

## Objectius

En aquesta quinzena aprendreu a:

- Distingir els conceptes de població i mostra.
- Diferenciar els tres tipus de variables estadístiques.
- Fer recomptes i gràfiques.
- Calcular i interpretar les mesures estadístiques de centralització més importants.
- Calcular les principals mesures de dispersió
- Entendre la importància de l'elecció de la mostra perquè sigui representativa.
- Utilitzar i representar variables bidimensionals.
- Calcular el centre de gravetat, la covariància, el coeficient de correlació i la recta de regressió en una distribució bidimensional.

1. Estadística descriptiva .....	pàg. 4
Població i mostra	
Variables estadístiques	
Gràfiques variables qualitatives	
Gràfiques variables quantitatives discretes	
Gràfiques variables quantitatives contínues	
2. Mesures de centralització .....	pàg. 7
Mitjana, moda i mediana	
Evolució de la mitjana	
Evolució de la mediana	
Mitjana i mediana comparades	
Mesures de posició	
3. Mesures de dispersió .....	pàg. 10
Desviació típica i recorregut	
Càlcul de les mesures de dispersió	
La mitjana i la desviació típica	
4. Representativitat de les mostres ...	pàg. 12
Mostreig estratificat	
Mostreig aleatori. Biaix	
5. Estadística bidimensional .....	pàg. 14
Distribucions bidimensionals	
Correlació lineal	
Rectes de regressió	

Exercicis per practicar

Per saber-ne més

Resum

Autoavaluació



## Abans de començar

### Recordeu

El curs passat ja vau estudiar estadística, i en nombroses ocasions heu fet estadística encara que no us n'hagueu adonat. Vegem-ne alguns exemples

### Nota mitjana

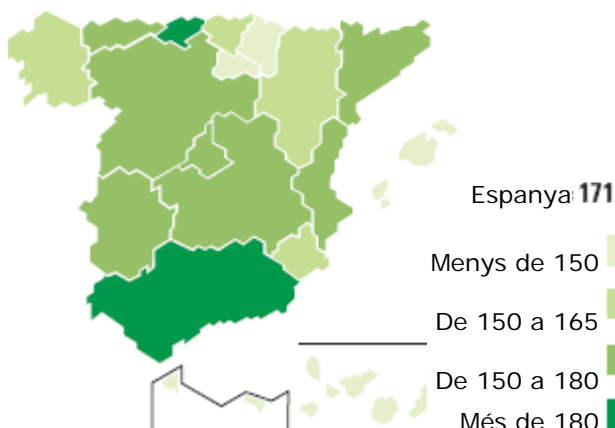
Al llarg d'un curs escolar tindreu moltes ocasions per calcular aquest valor. Si una nota depèn de dos exàmens i en un teniu un 4, intentareu treure almenys un 6 en l'altra.

Al final de l'institut, les mitjanes del batxillerat i de la prova de la selectivitat. Comparacions amb la mitjana local o nacional. Les notes de tall per a determinades carreres

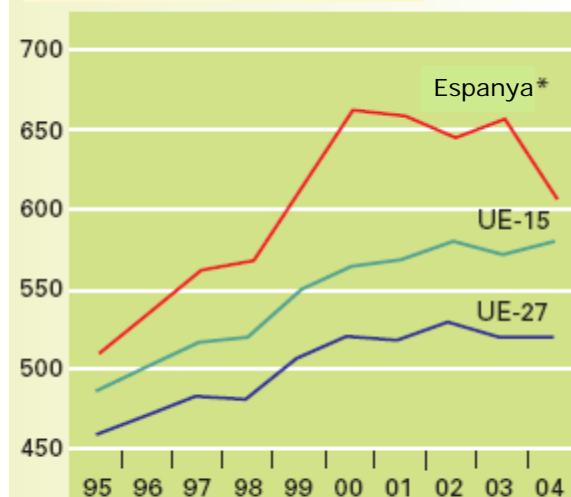
### Futbol

El jugador que més gols ha marcat, el porter que menys n'ha encaixat. La classificació de la lliga. La millor meitat de lliga. Les classificacions de competicions europees, les de descens, el nombre de vegades internacional, el nombre de fases finals, els minuts jugats, els tirs a porteria, les faltes.

### Consum mitjà d'aigua de les Ilars. 2004 (litres/hab./dia)



### Residus urbans (kg/hab./any)



\*El 2004 s'exclouen els residus de demolició, construcció i reparació d'habitatges, d'acord amb la metodologia d'Eurostat.

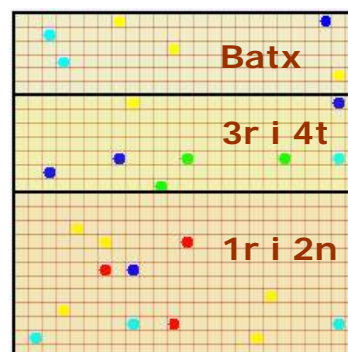
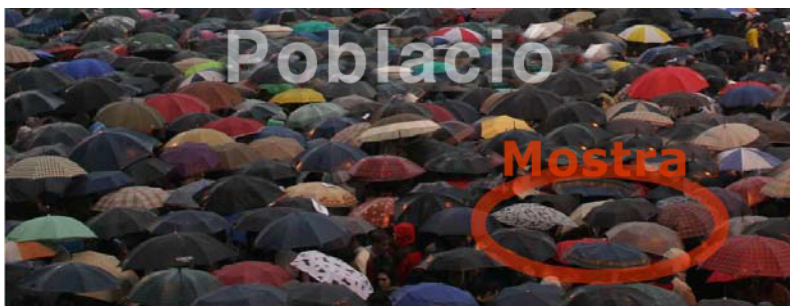
## 1. Estadística descriptiva

### Població i mostra.

**Població** és el conjunt d'individus, amb alguna característica comuna, sobre el qual es fa un estudi estadístic.

La **mostra** és un subconjunt de la població, seleccionada de manera que posi de manifest les característiques d'aquesta, i per això la propietat més important de les mostres és la seva representativitat.

El procés seguit en l'extracció de la mostra s'anomena **mostreig**



Si cada requadre representa cada un dels alumnes d'un institut fictici i se'ls pregunta sobre el seu color preferit, el total dels quadres és la població, 625 alumnes, i els 26 enquestats constitueixen la mostra.

### Variables estadístiques

La característica que s'estudia en una població és la **variable estadística**.

Les variables estadístiques poden ser essencialment de dos tipus: **qualitatives i quantitatives**.

Les variables qualitatives són les que no apareixen de forma numèrica, sinó com a una categoria o atribut.

Les variables quantitatives són les que poden expressar-se numèricament, i al seu torn poden ser:

- ✓ Quantitatives discretes, si només poden prendre un nombre finit de valors.
- ✓ Quantitatives contínues quan poden prendre qualsevol valor d'un interval.

- El color dels ulls, el formatge preferit, el continent on viu són **variables estadístiques qualitatives**.
- El nombre d'ordinadors a casa, o de televisors, i el nombre d'habitants per habitatge, per exemple, són **variables estadístiques quantitatives discretes**.
- El pes, l'altura, la velocitat, la densitat, la pressió són **variables estadístiques quantitatives contínues**.

Les dades:

xi fi  
 ● 7  
 ● 3  
 ● 1  
 ● 6  
 ● 5  
 Total 22



Tenen aquest diagrama de sectors

## Gràfiques en variables qualitatives.

**El diagrama de sectors** és el més indicat per a aquest tipus d'informació. El percentatge de dades de cada valor en una mostra es correspon amb el mateix percentatge de sector d'un cercle. Així, per exemple, si les dades són A, A, A, A, A, B, B, B, C i C. Les freqüències són (A,5), (B,3) i (C,2), els percentatges seran (A,50%), (B,30%) i (C,20%), que corresponen a un gràfic de sectors amb (A, 180°), (B,108°) i (C, 72°).

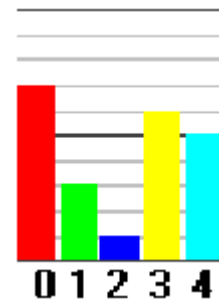
$$\frac{\text{freqüència}}{\text{nre. total de dades}} = \frac{\text{graus del sector}}{360}$$

## Gràfiques en variables discretes.

**Diagrama de barres.** N'hi ha prou que n'observeu un exemple.

A les dades,

1 2 4 4 3  
 3 3 3 0 0  
 0 4 0 1 0  
 0 3 4 1 3  
 0 4



els correspon el gràfic de la dreta.

Intervals	Recòmptes	fr.	Dens.
[0 5)		6	1,2
[5 6)		4	4
[6 7)		12	12
[7 8,5)		3	2
[8,5 10)		5	3,3

RECOMPTA DE LES NOTES EN 30 EXÀMENS

En el diagrama de freqüències, l'àrea major correspon a la columna vermella que no és la de més freqüència

$$\text{Densitat} = \frac{\text{Freqüència}}{\text{Longitud de l'interval}}$$

Les àrees de les barres-densitat són proporcionals a les freqüències en l'interval

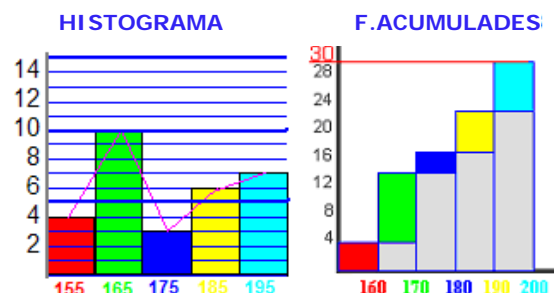
## Gràfiques en variables contínues.

**Histograma.** Les dades es representen per rectangles la base dels quals és l'amplitud de l'interval representat i amb l'altura que ens indica la freqüència absoluta, si tots els intervals són de la mateixa amplitud. Si no és el cas, les altures es calculen de manera que les àrees siguin proporcionals a les freqüències absolutes. A l'esquerra teniu un exemple fet

**Polígon de freqüències.** Unirem els centres de part superior de tots els rectangles per obtenir-lo.

També se sol dibuixar l'histograma de les **freqüències acumulades**, en cada dada s'acumula la freqüència de les dades anteriors

[150, 160] → 4  
 [160, 170] → 10  
 [170, 180] → 3  
 [180, 190] → 6  
 [190, 200] → 7



## EXERCICIS resolts

- Classifiqueu els exemples de variables estadístiques següents: longitud d'un camió, càrrega màxima, nombre de rodes, nombre d'eixos, tipus de camió, marques de pneumàtics, tipus de tapisseria, nombre de portes, altura màxima.

Qualitatives: Tipus de camió, marques de pneumàtics, tipus tapisseria

C. discretes: nombre de rodes, nombre d'eixos, nombre de portes

C. contínues: longitud d'un camió, càrrega màxima i altura màxima.

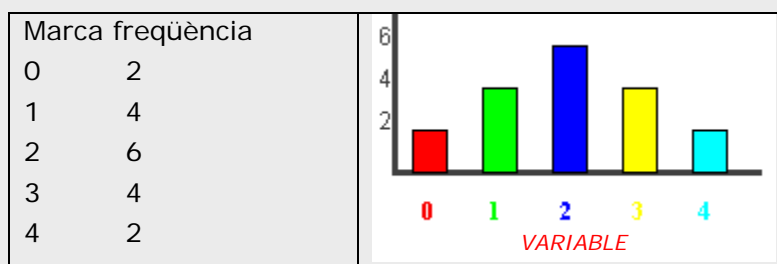
- Calculeu els graus que corresponen a cada valor en un gràfic de sectors fet a partir de les dades: R, R, V, V, V, V, V, V, A, A i A

Fem el recompte  $R \rightarrow 2$ ,  $V \rightarrow 5$  i  $A \rightarrow 3$ , i calculem

$$\frac{2}{10} = \frac{\text{Graus R}}{360}, \frac{5}{10} = \frac{\text{Graus V}}{360} \text{ i } \frac{3}{10} = \frac{\text{Graus A}}{360} \text{ i obtenim}$$

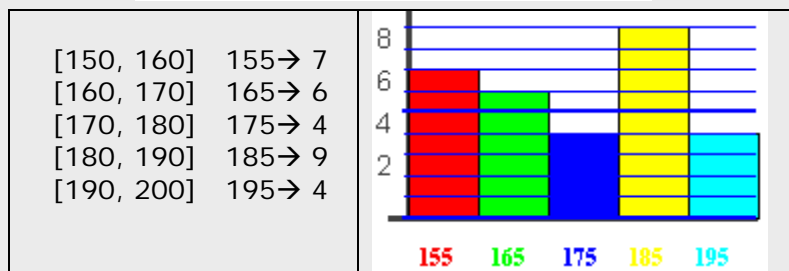
Graus R = 72, graus V = 180 i graus A = 108

- Agrupeu les dades següents i feu un diagrama de barres adequat. Dades = { 0 1 0 2 3 4 1 2 2 1 2 2 3 4 3 2 1 3 }



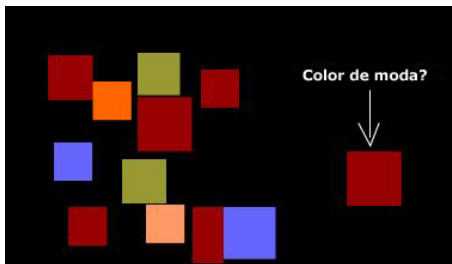
- Classifiqueu les dades en intervals i dibuixeu un histograma adequat.

180 197 154 181 189 162 152 162 167 190  
 189 160 166 197 187 194 152 181 173 154  
 177 184 186 174 177 159 158 189 160 150



## 2. Mesures de centralització

1ª EVALUACIÓ	
5	
6	
4	NOTA
1	MITJANA
9	5,5
7	
6	
6	



Per exemple, si tenim les observacions 6,7,8,6,7,6,8,6,9 i agrupem les dades veiem clarament que el valor 6 apareix més que cap altre. En aquest cas la **moda** és 6

xi → fr
6 → 4
7 → 2
8 → 2
9 → 1

Si ordenem les dades, i ja que el nombre de dades és imparell justament el 7 queda al centre.

6 6 6 6 7 7 8 8 1

Si les dades fossin 6,7,8,6,7,6,8,6,5 una vegada ordenades, i com que hi ha una quantitat parella de dades, dos d'aquests valors ocuparien el centre:

5 6 6 6 6 7 7 8 8 1

i la mediana serà  $(6 + 7)/2 = 6,5$

### Mitjana, mediana i moda.

Un conjunt N d'observacions, N nombres, pot ser que per si sol no ens digui res. En canvi, si a més ens diuen que estan situats al voltant d'un o diversos valors centrals ja tenim una referència que sintetitza la informació.

**Mitjana.** La suma dels N nombres dividida entre N. Per exemple, per a 3, 4 i 5,  $(3+4+5)/3 = 12/3 = 4$ ; per a 1, 1, 4, 8, 8 i 8,  $(1 \cdot 2 + 4 + 8 \cdot 3)/6 = 5$ .

$$\text{Mitjana} = \frac{X_1 \cdot f_1 + X_2 \cdot f_2 + \dots + X_n \cdot f_n}{f_1 + f_2 + \dots + f_n}$$

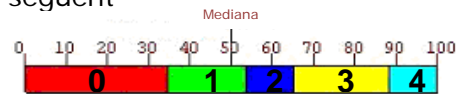
**Moda.** Si una observació es repeteix més que qualsevol altra serà considerada la moda d'aquestes dades. Per exemple, si tenim les observacions 6,7,8,6,7,6,8,6,9 i agrupem les dades  $64 \rightarrow$ ,  $72 \rightarrow$ ,  $8 \rightarrow 2$  i  $9 \rightarrow 1$  veiem clarament que el valor 6 apareix més que cap altre. En aquest cas la moda és 6

En el cas de variable contínua, considerarem per moda la marca de l'interval de major freqüència, quan passi això. També pot passar que hi hagi dues modes o que no n'hi hagi cap que destaquí.

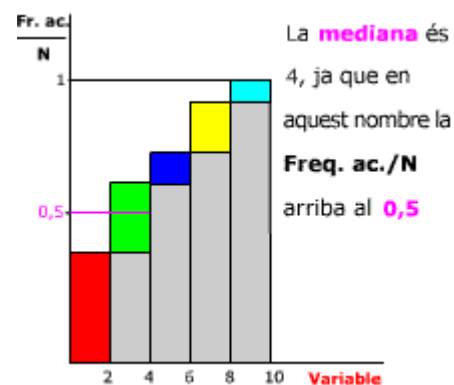
**Mediana.** El nombre tal que la meitat de les observacions són més grans que si mateix i l'altra meitat menors.

En general, per a poques dades el millor és procedir segons l'exemple de l'esquerra, segons sigui una quantitat parell o imparell.

Per a quantitats més grans, s'han d'agrupar les dades primer en una taula. I determinar segments de longitud proporcional a la seva freqüència, disposar-los de manera lineal i marcar el centre com mostra l'exemple següent



En aquest altre gràfic veiem indicada la mediana en un diagrama de freqüències relatives acumulades



## Mitjana. Evolució d'aquesta en afegir i/o canviar una dada.

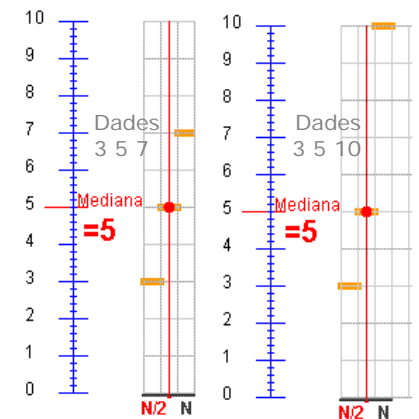
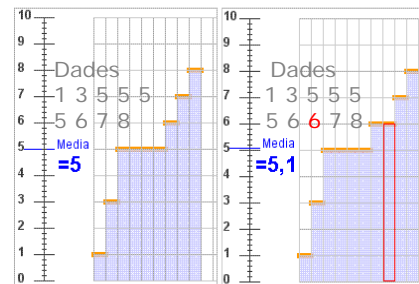
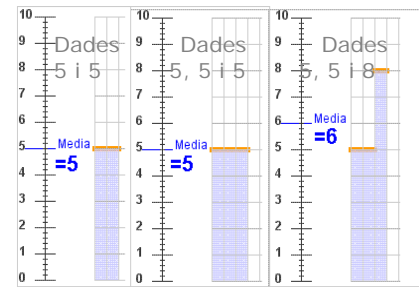
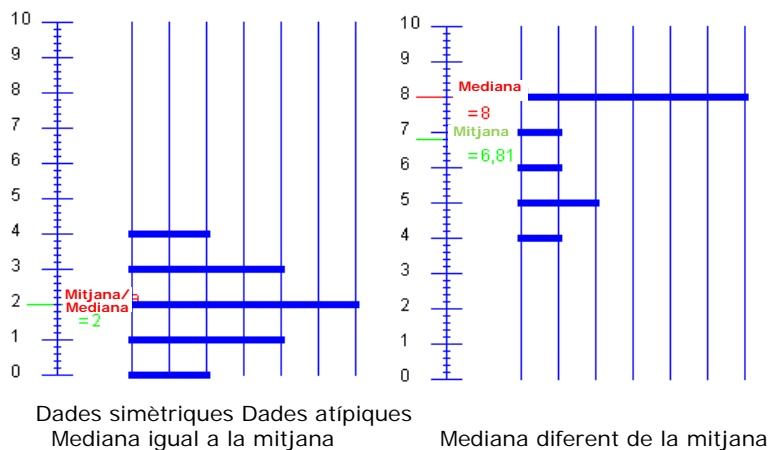
- 1 Per a les dades 5 i 5 la mitjana és 5. Si afegim un 5 es manté en 5. Si afegim un 8 la mitjana passa a ser 6. (Figura dreta)
- 2 Si tenim 9 dades amb mitjana 5, necessitem afegir un 6 perquè la mitjana passi a ser 5,1. Si tenim 19 dades amb una mitjana de 5, necessitem una dada de valor 7 perquè la mitjana pugui dins a 5,1. (Figura dreta)
- 3 Per a un conjunt de dades amb mitjana 5, si n'afegim un altre amb mitjana 5, per exemple 6 i 4, el nou conjunt conserva la mitjana.

## Mediana. Evolució d'aquesta en afegir i/o canviar una dada.

- 1 La mediana, per a les dades 2, 3 i 4 és  $Me=3$ . Si canviem el 4 per 5 o per 6 o per qualsevol altre valor major continua sent  $Me=3$ .
- 2 En canvi, si afegim una altra dada i tenim 2, 3, 4 i 4, per exemple, la  $Me=3,5$ . I si ara afegim un cinquè valor, un 4 o un 5 o un 6 o qualsevol altre major que 4, la mediana en 2, 3, 4, 4 i ?? passa a ser 4. És igual que el valor ?? sigui 5, 10 o 25.

## Mitjana i mediana comparades

Per a les dades 4 i 6 la mitjana i la mediana coincideixen en 5. Afegir un 8 o un 11 dona el mateix per a la mediana, que passa a ser en tots dos casos 6. Tanmateix la mitjana amb un 8 passa a ser 6 i amb un 11 passa a ser 7. Els valors 8 i 11 es consideren observacions atípiques, estan distanciats de la resta de valors, estiren la mitjana i no afecten la mediana. Si les dades estiguessin repartides simètricament respecte a un valor, aquest valor seria alhora la mitjana i la mediana. En canvi, si els valors en un costat de la mediana n'estan més allunyats que els de l'altre costat, la mitjana es desplaça cap a aquests valors allunyats que l'estiren. Hi ha una asimetria.



Per veure la mediana es traça una vertical des de l'eix horitzontal en  $N/2$

Per exemple, si tenim les observacions

1. 20, 24 y 28.

$$Me = 24$$

2. I per a 20, 24, 28 y 30

$$Me = (24+28)/2 = 26$$

3. Per a 20, 24, 28 y 100

$$Me = (24+28)/2 = 26$$

En canvi, la mitjana no es comporta de la mateixa manera per a les mateixes dades

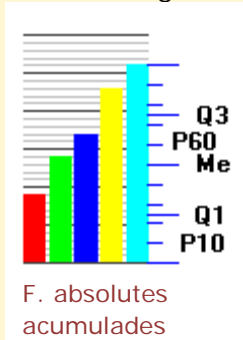
- 1  $\bar{X} = 24$

- 2  $\bar{X} = 25,5$

- 3  $\bar{X} = 43$



També podem fer un diagrama de freqüències acumulades i dividir-lo en parts iguals com mostra el gràfic

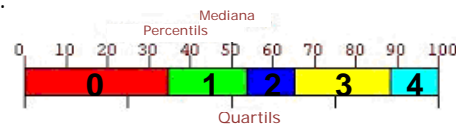


## Mesures de posició: quartils i percentils

Donat un conjunt de dades numèrics a més de la mediana podem considerar altres mesures de posició

- Si ens fixem en el primer valor que supera el 25% o el 75%, estem parlant del **primer i tercer quartil, Q<sub>1</sub> i Q<sub>3</sub>**.
- Per a altres valors com el 10%, o el 80% parlem de **percentils, P<sub>10</sub> y P<sub>80</sub>**.

**Exemple.** Per a la variable de valors 0, 1, 2, 3, 4, i freqüències 0, 9, 1, 5, 2, 3, 3, 6, 4, 3, dibuixem barres de longitud proporcional a les freqüències i dividim el total en parts iguals: en dues parts per a la mediana, quatre per als quartils i 10 per als percentils principals.



## Exercicis resolts

7. Calculeu la mitjana en cada cas:

1. 4, 6, 8
2. 4, 6, 8, 6
3. 100, 120, 180, 200

- Solucions: a)  $(4+6+8)/3 = 6$   
 b)  $(4+6+8+6)/4 = 24/4 = 6$   
 c)  $(100+120+180+200)/4 = 150$

8. Calculeu la mitjana en cada cas:

a	Marca	Fr	b	Marca	Fr
	10	2		100	2
	20	4		200	4
	30	3		300	3
	40	2		400	2

a)  $\bar{X} = \frac{10 \cdot 2 + 20 \cdot 4 + 30 \cdot 3 + 40 \cdot 2}{11} = 24,54$   
 b)  $\bar{X} = \frac{100 \cdot 2 + 200 \cdot 4 + 300 \cdot 3 + 400 \cdot 2}{11} = 245,45$

9. Determineu la moda i la mediana

- a) 5,6,6    c) 1,2,3,4,2    Solucions: a) Me=6, Mo=6    c) Me=2 Mo=2  
 b) 1,1,2,3    d) 3,2,3,2,2,2    b) Me=1,5 Mo=1    d) Me=2 Mo=2

10. Calcula la moda y la mediana en cada caso:

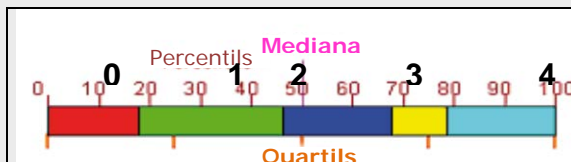
a	Marca	Fr	b	Marca	Fr
	10	2		100	2
	20	4		200	3
	30	3		300	4
	40	2		400	1

- Solucions:  
 a) Me=20 Mo=4  
 b) Me=250 Mo=300

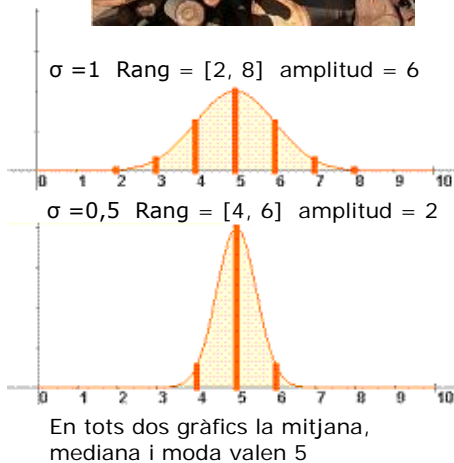
9. Calculeu la mediana, els quartils primer i tercer i el percentil 30 60 i 90 de les dades

4 1 3 3 2 3 1 3 3 4 0 0 0 4 4 3 0 3 0 3 2 1 0 0 4 3 0 1

Fem el recompte: 08→, 14→, 22→, 3→9 i 4→5 i barres de longitud proporcional a la freqüència per a cada valor. A més partim la longitud total de la barra en 2, 4 i 10 trossos per obtenir la mediana, els quartils i els percentils, tal com mostra la imatge.



Veiem que la mediana és entre el blau i el groc,  $(3 + 2)/2 = 2,5$ , Q1 entre vermell i verd, Q1 = 0, Q3 = 3, p30 = 1, P60 = 3 i P90 = 4



En la pràctica se sol usar la fórmula reduïda per al càlcul de la desviació típica.

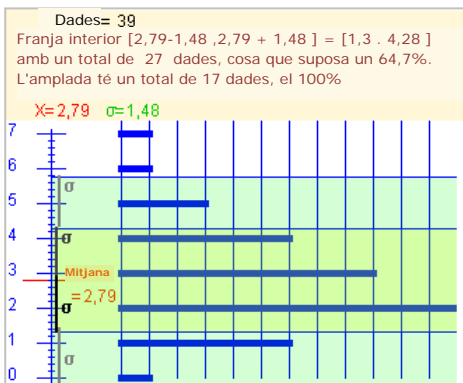
$$\sigma = \sqrt{\frac{\sum f_i X_i^2}{n} - \bar{X}^2}$$

Així, per a

Marqueu	Fr
4	3
5	3
6	2

Es té que la mitjana  $\bar{X} = 4,85$  y

$$\sigma = \sqrt{\frac{3 \cdot 4^2 + 3 \cdot 5^2 + 2 \cdot 6^2}{8} - 4,85^2}$$



## 3. Mesures de dispersió.

### Variància, desviació típica i rang

"L'estadística és una ciència segons la qual, si jo em menjo un pollastre i tu no en menges cap, ens hem menjat com a mitjana mig pollastre cada un".

L'estadística indicarà que tots mengen el mateix quan les mesures de dispersió siguin totes nul·les.

**Rang.** L'interval definit per la dada menor i la major. També es diu rang de la diferència entre la major i la menor de les dades.

**Variància.** La mitjana aritmètica dels quadrats de les diferències de les dades amb la mitjana.

$$\sigma^2 = \frac{\sum f_i (X_i - \bar{X})^2}{n} \text{ que equival a } \sigma^2 = \frac{\sum f_i (X_i)^2}{n} - (\bar{X})^2$$

**Desviació típica.** L'arrel quadrada positiva de la variància.

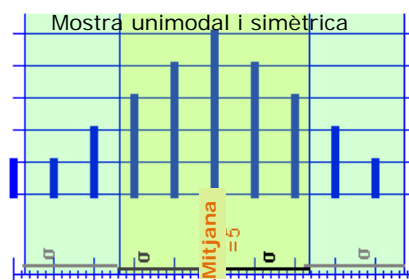
$$\sigma = \sqrt{\frac{\sum f_i (X_i - \bar{X})^2}{n}} \text{ o } \sigma = \sqrt{\frac{\sum f_i X_i^2}{n} - \bar{X}^2}$$

### Mesurar la dispersió

Aquest és l'objectiu d'aquestes mesures. Per exemple, les dades A = {20, 20}, B = {15, 20, 20, 25} tenen la mateixa mitjana, moda i mediana. En tots els casos igual a 20. Tanmateix, podeu comprovar que en cap de les tres mesures de dispersió definides a dalt no coincideixen.

### Mitjana i desviació típica.

Per a mostres unimodals (una sola moda) i gairebé simètriques, al voltant de la mitjana podem considerar un interval que contingui la majoria de les dades. Per exemple, per a una mostra amb mitjana 100 i desviació típica 10, la majoria de les dades seran entre 90 i 110, aproximadament el 68%; entre 80 i 120 hi haurà el 95% aproximadament. I gairebé tots entre 70 i 130. Hi ha una manera de distribució de dades anomenada **normal** que compleix amb l'anterior, i d'una manera o una altra, de totes les poblacions grans es poden extreure dades que s'hi ajusten. En cursos superiors veureu la importància d'aquestes distribucions.



**EXERCICIS resolta**

10. Calculeu la mitjana i la desviació típica en:

a) 200, 250

b) 175, 275

c) 250, 250

$$a) \bar{X} = \frac{250 + 200}{2} = 225 \quad \sigma = \sqrt{\frac{(250 - 225)^2 + (200 - 225)^2}{2}} = \sqrt{\frac{25^2 + 25^2}{2}} = 25$$

$$b) \bar{X} = \frac{175 + 275}{2} = 225 \quad \sigma = \sqrt{\frac{(175 - 225)^2 + (275 - 225)^2}{2}} = \sqrt{\frac{50^2 + 50^2}{2}} = 50$$

$$c) \bar{X} = \frac{250 + 250}{2} = 250 \quad \sigma = \sqrt{\frac{(250 - 250)^2 + (250 - 250)^2}{2}} = \sqrt{\frac{0^2 + 0^2}{2}} = 25$$

11. Calculeu la mitjana i la desviació típica en:

a) 7, 5, 3, 2, 4, 5

b) 20, 25, 20, 22, 21

$$a) \bar{X} = \frac{7 + 5 + 3 + 2 + 4 + 5}{6} = \frac{26}{6} = 4,33$$

$$\sigma = \sqrt{\frac{7^2 + 5^2 + 3^2 + 2^2 + 4^2 + 5^2}{6} - 4,33^2} = \sqrt{\frac{128}{6} - 18,75} = 1,59$$

$$b) \bar{X} = \frac{20 + 25 + 20 + 22 + 21}{5} = \frac{108}{5} = 21,6$$

$$\sigma = \sqrt{\frac{20^2 + 25^2 + 20^2 + 22^2 + 21^2}{5} - 21,6^2} = \sqrt{\frac{2350}{5} - 466,56} = 1,85$$

(Nota. - Observeu la fórmula utilitzada per a la desviació)

12. Organitzeu les dades següents en intervals de 10 cm des de 150 a 200. Amplieu la taula amb dues columnes, una per al producte de les marques amb les freqüències i una altra per al producte de les freqüències amb els quadrats de les diferències amb la mitjana. Calculeu la mitjana i la desviació típica.

**174 158 150 185 186 178 166 185 199**  
**183 175 173 175 164 173 178 179 164**  
**176 159 190 173 189 163 156 169**

	xi	fi	xi·fi	fi·(xi-X) <sup>2</sup>
[150,160)	155	5	775	1733,65
[160,170)	165	5	825	371,58
[170,180)	175	10	1750	19,02
[180,190)	185	7	1295	906,42
[190,200)	195	2	390	914,14
Total		29	5035	3944,82

Amb les dades de la taula és més fàcil, i es té:

**Mitjana i Desviació típica**

$$\bar{X} = \frac{5035}{29} = 173,62 \quad \sigma = \sqrt{\frac{3944,82}{29}} = 11,66$$

## 4. Representativitat

### Representativitat. Mostreig estratificat.

REPRESENTATIVITAT. Una mostra és representativa de la població quan hi podem trobar les mateixes proporcions de les característiques d'estudi que en el conjunt de la població. El procés d'escollir una mostra, a quins individus elegim com a representants de la població, és el punt important i d'això dependrà que l'estudi sigui útil o no (representatiu o no).

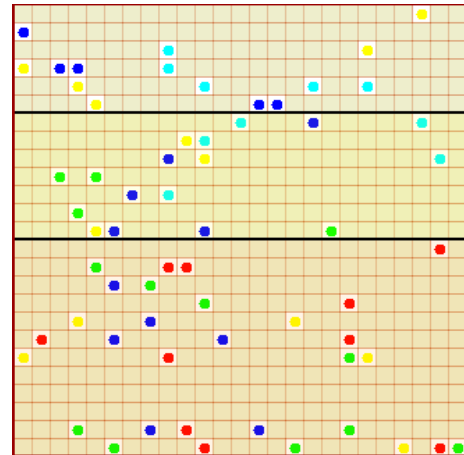
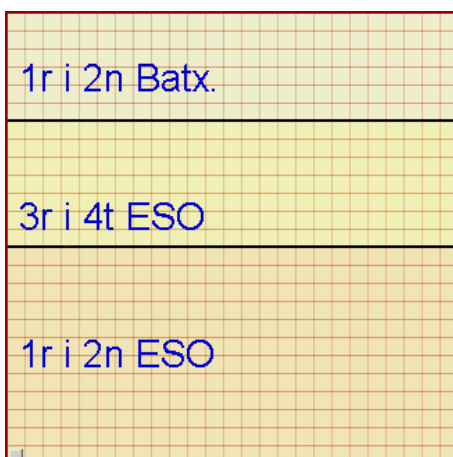
Escollir bé la mostra no és sinònim de representativitat, però escollir-la malament gairebé sí és sinònim de no representativitat.

Per exemple, si volem estudiar el poder adquisitiu d'una població, i només escollim individus d'una determinada zona, o principalment d'una determinada zona, la mostra amb tota seguretat no serà representativa. La mostra s'ha de triar prenent mostres d'individus proporcionals a la població de cada zona. Si hi ha tres zones amb 12.000, 18.000 i 20.000 habitants, la mostra haurà de tenir un 24% de la primera zona, 36% de la segona i 40% de l'última.

Aquest tipus de **mostreig**, escollint un repartiment proporcional als estrats, s'anomena **estratificat**.

### Exemple

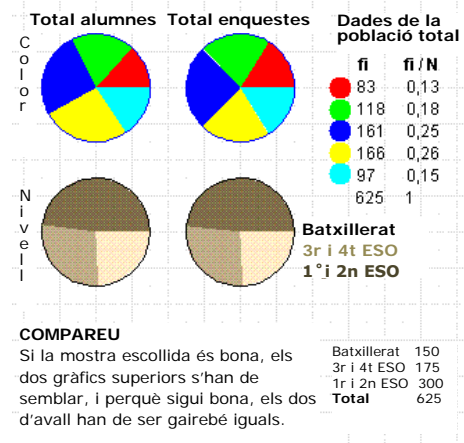
En la imatge teniu 625 quadres que representen l'alumnat d'un institut fictici



fi	fi / N	→ DADES DE LA MOSTRA
10	0,16	●
13	0,2	●
16	0,25	●
13	0,2	●
10	0,16	●
62	1	

**MOSTREIG**  
 Teniu en compte els alumnes que hi ha a cada nivell:  
 1r i 2n Batxillerat 150 alumnes  
 3r i 4t ESO 175 alumnes  
 1r i 2n ESO 300 alumnes

Batxillerat	15
2º cycle ESO	17
1º cycle ESO	30
Total	62
Percentatge	7,52%



A la dreta veiem la mostra estratificada que s'ha triat i el resultat de l'enquesta. Els últims diagrames de sectors comparen la realitat amb els resultats de l'enquesta

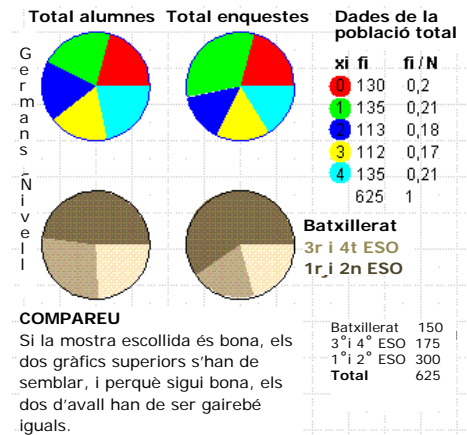
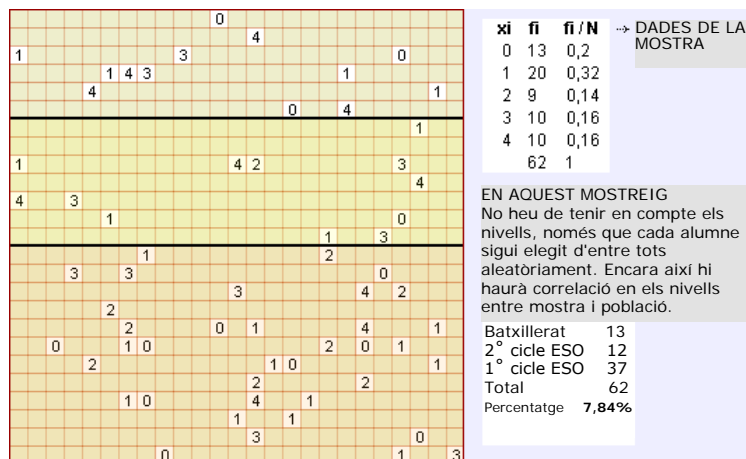
## Biaix. Mostreig aleatori

**Biaix.** Es diu que la mostra està esbiaixada quan hi ha diferència entre les dades de la mostra i les dades de tota la població.

Exemple: Trucades voluntàries. Aquestes enquestes tenen diverses fonts de biaix. Hi ha famílies que no tenen telèfon, i no tothom està disposat a assumir el cost de la trucada. Però, sobretot, el factor de desposta voluntària, els enquestats s'autoseleccionen. Solen contestar els que tenen una forta opinió negativa sobre el tema. L'enuig els anima a participar.

**Mostreig aleatori total.** A diferència de l'estratificat, que guarda les proporcions, aquesta forma d'escollir la mostra considera tota la població i tria individus aleatòriament. Es considera una bona manera de procedir.

En l'exemple següent s'ha escollit amb ordinador una mostra aleatòria total entre els 625 alumnes d'un institut, aquest mostreig pot sortir estratificat o no, en l'exemple no ha sortit gaire ben estratificat.



## EXERCICIS resolt

13. Una gran empresa té treballadors en quatre àrees. Operaris, representants, administració i direcció. Les condicions de feina són bastantes de diferents en cada àrea, per la qual cosa el grau de satisfacció no és igual en cada cas. Per esbrinar-ho, si hi ha 1.000, 500, 300 i 200 treballadors en les àrees d'operaris, representants, administratius i directius, quants en cal seleccionar de cada àrea per a una mostra de mida?
- 200
  - 100
  - 300
- a) D'un total de 2.000 empleats, els percentatges per a operaris, repartidors, administratius i directius són del 50%, 25%, 15% i 10%, que fa que la mostra prengui 100 operaris, 50 repartidors, 30 administratius i 20 directius
- 50, 25, 15 y 10.
  - 150, 75, 45 y 30

## 5. Estadística bidimensional

### Distribucions bidimensionals

Una distribució bidimensional és aquella en la qual intervenen dues variables X i Y que poden estar o no relacionades.

Podem representar conjuntament les dues variables en un **diagrama de dispersió** o **núvol de punts**, simplement fent correspondre un punt del pla a cada parell  $(x_i, y_i)$ .

Quan hi ha moltes dades i els parells de valors es repeteixen utilitzem una **taula de contingència**, com la de la dreta. En aquest cas la representació gràfica es fa mitjançant un gràfic tridimensional anomenat prismograma, o més sovint posant punts de mida proporcional a la freqüència.

### Correlació lineal

L'objectiu d'un estudi bidimensional és observar si hi ha algun tipus de relació entre la dues variables. Aquesta relació, que anomenarem **correlació** es pot apreciar mirant si el núvol de punts s'acosta o no a la gràfica d'una funció, en el nostre cas a una recta, per això parlarem de **correlació lineal**.

Com més s'acosti el núvol de punts a una recta més forta serà la correlació lineal, a més a més serà positiva o directa si la recta és creixent (si creix X creix Y) i negativa o inversa en cas contrari (si creix X decreix Y o viceversa).

Per quantificar aquesta relació emprarem un paràmetre **r**, el **coeficient de correlació lineal**, que pren valors entre -1 i 1. Com més s'acosti r als valors 1 o -1 més forta serà la correlació.

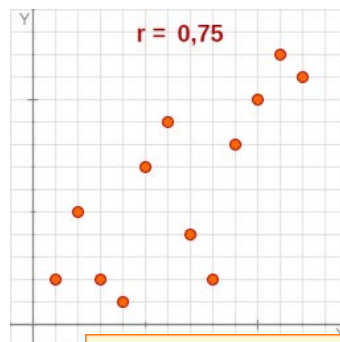
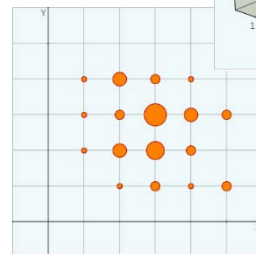
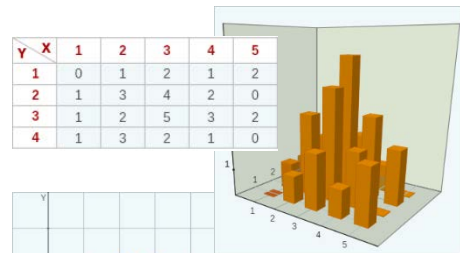
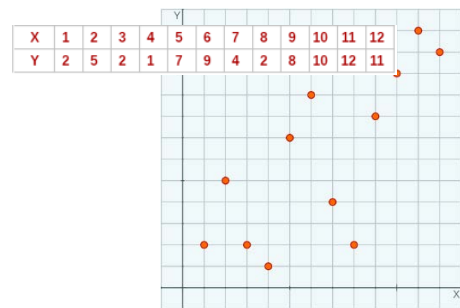
- Si **r=1** o **r=-1**, hi ha dependència funcional, els punts estan sobre una recta.
- Si **0,5 < r < 1** considerarem que la correlació és forta i directa o forta i inversa si **-1 < r < -0,5**.
- Si **r=0** o molt proper a 0, no hi ha correlació lineal entre les dues variables.

Disposem les dades en columnes i calculem la mitjana i la desviació típica de les dues distribucions X i Y. Per calcular r hem de calcular abans un altre paràmetre, la **covariància**, definida així:

$$\sigma_{XY} = \frac{\sum(x_i - \bar{x}) \cdot (y_i - \bar{y})}{n} = \frac{\sum x_i \cdot y_i}{n} - \bar{x} \cdot \bar{y}$$

I el **coeficient de correlació lineal**:

$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \quad (-1 \leq r \leq 1)$$



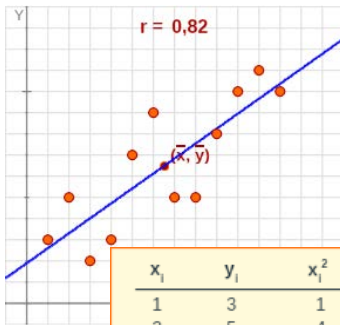
$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i \cdot y_i$
1	2	1	4	2
2	5	4	25	10
3	2	9	4	6
4	1	16	1	4
5	7	25	49	35
6	9	36	81	54
7	4	49	16	28
8	2	64	4	16
9	8	81	64	72
10	10	100	100	100
11	12	121	144	132
12	11	144	121	132
<b>78</b>	<b>73</b>	<b>650</b>	<b>613</b>	<b>591</b>

$$\bar{x} = \frac{78}{12} = 6,5 \quad \sigma_x = \sqrt{\frac{650}{12} - 6,5^2} = 3,45$$

$$\bar{y} = \frac{73}{12} = 6,08 \quad \sigma_y = \sqrt{\frac{613}{12} - 6,08^2} = 3,75$$

$$\sigma_{XY} = \frac{591}{12} - 6,5 \cdot 6,08 = 9,71$$

$$r = \frac{9,71}{3,45 \cdot 3,75} = 0,75$$



$x_i$	$y_i$	$x_i^2$	$x_i \cdot y_i$
1	3	1	3
2	5	4	10
3	2	9	6
4	3	16	12
5	7	25	35
6	9	36	54
7	5	49	35
8	5	64	40
9	8	81	72
10	10	100	100
11	11	121	121
12	10	144	120
78	78	650	608

$\bar{x} = 6,5$     $\bar{y} = 6,5$     $\sigma_x = 3,45$     $\sigma_{xy} = 8,42$

### Recta de regressió de Y sobre X

$$y = 6,5 + \frac{8,42}{3,45^2} (x - 6,5)$$

$$y = 0,7x + 1,9$$

## Rectes de regressió

Quan s'aprecia un cert grau de correlació entre les dues variables d'una distribució bidimensional, es busca la recta que millor s'ajusta al núvol de punts.

La recta de regressió de Y sobre X passa pel punt  $(\bar{x}, \bar{y})$ , centre de gravetat del núvol de punts. La seva equació és:

$$y = \bar{y} + \frac{\sigma_{XY}}{\sigma_X^2} (x - \bar{x})$$

El pendent  $\frac{\sigma_{XY}}{\sigma_X^2}$  és el **coeficient de regressió**.

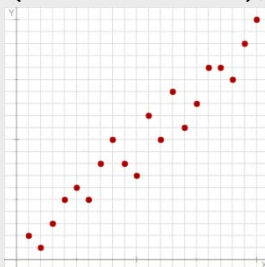
Aquesta recta de regressió serveix per estimar el valor que prendrà la variable Y per a un determinat valor de X. El valor d'aquesta estimació serà tant més fiable quant:

- Més s'acosti el coeficient de correlació a 1 o a -1.
- El valor quedi dins del rang de valors de X i més a prop estigui del centre de gravetat.

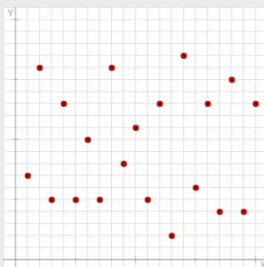
Si es pretén estimar valors de X a partir dels de Y haurem d'emprar una altra recta anàloga, la recta de regressió de X sobre Y.

## EXERCICIS resoltos

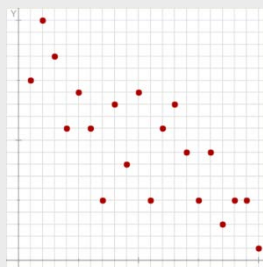
3. A la vista del núvol de punts indica si creus que la correlació és molt forta, forta (directa o inversa), dèbil o molt dèbil.



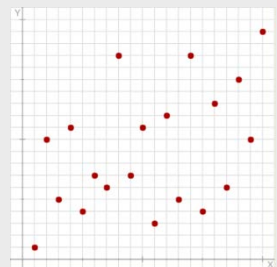
Molt forta i directa



Molt dèbil

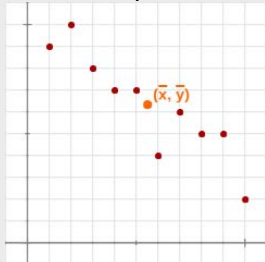


Forta i inversa



Dèbil

4. Una de les quatre equacions correspon a la recta de regressió de Y sobre X del núvol de punts. Indica la correcta.



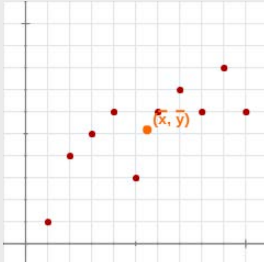
13.  $y = 0,7x + 2,4$

14.  $y = -0,7x + 10,2$

15.  $y = 0,7x + 10,2$

16.  $y = -0,7x + 2,4$

Sol: (b)



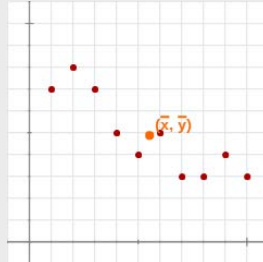
9.  $y = -0,5x + 2,5$

10.  $y = 0,5x + 8$

11.  $y = -0,5x + 8$

12.  $y = 0,5x + 2,5$

Sol: (d)



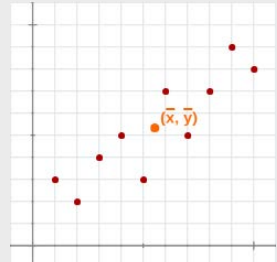
5.  $y = -0,5x + 2,2$

6.  $y = 0,5x + 2,2$

7.  $y = -0,5x + 7,7$

8.  $y = 0,5x + 7,7$

Sol: (c)



1.  $y = -0,7x + 1,4$

2.  $y = 0,7x + 9,2$

3.  $y = -0,7x + 9,2$

4.  $y = 0,7x + 1,4$

Sol: (d)

Alguns dels exercicis proposats a continuació estan elaborats a partir d'aquesta publicació d'INE. Podeu veure articles similars en

<http://www.ine.es/prodyser/pubfolletos.htm>

4/2007

# cifras INE

Boletín Informativo del Instituto Nacional de Estadística

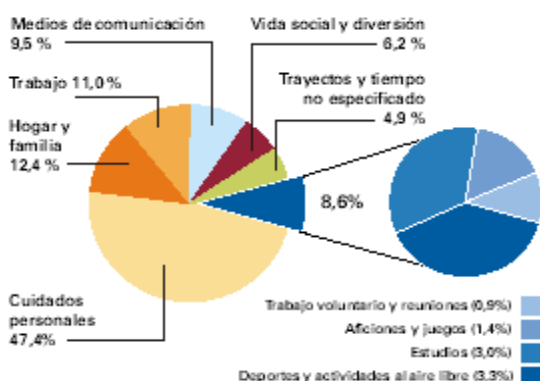


## Encuesta de Empleo del Tiempo

### Qué hacemos y durante cuánto tiempo



#### Distribución del tiempo por actividades



**NOTA:** Los informantes de 10 y más años han anotado las actividades realizadas en un día concreto (de lunes a domingo) elegido al azar. El tiempo así estimado se refiere a un "día promedio" obtenido al concentrar todas las actividades de todos los informantes en un solo día. Los datos que aquí se presentan se refieren a toda la población investigada, salvo que se indique expresamente lo contrario.

El Instituto Nacional de Estadística (INE) presenta en esta publicación algunos de los principales resultados de la **Encuesta de Empleo del Tiempo, primera y única encuesta de ámbito nacional sobre la utilización del tiempo**. Se realizó en España entre los años 2002 y 2003 de manera armonizada con las de otros países europeos, siguiendo las recomendaciones de la Oficina Estadística de la Unión Europea (Eurostat). Entre los años 1998 y 2004 otros países de la Unión llevaron a cabo investigaciones similares.

La encuesta facilita información, entre otras cosas, del **porcentaje de personas que realizan una determinada actividad en el transcurso del día y la duración media diaria dedicada a esa actividad por dichas personas**. Esta información primaria nos permite analizar con rigor la dimensión del trabajo no remunerado realizado por los hogares, la distribución de las responsabilidades familiares en el hogar, la participación de la población en actividades culturales y de ocio, etc. Por otra parte, la información recogida también permite comparar **datos nacionales de uso del tiempo en relación con los demás países europeos** que han realizado la encuesta.

Como principales resultados, cabe destacar el dato de que **las tareas domésticas y el cuidado de niños y ancianos son tareas eminentemente femeninas, ya que el 93% de las mujeres las realizan, frente al 70% de los varones**. En el contexto europeo, es de señalar la **primera posición de España en tiempo dedicado a caminar y pasear; pero también el último lugar por lo que se refiere a tiempo dedicado a la lectura**.

El INE quiere aprovechar esta ocasión para expresar su **agradecimiento a los cerca de 24.000 hogares de la muestra**, y pone a su disposición los resultados obtenidos.

Más información en: [www.ine.es](http://www.ine.es)

DEPÓSITO LEGAL: M-15643-2007

ISSN: 1578-2207

Nº P: 005-07-005-1

Fuentes estadísticas utilizadas:

Procedentes del INE: Encuesta de Empleo del Tiempo. La información internacional procede de Eurostat.





Per practicar

- Agrupa les següents variables: a) Pes, b) densitat, c) nre. de plantes dels edificis, d) Tipus de façana dels edificis, e) nre. de finestres, f) metres de façana, g) nre. d'habitants per edifici, h) tipus de porta principal.
- Escriu tres variables qualitatives que estiguin relacionades amb embarcacions.
- Escriu tres variables quantitatives discretes que estiguin relacionades amb avions.
- Escriu tres variables quantitatives contínues que estiguin relacionades amb trens.
- Si les freqüències per a R, V, A i T són  $R \rightarrow 3, V \rightarrow 2, A \rightarrow 4$  y  $T \rightarrow 1$  Quants graus li correspon a cada lletra en un gràfic de sectors?
- Fes una taula i un gràfic de sectors de les dades: R R A A R A R V N V R N
- Fes una taula i un gràfic de barres amb les dades:  
3 3 4 5 4 5 3 2 1 2 3 4 5 4 5 4 3 3 4 4
- Agrupa les dades següents en intervals y fes un histograma.

195 194 194 182 168 179 191 154 177 189  
184 187 155 167 177 187 161 171 190 162  
190 152 166 180 156 186 184 167 184 162

- Calcula la mitjana en cada cas:
  - 4, 6, 8
  - 4, 6, 8, 6
  - 100, 120, 180, 200

- Calcula la mitjana en cada cas:

a)		b)	
Marca	Fr	Marca	Fr
1	3	1000	3
2	5	2000	5
3	3	3000	3
4	2	4000	2

- Determina la moda i la mediana
  - 50,60,60
  - 12,12,22,32
  - 10,20,30,40,20
  - 35,25,35,25,25,25

- Calcula la moda i la mediana en cada cas:

a)		b)	
Marca	Fr	Marca	Fr
100	5	100	2
200	4	200	7
300	6	300	9
400	3	400	2

- Quina o quines de les dades següents es pot considerar una observació atípica en cadascuna de les dues sèries?
  - 4 5 6 5 7 8 4 5 8 7 5 12 6 7 6 5 4
  - 8 9 1 9 8 9 7 9 6 7 8
- Calcula la mediana, primer i tercer quartil i el percentil 90 de  
1 1 4 3 3 4 2 2 5 3 1 2 1 2 2 4 2 2 4 3 1
- Calcula la mediana, primer i tercer quartil i el percentil 20 de  
3 1 1 1 4 1 5 3 1 3 3 4 5 5 4 4 2 1 4 4
- Calcula la mitjana i la desviació típica en cadascun dels següents casos:  
100 i 100, 99 i 101, 110 i 90, 120 i 80

- Completa la taula amb les dades:

190 151 193 187 158 175 165 158 184 172  
197 161 157 157 183 180 150 161 182 169  
162 177 160 155 188 157 189 167 186 157

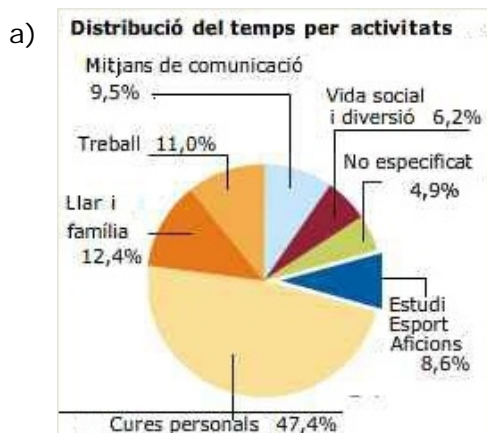
Interval	Marca Freq.			
	$x_i$	$f_i$	$f_i \cdot x_i$	$f_i(X-x_i)^2$
[150,160)	155			
[160,170)	165			
[170,180)	175			
[180,190)	185			
[190,200)	195			

- Determina la mitjana i la desviació típica, de les dades de la taula anterior.

- Determina els intervals  $(\bar{X} - \sigma, \bar{X} + \sigma)$  i  $(\bar{X} - 2\sigma, \bar{X} + 2\sigma)$  i el nombre d'elements que hi ha en cadascú.

Marca	Fr
0	5
1	4
2	7
3	3
4	2

20. Observa els següents gràfics i respon a las preguntes de cada un

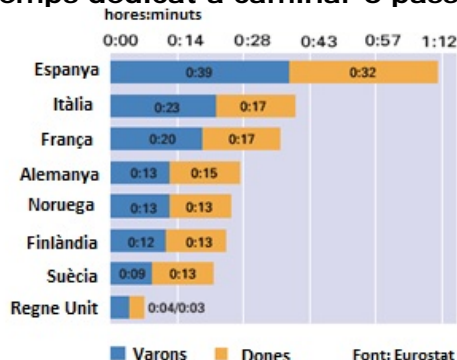


a1. Quina és la variable estudiada? i la freqüència?

a2. A quin grup d'activitats dediquem més temps els espanyols?

a3. Calcula quant de temps dediquem a la llar i la família quants graus ocupa aquest sector en el diagrama?

b) **Temps dedicat a caminar o passejar**



b1. En quins països passegen més les dones que els homes?

b2. Calcula el temps mitjà que es dedica a cada país a passejar.

b3. Quin país està en el percentil 50?

c)



c1. Creus que el dormir s'ha considerat com activitat de cura personal?

c2. A les 15:00 hi ha un màxim local en la gràfica A què és degut?

c3. A l'hora del dinar el 38% de les persones es dedica a la cura personal. Significa això que un 62% de les persones no dina?



d1. Quines són les comunitats en les quals es dedica menys temps a la vida social i a la diversió?

d2. Quant de temps dediquen a la diversió o a la vida social la major part de les comunitats?

d3. Quin és el temps mitjà que es dedica a Espanya a aquesta activitat?

21. Les notes de 8 alumnes en Llengua i Anglès van a ser:

Llengua (X)	4	4	4	5	7	8	9	9
Anglès (Y)	3	5	6	5	8	9	10	9

Dibuixa el núvol de punts i calcula el coeficient de correlació lineal.

22. Les hores setmanals que dediquen 10 personaes a fer esport i veure TV són:

Dep (X)	1	3	5	6	7	8	9	10	11	12
TV (Y)	14	14	13	10	8	9	4	8	5	5

Dibuixa el núvol de punts i calcula el coeficient de correlació lineal.

23. D'una distribució bidimensional coneixem  $\bar{x} = 8, \bar{y} = 7, \sigma_x = 1,5, \sigma_y = 2,7$  i  $\sigma_{xy} = 3,28$ . Calcula el coeficient de correlació lineal, la recta de regressió de Y sobre X i el valor estimat de y per  $x = 8$ .

24. D'una distribució bidimensional coneixem  $\bar{x} = 8, \bar{y} = 5, \sigma_x = 1,9, \sigma_y = 2,5$  i  $r = 0,83$ . Calcula la recta de regressió de Y sobre X i el valor estimat de i per  $x = 10$ . És fiable aquesta predicció?



## Per saber-ne més

### La professió d'infermeria.

Florence Nightingale (1820-1910), coneguda per ser la fundadora de la professió d'infermeria. Durant la guerra de Crimea es va adonar que la causa principal de les morts de ferits en combat era la falta de mesures sanitàries. En aplicar-les, la taxa de mortalitat va passar d'un 42,7% a un 2,2%. Gràcies a un ús eficaç de les dades va aconseguir modificar el sistema d'atenció sanitària un cop va tornar a la Gran Bretanya. Va canviar el sistema de registre de dades i va ser una de les primeres persones a utilitzar les gràfiques estadístiques per representar les dades d'una forma senzilla de manera que fins i tot els parlamentaris i els generals poguessin entendre-les.

Per a Florence, les dades no eren una cosa abstracta, eren una forma de poder salvar vides humanes.

### El pare de l'estadística.

Sir Ronald A. Fisher (1890-1962) està considerat el pare de l'estadística. Els escrits de Fisher van ajudar a organitzar l'estadística com a camp d'estudi precís els mètodes del qual s'apliquen a problemes pràctics de moltes disciplines. Com gairebé tots els pioners en estadística, els seus treballs van néixer de la necessitat de resoldre problemes pràctics.

### Inferència estadística

L'estadística desenvolupada en aquest tema és el que es coneix com a estadística descriptiva, s'hi recull informació i es fan càlculs que descriuen com estan repartits. Posem el cas que una mostra escollida a l'atzar ens dona una mitjana. La verdadera mitjana és pròxima a la de la mostra? Si considero un interval al voltant de la mitjana mostral, la verdadera, amb quina probabilitat hi serà o no? D'aquestes preguntes i d'altres s'encarrega la inferència estadística.

### Principals camps d'aplicació de l'estadística



L'estadística s'aplica en molts camps, com a la **indústria i les empreses**. Per al control de qualitat en la producció en cadena, per a l'anàlisi de mercats, per a l'estudi del preu de venda al públic dels articles fabricats, en gestió financera...

En la part dreta s'esmenten algunes altres aplicacions

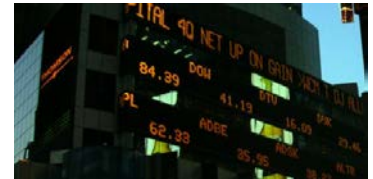
### Alguns camps d'aplicació de l'estadística

#### Administració pública



A través de les delegacions territorials i provincials, es recullen dades per analitzar-les i sotmetre-les a processos estadístics. D'aquesta manera se saben dades referides a naixements, defuncions, matrimonis, preus, salaris, feina, ensenyament, sanitat... Totes aquestes dades se solen publicar per l'INE

#### Economia.



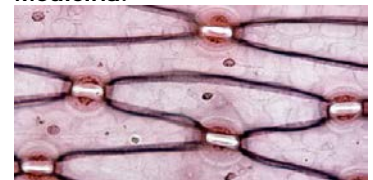
En aquest camp és imprescindible, sobretot en macromagnituds.

#### Psicologia.



La majoria dels treballs científics en psicologia experimental tenen com a principal eina d'estudi l'estadística.

#### Medicina.



En qualsevol estudi experimental d'aquestes àrees hi ha una assignatura específica anomenada bioestadística per cobrir aquests estudis experimentals. En genètica i antropometria trobem dos dels camps de més aplicació.



## Recordeu el més important

**Població.** Alumnes d'un institut fictici.

**Mostra.** Alumnes enquestats

**Variables estadístiques:** Qualitativa, color preferit; Quantitativa discreta, nre. de germans i quantitativa contínua, altura.

Considerem les dues mostres següents:

Nre. de germans: 4 3 2 3 1 2 0 2 0 1 2 3  
1 2 4 0 1 1 4 1 1 4 0 4 2 0 4 1

Altura: 182 172 157 194 150 166 163 196  
167 199 172 185 172 168 173 160 162 173  
161 192 156 164 173 180 193 172

**Recompte de dades:**

$x_i$	$f$	Intervalo	$x_i$	$f_i$
0	5	[150,160)	155	3
1	8	[160,170)	165	8
2	6	[170,180)	175	7
3	3	[180,190)	185	3
4	6	[190,200)	195	5
	28		Total	26

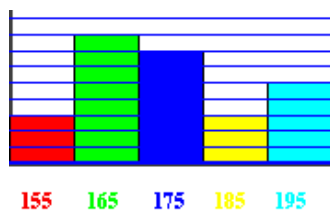
Gràfics de sectors i barres

Nre. de germans



Altura.

**Histograma**



**Mitjana i moda i desviació típica**

$x_i$	$f_i$	$x_i \cdot f_i$	$f_i \cdot (x_i - \bar{x})^2$
0	5	0	0
1	8	8	6,37
2	6	12	0,06
3	3	9	3,67
4	6	24	26,64
Total	28	53	54,67

Mitjana:

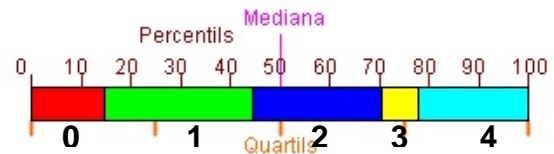
$$\bar{x} = \frac{53}{28} = 1,89$$

Moda = 1

Desviació típica:

$$\sigma = \sqrt{\frac{54,67}{28}} = 1,39$$

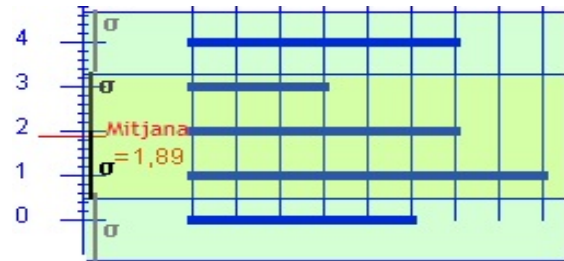
**Quartil, mediana, percentil**



Me=2, Q1=1, Q3=3, P20=1, P60=2, P90=4

Recorregut. De 0 a 4, d'amplitud 4

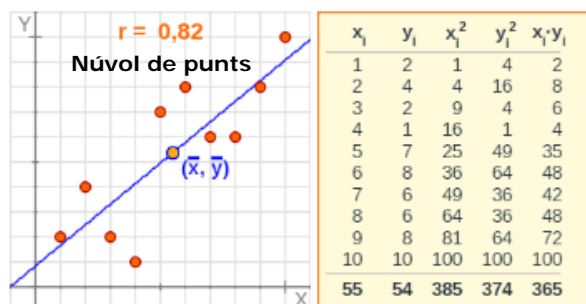
**Mitjana i desviació** En el nostre exemple, 17 de 28 dades no s'allunyen de la mitjana més de la desviació típica, són el 60,7%, i el 100% no s'allunyen de la mitjana més de dues vegades la desviació.



**Representativitat**

Una mostra és representativa de la població quan en ella podem trobar les mateixes proporcions de les característiques d'estudi que en el conjunt de la població.

**Distribucions bidimensionals**



Covariància:  $\sigma_{xy} = \frac{\sum x_i \cdot y_i}{n} - \bar{x} \cdot \bar{y}$   $x = 5,5$   $\sigma_x = 2,87$   $\sigma_{xy} = 6,8$

Coefficient de correlació lineal  $r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$   $y = 5,4$   $\sigma_y = 2,87$

Recta de regressió de Y sobre X

$$y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x})$$

$$r = \frac{6,8}{2,87 \cdot 2,87} = 0,82$$

$$-1 \leq r \leq 1$$

$$y = 0,82x + 0,87$$

Autoavaluació



1. Quants graus corresponen al valor de freqüència 3?

$x_i$	$f_i$
0	1
1	2
2	3
3	8
4	2

2. La mediana de la distribució anterior és?

3. Quina és la moda ?

$x_i$	$f_i$
3	1
4	3
5	4
6	2

4. Quin percentatge de la mostra correspon a les dues primeres marques ?

$x_i$	$f_i$
0	1
1	2
2	3
3	8
4	2

5. Quin és el percentil més petit que deixa per sota els valors menors a 3?

$x_i$	$f_i$
0	1
1	2
2	3
3	8
4	2

6. Quina és la mitjana?

$x_i$	$f_i$
155	2
165	2
175	8
185	1
195	5

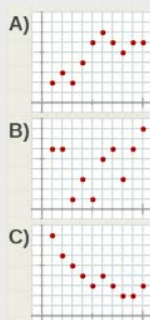
7. Quina és la desviació típica de les dades anteriors?

8. Associa cada núvol de punts amb el seu corresponent coeficient de correlació,  $r$ .

- 1) 0.32    2) 0.79    3) -0,88

9. Calculeu la covariància.

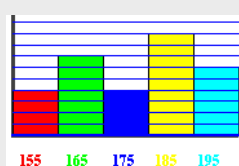
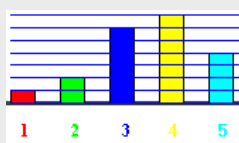
10. El centre de gravetat de una distribució bidimensional és  $(4,5, 3,75)$  i el pendent de la recta de regressió de  $Y$  sobre  $X$  és 0,57. Estima el valor de  $y$  para  $x=7$ .



$x_i$	$y_i$	$x_i \cdot y_i$
1	1	1
2	2	4
3	3	9
4	5	20
5	4	20
6	5	30
7	5	35
8	5	40

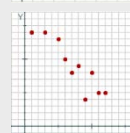
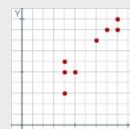
## Solucions dels exercicis per practicar

- Qualitatius: d) h)  
Quantitatius discretes c) e) g)  
C. contínues: a) b) f)
- Propulsió, Càrrega, Tipus de travessa
- Nre. de passatgers, nre. rodes, nre. finestres
- Velocitat màxima, càrrega màxima, potència.
- $R \rightarrow 108^\circ$ ,  $V \rightarrow 72^\circ$ ,  $A \rightarrow 144^\circ$  i  $T \rightarrow 36^\circ$
- $R \rightarrow 5$ ,  
 $A \rightarrow 3$ ,  
 $V \rightarrow 2$ ,  
 $N \rightarrow 2$
- $1 \rightarrow 1$ ,  $2 \rightarrow 2$ ,  $3 \rightarrow 6$ ,  
 $4 \rightarrow 7$ ,  $5 \rightarrow 4$
- Interval  $x_i$   $f_i$   
[150,160) 155 4  
[160,170) 165 7  
[170,180) 175 4  
[180,190) 185 9  
[190,200) 195 6
- a) 6 b) 6 c) 150
- a) 2.3 b) 2307
- a)  $Mo=60$ ,  $Me=60$  b)  $Mo=12$ ,  $Me=17$   
c)  $Mo=20$ ,  $Me=20$   
d)  $Mo=25$   $Me=25$
- a)  $Mo=300$ ,  $Me=250$  b)  $Mo=300$ ,  
 $Me=300$
- a) 12 b) 1
- $Me=2$ ,  $Q1=2$ ,  $Q3=3$ ,  $P90=4$
- $Me=3$ ,  $Q1=1$ ,  $Q3=4$  y  $P20=1$
- La mitjana és 100 en els 4, i la desviació 0, 1, 10 i 20.



Interval	Marca	Frec.		
	$x_i$	$f_i$	$f_i \cdot x_i$	$f_i(X-x_i)^2$
[150,160)	155	9	1395	2401
[160,170)	165	7	1155	280,77
[170,180)	175	3	525	40,33
[180,190)	185	8	1480	1494,22
[190,200)	195	3	585	1680,33
		30	5140	5896,66

- $\bar{x} = 171$ ,  $\hat{\sigma} = 14,02$
- A (0.42, 2.9) hi ha 11,  
i a (-0.88, 4.14) tots
- a1) variable: activitats. Fr: percentatge de temps diari que es dedica a cada activitat  
a2) cures personals  
a3) 2h 58m 34s 44,64 graus  
b1) Alemanya, Suècia i Finlàndia  
b2) E35,5 I20, F18,5 A14 N13 F12,5 S11 R3,5 en minuts  
b3) França  
c1) Sí. c2) menjar i descans  
c3) No, aquesta punxa ocupa dues hores i alguns mengen en mitja hora  
d1) País Basc, Catalunya i Madrid  
d2) entre 1:30 i 1:40 hores: minuts  
d3) 1:29
- $r = 0,93$
- $r = -0,91$
- $r = 0,81$   
 $y = 1,46x - 1,66$   
 $x = 8$   $y' = 7$
- $y = 1,09x - 3,73$   
 $x = 10$ ,  $y' = 7,18$



## Solucions AUTOAVALUACIÓ

- |               |                |
|---------------|----------------|
| 1. Sol 67,5°  | 6. Sol 177,78  |
| 2. Sol 3      | 7. Sol 12,83   |
| 3. Sol 5      | 8. 1-B 2-A 3-C |
| 4. Sol 18,75% | 9. Sol 3       |
| 5. Sol 37     | 10. Sol 5,18   |