

ESTADÍSTICA



GRAU EN TURISME

TEMA 2: ANÁLISI DE DADES TURÍSTIQUES UNIDIMENSIONALS

Prof: Rosario Martínez
Versió en valencià: Prof. Rosa Yagüe



TEMA 2: ANALISI DE DADES TURÍSTIQUES UNIDIMENSIONALS

1. Presentació de les dades: distribucions de freqüències, taules estadístiques i representacions gràfiques.
2. Mesures de posició.
3. Mesures de dispersió i de forma.
4. Valors atípics i diagrama de caixa
5. Mesures de concentració.

ETAPES DE L 'ANÁLISI DESCRIPTIVA DE LES DADES TURÍSTIQUES UNIDIMENSIONALS.

- **1a ETAPA:** Presentació de les dades
 - Classificar i ordenar les dades en una Taula estadística o de Freqüències.
 - Representar-los Gràficament (diagrama de barres, histograma, etc.).
- **2a ETAPA:** Resumir, sintetitzar i representar tota la distribució de freqüències per mitjà d'una sèrie de mesures que intenten arregar els aspectes més rellevants de la distribució. Interessa calcular i interpretar el valor d'eixes mesures
 - Mesures de posició,
 - Mesures de dispersió
 - Mesures de forma o perfil.



Font: www.gaturro.com

Reproduït amb permís de l'Autor: Nik

1.- Presentació de les dades: distribucions de freqüències, taules estadístiques i representacions gràfiques.

TIPUS DE DISTRIBUCIONS DE FREQUÈNCIES UNIDIMENSIONALS.

- **Distribucions de freqüències unitàries**

Aquelles en què cada valor de la variable apareix una sola vegada, no hi ha repeticions, per això les freqüències són totes 1.

- **Distribucions de freqüències no unitàries**

Hi ha valors repetits. Es classifiquen en:

	Nombre valors diferents de la variable	pocs	Molts
Nombre d'observacions			
Moltes		1) De dades sense agrupar	2) De dades agrupades en intervals

a) Intervals de mateixa amplària

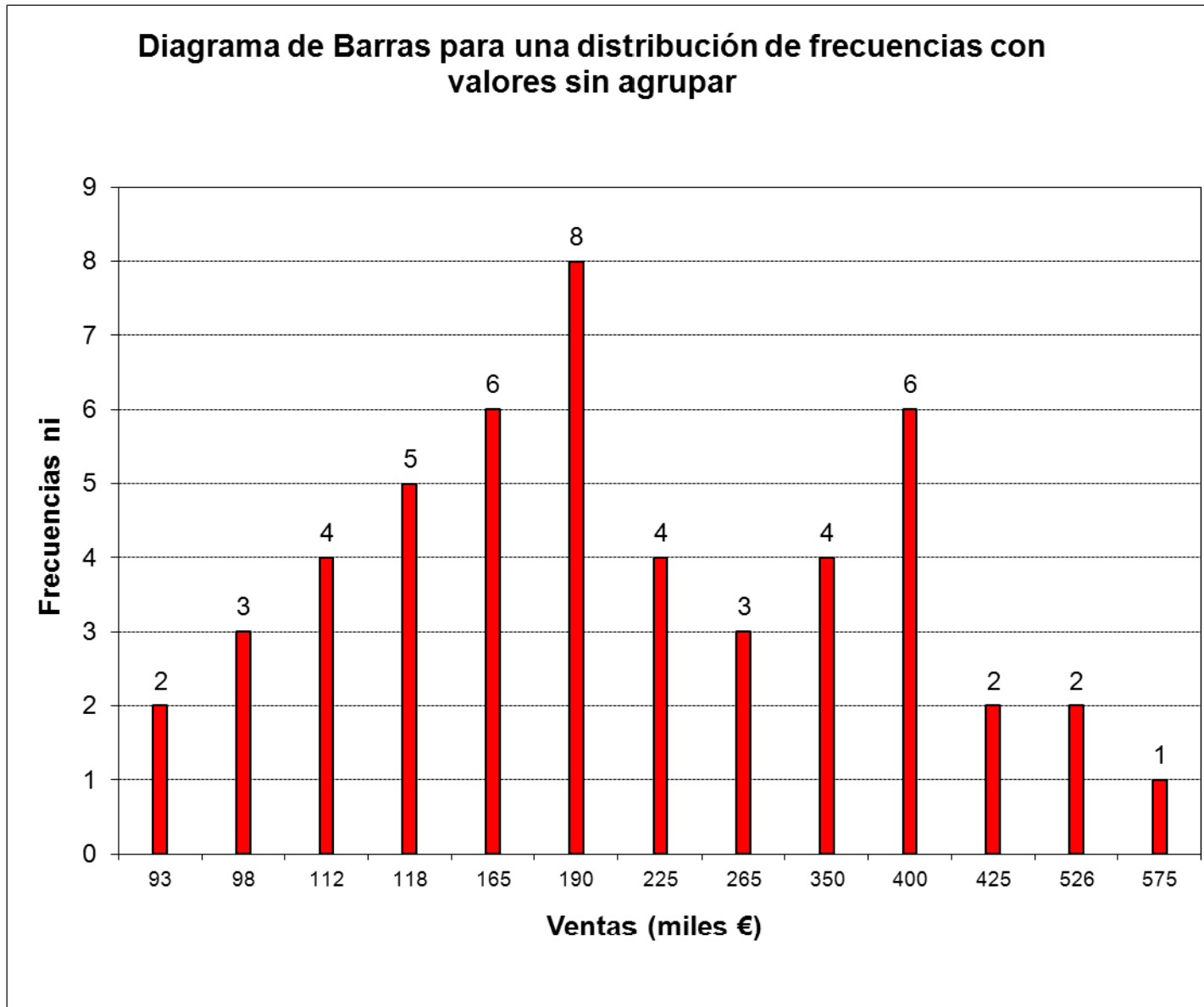
b) Intervals de diferent amplària

1) Exemple de Distribució de Freqüències amb Dades sense Agrupar

Taula de freqüències o Taula estadística amb els diferents tipus de freqüències:

Vendes (milers €) d'una cadena de agències de viatges	Freqüències			
	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	50	1
Total. N	50	1		

Representació Gráfica:



2a) Distribució de Freqüències amb Dades Agrupades en Intervalos mateixa amplària

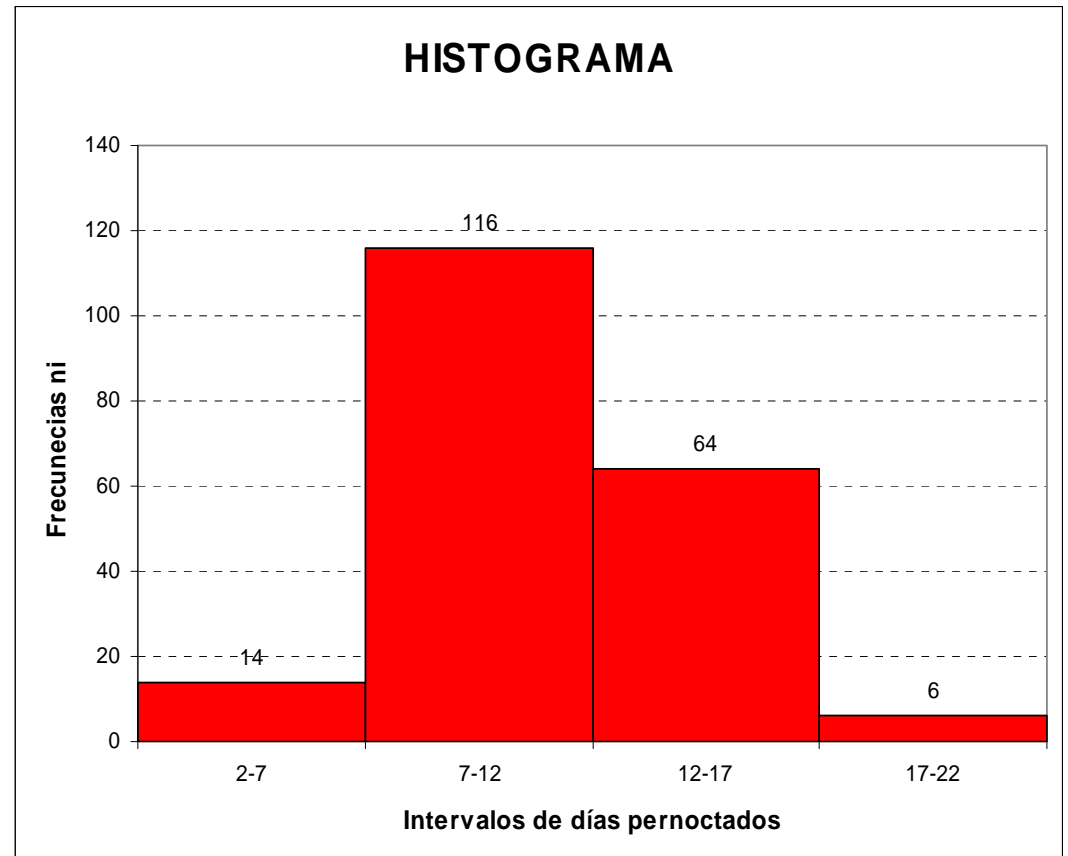
Colectiu: 200 turistes que visitaren la Comunitat Valenciana i l' estada no superà els 22 dies. **N = 200**

Variable X: dies pernoctats, agrupats en intervalos

**Taula de freqüències
amb dades agrupades en intervalos
de mateixa amplària:**

Representació Gráfica:

intervalos	marcas de classe x_i	freqüències absolutes		freqüències relatives	
		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
		N=200		1	



2b) Distribució de Freqüències amb Dades Agrupades en Intervalos diferent amplària

Colectiu: 524 turistes

$N = 524$

Variable **X**: dies pernoctats en allotjaments de turisme rural (exemple 4 tema 1)

Valors observats: s'han agrupat en intervalos de distinta amplària c_i

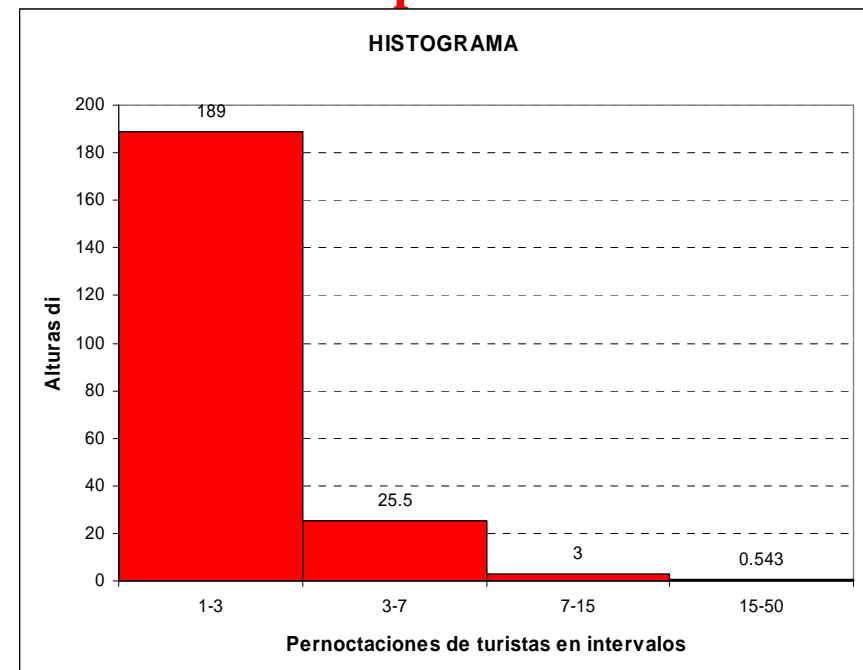
Intervalos	Marcas de classe	Freqüències				Amplitud Interval	Altures
		absolutes		relatives			
		n_i	N_i	$f_i = n_i/N$	$F_i = N_i/N$		
$L_{i-1} - L_i$	x_i					$c_i = L_i - L_{i-1}$	$d_i = n_i/c_i$
1-3	2	378	378	0,72	0,72	2	189
3-7	5	102	480	0,19	0,92	4	25,5
7-15	11	25	505	0,05	0,96	8	3,125
15-50	32,5	19	524	0,04	1	35	0,543
	Σ	$N=524$		1			

o densitat de freqüència d_i

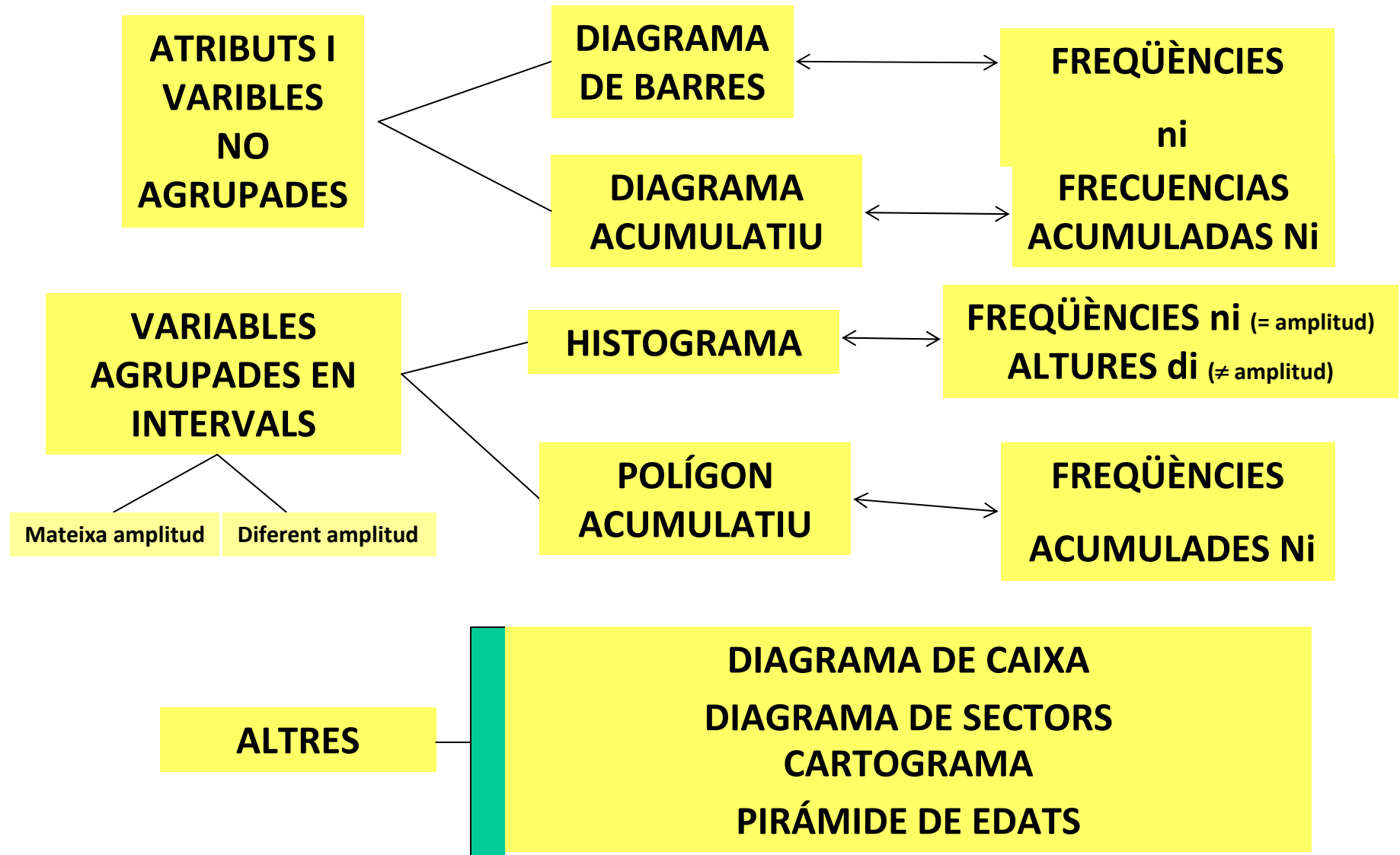
Representació Gràfica

Taula de freqüències amb dades agrupades en intervalos

Fuente: Elaboració pròpia a partir de http://www.euskadi.net/turismo/estudios/docu5_c.htm, "Encuesta de comportamientos y tipologías de visitantes en Euskadi".



REPRESENTACIONES GRÁFIQUES



Ejemplo 1 de Variable cualitativa nominal:

La sigüent taula mostra els turistes que pernocten en establiments públics por tipus d'establiment de pernoctació.

Establiment públic de pernoctació	Nombre d'establiments
Establiment hotelers	2697
Turisme Rural	535
Camping	678

Font: http://www.euskadi.net/turismo/estudios/docu5_c.htm, "Encuesta de comportamientos y tipologías de visitantes en Euskadi".

DIAGRAMA DE BARRES

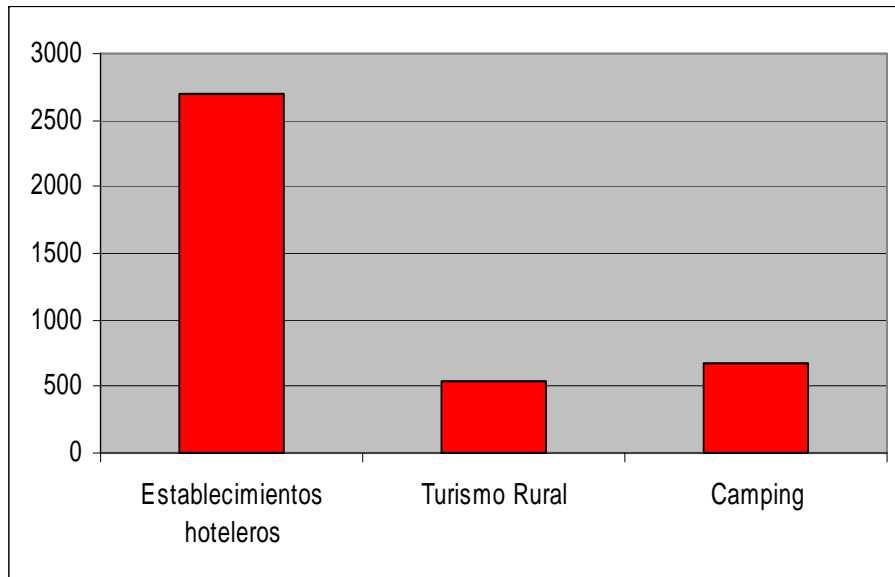
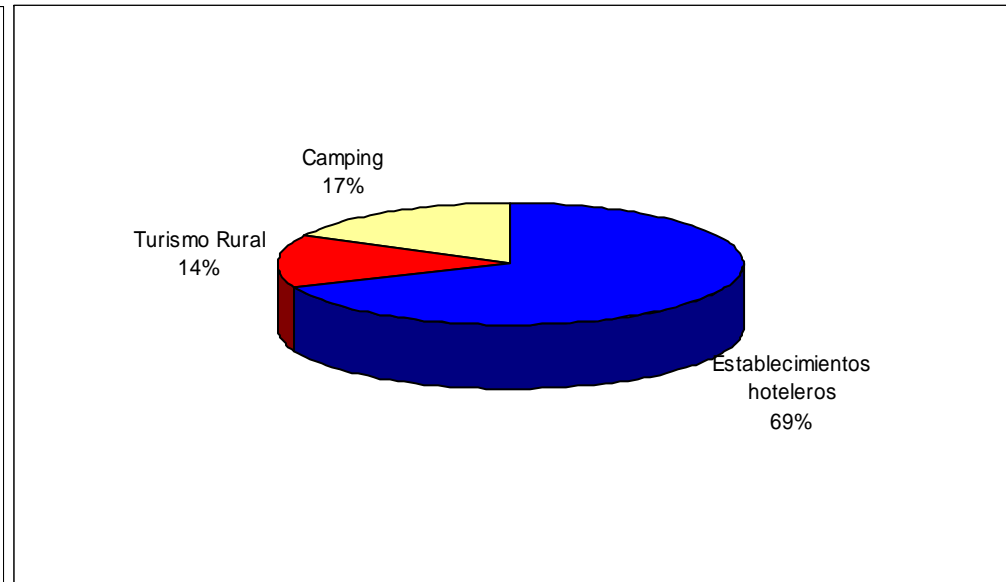
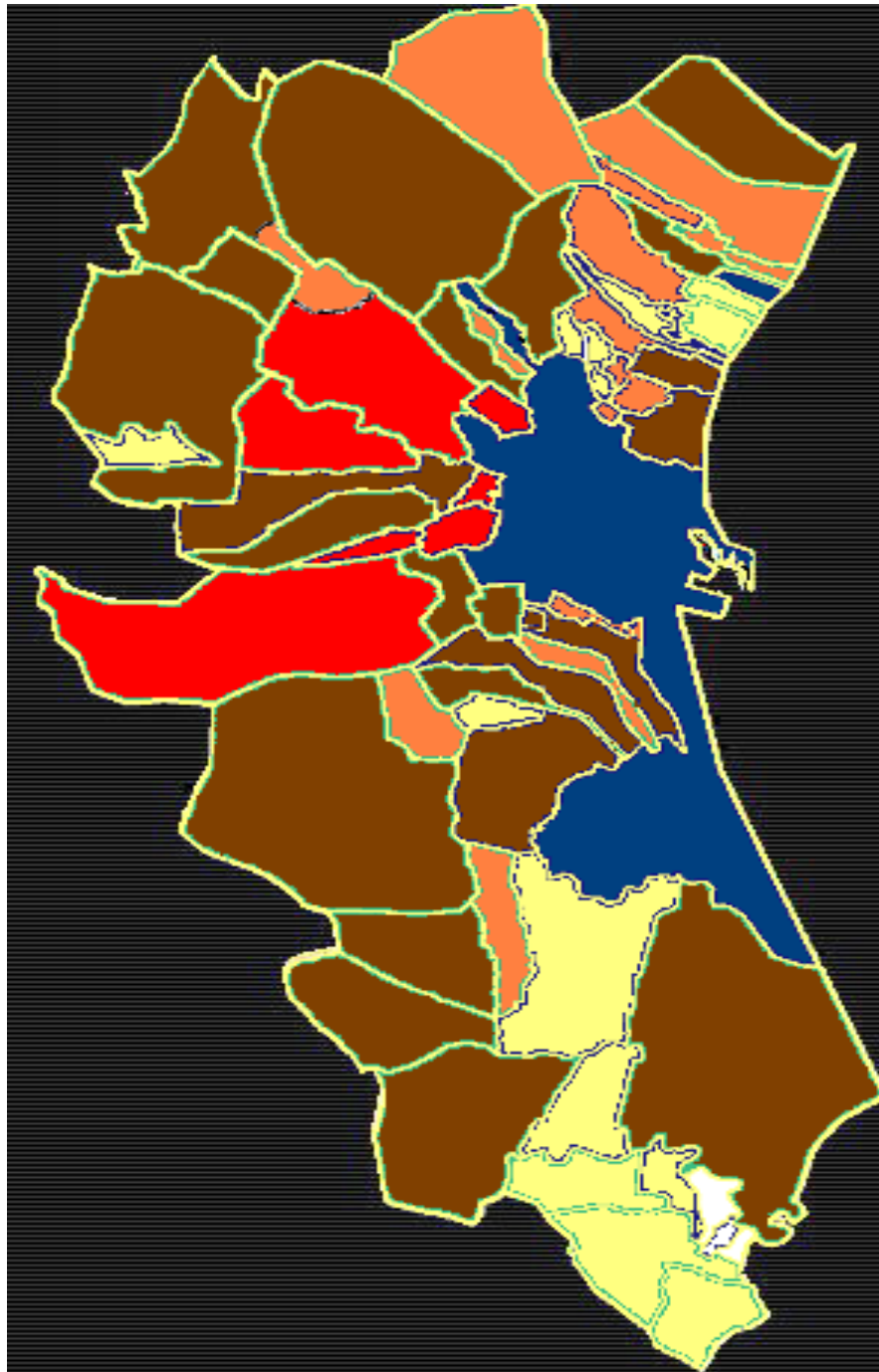


DIAGRAMA DE SECTORS



**Cartograma de la Població 2008 dels
municipis de l'Àrea Metropolitana de
València agrupada en intervals**



Intervals de Població:

- 0 - 5.000
- 5.000-10.000
- 10.000-30.000
- 30.000-100.000
- València

2.- Mesures de posició

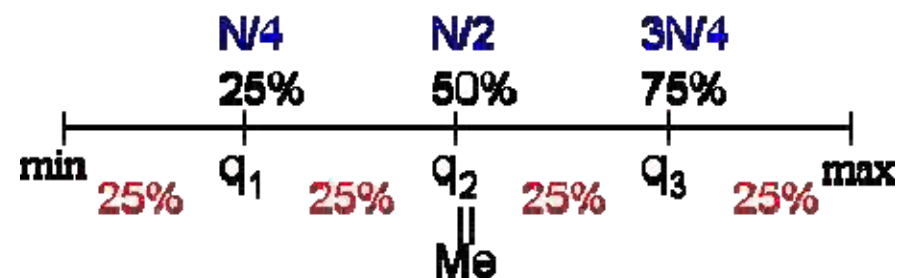
\bar{X} Me Mo

- De Tendència central: Mitjana, Mediana, Moda

- Quantils: **Quartils**, Decils, Percentils

3 valors que dividixen a la
distribució en 4 parts iguals

↓
 q_1, q_2, q_3



Rang interquartílic RIC = $q_3 - q_1$

Entre q_1 i q_3 es troba el 50% de les observacions que estan en la zona central de la distribució.

3a.- Mesures de dispersió

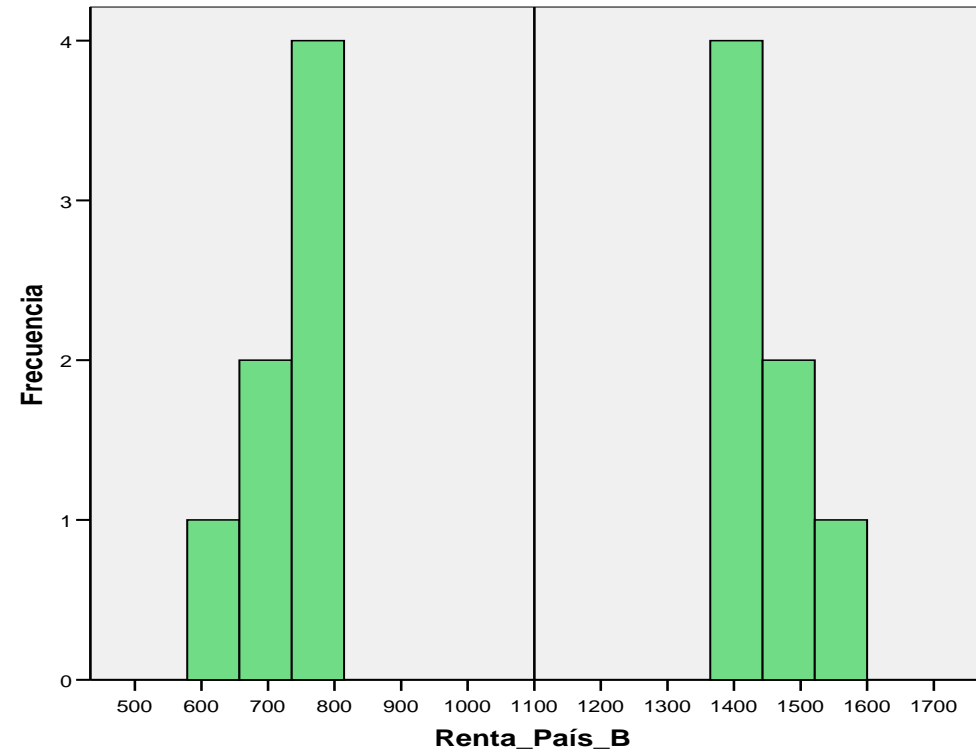
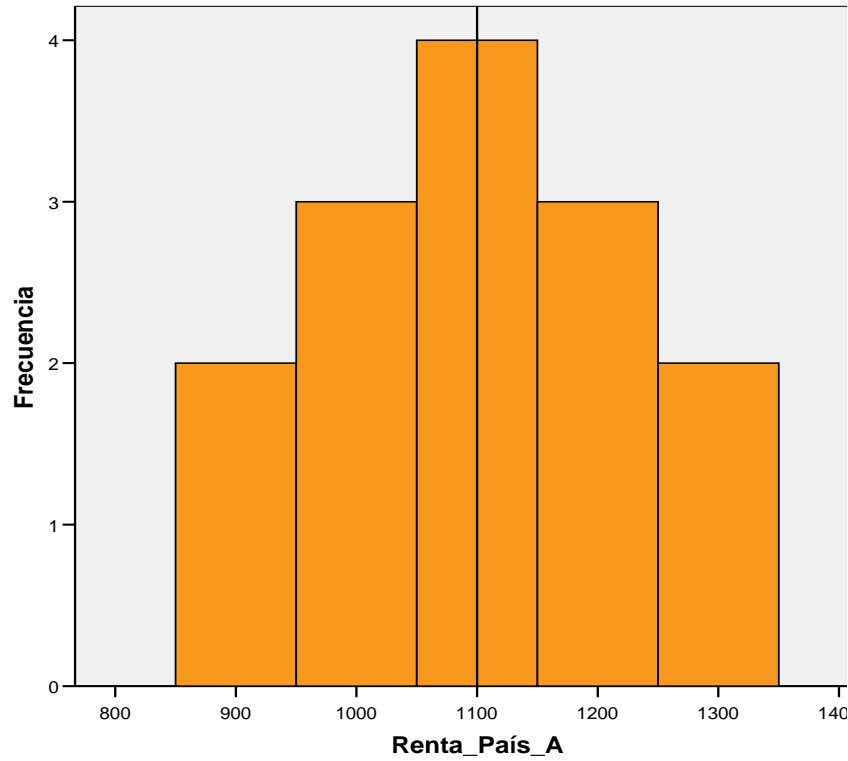
- Absolutes: Variància S^2 , Desviació Típica $S = \sqrt{S^2}$
- Relatives: Coeficient de variació de Pearson $g_0(x) = \frac{S}{\bar{X}}$

3b.- Mesures de forma o perfil

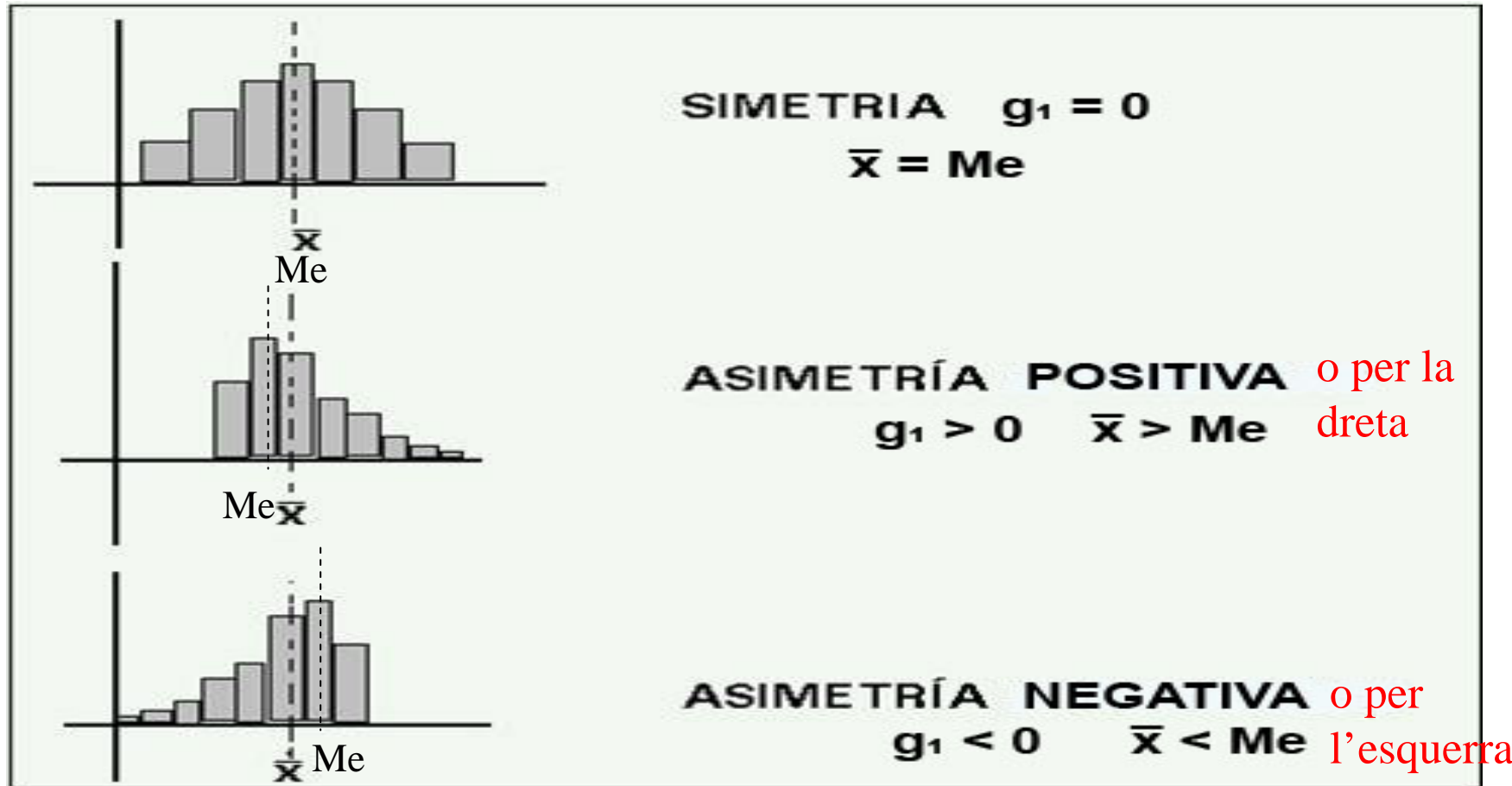
- Mesures de simetria-asimetria: Coeficient d'asimetria $g_1(x)$
- Mesures d'apuntament: Coeficient de Curtosis $g_2(x)$

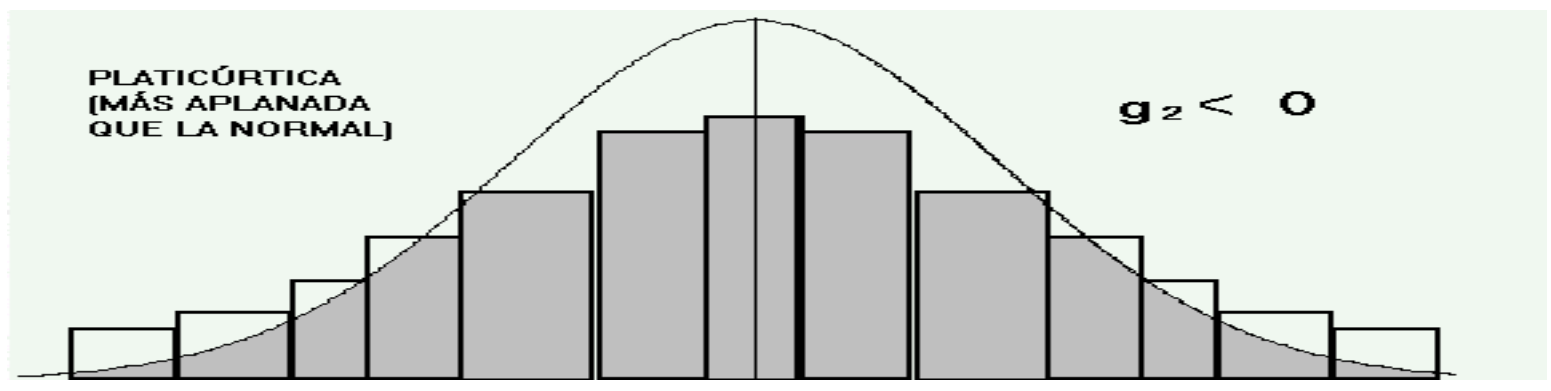
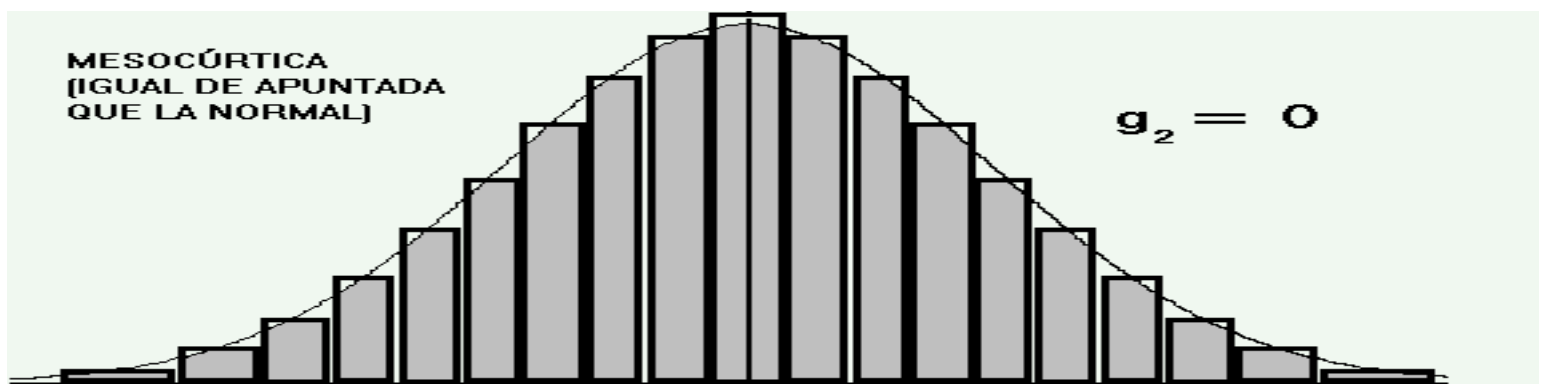
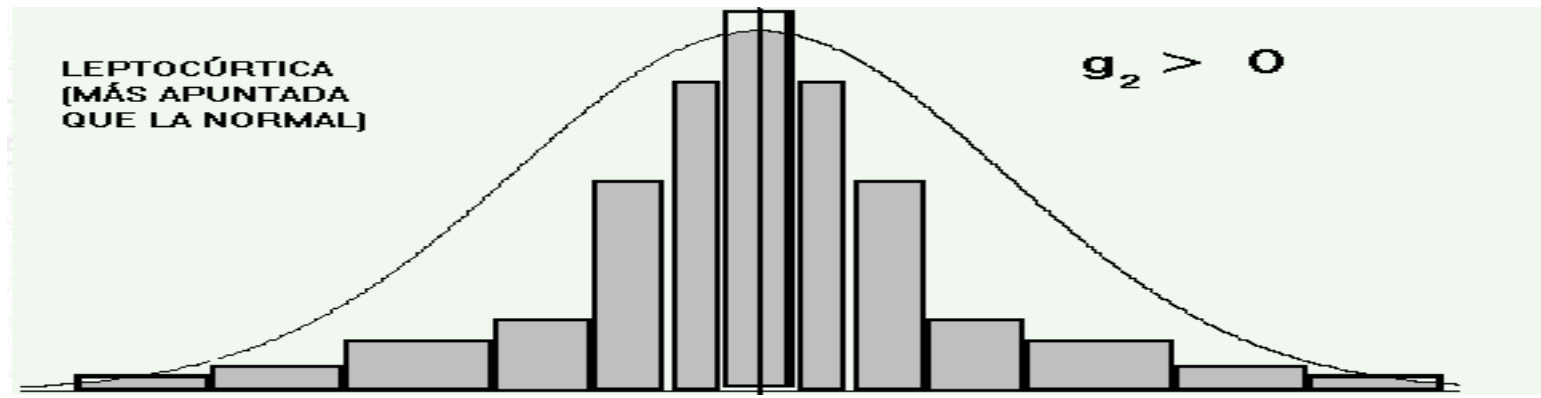
EXEMPLE DE REPRESENTATIVITAT DE LA MITJANA

Compara la mitjana i la dispersió de la distribució de la renda d'estos dos països:



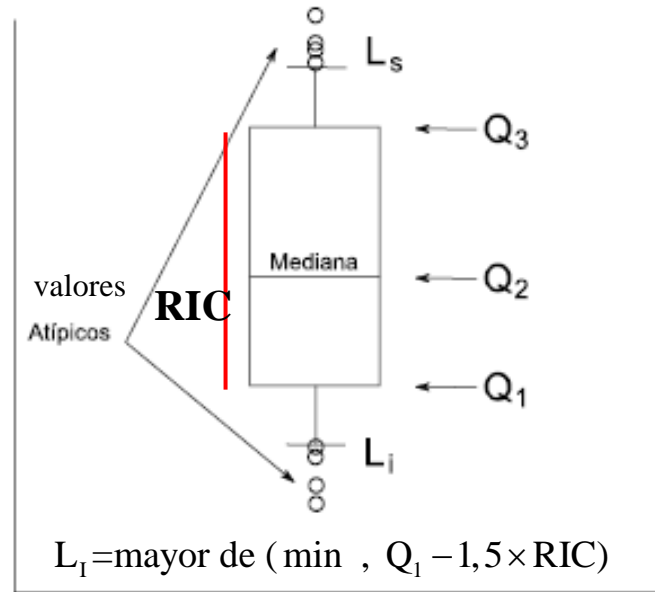
(COEFICIENT D' ASIMETRIA g_1)





4.- VALORS ATÍPICS I DIAGRAMA DE CAIXA

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



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

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

Gràfic basat en els 3 quartils que està compost per un rectangle, la "caixa", i dos braços, els "bigots".

- Proporciona informació sobre la simetria de la distribució; si la mediana no està en el centre del rectangle, la distribució no és simètrica.
- Útil per detectar la presència de valors atípics o outliers en la distribució.

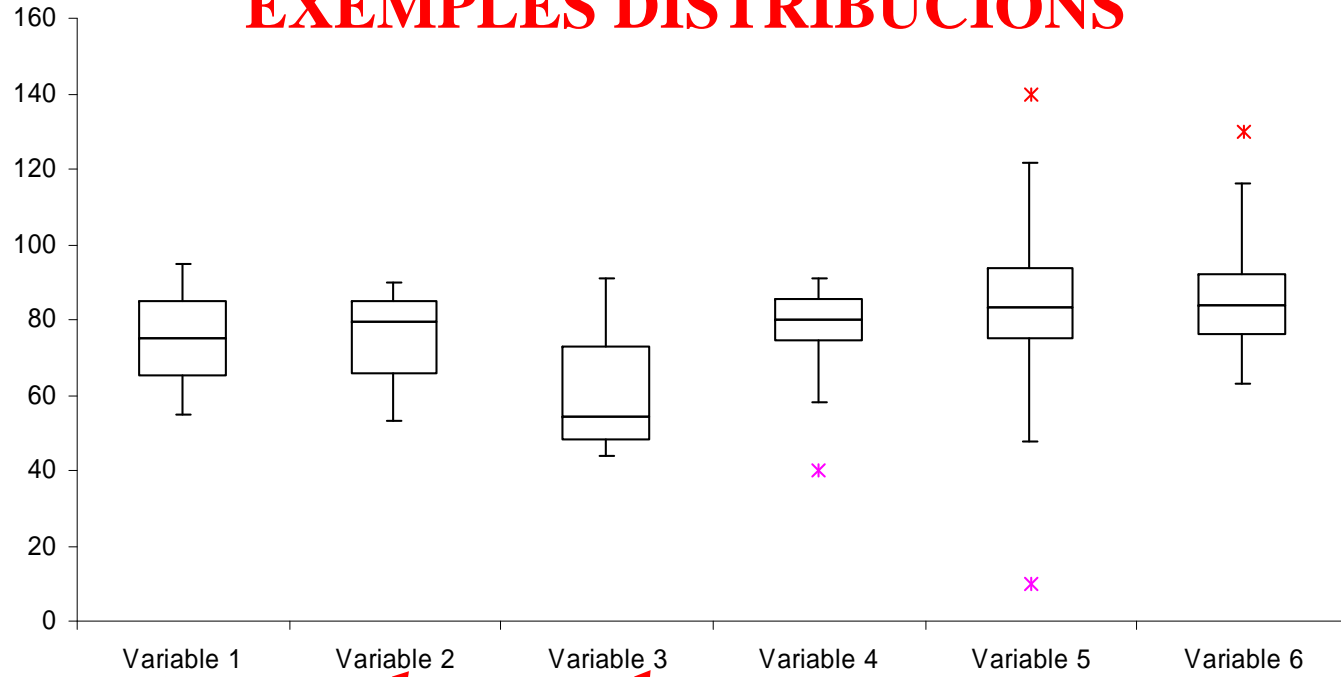
• Un valor serà considerat valor atípic lleu si:

$$o \begin{cases} \text{valor} < Q_1 - 1,5 \text{ RIC} \\ \text{valor} > Q_3 + 1,5 \text{ RIC} \end{cases}$$

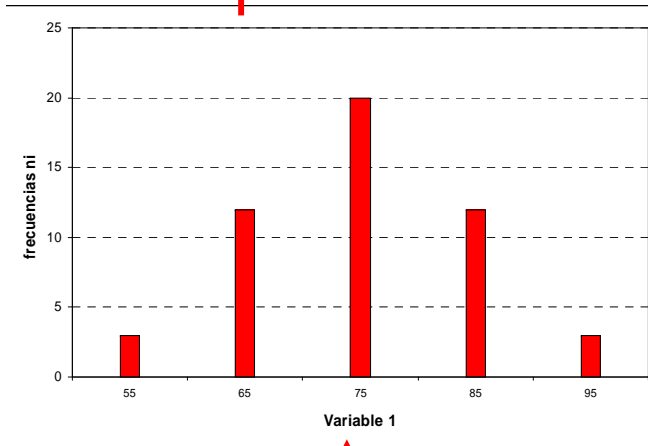
• Un valor serà considerat valor atípic extrem si:

$$o \begin{cases} \text{valor} < Q_1 - 3 \text{ RIC} \\ \text{valor} > Q_3 + 3 \text{ RIC} \end{cases}$$

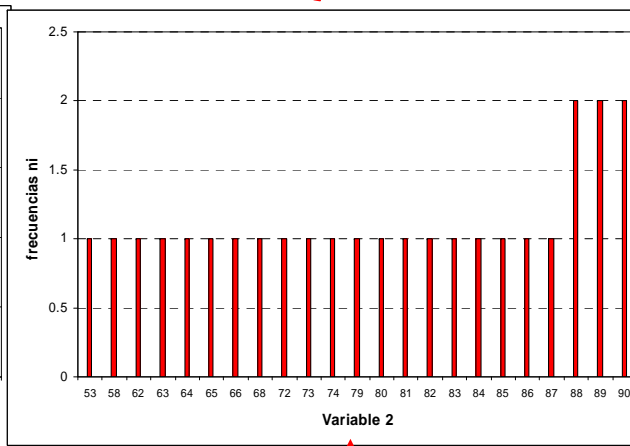
EXEMPLES DISTRIBUCIONS



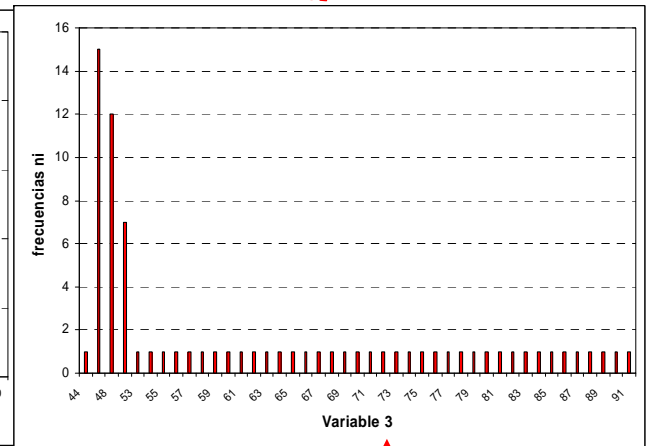
* Min. Outlier * Max. Outlier



Coef. simetria $g_1=0$
Distribució Simétrica



Coef. simetria $g_1=-0,52 < 0$
Distrib. Asimétrica per l'esquerra



Coef. simetria $g_1=0,61 > 0$
Distrib. Asimétrica per la dreta

5.- MESURES DE CONCENTRACIÓ

Corba de Lorenz
Index de Gini (IG)

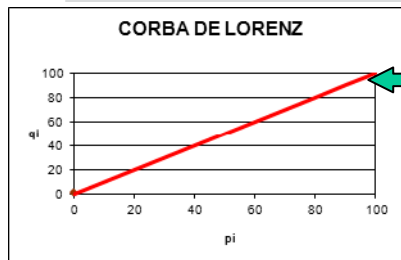
Colectiu: 10 treballadors d'una empresa turística.

Variable X: salari mensual en euros.

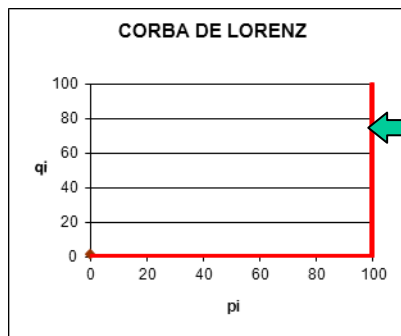
xi	ni	xi ni	freq. relatives fi en %	% acumulatiu xi ni en %	% acumulatiu pi	% acumulatiu qi
600	6	3600	60	42.86	60	42.86
900	2	1800	20	21.43	80	64.29
1200	1	1200	10	14.29	90	78.57
1800	1	1800	10	21.43	100	100
	10	8400				

Interpretació
pi i qi

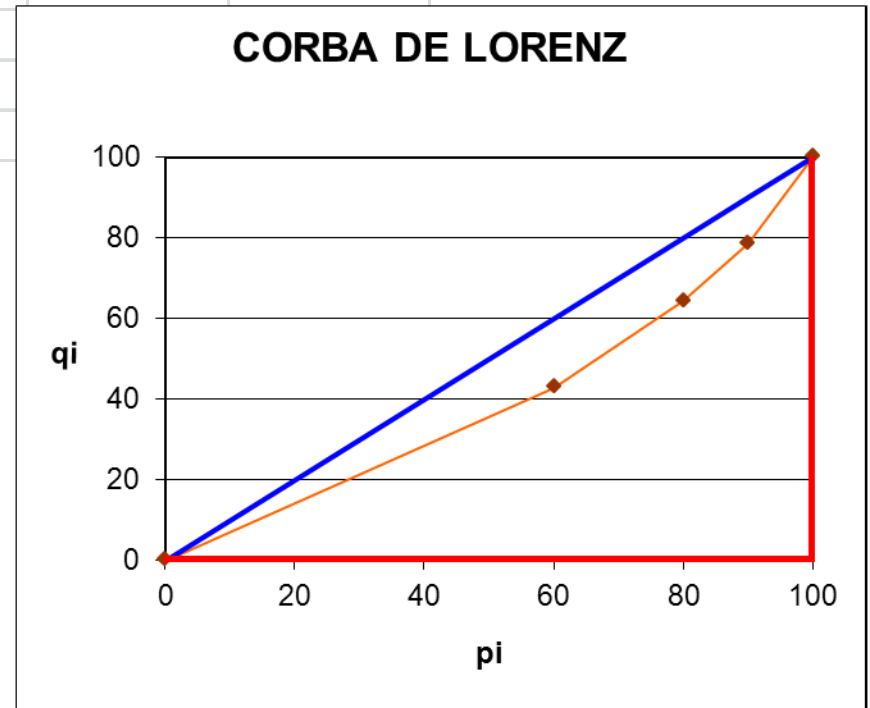
Objectiu: analitzar grau total de salaris de concentració en el pagats per l' empresa



• **Concentració mínima:** $pi=qi$



• **Concentració màxima:** tots $qi=0$ menys l'últim, que és 100



INDEX DE GINI

pi	qi	pi-qi
60	42.86	17.14
80	64.29	15.71
90	78.57	11.43
230		44.29

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

• **Concentració mínima:** $p_i = q_i \Rightarrow IG = 0$

• **Concentració máxima:** tots $q_i = 0 \Rightarrow IG = 1$

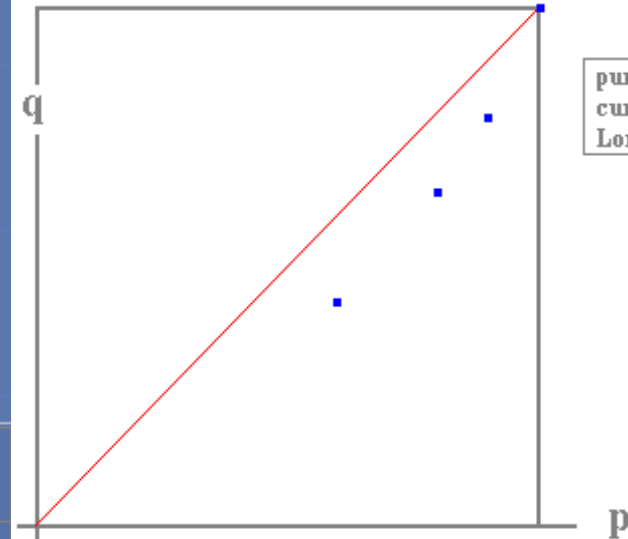
• **Casos intermedis:**

$$\Rightarrow 0 \leq IG \leq 1$$

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

MITJANA CONJUNTA D'UN CONJUNT DE DADES

PER A CALCULAR LA MITJANA CONJUNTA D'UN CONJUNT DE DADES ES NECESARI APLICAR ESTA FÓRMULA DE LA MITJANA PONDERADA:

Conjunt de datos	1º	2º	kº	Global (agregat dels k conjunts)
Nombre d'observacions de cada conjunt de dades	N_1	N_2	N_k	$N_1 + N_2 + \dots + N_k$
Mitjana de cada conjunt de dades	\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 mitjana de l'agrupació de 2 o més conjunts de dades és la mitjana de les mitjanes dels diferents conjunts, però cadascuna ponderada pel nombre d'observacions de cada conjunt de dades.

EXEMPLE


A partir de la següent informació, comprova que la mitjana de la Població 2009 del conjunt dels 183 municipis de la província de València és 13849,22 habitants

Conjunt de dades	Municipis Turístics	Municipis No Turístics
Nombre d'observacions de cadascún del conjunt de dades	49	134
Mitjana de la Població 2009 de cadascún del conjunt de dades	$\bar{x}_1 = 31180,12$	$\bar{x}_2 = 7511,81$

TIPIFICACIÓ DE VARIABLES

- Es tracta de un procediment que facilita la comparació entre observacions provinents de dos conjunts de dades diferents quan estes tenen mitjanes i desviacions típiques diferents.

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


- Un valor que s' ha tipificat, indica el nombre de desviacions típiques que eixe valor está per damunt o per sota de la mitjana.
- S'utilitza per comparar la posició relativa de dues observacions que provenen cadascuna d'una distribució diferent (exemple) 
- La nova variable tipificada Z que ha sorgit d'aplicar esta transformació a una variable X té una propietat especial: la seua mitjana és zero i la seua desviació típica és 1.

EXEMPLE DE TIPIFICACIÓ I DE DISPERSIÓ

En l'assignatura d' Incorporació, la nota mitjana dels estudiants ha estat de 7 amb una desviació típica de 2. En l'assignatura d'Estadística la nota mitjana dels estudiants ha estat de 5 amb una desviació típica de 1.

- a) Un estudiant ha obtingut una nota en Incorporació de 6, mentre que la seua nota en Estadística ha estat de 5,5. En quina assignatura ha obtingut l'estudiant un resultat relativament millor?
- b) En quina assignatura la dispersió relativa és menor?

Més informació 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