

# ESTADÍSTICA



## GRADO TURISMO

### TEMA 2: ANÁLISIS DE DATOS TURÍSTICOS UNIDIMENSIONALES

Prof. Rosario Martínez Verdú



# TEMA 2: ANÁLISIS DE DATOS TURÍSTICOS UNIDIMENSIONALES

1. Presentación de los datos: distribuciones de frecuencias, tablas estadísticas y representaciones gráficas.
2. Medidas de posición.
3. Medidas de dispersión y de forma.
4. Valores atípicos y diagrama de caja.
5. Medidas de concentración.

## ETAPAS DEL ANÁLISIS DESCRIPTIVO DE LOS DATOS TURÍSTICOS UNIDIMENSIONALES.

- **1ª ETAPA:** Presentación de los datos (apartado 1)
  - Clasificar y ordenar los datos en una Tabla estadística o de Frecuencias.
  - Representarlos Gráficamente (diagrama de barras, histograma, etc.).
- **2ª ETAPA:** Resumir, sintetizar y representar toda la distribución de frecuencias mediante una serie de medias que intenten recoger los aspectos más relevantes de la distribución. Interesa calcular e interpretar el valor de esas medidas (apartados 2 y 3)
  - Medidas de posición,
  - Medidas de dispersión
  - Medidas de forma o perfil.



Fuente: [www.gaturro.com](http://www.gaturro.com)

Reproducido con permiso del Autor: Nik

# 1.- Presentación de los datos: distribuciones de frecuencias, tablas estadísticas y representaciones gráficas.

## TIPOS DE DISTRIBUCIONES DE FRECUENCIAS UNIDIMENSIONALES.

- **Distribuciones de frecuencias unitarias**

Aquellas en el que cada valor de la variable aparece una sola vez, no hay repeticiones, por eso las frecuencias son todas 1.

- **Distribuciones de frecuencias no unitarias**

Hay valores repetidos. Se clasifican en:

n° de observaciones	n° valores distintos de la variable	pocos	Muchos
		1) De datos sin agrupar	2) De datos agrupados en intervalos

a) Intervalos de misma amplitud

b) Intervalos de distinta amplitud

# 1) Ejemplo de Distribución de Frecuencias con Datos sin Agrupar

## (Ejemplo 3 de Variable cuantitativa discreta del tema 1)

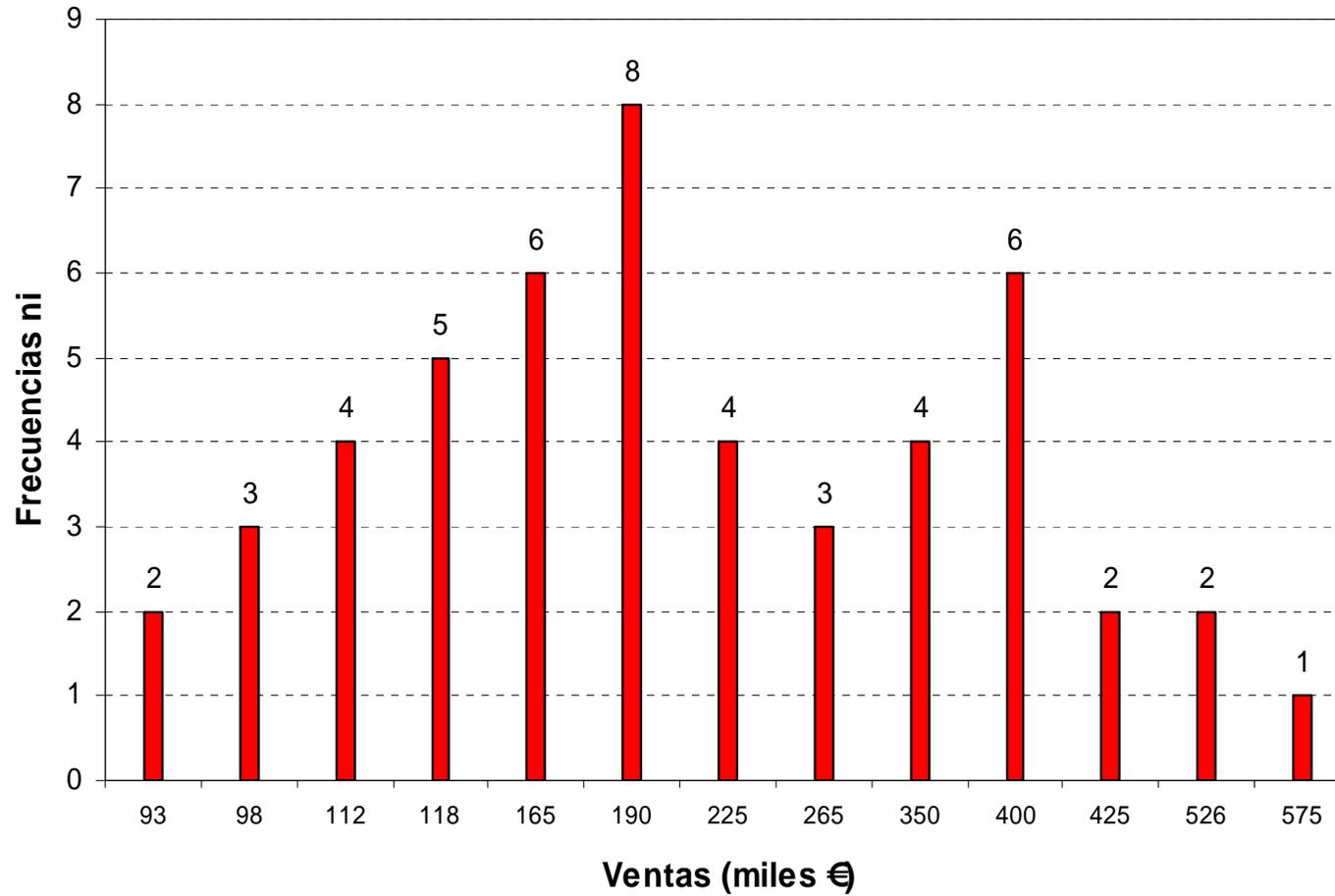
**Tabla de frecuencias o Tabla estadística con los distintos tipos de frecuencias:**

Ventas (miles €) de una cadena de agencias de viajes	Frecuencias			
	$n_i$	$f_i = n_i / N$	$N_i$	$F_i = N_i / N$
	Absoluta	relativa	absoluta acumulada	relativa acumulada
93	2	0,04	2	0,04
98	3	0,06	5	0,10
112	4	0,08	9	0,18
118	5	0,10	14	0,28
165	6	0,12	20	0,40
190	8	0,16	28	0,56
225	4	0,08	32	0,64
265	3	0,06	35	0,70
350	4	0,08	39	0,78
400	6	0,12	45	0,90
425	2	0,04	47	0,94
526	2	0,04	49	0,98
575	1	0,02	<b>50</b>	<b>1</b>
<b>Total. N</b>	<b>50</b>	<b>1</b>		

Fuente: [http://www.innova.uned.es/webpages/Ilde/Fund\\_Esta\\_Turis/04\\_jun\\_2.pdf](http://www.innova.uned.es/webpages/Ilde/Fund_Esta_Turis/04_jun_2.pdf)

## Representación Gráfica:

Diagrama de Barras para una distribución de frecuencias con valores sin agrupar



## 2a) Distribución de Frecuencias con Datos Agrupados en Intervalos de misma amplitud

**Colectivo: 200 turistas que visitaron la Comunidad Valenciana y cuya estancia no superó los 22 días.**

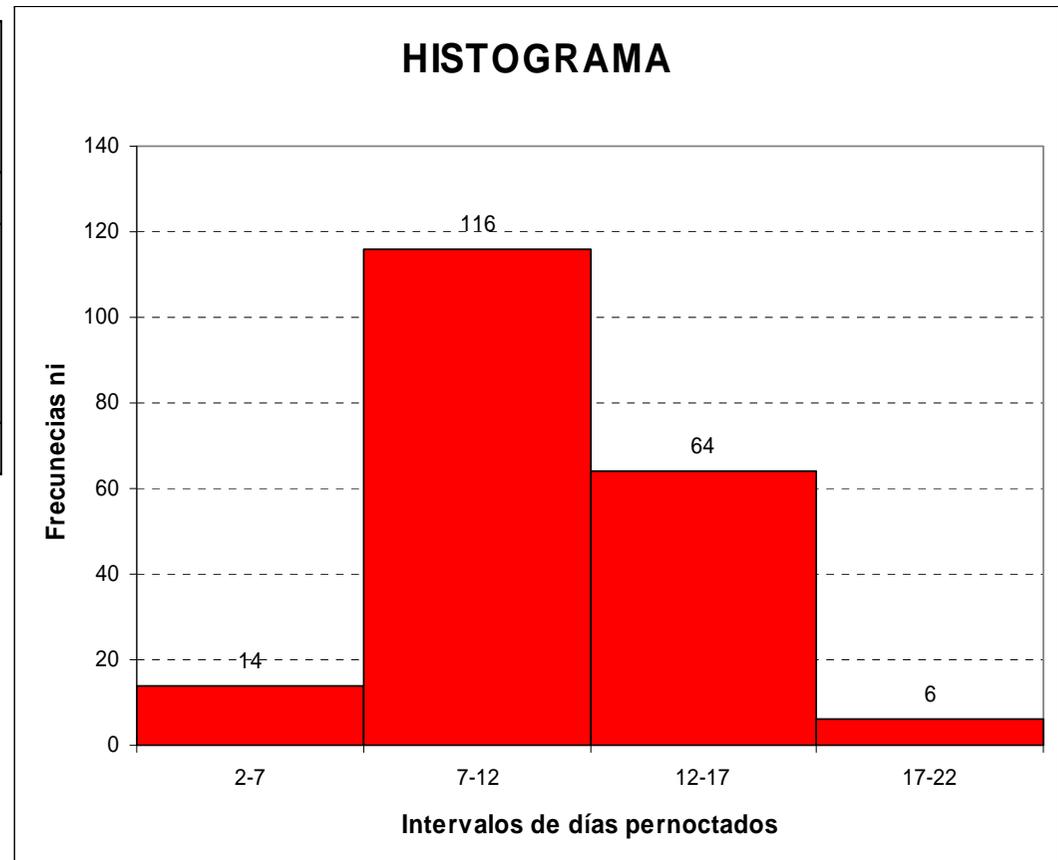
**N = 200**

**Variable X: días pernoctados, agrupados en intervalos**

**Tabla de frecuencias  
con datos agrupados en intervalos  
de misma amplitud:**

**Representación Gráfica:**

intervalos	marcas de clase $x_i$	frecuencias absolutas		frecuencias relativas	
		$n_i$	$N_i$	$f_i$	$F_i$
[2-7]	3,5	14	14	0,07	0,08
]7-12]	9,5	116	130	0,58	0,65
]12-17]	14,5	64	194	0,32	0,97
]17-22]	19,5	6	200	0,03	1
		<b>N=200</b>		1	



## 2b) Distribución de Frecuencias con Datos Agrupados en Intervalos de distinta amplitud

Colectivo: 524 turistas

$N = 524$

Variable **X**: días pernoctados en alojamientos de turismo rural (ejemplo 4 tema 1)

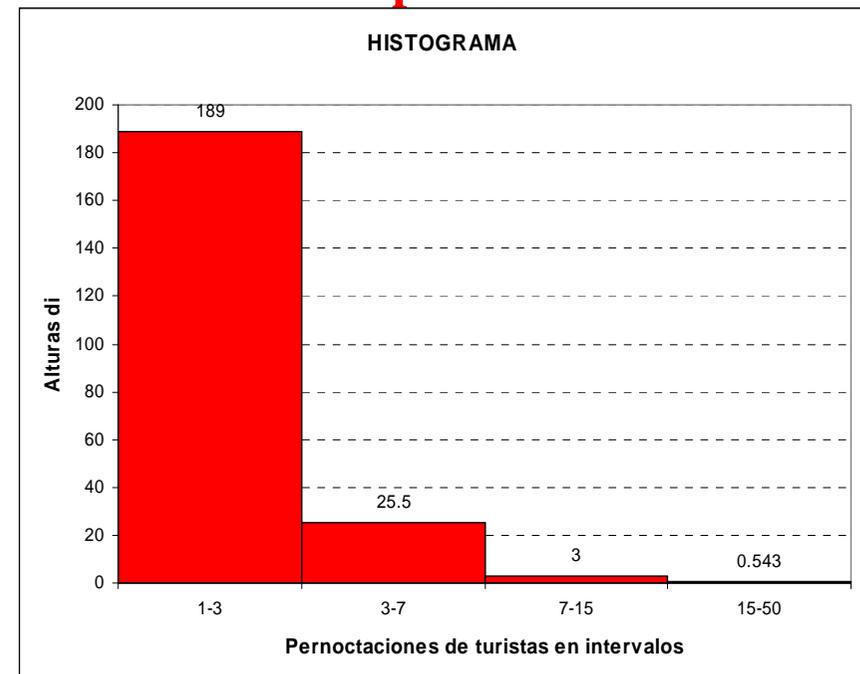
Valores observados: se han agrupado en intervalos de distinta amplitud  $c_i$

Intervalos $L_{i-1} - L_i$	Marcas de clase $x_i$	Frecuencias				Amplitud Intervalo $c_i = L_i - L_{i-1}$	Alturas o densidad de frecuencia $d_i$ $d_i = n_i / c_i$
		absolutas		relativas			
		$n_i$	$N_i$	$f_i = n_i / N$	$F_i = N_i / N$		
1-3	2	378	378	0,72	0,72	2	189
3-7	5	102	480	0,19	0,92	4	25,5
7-15	11	25	505	0,05	0,96	8	3,125
15-50	32,5	19	524	0,04	1	35	0,543
	$\Sigma$	N=524		1			

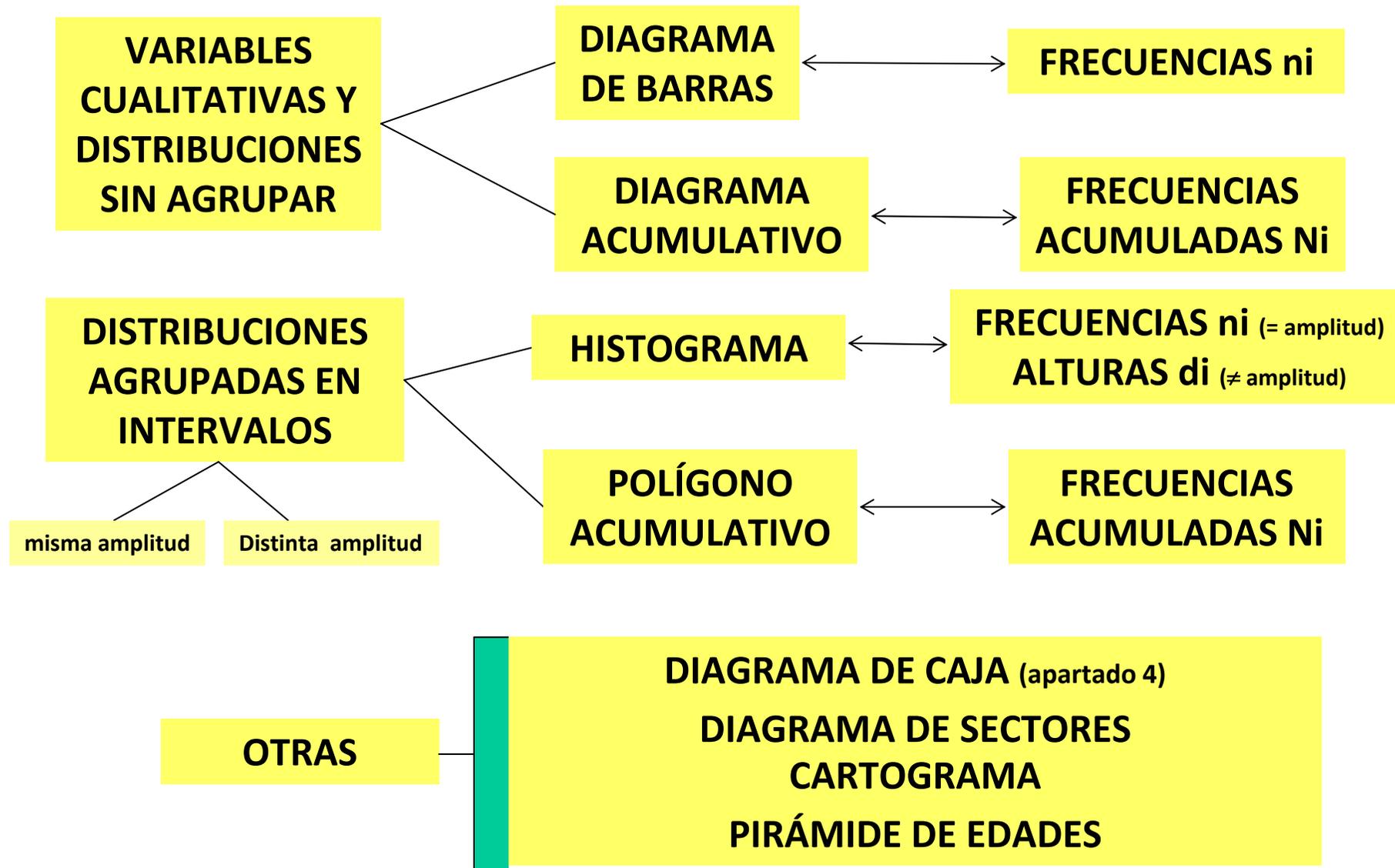
### Representación Gráfica

### Tabla de frecuencias con datos agrupados en intervalos

Fuente: Elaboración propia en base a [http://www.euskadi.net/turismo/estudios/docu5\\_c.htm](http://www.euskadi.net/turismo/estudios/docu5_c.htm), "Encuesta de comportamientos y tipologías de visitantes en Euskadi".



# REPRESENTACIONES GRÁFICAS



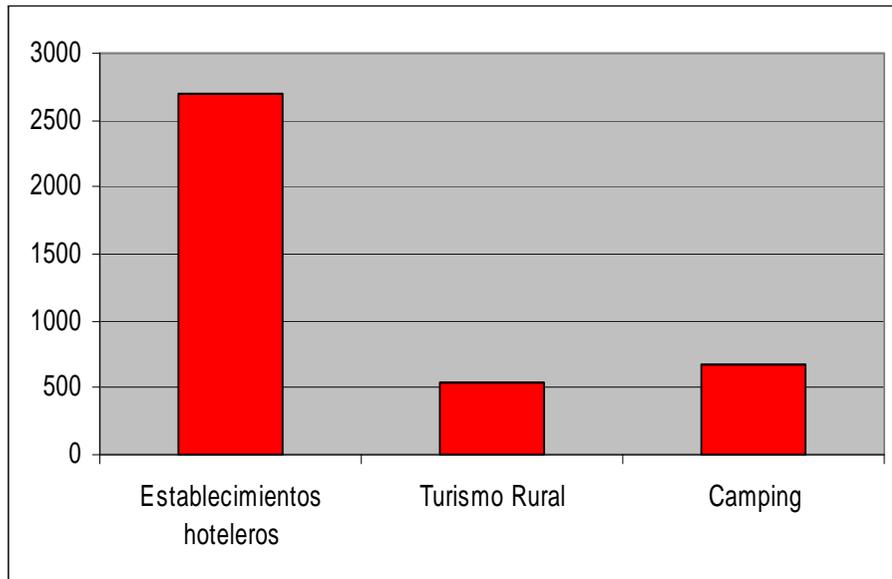
# Ejemplo 1 de Variable cualitativa nominal:

La siguiente tabla muestra los turistas que pernoctan en establecimientos públicos por tipo de establecimiento de pernoctación.

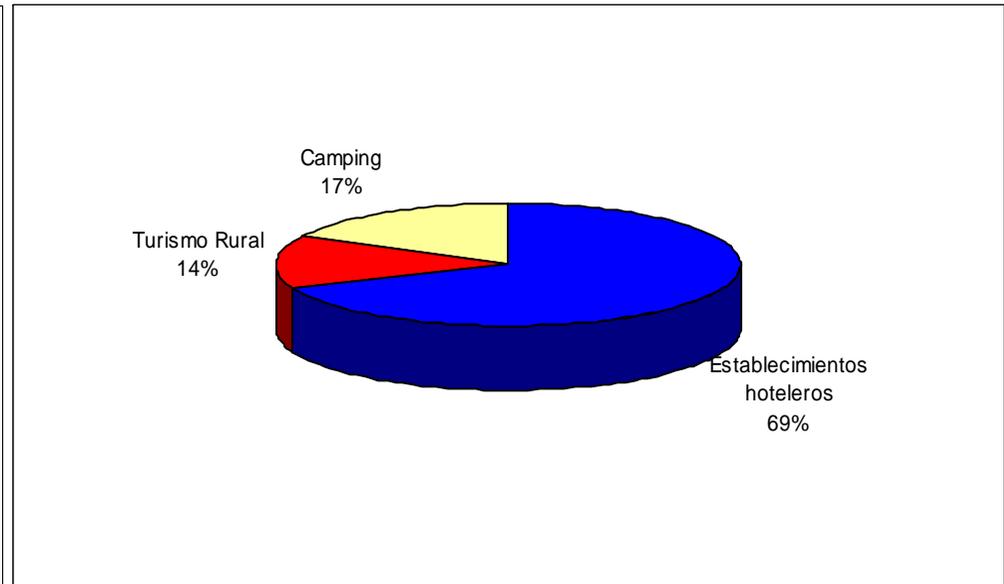
Establecimientos públicos de pernoctación	Nº de establecimientos
Establecimientos hoteleros	2697
Turismo Rural	535
Camping	678

Fuente: [http://www.euskadi.net/turismo/estudios/docu5\\_c.htm](http://www.euskadi.net/turismo/estudios/docu5_c.htm), "Encuesta de comportamientos y tipologías de visitantes en Euskadi".

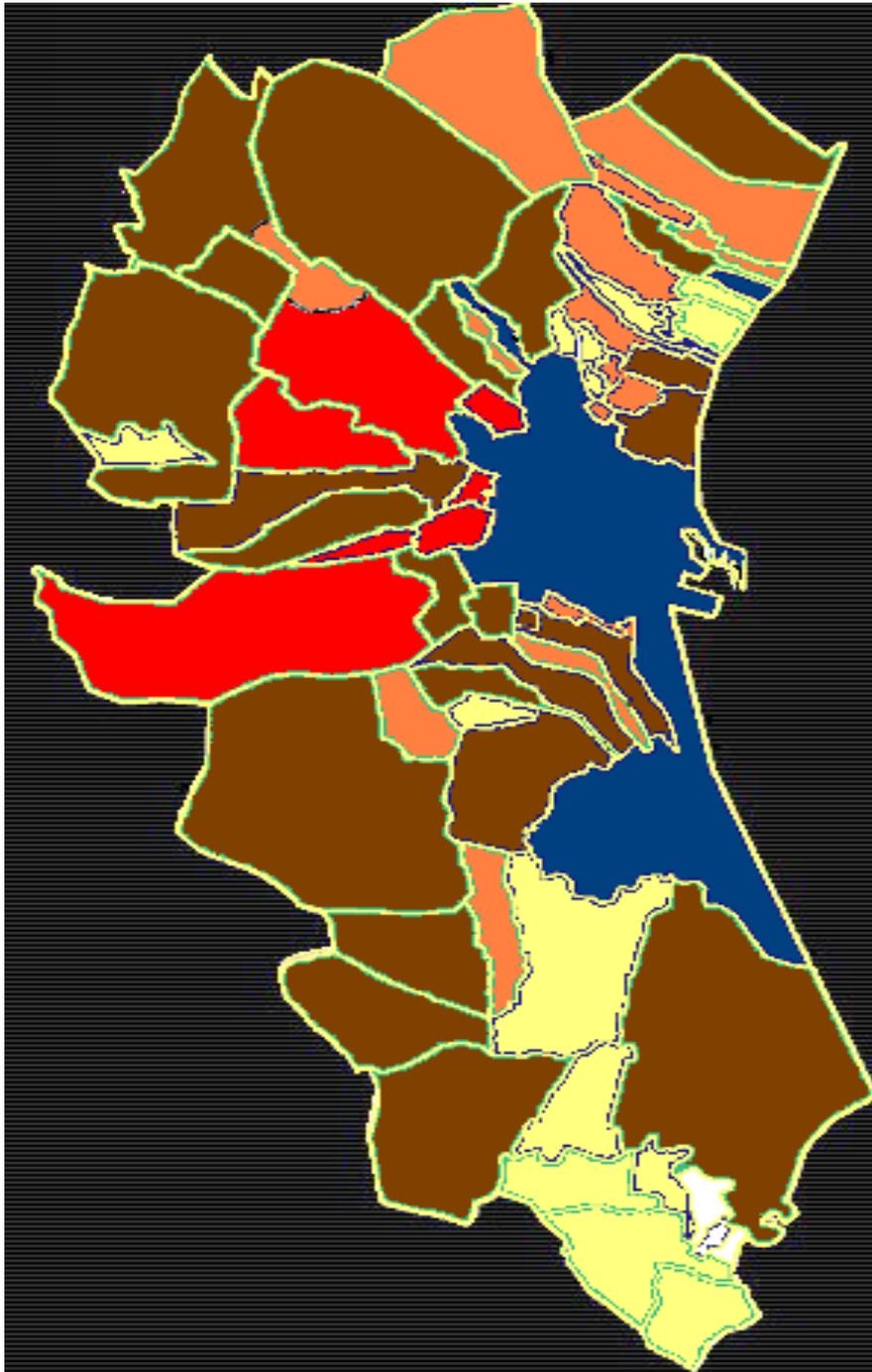
**DIAGRAMA DE BARRAS**



**DIAGRAMA DE SECTORES**



**Cartograma de la Población 2008 de los municipios del Área Metropolitana de Valencia agrupada en intervalos**



Intervalos de Población:

- 0 - 5.000
- 5.000-10.000
- 10.000-30.000
- 30.000-100.000
- Valencia

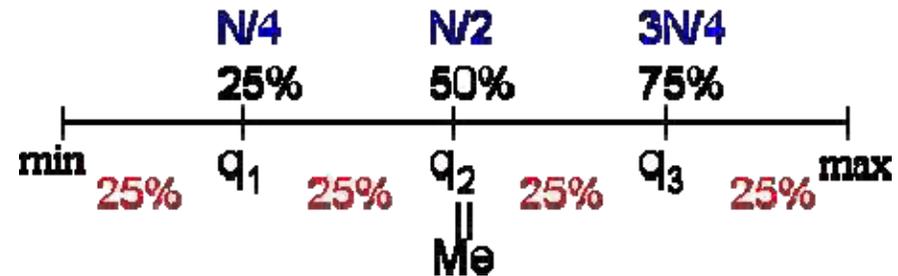
## 2.- Medidas de posición

$\bar{X}$       Me      Mo

- De Tendencia central: Media, Mediana, Moda
- Cuantiles: **Cuartiles**, Deciles, Percentiles

3 valores que dividen a la  
distribución en 4 partes iguales

$q_1, q_2, q_3$



Rango intercuartílico  $RIC = q_3 - q_1$

Entre  $q_1$  y  $q_3$  se encuentra el 50% de las observaciones que están en la zona central de la distribución.

### 3a.- Medidas de dispersión

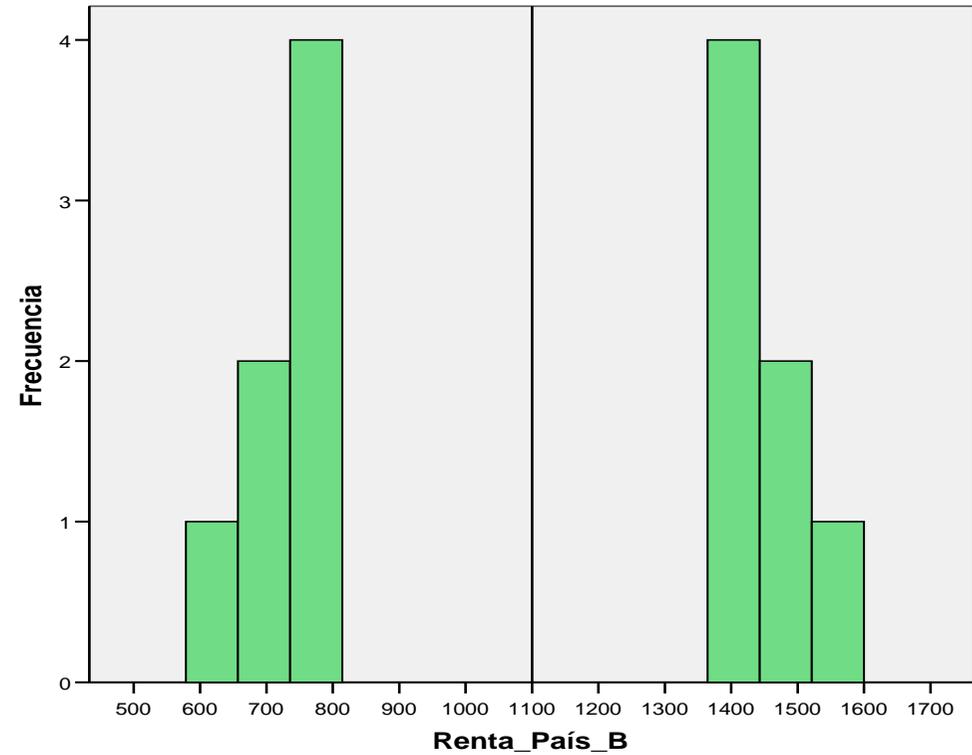
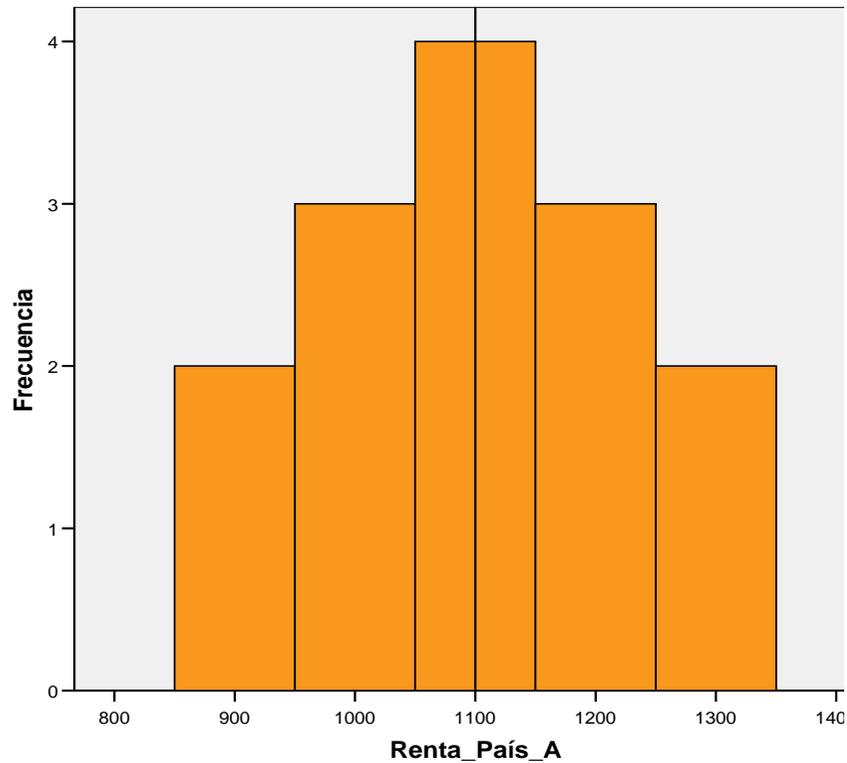
- Absolutas: Varianza  $S^2$ , Desviación Típica  $S = \sqrt{S^2}$
- Relativas: Coeficiente de variación de Pearson  $g_0(x) = \frac{S}{\bar{X}}$

### 3b.- Medidas de forma o perfil

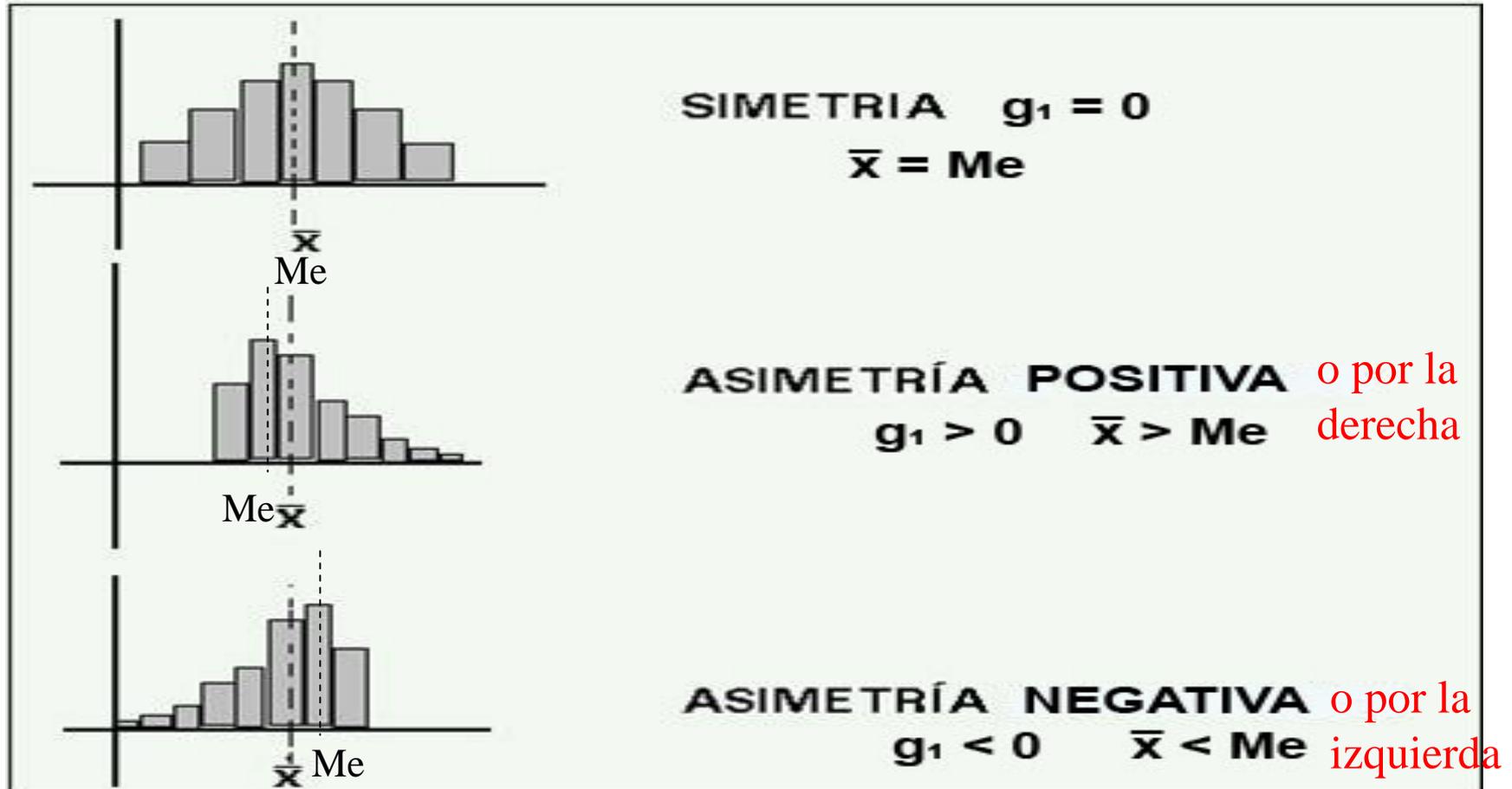
- Medidas de simetría-asimetría: Coeficiente de asimetría  $g_1(x)$
- Medidas de apuntamiento-aplastamiento: Coeficiente de Curtosis  $g_2(x)$

# EJEMPLO DE REPRESENTATIVIDAD DE LA MEDIA

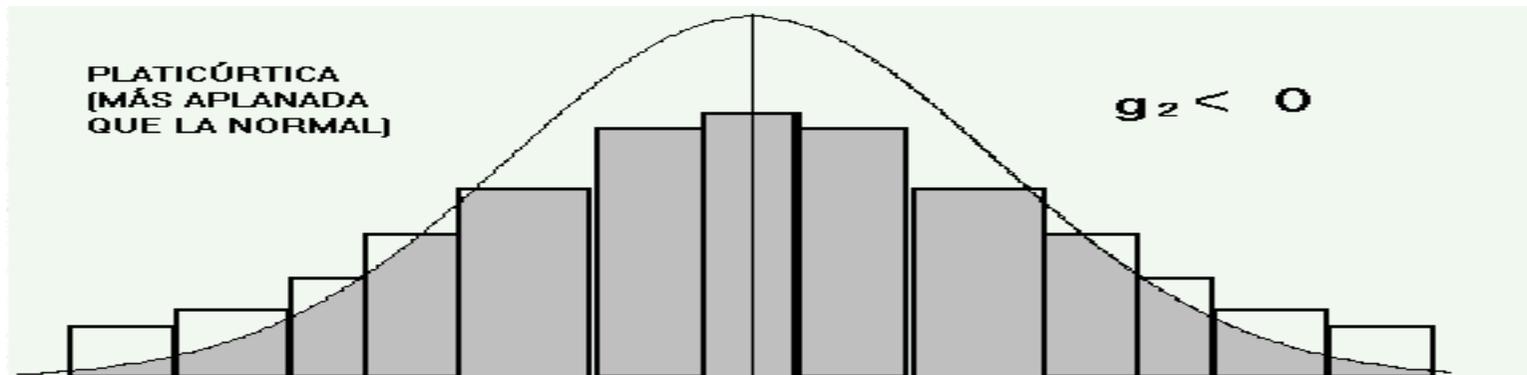
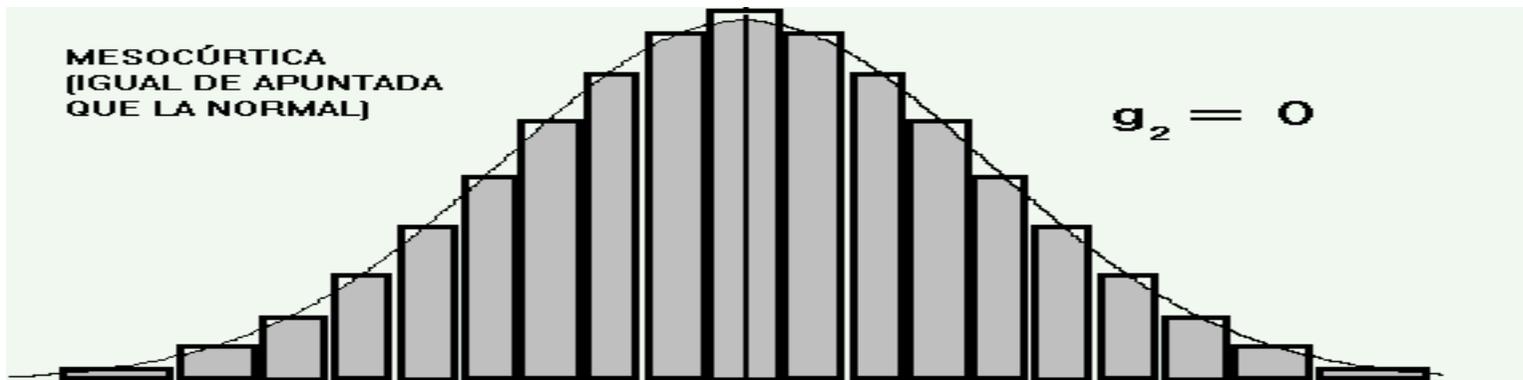
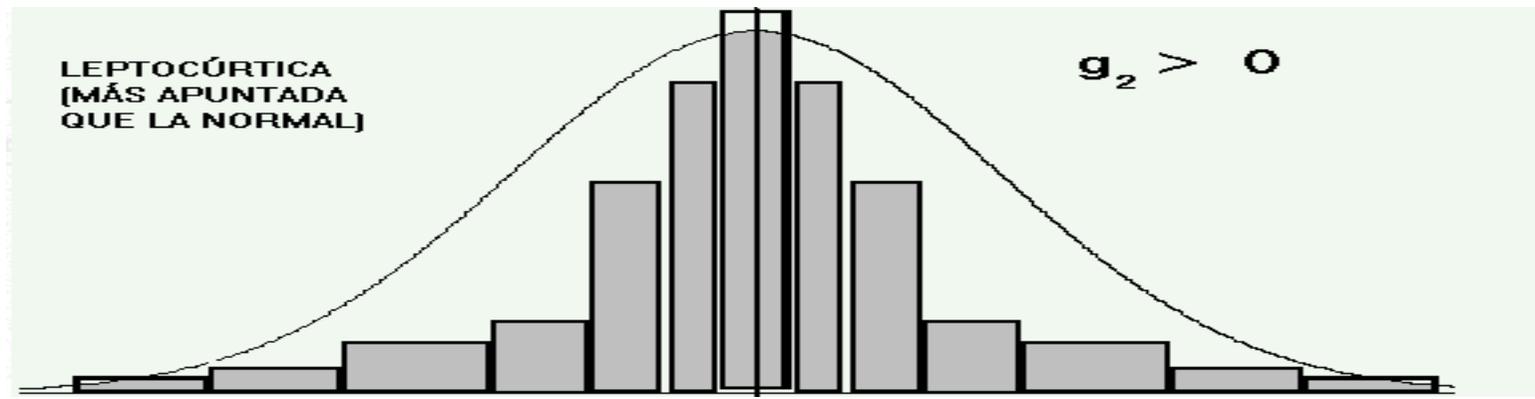
Compara la media y la dispersión de la distribución de la renta de estos dos países:



**MEDIDAS DE ASIMETRÍA**  
**(COEFICIENTE DE ASIMETRÍA  $g_1$ )**

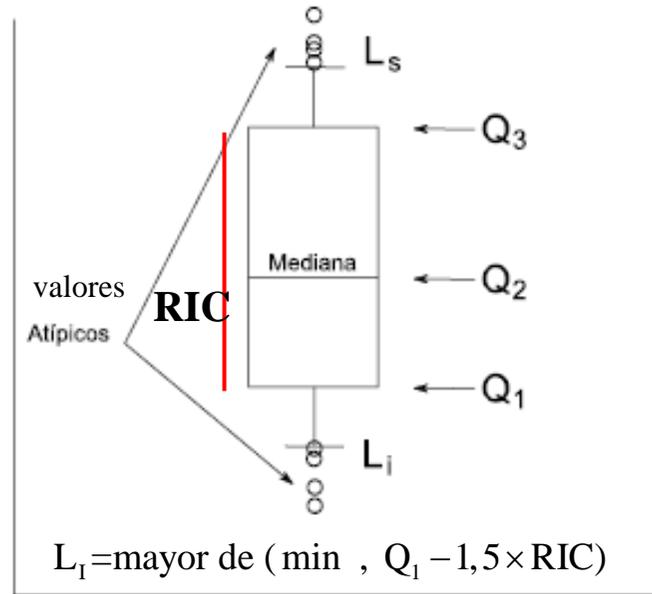


## MEDIDAS DE CURTOSIS.( COEFICIENTE DE CURTOSIS)



# 4.- VALORES ATÍPICOS Y DIAGRAMA DE CAJA

$$L_s = \text{menor de } (\max, Q_3 + 1,5 \times \text{RIC})$$



$$L_1 = \text{mayor de } (\min, Q_1 - 1,5 \times \text{RIC})$$

$$\text{RIC} = Q_3 - Q_1 \text{ rango intercuartílico}$$

Gráfico basado en los 3 cuartiles que está compuesto por un rectángulo, la "caja", y dos brazos, los "bigotes".

- Proporciona información sobre la simetría de la distribución; si la mediana no está en el centro del rectángulo, la distribución no es simétrica.
- Sirve para detectar la presencia de valores atípicos o outliers en la distribución.

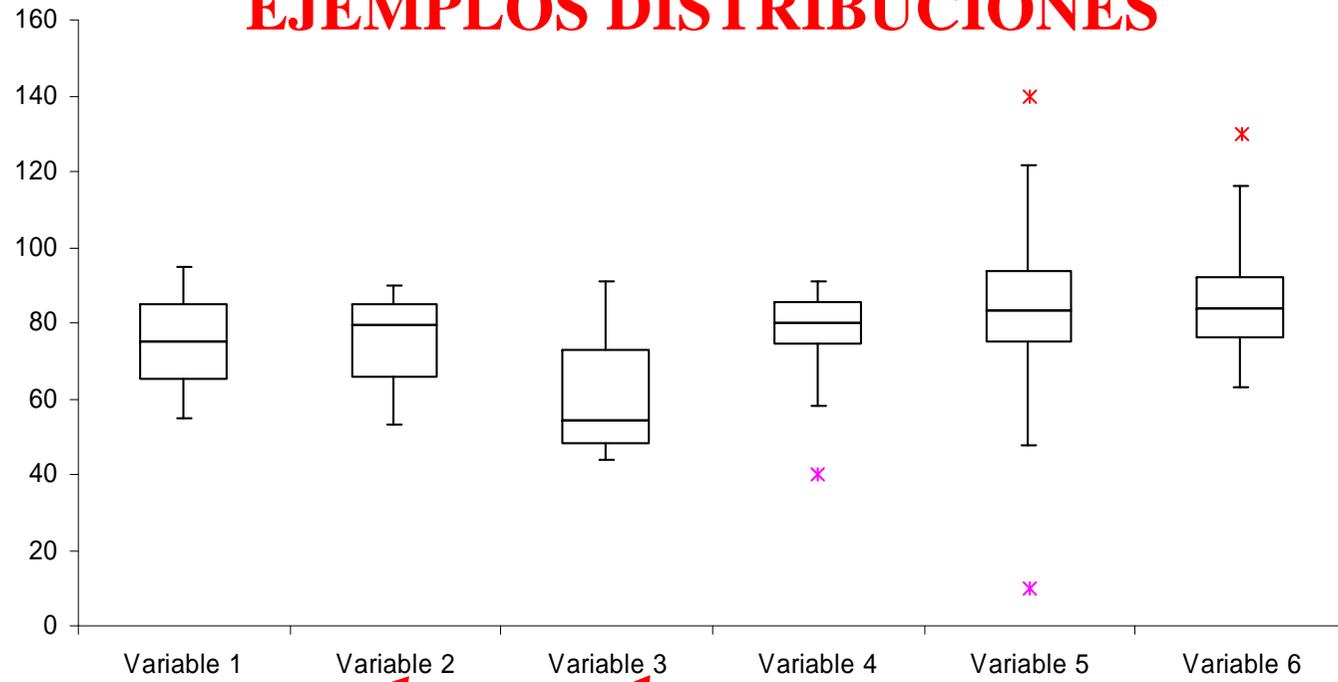
• Un valor será considerado valor atípico leve si:

$$o \left\{ \begin{array}{l} \text{valor} < Q_1 - 1,5 \text{ RIC} \\ \text{valor} > Q_3 + 1,5 \text{ RIC} \end{array} \right.$$

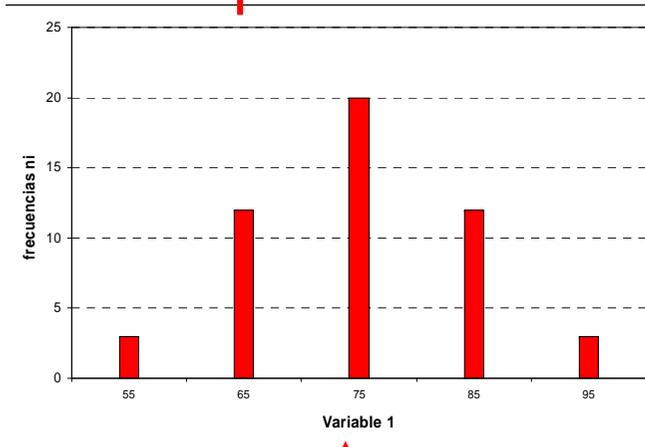
• Un valor será considerado valor atípico extremo si:

$$o \left\{ \begin{array}{l} \text{valor} < Q_1 - 3 \text{ RIC} \\ \text{valor} > Q_3 + 3 \text{ RIC} \end{array} \right.$$

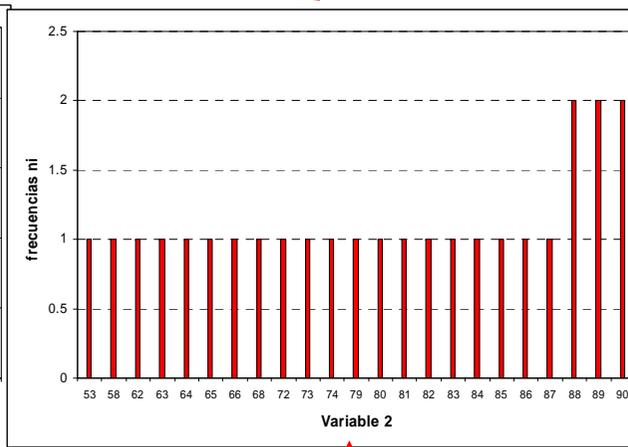
# EJEMPLOS DISTRIBUCIONES



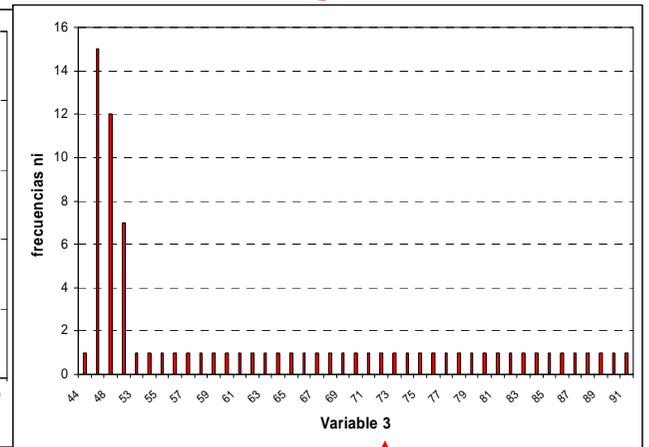
\* Min. Outlier \* Max Outlier



Coef. simetría  $g_1=0$   
**Distribución Simétrica**



Coef. simetría  $g_1=-0,52<0$   
**Distrib. Asimétrica por la izquierda**



Coef. simetría  $g_1=0,61>0$   
**Distrib. Asimétrica por la derecha**

# 5.- MEDIDAS DE CONCENTRACIÓN

Curva de Lorenz

Indice de Gini (IG)

Colectivo: 10 trabajadores de una empresa turística.

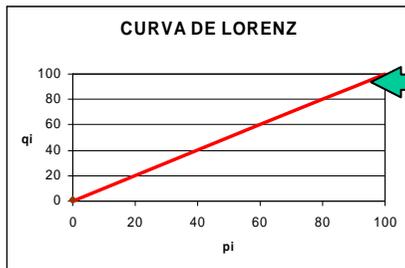
Variable X: salario mensual en euros.

xi	ni	xi ni	frec. relativas		% acumulativo	
			fi en %	xi ni en %	pi	qi
600	6	3600	60	42.86	60	42.86
900	2	1800	20	21.43	80	64.29
1200	1	1200	10	14.29	90	78.57
1800	1	1800	10	21.43	100	100
	10	<b>8400</b>				

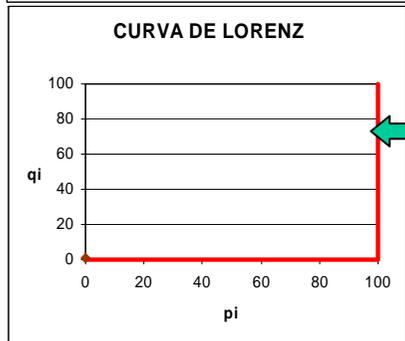
N

Interpretación pi y qi

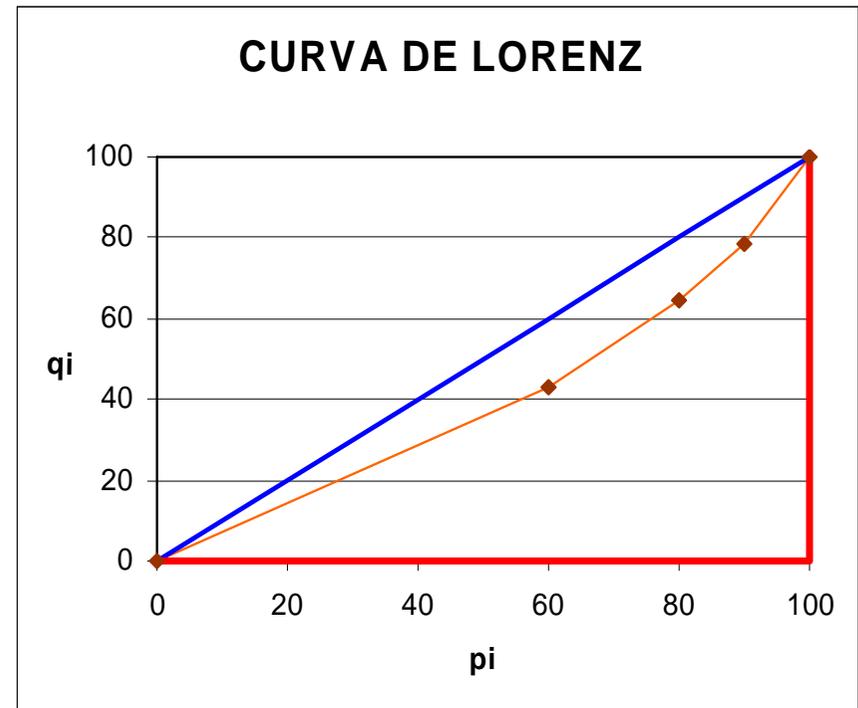
Objetivo: analizar grado de concentración en el total de salarios pagados por la empresa



• **Concentración mínima:**  
pi=qi



• **Concentración máxima:** todos qi=0 menos el último, que es 100



## INDICE DE GINI

$p_i$	$q_i$	$p_i - q_i$
60	42.86	17.14
80	64.29	15.71
90	78.57	11.43
230		44.29

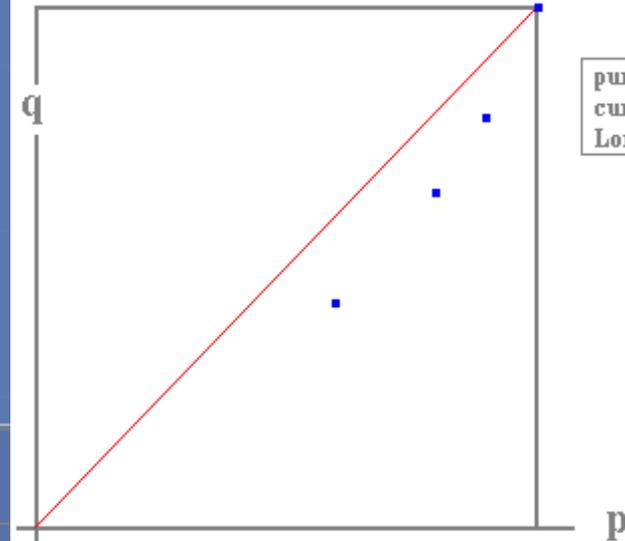
$$IG = \frac{\sum_{i=1}^{I-1} (p_i - q_i)}{\sum_{i=1}^{I-1} p_i} = \frac{44.29}{230} = 0.193$$

- **Concentración mínima:**  $p_i = q_i \Rightarrow IG = 0$
  - **Concentración máxima:** todos  $q_i = 0 \Rightarrow IG = 1$
- }  $\Rightarrow$  • **Casos intermedios:**  
 $0 \leq IG \leq 1$

# Índice de GINI

Pegar o escribir los valores (los de X ordenados)

X	frecuencias
600	6
900	2
1200	1
1800	1



índice de GINI

Número de decimales; por defecto...

 Juan Mtnez. de Lejarza

Valores	frecuencia	frecuencia acumulada	monto	monto acumulado	p	q
600	6	6	3600	3600	60	42.857
900	2	8	1800	5400	80	64.286
1200	1	9	1200	6600	90	78.571
1800	1	10	1800	8400	100	100

## MEDIA CONJUNTA DE VARIOS CONJUNTOS DE DATOS

PARA CALCULAR LA MEDIA CONJUNTA DE VARIOS CONJUNTOS DE DATOS ES NECESARIO APLICAR ESTA FORMULA DE LA MEDIA PONDERADA:

Conjunto de datos	1°	2°	.....	k°	Global (agregado de los k conjuntos)
N° de observaciones de cada conjunto de datos	$N_1$	$N_2$	.....	$N_k$	$N_1 + N_2 + \dots + N_k$
Media de cada conjunto de datos	$\bar{X}_1$	$\bar{X}_2$	.....	$\bar{X}_k$	$\bar{X} = \frac{\bar{X}_1 N_1 + \bar{X}_2 N_2 + \dots + \bar{X}_k N_k}{N_1 + N_2 + \dots + N_k}$

La media de la agrupación de 2 o más conjuntos de datos es la media de las medias de los distintos conjuntos, pero cada una ponderada por el n° de observaciones de cada conjunto de datos.

## EJEMPLO

A partir de la siguiente información, comprueba que la media de la Población 2009 del conjunto de los 183 municipios de la provincia de Valencia es 13849,22 habitantes

Conjunto de datos	Municipios Turísticos	Municipios No Turísticos
Nº de observaciones de cada conjunto de datos	49	134
Media de la Población 2009 de cada conjunto de datos	$\bar{x}_1 = 31180,12$	$\bar{x}_2 = 7511,81$

# TIPIFICACIÓN DE VARIABLES

- Se trata de un procedimiento que facilita la comparación entre observaciones provenientes de dos conjuntos de datos distintos cuando éstos tienen medias y desviaciones típicas diferentes.

Para Tipificar:  $Z = \frac{X - \bar{X}}{S}$  (Z es la variable X tipificada)

- Un valor que se ha tipificado, indica el número de desviaciones típicas que ese valor está por encima o por debajo de la media.
- Sirve para comparar la posición relativa de dos observaciones que provienen cada una de una distribución distinta (ver ejemplo). 
- La nueva variable tipificada Z que ha surgido de aplicar esta transformación a una variable X tiene una propiedad especial: su media es cero y su desviación típica es 1.

## EJEMPLO DE TIPIFICACIÓN Y DE DISPERSIÓN

En la asignatura de Incorporación, la nota media de los estudiantes ha sido de 7 con una desviación típica de 2. En la asignatura de Estadística la nota media de los estudiantes ha sido de 5 con una desviación típica de 1.

- a) Un estudiante ha obtenido una nota en Incorporación de 6, mientras que su nota en Estadística ha sido de 5,5. ¿En qué asignatura ha obtenido el estudiante un resultado relativamente mejor?
- b) ¿En qué asignatura la dispersión relativa es menor?

# Más información sobre este tema en:

- PARRA, E; CALERO, F.J.: Estadística para Turismo. Ed. McGraw-Hill, Madrid, 2007. Capítulo 4.
- ESTEBAN, J.; y otros.: “Estadística Descriptiva y nociones de Probabilidad”, Ed. Thomson, segunda impresión 2006. Capítulo 2.
- MONTIEL, A.M.; RIUS, F.; BARÓN F.J.: *Elementos básicos de Estadística Económica y Empresarial*. Ed. Prentice Hall, Madrid, 1997. Capítulos 3 y 4.
- RONQUILLO, A: Estadística Aplicada al Sector Turístico, Ed Ramón Areces, Madrid, 1997. Capítulos 2, 3, 4 y 5.
-  <http://www.uv.es/ceaces/descriptiva/simple.htm>
- [http://webpersonal.uma.es/de/J\\_SANCHEZ/Capitulo2.PDF](http://webpersonal.uma.es/de/J_SANCHEZ/Capitulo2.PDF)