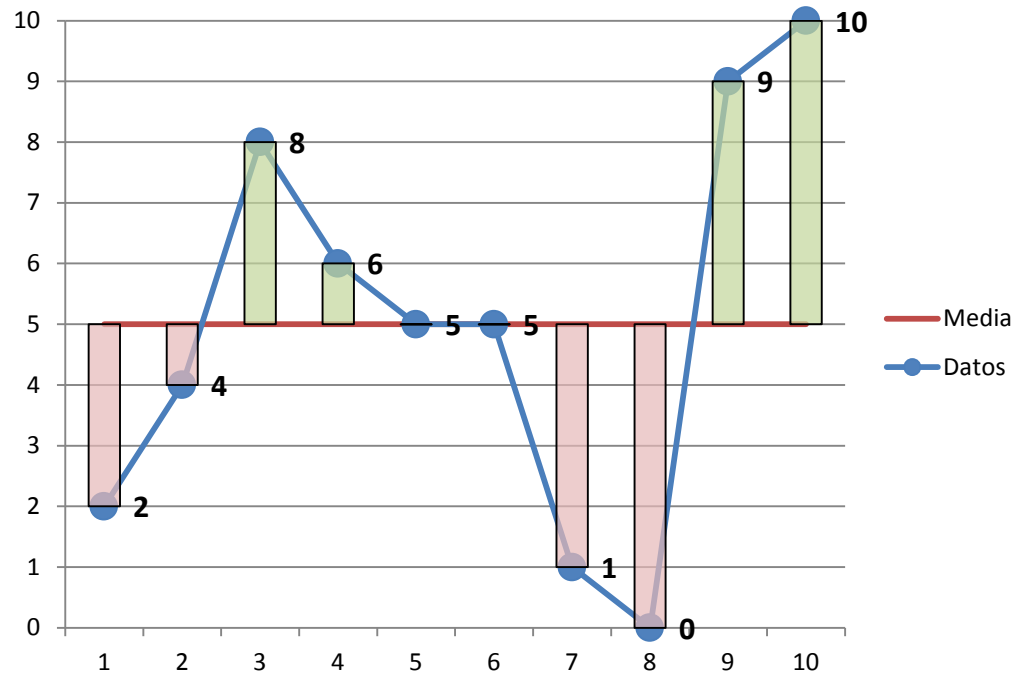
...si la D.M. es pequeña significa que las puntuaciones del grupo están próximas a la media aritmética (grupo homogéneo).

$$D.M. = \frac{\sum_{i=1}^n |(X_i - \bar{X})|}{N}$$

n	$X_i$	$X_i - X_m$	$ X_i - X_m $
1	2	-3	3
2	4	-1	1
3	8	3	3
4	6	1	1
5	5	0	0
6	5	0	0
7	1	-4	4
8	0	-5	5
9	9	4	4
10	10	5	5
Suma	50	0	26
Media=5		DM=2,6	

## 5.3.1 La desviación media.

$$D.M. = \frac{\sum_{i=1}^n |(X_i - \bar{X})|}{N}$$



n	$X_i$	$X_i - \bar{X}$	$ X_i - \bar{X} $
1	2	-3	3
2	4	-1	1
3	8	3	3
4	6	1	1
5	5	0	0
6	5	0	0
7	1	-4	4
8	0	-5	5
9	9	4	4
10	10	5	5
Suma	50	0	26

Media=5      DM=2,6



### 5.3.2 La desviación típica.

Los índices de variabilidad más conocidos y utilizados son la desviación típica ( $s$ ) y la varianza ( $s^2$ )

Al elaborar un informe descriptivo, lo usual es incluir la desviación típica como medida de dispersión. No siempre es fácil interpretar una desviación típica, decir si es grande o pequeña. Si tenemos un solo grupo, los valores que podemos tomar de referencia son:

- **La desviación típica mínima:** siempre cero (todos los sujetos tienen la misma puntuación *(p.ej.: 5, 5, 5, 5, 5, 5)*)
- **La desviación típica máxima:** la mitad de los sujetos con la puntuación máxima de la escala, y la otra mitad de los sujetos con la puntuación mínima de la escala *(p.ej.: 0, 0, 0, 10, 10, 10)*. *La desviación típica es la puntuación mayor menos la puntuación menor entre dos.*

**inseguro**

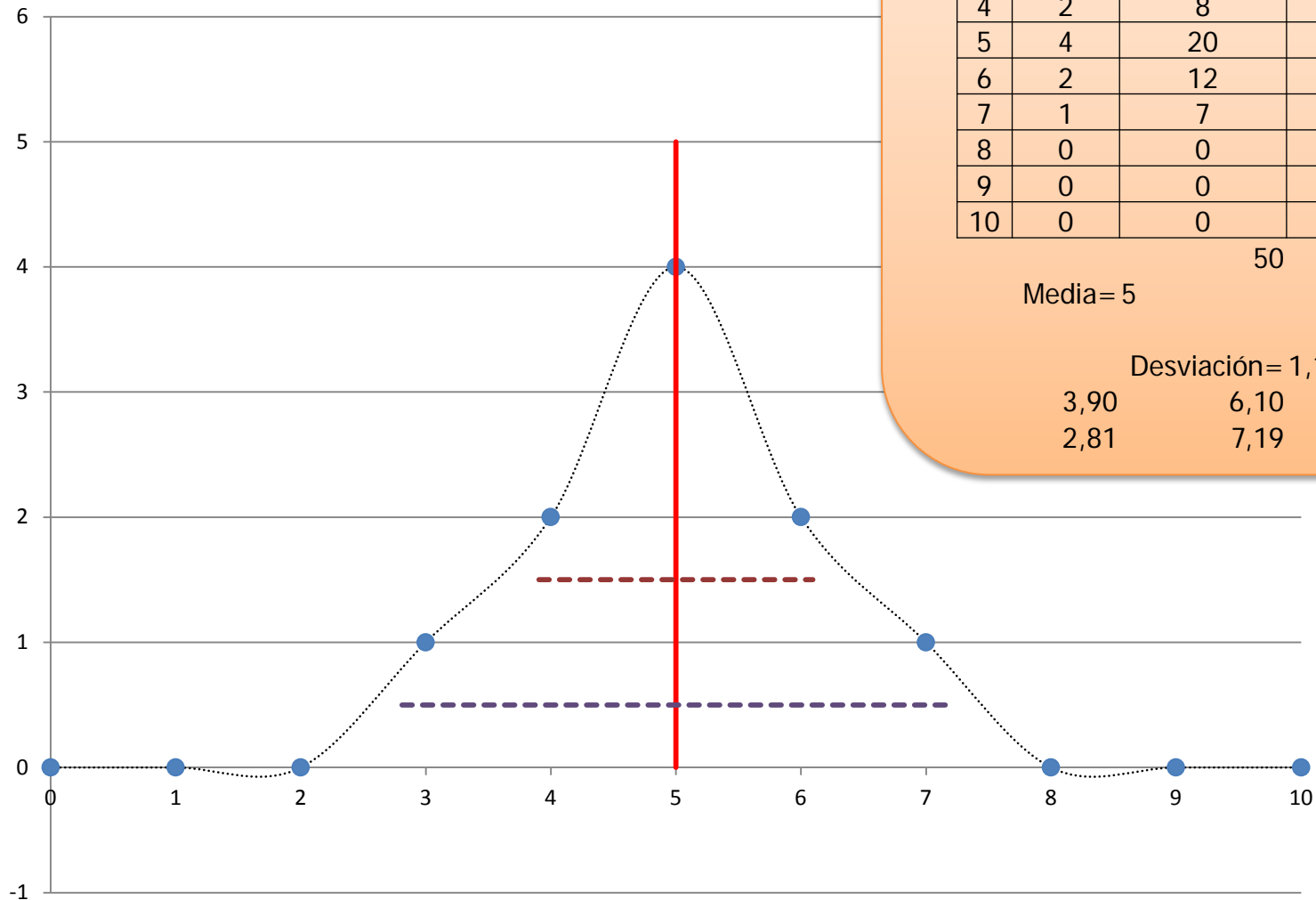
$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{N} \quad s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{N}}$$

$\sigma_n$

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{N-1} \quad s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{N-1}}$$

$\sigma_{n-1}$

## 5.3.2 La desviación típica.



$X_i$	$f_i$	$X_i + f_i$	$(X_i - M_d)^2 \cdot f_i$
1	0	0	0
2	0	0	0
3	1	3	4
4	2	8	2
5	4	20	0
6	2	12	2
7	1	7	4
8	0	0	0
9	0	0	0
10	0	0	0

50 12  
 Media= 5 1,2

Desviación= 1,10  
 3,90 6,10  
 2,81 7,19

- .....●..... Notas
- Media
- - - - -  $X_m + s$
- - - - -  $X_m + 2s$

## 5.3.3 La amplitud o recorrido.

La medida de variabilidad más básica es la **amplitud** o **recorrido** de una variable

$$A = X_{i_{mayor}} - X_{i_{menor}} + 1$$

La amplitud se utiliza como medida de dispersión únicamente cuando no es posible calcular otra o como complemento de la moda (cuando el nivel de medida de la variable es nominal)...

...al igual que sucedía con la moda como índice de tendencia central, es un índice muy simple, basado en tan sólo dos puntuaciones, lo que puede dar lugar a malas interpretaciones si no se valora con precaución, sobre todo cuando existen puntuaciones extremas.

## 5.3.4 La desviación semi-intercuartílica.

La **desviación semi-intercuartílica (Q)** es una medida que indica la dispersión en el 50% central de la distribución...

...es adecuada cuando el nivel de medida de la variable es ordinal (en este caso será el complemento de la mediana). También es adecuado su uso cuando la existencia de puntuaciones extremas pueda distorsionar en exceso la desviación típica...

$$Q = \frac{Q_3 - Q_1}{2}$$

...la Q prescinde del 25% inferior y del 25% superior de las puntuaciones, calculando la dispersión en el 50% central, entre los percentiles 25 y 75 o, lo que es lo mismo, entre los cuartiles 1 y 3.

$$Q_1 = 25\%$$

$$Q_2 = 50\% = \textit{Mediana}$$

$$Q_3 = 75\%$$

$$Q_4 = 100\% = N$$

$$Q_k = \frac{kN}{4}$$

donde  $k = 1, 2, 3, 4$

## 5.3.5 El cociente de variación.

El coeficiente de variación (V), a diferencia de los índices anteriores, nos permite comparar la variabilidad entre variables que tienen distinta amplitud.

$$V = \frac{s}{\bar{X}} (100)$$

Y es que dos desviaciones típicas procedentes de instrumentos con distinto recorrido o distinta escala de medida no son directamente comparables

## 5.4 Media y desviación típica para variables dicotómicas.

Variables dicotómicas: son variables que sólo pueden tomar 2 valores. Normalmente se codifican con ceros y unos:

- 0 = "noes" o "respuestas incorrectas"
- 1 = "síes" o "respuestas correctas"

La media aritmética representa la proporción de unos.

$$p = \frac{\text{n}^\circ \text{ de "1"}}{\text{n}^\circ \text{ total de respuestas}} \quad p + q = 1 \quad q = \frac{\text{n}^\circ \text{ de "0"}}{\text{n}^\circ \text{ total de respuestas}}$$

Las medias con variables dicotómicas son muy usadas en pruebas objetivas con preguntas de acierto/error. En este caso, la media nos dirá la proporción de sujetos que ha contestado correctamente a la pregunta, que viene a denominarse **"índice de dificultad del ítem"**

$$s^2 = p \cdot q$$

$$s = \sqrt{p \cdot q}$$

La desviación típica variará entre 0 (ausencia de variabilidad) y 0,5 (máxima variabilidad, es decir, cuando la mitad contesta que sí y la otra mitad contesta que no)

## 5.5 Asimetría y apuntamiento: relación con la curva normal.

Muchas variables educativas, psicológicas y biológicas se distribuyen según la llamada **distribución normal = curva normal = campana de Gauss** (*rendimiento académico, inteligencia, aptitud, motivación, peso, altura*)

La distribución viene a indicar cómo la mayoría de las personas nos encontramos en torno a los valores medios de la distribución y según nos alejamos hacia valores extremos el número de sujetos disminuye progresivamente

La **curva normal** es una distribución teórica, **simétrica** y **asintótica**, donde: la **media**, la **mediana** y la **moda** coinciden.

(Tiene índice de **asimetría** igual a cero y **apuntamiento** igual a cero)

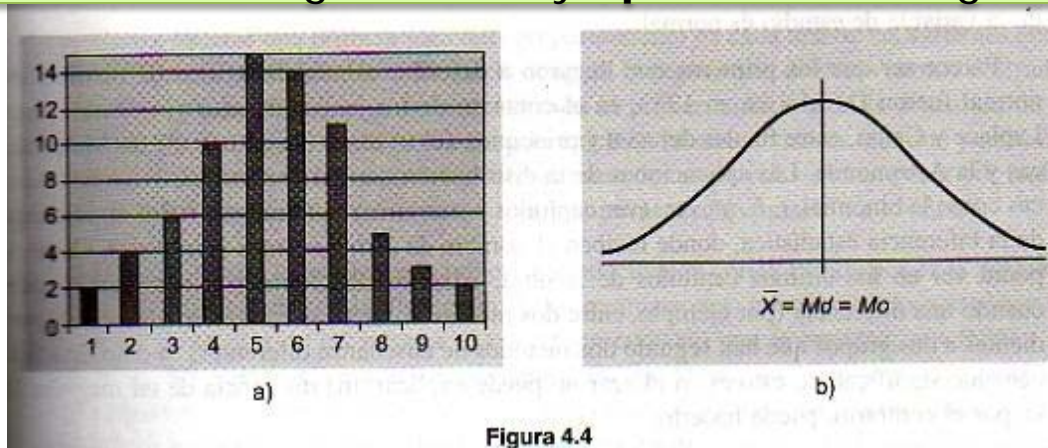


Figura 4.4

## 5.5 Asimetría y apuntamiento: relación con la curva normal.

La **asimetría positiva** indica que la mayoría de los sujetos tienden a concentrarse en la parte baja de las puntuaciones de la distribución de frecuencias. La cola de la distribución está a la derecha...

...la **asimetría negativa** indica lo contrario: los sujetos tienden a agruparse en torno a las puntuaciones altas de la distribución, La cola de distribución está a la izquierda.

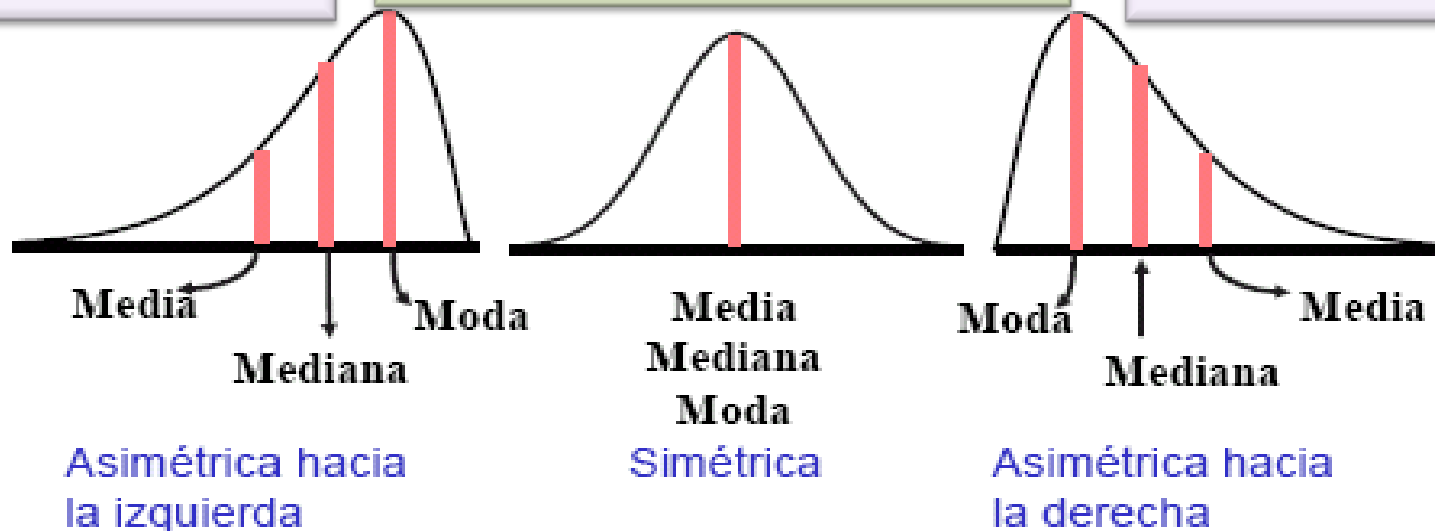
### ASIMETRÍA NEGATIVA

Asimétrica Izquierda →  
Media < Mediana < Moda

Media = Mediana = Moda  
**Simétrica**

### ASIMETRÍA POSITIVA

Asimétrica Derecha →  
Moda < Mediana < Media



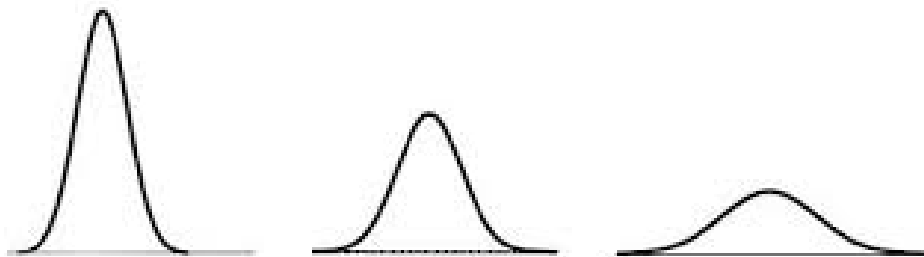




## 5.5 Asimetría y apuntamiento: relación con la curva normal.

El apuntamiento o curtosis indica el grado en el que la distribución es más o menos picuda (más o menos dispersa respecto a la media):

- **Mesocúrtica**: la curva normal. **Curtosis = 0**
- **Leptocúrtica**: más apuntada. Concentración de las puntuaciones alrededor de la media. **Curtosis > 0**
- **Platicúrtica**: más achatada. Dispersión de las puntuaciones alrededor de la media. **Curtosis < 0**



Leptocúrtica

Mesocúrtica

Platicúrtica

### Curtosis

$$g_2 = \frac{1}{N} \frac{\sum_{i=1}^n (X_i - X)^4 \cdot f_i}{\sigma^4} - 3$$

## 5.6 Representaciones gráficas.

Las representaciones gráficas son un buen complemento para los índices de tendencia central y variabilidad, y ayudan a comprender rápidamente y de manera intuitiva la información descriptiva.

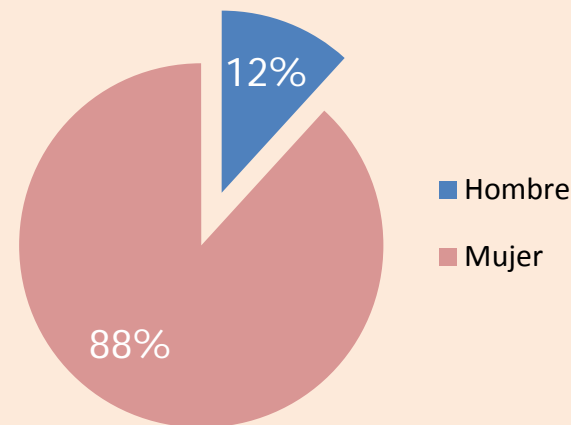
## 5.6.1 Gráfico de sectores. (pie chart)

El **gráfico de sectores** (o “ciclograma” o “gráfico de tarta”) se utiliza para representar variables categóricas (**nivel de medida nominal**).

Tiene forma circular, y está dividida en **porciones** que representan **proporcionalmente** la presencia de cada una de las categorías de la variable.

Parla	frecuencia	Porcentaje
Hombre	6	11,76%
Mujer	45	88,24%
	<b>51</b>	<b>100,00%</b>

Alumnos AU Parla



## 5.6.2 Gráfico de barras. (bar chart)

El **gráfico de barras** se utiliza para variables nominales u ordinales (cuantitativas discretas).

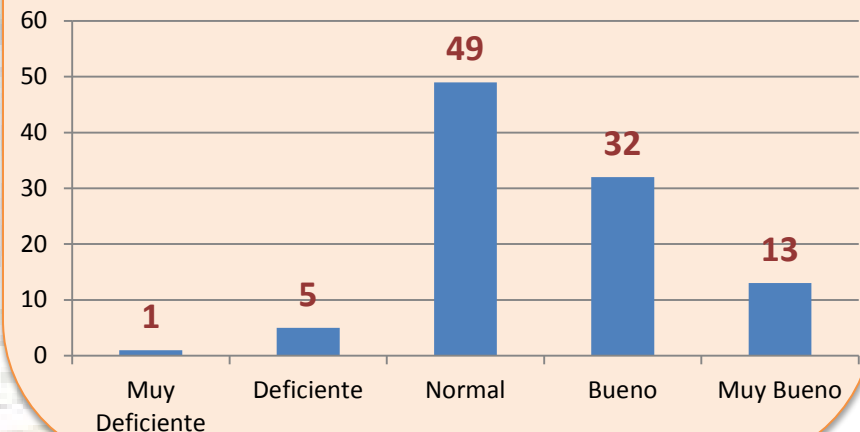
**Eje horizontal** = categorías o niveles de la variable.

**Eje vertical** = frecuencias o porcentajes.

Cuando las variables son cuantitativas continuas (**con niveles de intervalo o de razón**), las barras en vez de separadas están juntas y se forma un Histograma, en el que se suelen agrupar las puntuaciones por intervalos.

Percepción	frecuencia
Muy Deficiente	1
Deficiente	5
Normal	49
Bueno	32
Muy Bueno	13
	<b>100</b>

**Percepción del rendimiento académico de los alumnos**



## 5.6.3 Histograma.

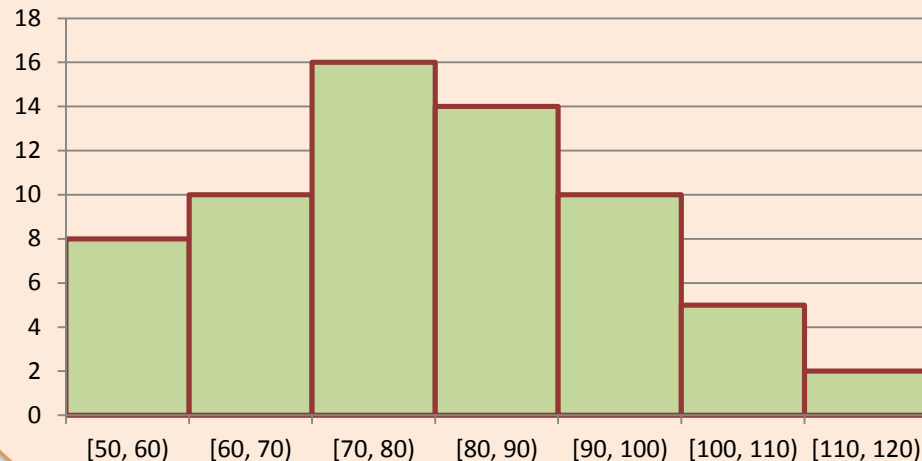
Cuando las variables son cuantitativas continuas (con niveles de intervalo o de razón), las barras en vez de separadas están juntas y se forma un **Histograma**, en el que se suelen agrupar las puntuaciones por intervalos.

El peso de 65 personas adultas viene dado por la siguiente tabla:

	$c_i$	$f_i$	$F_i$
[50, 60)	55	8	8
[60, 70)	65	10	18
[70, 80)	75	16	34
[80, 90)	85	14	48
[90, 100)	95	10	58
[100, 110)	105	5	63
[110, 120)	115	2	65

65

**Peso personas adultas**

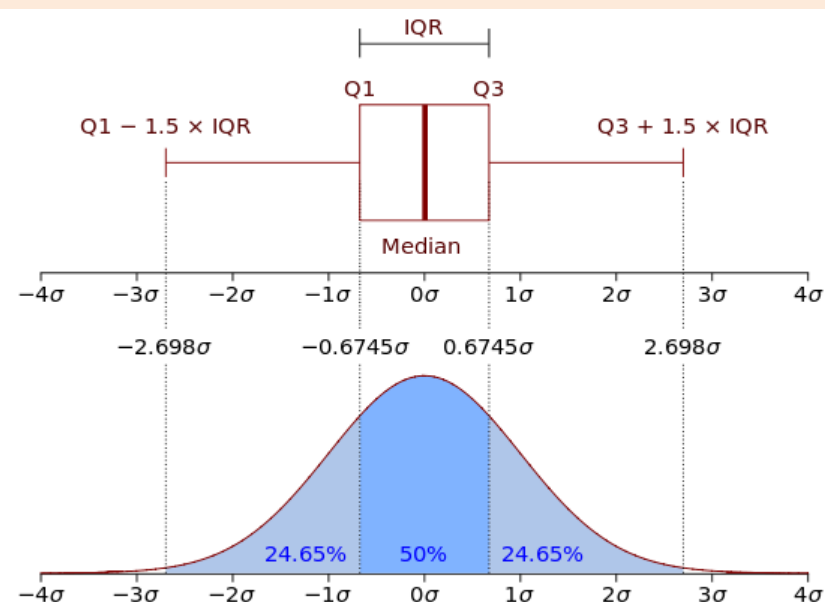
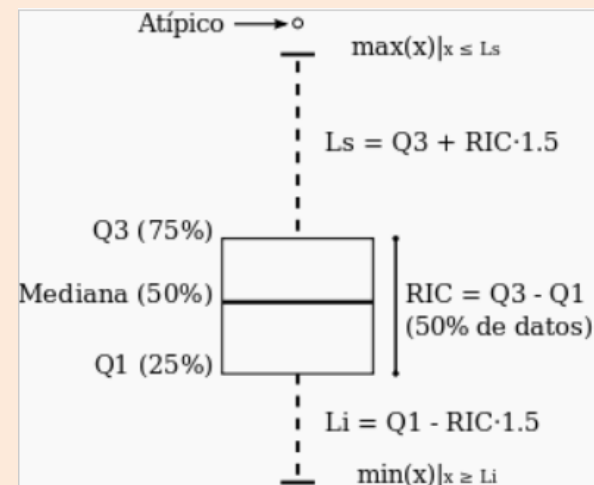


## 5.6.4 Gráfico de Caja. (box-plot)

El diagrama de caja (o de caja y patillas, o de caja y bigotes) es un gráfico muy práctico para hacerse una idea de la distribución de las puntuaciones en la zona central (zona entre Q1 y Q3) y en los extremos. Las patillas representan las puntuaciones hasta los extremos de la distribución.

Las puntuaciones raras, atípicas o extremas (outlier) son aquellas más allá de:

$$Md \pm 1,5 (Q3 - Q1)$$



## 5.6.5 Gráfico de tallo y hojas. (Steam-and-leaf Plot)

El gráfico de tallo y hojas combina la representación numérica y gráfica.

Es una especie de histograma horizontal cuyas barras están construidas con los números correspondientes a las puntuaciones.

**Hojas:** último dígito de la puntuación.

**Tallo:** el resto de dígitos

15 16 21 23 23 26 26 30 32 41

15	1	5
16	1	6
21	2	1
23	2	3
23	2	3
26	2	6
26	2	6
30	3	0
32	3	2
41	4	1

156
213366
302
41



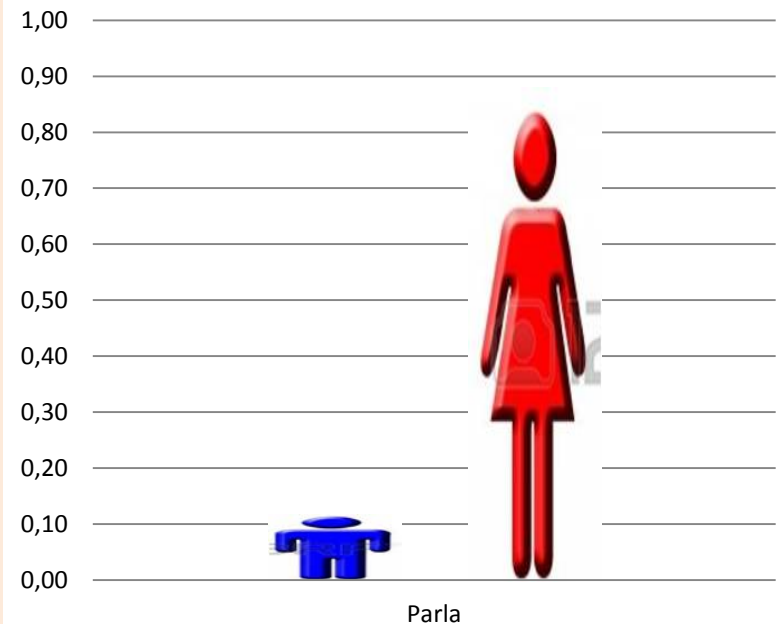
## Pictograma. (pictograph)

Es un diagrama que utiliza imágenes o símbolos para mostrar datos para una rápida comprensión. En un pictograma, se utilizan imágenes o símbolos para representar una cantidad específica y su tamaño o cantidad es proporcional a la frecuencia que representa.

Para realizarlo primero se escogen figuras alusiva al tema y se le asigna una imagen.

Parla	frecuencia	Porcentaje
Hombre	6	11,76%
Mujer	45	88,24%
	<b>51</b>	<b>100,00%</b>

**Distribución Alumnos AU Parla**



## Nomenclatura

Datos ( $X_i$ )

Frecuencia Absoluta ( $f_i$ )

Frecuencia Relativa ( $f_i / N$ )

Frecuencia Absoluta Acumulada ( $f_a$ )

Frecuencia Relativa Acumulada ( $f_a / N$ )

Media ( $\bar{X}$ )

Moda ( $Mo$ )

Mediana ( $Md$ )

Desviación Media ( $DM$ )

Varianza ( $s^2$ )

Desviación Típica ( $s$ )

*Media aritmética:*

$$\bar{X} = \frac{\text{Suma de todas las puntuaciones}}{\text{Número total de puntuaciones}}$$

$$\bar{X} = \frac{\sum X_i}{N}$$

*Desviación Media:*

$$\text{D.M.} = \frac{\text{Valor absoluto de la suma de las diferencias entre cada puntuación directa y la media}}{\text{Número total de sujetos}}$$

$$\text{D.M.} = \frac{\sum |X_i - \bar{X}|}{N}$$

*Desviación típica sesgada:*

$$S = \sqrt{\frac{\text{(Suma de las diferencias entre cada puntuación directa y la media aritmética)}^2}{\text{Número total de sujetos}}}$$

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

*Varianza:*

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{N}$$

*Desviación típica insesgada:*

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N-1}}$$

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{N-1}$$

*Amplitud o recorrido:*

$$A = X_{\text{mayor}} - X_{\text{menor}} + 1$$

*Coefficiente de variación:*

$$V = \frac{s}{\bar{X}} (100)$$

*Desviación típica y varianza en variables dicotómicas:*

$$s = \sqrt{p \cdot q} ; s^2 = p \cdot q$$

*Asimetría: Índice de Pearson:*

$$A_s = \frac{\bar{X} - M_0}{s}$$

*Curtosis:*

$$g_2 = \frac{1}{N} \cdot \frac{\sum_{i=1}^n (X_i - \bar{X})^4 \cdot f_i}{\sigma^4} - 3$$

## Resumen

Estadística Descriptiva.

Medidas de tendencia central.

Medidas de dispersión o variabilidad.

Media aritmética.

Puntuación directa.

Moda.

Mediana.

Distribución bimodal.

Distribución plurimodal.

Distribución de frecuencias.

Desviación Media (DM).

Desviación típica.

Varianza.

Desviación típica insesgada.

Varianza insesgada.

Amplitud o recorrido.

Puntuaciones extremas (outliers).

Desviación semi-intercuartílica.

Coefficiente de Variación (V).

Índice de dificultad del ítem-

Asimetría. (positiva o negativa).

Grado de asimetría.

Apuntamiento o curtosis.

Platicúrtica, Mesocúrtica, Leptocúrtica.

Gráficos:

de Sectores, de Barras, Histogramas, de caja,

de tallo y hojas.

## Fe de erratas

**Pág. 98.** Línea 7, Donde dice: “...frecuencias absolutas **(f)**...” debe decir: “...frecuencias absolutas **(fi)**...”. Asimismo, en el encabezado de la tabla 5.1., segunda columna, debe poner **fi** en vez de **f**.

**Pág. 103.** En la fórmula de la varianza, debe poner: 
$$s^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}$$

**Pág. 107,** penúltima frase del primer párrafo, cuando dice: “...sin embargo, podría darse asimetría **negativa** (por ejemplo la mayoría...” Debe decir: “...sin embargo, podría darse asimetría **positiva** (por ejemplo la mayoría....”

**Pág. 112,** último párrafo. **Donde dice:**

*“El diagrama de caja, también conocido como caja y patillas o caja y bigotes, es un gráfico muy práctico porque permite hacerse una idea rápida de la distribución de las puntuaciones en la zona central (el espacio en rojo, que comprende desde el cuartil 1 o percentil 25 hasta el cuartil 3 o percentil 75) y en los extremos. Las patillas representan las puntuaciones hasta los extremos de la distribución.*

*Este gráfico corresponde a la misma variable presentada en el histograma anterior, y podemos ver que la mitad del grupo se sitúa entre las puntuaciones 65 y 84, estando la mediana en 75. Según se defina en el programa, las puntuaciones que se alejen de la media del grupo el número de desviaciones típicas que especifiquemos, son consideradas valores atípicos, raros o extremos (outliers). Aunque en el gráfico vemos que la puntuación menor sería el outlier, por ser un caso raro se le dibuja fuera de la patilla*

## Fe de erratas

*inferior, indicando que dicho caso se podría descartar en posteriores análisis”.*

### Debe decir:

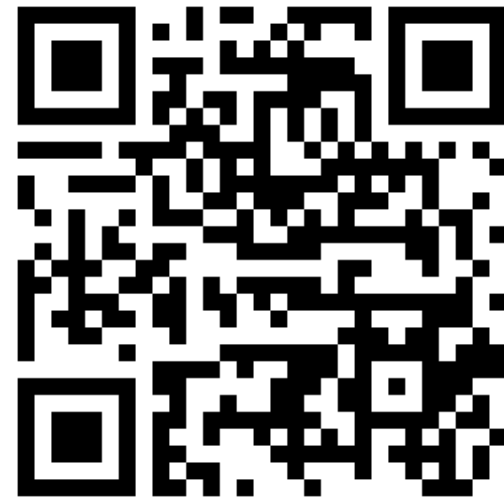
*“El diagrama de caja, también conocido como caja y patillas o caja y bigotes, es un gráfico muy práctico porque permite hacerse una idea rápida de la distribución de las puntuaciones en la zona central (el espacio sombreado, que comprende desde el cuartil 1 o percentil 25 hasta el cuartil 3 o percentil 75) y en los extremos. Otra de sus utilidades es que nos informa gráficamente del grado de asimetría de la distribución y presenta los valores atípicos (outliers). Las patillas o bigotes tienen como extremos posibles los valores máximo y mínimo de la variable. Sin embargo, las patillas tienen un máximo de prolongación; para determinar los límites de las patillas suele utilizarse la fórmula de multiplicar el intervalo intercuartil –que es la simple diferencia entre el tercer y el primer cuartil– por  $\pm 1,5$ . Si hay valores que se alejan de los límites de este producto, se clasifican como atípicos. También puede especificarse la existencia de valores aún más extremos: Si hay puntuaciones que se alejan más del producto del intervalo intercuartil por  $\pm 3$ , entonces se habla de puntuaciones extremas. Esto nos ayuda a identificar valores que pueden distorsionar el comportamiento de la variable al calcular valores como la media aritmética, la desviación típica, una correlación, etc. En ciertos casos, puede estar justificado eliminar de algunos análisis estadísticos a los sujetos con puntuaciones atípicas o extremas.*

*Este gráfico corresponde a la misma variable presentada en el histograma anterior, y podemos ver que la mitad del grupo se sitúa entre las puntuaciones 65 y 84, estando la mediana en 75. En este caso, el intervalo intercuartil sería  $84-65=19$ .  $19*1,5=28,5$ . Entonces, si le sumamos y restamos este valor a la mediana, obtenemos los extremos de las patillas:  $75\pm 28,5=(46,5, 103,5)$ . En el gráfico vemos que la puntuación menor sería el outlier, y por ser un caso raro se le dibuja fuera de la patilla inferior, indicando que dicho caso debería ser analizado antes de incluirlo en posteriores análisis”.*

**Pág. 113**, última línea, donde dice “La mediana más alta corresponde al centro **10**”, debe decir: “La mediana más alta corresponde al centro **42**”.

# PREGUNTAS

Exámenes  
anteriores





1 La mediana es útil para:

Seleccione una:

- a. Dividir a un grupo en dos mitades con el mismo número de sujetos.
- b. Medir la heterogeneidad de un grupo.
- c. Conocer la tendencia central de un grupo atendiendo a sus puntuaciones extremas.

2

El cálculo de la desviación típica de una distribución de datos nos permite asegurar que su valor será siempre positivo porque los valores en que se basa:

Seleccione una:

- a. Al ser elevados al cuadrado, son positivos.
- b. Se toma en valor absoluto.
- c. Son positivos.

1

La mediana es útil para:

Seleccione una:

- a. Dividir a un grupo en dos mitades con el mismo número de sujetos. ✓
- b. Medir la heterogeneidad de un grupo.
- c. Conocer la tendencia central de un grupo atendiendo a sus puntuaciones extremas.

La respuesta correcta es: Dividir a un grupo en dos mitades con el mismo número de sujetos.

2

El cálculo de la desviación típica de una distribución de datos nos permite asegurar que su valor será siempre positivo porque los valores en que se basa:

Seleccione una:

- a. Al ser elevados al cuadrado, son positivos. ✓
- b. Se toma en valor absoluto.
- c. Son positivos.

La respuesta correcta es: Al ser elevados al cuadrado, son positivos.

3

La moda y la amplitud son:

Seleccione una:

- a. Medidas de dispersión.
- b. Medidas de tendencia central.
- c. Estadísticos descriptivos.

4

En un gráfico de cajas y bigotes, la parte central del mismo o cajo, recoge el 50% de los valores de la distribución dado que está limitado por:

Seleccione una:

- a.  $\pm\sigma$
- b.  $\pm 2\sigma$
- c. Q1 y Q3

3

La moda y la amplitud son:

Seleccione una:

- a. Medidas de dispersión.
- b. Medidas de tendencia central.
- c. Estadísticos descriptivos. ✓

La respuesta correcta es: Estadísticos descriptivos.

4

En un gráfico de cajas y bigotes, la parte central del mismo o cajo, recoge el 50% de los valores de la distribución dado que está limitado por:

Seleccione una:

- a.  $\pm\sigma$
- b.  $\pm 2\sigma$
- c. Q1 y Q3 ✓

La respuesta correcta es: Q1 y Q3

5

Los diagramas de barras Se utilizarán:

Seleccione una:

- a. En sustitución de los diagramas de barras y hojas.
- b. Para representar la distribución de los datos en una variable nominal.
- c. Para dibujar la distribución de los datos en una variable continua.

6

El gráfico de caja y patillas resulta inadecuado para una escala. ...

Seleccione una:

- a. De razón.
- b. De intervalo.
- c. Nominal.

5

Los diagramas de barras Se utilizarán:

Seleccione una:

- a. En sustitución de los diagramas detallo y hojas.
- b. Para representar la distribución de los datos en una variable nominal. ✓
- c. Para dibujar la distribución de los datos en una variable continúa.

La respuesta correcta es: Para representar la distribución de los datos en una variable nominal.

6

El gráfico de caja y patillas resulta inadecuado para una escala. ...

Seleccione una:

- a. De razón.
- b. De intervalo.
- c. Nominal. ✓

La respuesta correcta es: Nominal.



Estadística aplicada a la Educación

EJERCICIOS

**TEMA : 5.**

## ACTIVIDADES RECOMENDADAS TEMA 5

A partir de la matriz de datos construida en el tema anterior, calcule los índices de tendencia central y variabilidad para cada una de las variables e interprete los resultados. Realice una representación gráfica de una variable cualitativa y de una variable cuantitativa.

Si desea hacerlo con Excel, vaya a opciones de Excel--complementos e instale el complemento “Herramientas para análisis”. Le aparecerá entonces, en el menú de Datos la opción Análisis de datos. Utilice aquí la opción “Estadística Descriptiva”. Puede recurrir igualmente a otros programas de análisis estadístico específico como SPSS.

