

Estadística econòmica i empresarial I: teoria i exercicis

Miquel Clar (coord.)

Ramon Alemany
Miquel Clar
Raül Ramos

Índex

INTRODUCCIÓ	9
 PRIMERA PART ANÀLISI I DESCRIPCIÓ DE DADES	
BLOC TEMÀTIC I INTRODUCCIÓ A L'ESTADÍSTICA	
1 CONTINGUT	17
2 OBJECTIUS	17
3 CONCEPTES CLAU	17
4 FITXA RESUM	18
4.1 Objecte de l'estadística	18
4.1.1 Ciència empírica	18
4.1.2 Fenòmens aleatoris «versus» deterministes	18
4.1.3 Estadística descriptiva, teoria de la probabilitat (estadística teòrica) i inferència estadística	19
4.1.4 Població «versus» mostra	19
4.2 Tipologia de dades	20
4.2.1 Qualitatives «versus» quantitatives (discretes i contínues)	20
4.2.2 Unidimensionals, bidimensionals i multidimensionals	20
4.2.3 Tall transversal («cross-section») «versus» sèrie temporal	21
5 EXERCICIS	22
5.1 Enunciats	22
5.2 Respostes	23
6 BIBLIOGRAFIA	23
 BLOC TEMÀTIC II ESTADÍSTICA DESCRIPTIVA UNIDIMENSIONAL	
1 CONTINGUT	25
2 OBJECTIUS	25
3 CONCEPTES CLAU	26
4 FITXA RESUM	26
4.1 Tabulació	26
4.1.1 Descripció numèrica	26
4.1.2 Elaboració d'una taula de freqüències amb dades sense agrupar	27

4.1.3	Elaboració d'una taula de freqüències amb dades agrupades en intervals.....	28
4.1.4	Mesures de síntesi	30
4.1.5	Transformacions lineals	42
4.1.6	Representació gràfica	42

5 EXERCICIS

5.1	Enunciats	46
5.2	Respostes	50

6 BIBLIOGRAFIA

BLOC TEMÀTIC III ESTADÍSTICA DESCRIPTIVA BIDIMENSIONAL

1 CONTINGUT

2 OBJECTIUS

3 CONCEPTES CLAU

4 FITXA RESUM

4.1	Tabulació.....	55
4.2	Independència estadística i associació lineal: covariància i correlació	57

5 EXERCICIS

5.1	Enunciats	63
5.2	Respostes	65

6 BIBLIOGRAFIA

SEGONA PART POBLACIÓ I MOSTRA: CONCEPTUALITZACIÓ I MARC TEÒRIC

BLOC TEMÀTIC IV PROBABILITAT I POBLACIONS ESTADÍSTIQUES UNIVARIANTS

1 CONTINGUT

2 OBJECTIUS

3 CONCEPTES CLAU

4 FITXA RESUM

4.1	Probabilitat	73
4.1.1	Fenomen aleatori «versus» determinista	73
4.1.2	Experiment aleatori	73
4.1.3	Definició de probabilitat. Axiomàtica i propietats	75
4.1.4	Assignació de la probabilitat d'un esdeveniment	76
4.1.5	Probabilitat condicionada. Independència	77
4.1.6	Teorema de la intersecció	79
4.1.7	Teorema de la probabilitat total	79
4.1.8	Teorema de Bayes	80
4.2	Variable aleatòria. Definició i tipologia. Distribució de probabilitat	81
4.3	Variable aleatòria discreta: funció de quantia. Variable aleatòria contínua: funció de densitat. Funció de distribució. Esperança matemàtica i variància. Moments ..	81

4.4	Distribucions de probabilitat per a variables aleatòries discretes	87
4.5	Distribucions de probabilitat per a variables aleatòries contínues	88
4.6	La llei normal. Naturalesa, significat i tabulació	89
4.6.1	Descripció: funcions de densitat i distribució	89
4.6.2	La normal estandarditzada o tipificada: $N(0,1)$	95
4.6.3	Càlcul de probabilitats	95
4.6.4	Aproximació de les distribucions binomial i de Poisson per la normal. Correcció de continuïtat	96
4.7	Distribucions associades a la llei normal. Tabulació	99
4.7.1	Distribució khi quadrat	99
4.7.2	Distribució t de Student	100
4.7.3	Distribució F de Snedecor	101
4.8	La desigualtat de Tchebitcheff	101
4.9	Suma de variables aleatòries	102
4.10	Teorema central del límit	103
5	EXERCICIS	103
5.1	Enunciats	103
5.2	Respostes	108
6	BIBLIOGRAFIA	119

TERCERA PART

TRACTAMENT INFORMÀTIC AMB MICROSOFT EXCEL, «R» I «R-COMMANDER»

CAPÍTOL I

INTRODUCCIÓ A MICROSOFT EXCEL 14.0 (OFFICE 2010)

1	INTRODUCCIÓ	123
2	UTILITZACIÓ DE LES PRINCIPALS FUNCIONS ESTADÍSTIQUES DISPONIBLES EN MICROSOFT EXCEL	124
2.1	Funcions relacionades amb l'anàlisi estadística descriptiva	125
2.2	Funcions relacionades amb distribucions estadístiques	126
3	INSTAL·LACIÓ I BREU DESCRIPCIÓ DEL COMPLEMENT D'EINES PER A L'ANÀLISI ESTADÍSTICA	128
4	ALTRES POSSIBILITATS DE MICROSOFT EXCEL	131

CAPÍTOL II

INTRODUCCIÓ A «R» I «R-COMMANDER»

1	INTRODUCCIÓ	137
2	COM UTILITZAR «R-COMMANDER» EN L'ÀMBIT DE L'ESTADÍSTICA DESCRIPTIVA	145
3	COM UTILITZAR «R-COMMANDER» PER ANALITZAR DISTRIBUCIONS DISCRETES	147
4	COM UTILITZAR «R-COMMANDER» PER ANALITZAR DISTRIBUCIONS CONTÍNUES ...	149
5	COM UTILITZAR «R-COMMANDER» PER GENERAR NOMBRES ALEATORIS	153

Bloc temàtic II

ESTADÍSTICA DESCRIPTIVA UNIDIMENSIONAL

1. Contingut

a) *Tabulació*

- Presentació de les dades
- Ordenació de les dades i concepte de freqüència absoluta
- Elaboració d'una taula de freqüències amb dades agrupades en intervals i amb dades sense agrupar
- Interval: límit inferior i superior, amplada, marca de classe i alçada
- Freqüència relativa
- Freqüències absoluta i relativa cumulatives

b) *Representació gràfica*

- Dades quantitatives sense agrupar: sectors, barres, escala, de caixa (*box-plot*) i tija i fulles
- Dades quantitatives agrupades en intervals: histograma, polígon de freqüències no cumulatives i polígon de freqüències cumulatives

c) *Mesures de síntesi*

- Concepte d'estadístic
- Mesures de posició o localització: mitjana aritmètica simple, mitjana aritmètica ponderada, mediana, moda (relació entre mitjana aritmètica simple, mediana i moda), quantils (quartils, decils i percentils)
- Moments d'una distribució de freqüències
- Moments ordinaris o respecte a l'origen
- Moments centrals o respecte a la mitjana
- Mesures de dispersió absolutes i relatives: recorregut o rang, recorregut interquartílic, variància, desviació típica o estàndard, coeficient de variació de Pearson
- Mesures de forma: asimetria *versus* simetria (coeficients de Fisher i de Pearson), apunament o curtosi
- Concentració *versus* equidistribució. Corba de Lorenz i índex de Gini

d) *Transformacions lineals*: canvi d'origen, canvi d'escala, canvi d'origen i d'escala, i estandarització o tipificació d'una variable

2. Objectius

Els objectius del segon bloc temàtic es poden sintetitzar en els quatre punts següents:

- Saber tabular i construir una taula de freqüències a partir d'un conjunt de dades unidimensionals.

- Conèixer les diferents possibilitats de representació gràfica de les dades unidimensionals.
- Conèixer les diferents mesures descriptives unidimensionals per sintetitzar la informació, així com els seus avantatges i inconvenients.
- Concentració *versus* equidistribució.

3. Conceptes clau

Els conceptes clau del bloc temàtic són els següents:

Dades agrupades, dades sense agrupar, distribució de freqüències, taula de freqüències, freqüència absoluta, relativa, absoluta cumulativa i relativa cumulativa, interval, límit inferior, límit superior, marca de classe, amplada, alçada, estadístic, mesura de posició, moda, mediana, mitjana aritmètica simple, mitjana aritmètica ponderada, quantils, quartils, decils, percentils, mesura de dispersió, recorregut o rang, recorregut interquartílic, variància, desviació típica o estàndard, coeficient de variació de Pearson, moments d'una distribució (ordinaris i centrals), gràfic de sectors, de barres, d'escala, de caixa (*box-plot*), de tija i fulles, histograma, mesura de forma, coeficient de simetria, coeficient d'apuntament, coeficient de curtosi, concentració, equidistribució, corba de Lorenz, índex de Gini, transformacions lineals, canvi d'origen, canvi d'escala, estandardització o tipificació.

4. Fitxa resum

4.1. Tabulació

4.1.1. Descripció numèrica

- a) *Freqüència absoluta*, n_i , és el nombre d'individus pels quals s'ha observat el valor X_i de la variable X , de manera que es compleix que:

$$\sum_{i=1}^k n_i = N,$$

on k és el nombre de valors diferents que pren (que s'han observat de) la variable X i N és el nombre total d'observacions (individus) disponibles.

- b) *Freqüència relativa*, f_i , és la proporció d'individus pels quals s'ha observat el valor X_i de la variable X . S'obté a partir de l'expressió següent:

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

de manera que es compleix que $\sum_{i=1}^k f_i = 1$, on k és el nombre de valors diferents que pren (que s'han observat de) la variable X .

- c) *Freqüència absoluta cumulativa*, N_j , és el nombre d'individus observats fins a un determinat valor X_j de la variable X , de manera que es compleix que:

per al primer valor observat de la variable, X_1 , $N_1 = n_1$; i,
per al darrer valor observat de la variable, X_k , $N_k = N$.

S'obté a partir de l'expressió següent:

$$N_j = \sum_{i=1}^j n_i.$$

- d) *Freqüència relativa cumulativa*, F_j , és la proporció d'individus observats fins a un determinat valor X_j de la variable X , de manera que es compleix que:

per al primer valor observat de la variable, X_1 , $F_1 = f_1$; i,
per al darrer valor observat de la variable, X_k , $F_k = 1$.

S'obté a partir de l'expressió següent:

$$F_j = \sum_{i=1}^j f_i = \frac{N_j}{N}.$$

- e) *Interval*, I_i , es defineix com el conjunt de valors continguts entre dos valors: el límit inferior (L_{i-1}) —valor mínim de l'interval— i el límit superior (L_i) —valor màxim de l'interval.

- f) *Amplada de l'interval*, a_i , es defineix com segueix:

$$a_i = L_i - L_{i-1},$$

on, tal com s'ha esmentat anteriorment, L_i representa l'extrem superior de l'interval i L_{i-1} l'inferior.

- g) *Marca de classe*, c_i , és el punt mitjà de l'interval:

$$c_i = \frac{L_i + L_{i-1}}{2}.$$

- h) *Alçada de l'interval*, h_i , és una mesura de la densitat de l'interval que es calcula com:

$$h_i = \frac{n_i}{a_i}.$$

4.1.2. Elaboració d'una taula de freqüències amb dades sense agrupar
Els passos que cal seguir són els següents:

- Ordenar els valors observats de la variable de menor a major.
- Calcular la freqüència absoluta.
- Calcular la freqüència relativa.
- Calcular la freqüència absoluta cumulativa.
- Calcular la freqüència relativa cumulativa.

EXEMPLE 2.1.

Suposeu que es disposa de les següents dades corresponents a l'observació de la variable *nombre d'assignatures matriculades* per a un col·lectiu de vint estudiants:

6	7	6	8
5	6	8	9
9	10	10	7
8	10	6	5
7	6	5	8

La taula de freqüències és:

X_i	n_i	f_i	N_i	F_i
5	3	0,15	3	0,15
6	5	0,25	8	0,4
7	3	0,15	11	0,55
8	4	0,2	15	0,75
9	2	0,1	17	0,85
10	3	0,15	20	1
Suma	20	1		

A partir de la informació recollida a la taula de freqüències anterior, es poden contestar preguntes com ara:

- a) Quants estudiants s'han matriculat de 6 assignatures? → 5.
- b) Quin percentatge d'estudiants s'han matriculat de 9 assignatures? → 10%.
- c) Quants estudiants s'han matriculat de més de vuit assignatures? → 5.
- d) Quin percentatge d'estudiants s'han matriculat de 6 o menys assignatures? → 40%.
- e) Quina és la probabilitat que, en triar a l'atzar un estudiant d'aquest col·lectiu, s'hagi matriculat de 8 assignatures? → 20%.
- f) Quina és la probabilitat que, en triar a l'atzar un estudiant d'aquest col·lectiu, s'hagi matriculat de 7 o més assignatures? → 60%.

4.1.3. Elaboració d'una taula de freqüències amb dades agrupades en intervals

Els passos que cal seguir són els següents:

- a) Determinar el nombre d'intervals o, alternativament, l'amplada dels intervals a partir de l'expressió:

$$a_i = \frac{\text{Recorregut de la variable}}{\text{Nombre d'intervals}},$$

sent el recorregut de la variable la diferència entre el major i el menor valor observats de la variable.

- b) Ordenar els intervals de menor a major.

- c) Calcular la freqüència absoluta.
- d) Calcular la freqüència relativa.
- e) Calcular la freqüència absoluta cumulativa.
- f) Calcular la freqüència relativa cumulativa.
- g) Calcular les marques de classe dels intervals.
- h) Calcular l'amplada dels intervals.
- i) Calcular l'alçada dels intervals.

EXEMPLE 2.2.

A partir de la següent informació referent a les vendes d'una empresa amb establiments a les diferents comunitats autònomes espanyoles, elaboreu una taula de freqüències amb les dades agrupades en intervals.

Comunitat Autònoma	Vendes
Andalusia	11.049.398
Aragó	2.726.691
Astúries	1.994.784
Balears	2.511.753
Canàries	3.538.855
Cantàbria	1.032.313
Castella i Lleó	4.812.911
Castella-La Manxa	2.889.848
Catalunya	16.091.029
Comunitat Valenciana	8.471.142
Extremadura	1.643.618
Galícia	4.873.121
Madrid	13.867.322
Múrcia	1.864.375
Navarra	1.323.886
País Basc	5.105.780
La Rioja	630.820

La taula de freqüències amb els intervals de la mateixa amplada és la següent:

X_i	n_i	f_i	N_i	F_i	c_i	a_i	h_i
0-4.000.000	10	0,588	10	0,588	2.000.000	4.000.000	0,0000025
4.000.000-8.000.000	3	0,176	13	0,765	6.000.000	4.000.000	0,00000075
8.000.000-12.000.000	2	0,118	15	0,882	10.000.000	4.000.000	0,0000005
12.000.000-16.000.000	1	0,059	16	0,941	14.000.000	4.000.000	0,00000025
16.000.000-20.000.000	1	0,059	17	1	18.000.000	4.000.000	0,00000025
Suma	17	1					

La taula de freqüències amb els intervals de diferent amplada és la següent:

X_i	n_i	f_i	N_i	F_i	c_i	a_i	h_i
0-2.000.000	6	0,353	6	0,353	1.000.000	2.000.000	0,000003
2.000.000-4.000.000	4	0,235	10	0,588	3.000.000	2.000.000	0,000002
4.000.000-6.000.000	3	0,176	13	0,765	5.000.000	2.000.000	0,0000015
6.000.000-12.000.000	2	0,118	15	0,882	9.000.000	6.000.000	0,0000003
12.000.000-20.000.000	2	0,118	17	1	16.000.000	8.000.000	0,00000025
Suma	17	1					

4.1.4. Mesures de síntesi

Concepte d'estadístic: És qualsevol mesura que podem calcular a partir de les observacions corresponents a una determinada variable.

Mesures de posició o localització

a) Mitjana aritmètica simple

a1) Dades no agrupades en intervals:

$$\bar{X} = \underbrace{\frac{\sum_{i=1}^N X_i}{N}}_{\text{emprant dades observades}} = \underbrace{\frac{\sum_{i=1}^k X_i \cdot n_i}{N}}_{\text{emprant freqüències absolutes}} = \underbrace{\sum_{i=1}^k X_i \cdot f_i}_{\text{emprant freqüències relatives}}$$

on k és el nombre de valors diferents que pren (que s'han observat de) la variable X .

a2) Dades agrupades en m intervals:

$$\bar{X} = \underbrace{\frac{\sum_{i=1}^m c_i \cdot n_i}{N}}_{\text{emprant freqüències absolutes}} = \underbrace{\sum_{i=1}^m c_i \cdot f_i}_{\text{emprant freqüències relatives}}$$

Inconvenient de la mitjana aritmètica simple



Esbiaixada cap als valors extrems

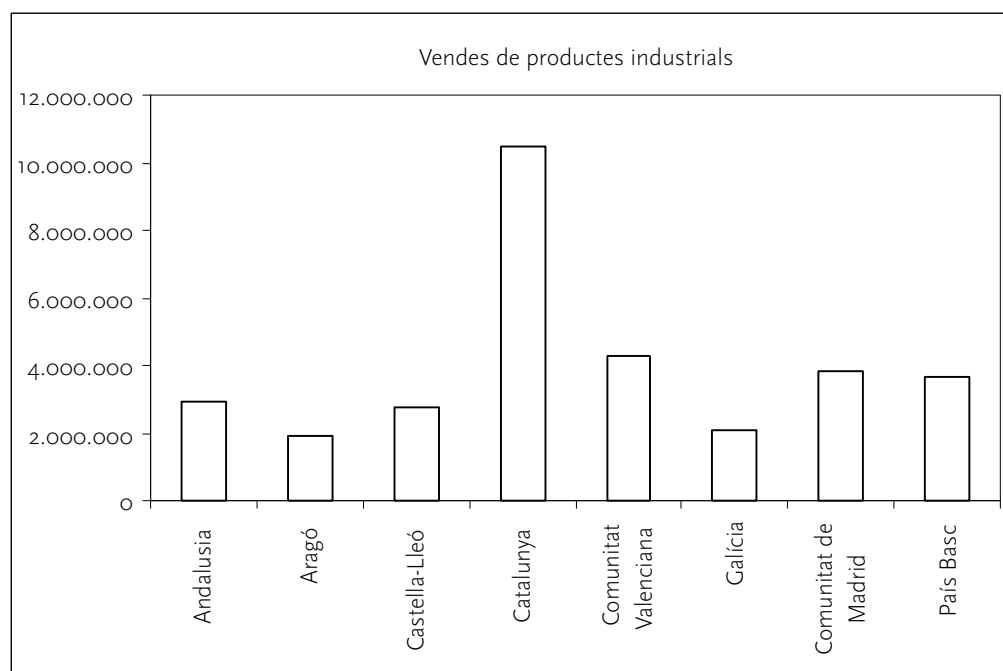
EXEMPLE 2.3.

A la taula següent es presenten les vendes de productes d'una determinada empresa industrial a vuit comunitats autònomes espanyoles on té seu:

Vendes de productes industrials

Andalusia	2.932.477
Aragó	1.900.761
Castella-Lleó	2.745.274
Catalunya	10.482.782
Comunitat Valenciana	4.294.412
Galícia	2.085.659
Comunitat de Madrid	3.832.775
País Basc	3.668.374

Si es calcula la mitjana aritmètica simple s'obté la quantitat 3.992.814,25. Si s'analitza aquest resultat es veu que, de les vuit regions considerades, totes (tret de Catalunya i de la Comunitat Valenciana) estan per sota de la mitjana. Si es torna a calcular la mitjana aritmètica simple sense considerar Catalunya, s'obté la quantitat 3.065.676,00. Aquest nou valor sembla molt més representatiu que l'anterior. Aquest fet es produeix perquè, en el primer cas, la mitjana aritmètica simple està esbiaixada a l'alça per «l'efecte de Catalunya», la qual presenta un valor clarament molt més gran que la resta de regions considerades, tal com es pot veure en el gràfic següent:



b) Mitjana aritmètica ponderada

b1) Dades no agrupades en intervals:

$$\bar{X}_{ponderada} = \frac{\sum_{i=1}^N X_i \cdot \omega_i}{\underbrace{\sum_{i=1}^N \omega_i}_{\text{emprant dades observades}}} = \frac{\sum_{i=1}^k X_i \cdot \omega_i}{\underbrace{\sum_{i=1}^k \omega_i}_{\text{emprant dades no agrupades en forma de taula}}}$$

on ω_i són les diferents ponderacions de cada valor observat de X .

b2) Dades agrupades en m intervals:

$$\bar{X}_{\text{ponderada}} = \frac{\sum_{i=1}^m c_i \cdot \omega_i}{\sum_{i=1}^m \omega_i},$$

on ω_i són les diferents ponderacions de cadascun dels m intervals en què està dividit el conjunt de valors observats de X .

Si $\omega_i = n_i$, aleshores $\bar{X}_{\text{ponderada}} = \bar{X}$.

La mitjana aritmètica és un cas particular de la mitjana aritmètica ponderada.

c) Mediana (M_e)

c1) Dades no agrupades en intervals. Valor de la variable que té associada una freqüència absoluta cumulativa de $\frac{N}{2}$.

c2) Dades agrupades en m intervals. En primer lloc cal determinar l'interval medià (aquell que té associada una freqüència absoluta cumulativa de $\frac{N}{2}$). Sota el supòsit que en l'interval medià les observacions estan distribuïdes uniformement, la mediana és donada per:

$$L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} \cdot a_i,$$

on L_{i-1} és l'extrem inferior de l'interval medià, N_{i-1} és la freqüència absoluta cumulativa fins a l'interval anterior a l'interval medià i, n_i i a_i són, respectivament, la freqüència absoluta i l'amplada de l'interval medià.

Avantatge de la mediana



No l'afecten els valors extrems

EXEMPLE 2.4.

Reprenent les dades de l'exemple 2.3 relatives a les vendes de productes de l'empresa industrial a les vuit comunitats autònomes espanyoles on tenia seu, si es calcula la mediana considerant les vuit comunitats autònoms s'obté la quantitat 3.300.425,5 (mitjana aritmètica entre el valor de la comunitat autònoma que ocupa la quarta posició —Andalusia— i el de la cinquena —País Basc—, un cop ordenades de menor a major les vuit comunitats autònoms).

EXEMPLE 2.5.

Suposeu que es disposa de les següents dades corresponents a l'edat d'un col·lectiu d'onze individus:

18	18	19	19	19	21	22	22	23	23	23
----	----	----	----	----	----	----	----	----	----	----

Si es calcula la mitjana i la mediana s'obté, respectivament, 20, 64 i 21.

Suposeu ara que les dades de què es disposa són les següents:

18	18	19	19	19	21	22	22	23	23	62
----	----	----	----	----	----	----	----	----	----	----

En aquest cas, respecte al supòsit inicial, només hi ha una diferència: l'edat de l'onze individu és de 62 anys en lloc de 23. Si es calcula la mitjana i la mediana s'obté, respectivament, 24,18 i 21. Veiem, doncs, que, mentre que la mediana no varia, la mitjana pren un valor més gran a conseqüència de «l'efecte d'arrossegament» que ocasiona el valor (atípic dins el col·lectiu considerat) de l'onze individu.

d) *Moda (M_o)*

- d1) Dades no agrupades en intervals: Valor de la variable que té associada una major freqüència absoluta (o relativa).
- d2) Dades agrupades en m intervals de la mateixa amplada: En primer lloc cal determinar l'interval modal (aquell que té una major freqüència absoluta o relativa). Sota el supòsit que en l'interval modal les observacions estan distribuïdes uniformement, la moda és donada per:

$$L_{i-1} + \frac{n_{i+1}}{n_{i-1} + n_{i+1}} \cdot a_i,$$

on L_{i-1} és l'extrem inferior de l'interval modal, n_{i+1} és la freqüència absoluta de l'interval posterior a l'interval modal, n_{i-1} és la freqüència absoluta de l'interval anterior a l'interval modal i a_i és l'amplada de l'interval modal.

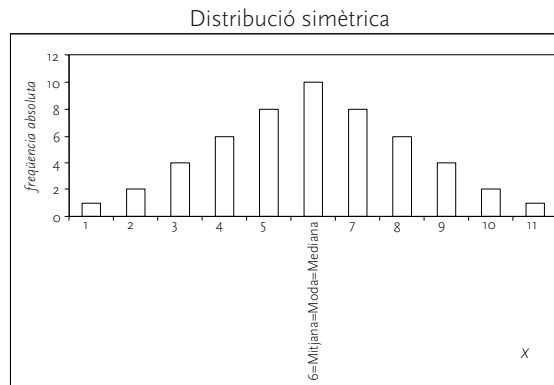
- d3) Dades agrupades en m intervals no tots de la mateixa amplada: En primer lloc cal determinar l'interval modal (aquell que té una major alçada). Sota el supòsit que en l'interval modal les observacions estan distribuïdes uniformement, la moda és donada per:

$$L_{i-1} + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} \cdot a_i,$$

on L_{i-1} és l'extrem inferior de l'interval modal, h_{i+1} és l'alçada de l'interval posterior a l'interval modal, h_{i-1} és l'alçada de l'interval anterior a l'interval modal i a_i és l'amplada de l'interval modal.

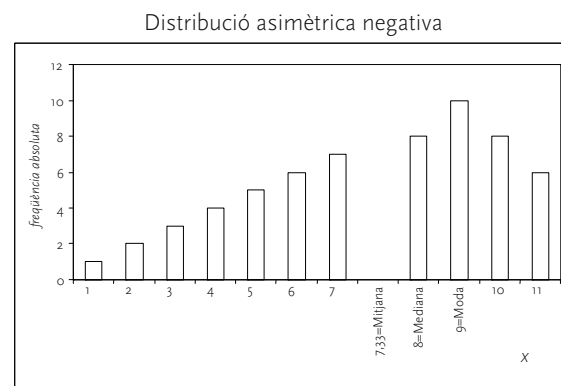
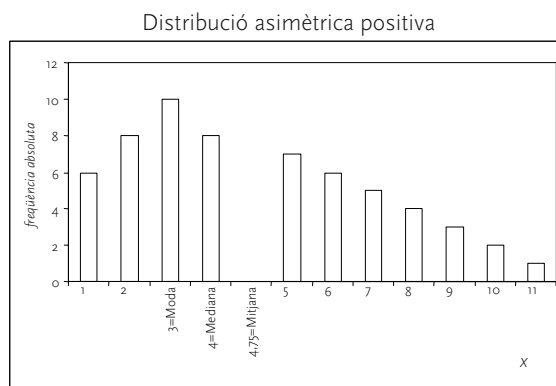
e) *Relació entre mitjana aritmètica, mediana i moda*

Si coincideixen $\Rightarrow$ el valor que més cops es repeteix (moda) i que deixa el 50% de les observacions a cada banda (mediana) és la mitjana aritmètica $\Rightarrow$ la distribució és simètrica.



Si no coincideixen $\Rightarrow$ la distribució no és simètrica.

- Si $\bar{X}$ i M_e estan a la dreta de (prenen un valor superior a) la M_o , la distribució és asimètrica per la dreta o positiva.
- Si $\bar{X}$ i M_e estan a l'esquerra de (prenen un valor inferior a) la M_o , la distribució és asimètrica per l'esquerra o negativa.



f) *Quantils*. Són els valors de la variable que divideixen els N valors observats en parts iguals, és a dir, en intervals cadascun dels quals conté el mateix percentatge d'observacions. Entre els quantils destaquem els estadístics següents:

f1) *Quartils*. Són els tres valors de la variable que divideixen els N valors observats en quatre parts iguals, és a dir, en intervals cadascun dels quals conté el 25% de les observacions:

Quartil	Observacions que acumula	Percentatge d'observacions que acumula
Primer, Q_1	$N/4$	25%
Segon, Q_2	$N/2$	50% $\rightarrow$ Mediana
Tercer, Q_3	$3N/4$	75%

L'expressió que permet trobar els quartils és la següent:

$$Q_r = L_{i-1} + \frac{\frac{r}{4} \cdot N - N_{i-1}}{n_i} \cdot a_i \quad r = 1, 2, 3.$$

A partir de l'expressió anterior, es pot comprovar que si $r = 2$, aleshores Q_2 coincideix amb Me .

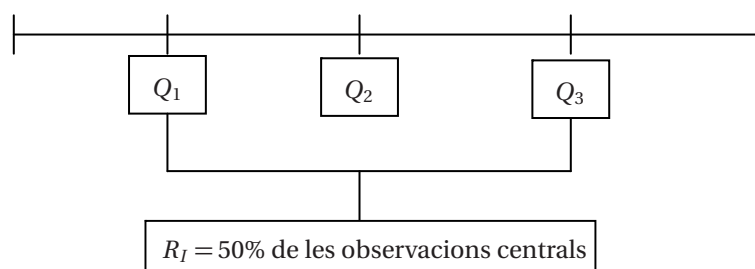
f2) *Decils* Són els nou valors de la variable que divideixen els N valors observats en deu parts iguals, és a dir, en intervals cadascun dels quals conté el 10% de les observacions. El cinquè decil coincideix amb la mediana.

f3) *Percentils* Són els noranta-nou valors de la variable que divideixen els N valors observats en cent parts iguals, és a dir, en intervals cadascun dels quals conté l'1% de les observacions. El cinquantè percentil coincideix amb la mediana.

Mesures de dispersió absolutes

a) *Recorregut o rang*: $R_X = \max(X_i) - \min(X_i)$.

b) *Recorregut interquartílic*: $R_I = Q_3 - Q_1$. Gràficament s'interpreta de la manera següent:



c) *Variància*

c1) Dades no agrupades en intervals:

$$\begin{aligned}
 S_X^2 &= \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N} = \frac{\left(\sum_{i=1}^N X_i^2 \right) - (N \cdot \bar{X}^2)}{N} = \frac{\sum_{i=1}^N X_i^2}{N} - \bar{X}^2 = \\
 &\quad \underbrace{\hspace{10em}}_{\text{emprant dades observades}} \\
 &= \frac{\sum_{i=1}^k (X_i - \bar{X})^2 \cdot n_i}{N} = \frac{\sum_{i=1}^k X_i^2 \cdot n_i}{N} - \bar{X}^2 = \underbrace{\sum_{i=1}^k (X_i - \bar{X})^2 \cdot f_i}_{\text{emprant freqüències relatives}}, \\
 &\quad \underbrace{\hspace{10em}}_{\text{emprant freqüències absolutes}}
 \end{aligned}$$

on k és el nombre de valors diferents que pren (que s'han observat de) la variable X .

c2) Dades agrupades en m intervals:

$$S_X^2 = \frac{\sum_{i=1}^m (c_i - \bar{X})^2 \cdot n_i}{N} = \sum_{i=1}^m (c_i - \bar{X})^2 \cdot f_i.$$

$\underbrace{\hspace{10em}}_{\text{emprant freqüències absolutes}}$
 $\underbrace{\hspace{10em}}_{\text{emprant freqüències relatives}}$

d) *Desviació estàndard o desviació típica*

$$S_X = +\sqrt{S_X^2}.$$