

Objectius

En aquesta quinzena aprendràs a:

- Distingir els conceptes de població i mostra.
- Diferenciar els tres tipus de variables estadístiques.
- Fer recomptes i gràfics.
- Calcular i interpretar les mesures estadístiques de centralització i de posició.
- Calcular les principals mesures de dispersió.
- Entendre la importància de l'elecció de la mostra per a què sigui representativa.

1. Estadística descriptiva	pàg. 204
Població i mostra	
Variables estadístiques	
Gràfics variables qualitatives	
Gràfics variables quantitatives discretes	
Gràfics variables quantitatives contínues	
2. Mesures de centralització	pàg. 207
Mitjana, moda i mediana	
Evolució de la mitjana	
Evolució de la mediana	
Mitjana i mediana comparades	
3. Mesures de posició	pàg. 210
Quartils i Percentils	
Diagrames de caixa	
4. Mesures de dispersió	pàg. 212
Desviació típica i recorregut	
Càlcul de les mesures de dispersió	
La mitjana i la desviació típica	
5. Representativitat de les mostres.....	pàg. 214
Mostreig aleatori	
Mostreig estratificat	

Exercicis per practicar

Per saber-ne més

Resum

Autoavaluació

Abans de començar

Recorda

El curs passat ja vas estudiar estadística, i en moltes ocasions has fet estadística encara que no te n'hagis donat compte. Anem a veure alguns exemples.

Nota mitjana

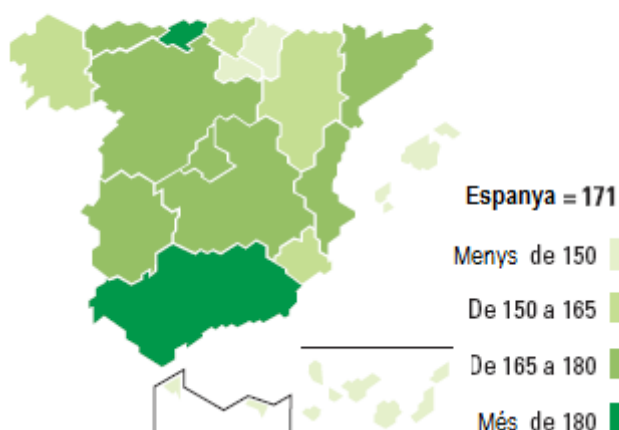
Al llarg d'un curs escolar tindràs moltes ocasions on calcular aquest valor. Si una nota depèn de dos exàmens i en un tens un 4, intentaràs treure almenys un 6 en l'altre.

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

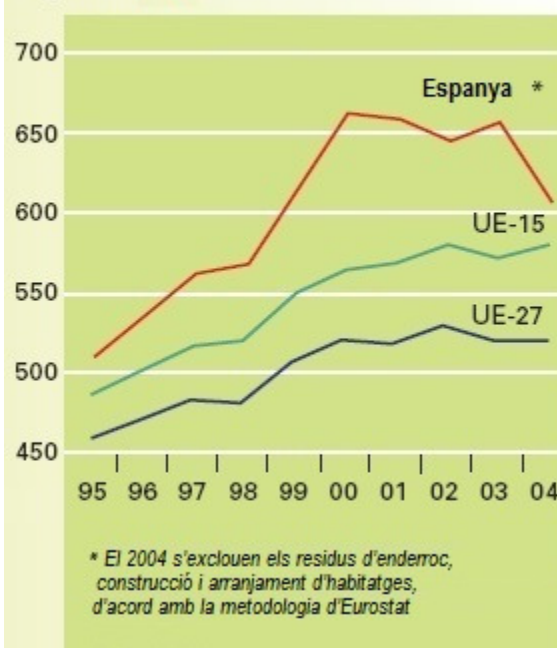
Futbol

El jugador que més gols ha marcat, el porter que menys ha encaixat. La classificació de la lliga. La millor meitat de lliga. Els llocs de competicions europees, els de descens, núm. de vegades internacional, núm. de fases finals, minuts jugats, llançaments a porta, faltes.

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



Residus urbans (kg/hab./any)



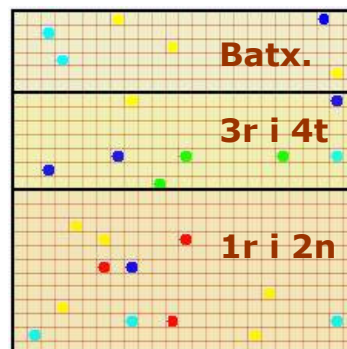
1. Estadística descriptiva

Població i mostra.

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

En la pràctica és freqüent haver de recórrer a una mostra per inferir dades de la població. La **mostra** és un subconjunt de la població, seleccionada de manera que posi de manifest les característiques de la mateixa, 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 quadrat representa cadascun dels alumnes d'un institut fictici i els preguntem sobre el seu color preferit, el total dels quadrats és la població, 625 alumnes, i els 26 enquestats constitueixen la mostra.

Variables estadístiques

La característica a estudiar 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 en forma numèrica sinó com una categoria o atribut.

Les variables quantitatives són les que poden expressar-se numèricament, i a la vegada poden ser:

- ✓ Quantitatives **discretes**, si sols 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 vius, són **variables estadístiques qualitatives**.
- El nre. d'ordinadors a casa, o de televisors i el nre. 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

Gràfics en variables qualitatives.

El diagrama de sectors és el més indicat per 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%) els que correspon a un gràfic de sectors amb (A, 180°), (B,108°) i (C, 72°).

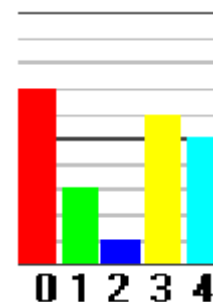
$$\frac{\text{freqüència}}{\text{nre. total de dades}} = \frac{\text{graus del sector}}{360}$$

Gràfics en variables discretes.

Diagrama de barres. N'hi haurà prou que observis 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



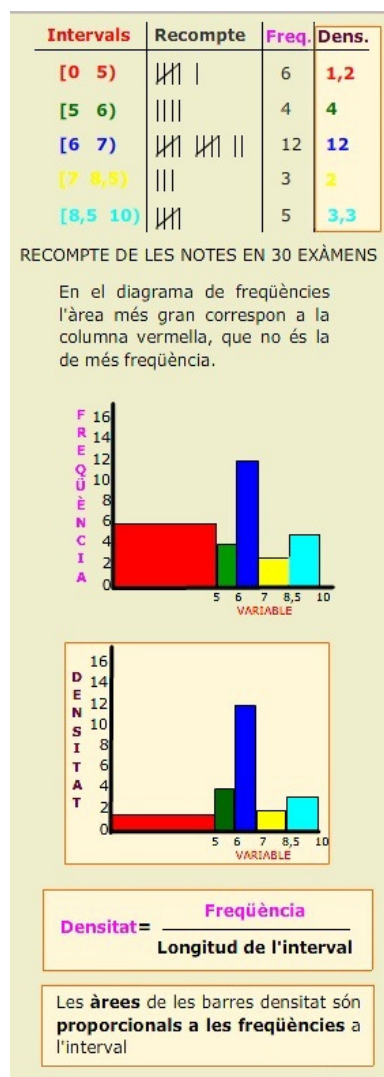
els correspon el gràfic de la dreta.

Gràfics en variables contínues.

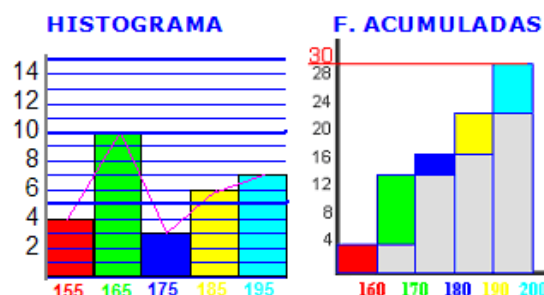
Histograma. Les dades es representen per rectangles de base l'amplitud de l'interval representat i amb l'altura que ens indica la freqüència absoluta, si tots els intervals són de la mateixa amplitud. Si no és així, les altures es calculen de manera que les àrees siguin proporcionals a les freqüències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

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

També s'acostuma a 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

- Classifica els següents exemples de variables estadístiques: Longitud d'un camió, Càrrega màxima, nre. de rodes, nre. d'eixos, tipus de camió, marques de pneumàtics, tipus de tapisseria, nre. de portes, altura màxima.

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

C. discretes: Nre. de rodes, nre. d'eixos, nre. de portes

C. contínues: Longitud d'un camió, Càrrega màxima i altura màxima.

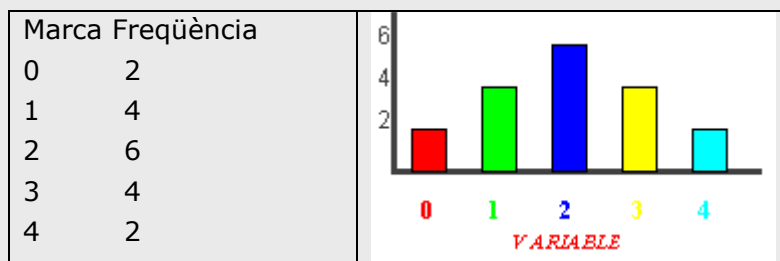
- Calcula els graus que corresponen a cada valor en un gràfic de sectors fet a partir de les dades: R, R, V, V, V, V, V, A, A y A

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

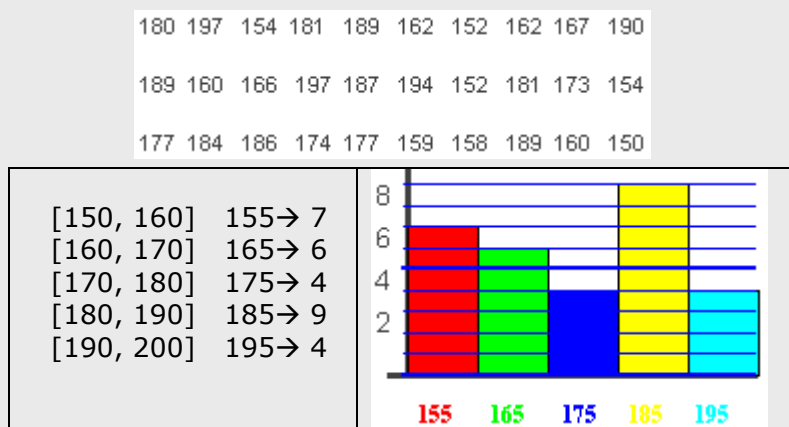
$$\frac{2}{10} = \frac{\text{Graus R}}{360}, \quad \frac{5}{10} = \frac{\text{Graus V}}{360} \quad \text{i} \quad \frac{3}{10} = \frac{\text{Graus A}}{360} \quad \text{i obtenim}$$

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

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

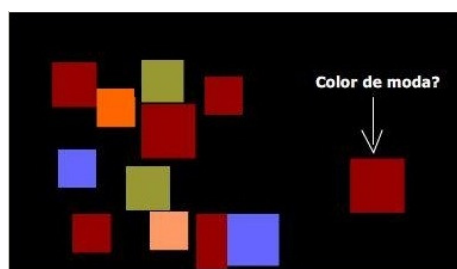


- Classifica les dades en intervals i dibuixa un histograma adequat.



2. Mesures de centralització

1. AVALUACIÓ	
5	NOTA MITJANA 5,5
6	
4	
1	
9	
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.

$x_i \rightarrow fr$
6 \rightarrow 4
7 \rightarrow 2
8 \rightarrow 2
9 \rightarrow 1

Si ordenem les dades, i com que el nre. de dades és senar, just el 7 queda en el centre.

6 6 6 6 7 7 8 8 1

Si les dades ordenades fossin 6,7,8,6,7,6,8,6,5 com que n'hi ha una quantitat parell, dues d'elles estarien en 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 **3, 4 y 5**, $(3+4+5)/3 = 12/3 = 4$;
per **1, 1, 4, 8, 8 y 8**, $(1+1+4+8+8+8)/6=5$.

$$\text{Mitjana} = \frac{x_1f_1 + x_2f_2 + \dots + x_nf_n}{N}$$

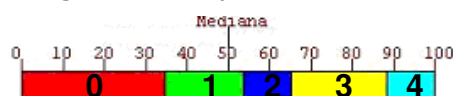
Moda. Si una observació es repeteix més que qualsevol altra, serà considerada la moda d'aquestes dades. Per exemple, si tenim les observacions 6,7,8,6,7,6,8,6,9 i agrupem les dades 6 \rightarrow 4, 7 \rightarrow 2, 8 \rightarrow 2 y 9 \rightarrow 1 veiem clarament que el valor 6 apareix més que cap altre. En aquest cas la moda és 6.

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

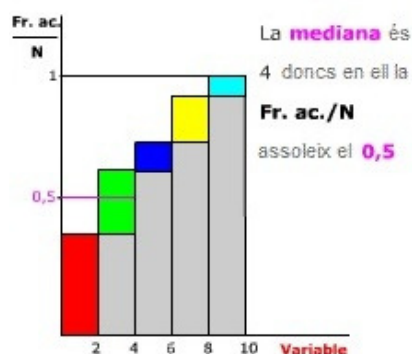
Mediana. El nombre tal que la meitat de les observacions són majors que ell i l'altra meitat menors.

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

Per a quantitats majors, haurem d'agrupar primer les dades en una taula. I determinar segments de longitud proporcional a la seva freqüència, disposar-los de forma lineal i marcar el centre com es pot veure en el següent exemple.



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



Evolució de la Mitjana.

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 de la 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 mitjana 5, ara necessitem una dada de valor 7 perquè la mitjana pugi a 5,1. (Figura de la dreta).

3 Per un conjunt de dades amb mitjana 5, si afegim un nou conjunt amb mitjana 5, per exemple 6 i 4, el conjunt resultant conserva la mitjana.

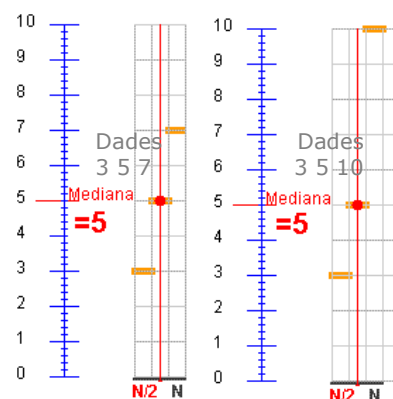
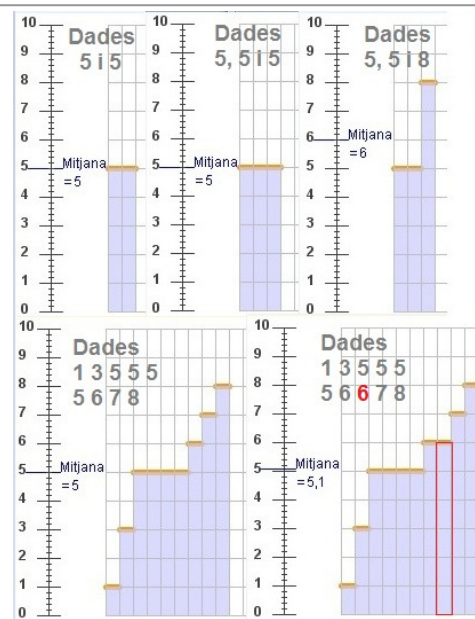
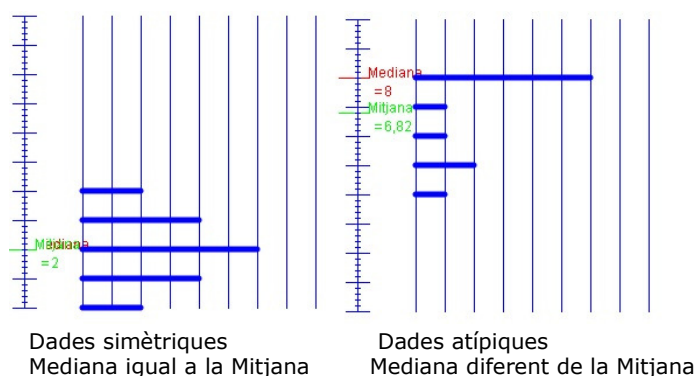
Evolució de la Mediana.

1 La mediana, per les dades 2, 3 i 4 és $Me=3$. Si canviem el 4 per 5 o per 6 o per qualsevol altre valor major segueix essent $Me=3$.

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 és el mateix per a la mediana, que passa a ser en ambdós casos 6. No obstant això 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, fan més gran la mitjana i no afecten a la mediana. Si les dades estiguessin repartides simètricament respecte a un valor, aquest valor seria a la vegada la mitjana i la mediana. En canvi, si els valors a un costat de la mediana estan més allunyats d'ella que els de l'altre costat, la mitjana es desplaça cap a aquests valors allunyats que la fan més gran. 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 i 28.

$Me = 24$

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

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

3. Per a 20, 24, 28 i 100

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

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

1 $\bar{X} = 24$

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

3 $\bar{X} = 43$

EXERCICIS resolts

5. Calcula la mitjana en cada cas:

a) 4, 6, 8

Solucions: a) $(4+6+8)/3 = 6$

b) 4, 6, 8, 6

b) $(4+6+8+6)/4 = 6$

c) 100, 120, 180, 200

c) $(100+120+180+200)/4 = 150$

6. Calcula la mitjana en cada cas:

a

Marca	Fr
10	2
20	4
30	3
40	2

b

Marca	Fr
100	2
200	4
300	3
400	2

a) $\bar{X} = \frac{10 \cdot 2 + 20 \cdot 4 + 30 \cdot 3 + 40 \cdot 2}{11} = 24,54$

b) $\bar{X} = \frac{100 \cdot 2 + 200 \cdot 4 + 300 \cdot 3 + 400 \cdot 2}{11} = 245,45$

7. Determina 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

8. Calcula la moda i la mediana en cada cas:

a

Marca	Fr
10	2
20	4
30	3
40	2

b

Marca	Fr
100	2
200	3
300	4
400	1

Solucions:

a) Me=20 Mo=4

b) Me=250 Mo=300

9. S'han mesurat les altures en cm d'un grup de 30 persones i s'han obtingut les dades següents:

Altura en cm	f_i
(150,160]	7
(160,170]	9
(170,180]	10
(180,190]	3
(190,200]	1

Calcula la mitjana, la moda i la mediana.

a) Completeu la taula afegint una columna per a x_i i dues més per a $x_i \cdot f_i$ i per a les freqüències acumulades.

Altura en cm	x_i	f_i	$x_i \cdot f_i$	F_i
(150,160]	155	7	1085	7
(160,170]	165	9	1485	16
(170,180]	175	10	1750	26
(180,190]	185	3	555	29
(190,200]	195	1	195	30
SUMA:		30	5070	

← Me

← Mo

$$\bar{x} = \frac{5070}{30} = 169 \quad \text{Me} = 165 \quad \text{Mo} = 175$$

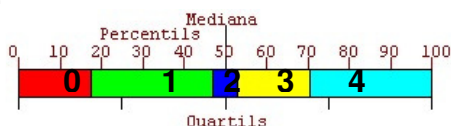
3. Mesures de posició

Quartils percentils

Donat un conjunt de dades numèriques corresponents a un estudi estadístic, si les ordenem de forma creixent i considerem la que estigui en el centre, ens estarem fixant en la **mediana**. És la primera que supera (o iguala) el 50% de valors, però també podem fixar-nos en altres posicions:

- Si ens fixem en el primer valor que supera o iguala el 25% o el 75%, estem parlant del **primer** o del **tercer quartil**, Q_1 i Q_3 .
- Per a altres valors com el 10%, o el 80% parlem de **percentils**, P_{10} i P_{80} .

Exemple. Per a la variable de valors 0, 1, 2, 3, 4, i freqüències $0 \rightarrow 9$, $1 \rightarrow 5$, $2 \rightarrow 3$, $3 \rightarrow 6$, $4 \rightarrow 3$, 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.

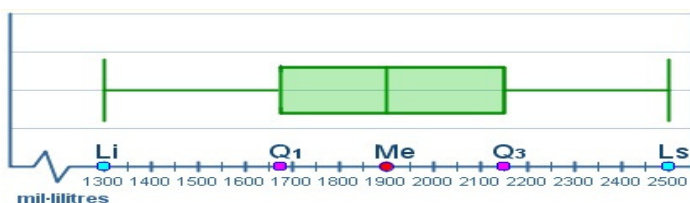


Diagrames de caixa

A partir del valor de la mediana i els quartils es poden representar les distribucions estadístiques mitjançant els anomenats "diagrames de caixa i bigotis".

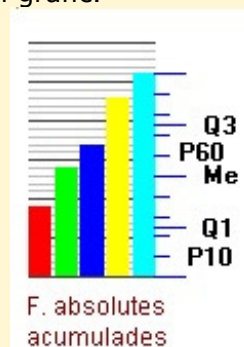
Observa com es construeix amb les dades de la taula de la dreta. Un cop ordenades les dades, es calculen els valors mínim i màxim, els quartils i la mediana.

mín=1300 Q_1 =1675 Me=1900 Q_3 =2150 màx=2500



Es situen aquests valors sobre l'eix d'abscisses i es dibuixa la "caixa" des del primer al tercer quartil (el recorregut *interquartílic*), i els "bigotis" com indica la figura.

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



La taula mostra el consum diari d'aigua, en ml, dels 20 alumnes d'una classe.

Joan	1650	Lluís	1300	Mín.
Lluís	1300	Tere	1500	
Alma	2400	Maya	1600	
Toni	2000	Marta	1650	
Rosa	2100	Juan	1650	Q_1
Lupe	1700	Lupe	1700	
Cesc	1900	David	1750	
Tere	1500	Pep	1850	
Iris	1900	Alex	1900	
Pep	1850	Iris	1900	Me
Marc	2000	Cesc	1900	
Lisa	2200	Marc	2000	
Juli	2300	Toni	2000	
Maya	1600	Omar	2100	
Alex	1900	Rosa	2100	Q_3
Albert	2500	Lisa	2200	
Rita	2200	Rita	2200	
Marta	1650	Juli	2300	
Omar	2100	Alma	2300	
David	1750	Albert	2500	Màx.

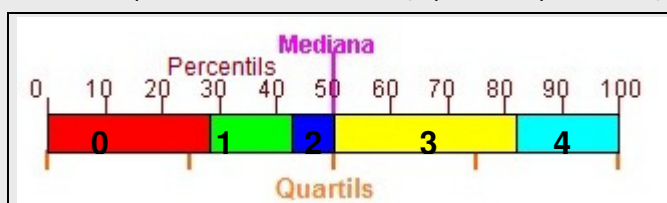
NOTA: La longitud dels bigotis no ha de passar una vegada i mitja la de la caixa, si hi ha valors extrems que superen aquesta mesura es dibuixen com punts aïllats.

EXERCICIS resoltos

10. Calcula la mediana, quartils primer i 3r, 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

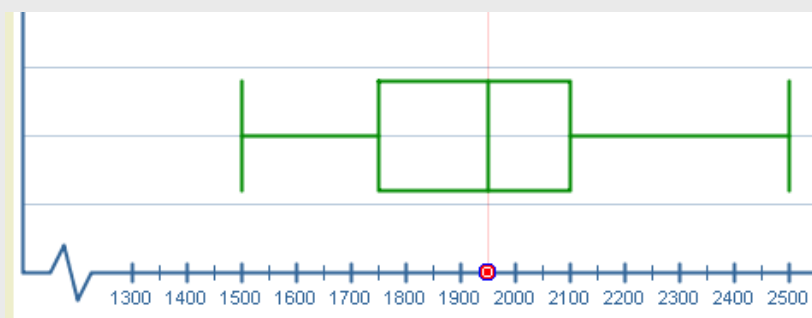
Fem el recompte: 0→8, 1→4, 2→2, 3→9 y 4→5 i barres de longitud proporcional a la freqüència per a cada valor. A més a més partim la longitud total de la barra en 2, 4 i 10 trossos per obtenir la mediana, quartils i percentils, tal com es pot veure a la imatge.



Es pot veure que la mediana està entre el blau i el groc, $(3+2)/2 = 2.5$, Q_1 en el vermell, Q_3 en groc.

$Q_1=0$ $Me=2,5$ $Q_3=3$
 $P_{30}=1$ $P_{60}=3$ i $P_{90}=4$

11. Analitza el següent diagrama de caixa i calcula, a partir d'ell, els valors màxim i mínim, la mediana i els quartils.



Mínim = 1500

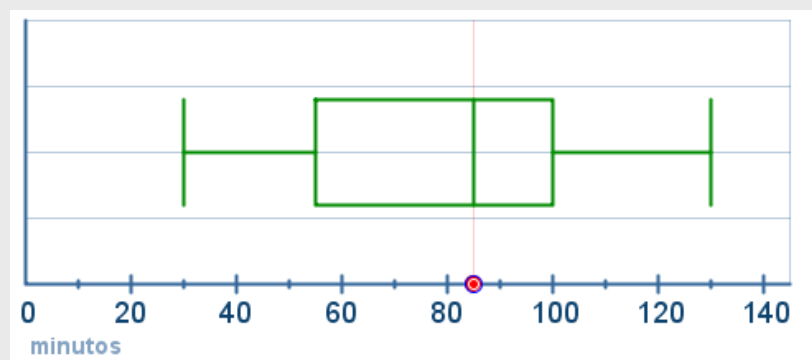
$Q_1 = 1750$

$Me = 1950$

$Q_3 = 2100$

Màxim = 2500

12. Analitza el següent diagrama de caixa. Mostra els minuts que tarda en fer efecte un medicament en una població. Interpreta la informació que presenta i respon les preguntes.



Mínim = 30

$Q_1 = 55$

$Me = 85$

$Q_3 = 100$

Màxim = 130

- A quin percentatge de la població li havia fet efecte després de 30 minuts?
 - Després de quants minuts havia fet efecte al 50 % de la població?
 - Quants minuts va tardar en fer efecte al 100% de la població?
 - A quin percentatge li havia fet efecte als 55 minuts?
- Quant va tardar en fer efecte a les tres quartes parts de la població?

RESPOSTES: a) Al 0%, 30 és el valor mínim. b) als 85 minuts (la mediana)
 c) 130 minuts (valor màxim) d) 55 és el primer quartil, al 25%
 e) 100 minuts, $\frac{3}{4}$ parts són el 75%

4. Mesures de dispersió.

Rang, variància i desviació típica

"L'estadística és una ciència segons la qual, si jo em menjo un pollastre i tu no te'n menges cap, ens hem menjat com a terme mitjà 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 menor i la major dada. També s'anomena rang a 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.

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

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

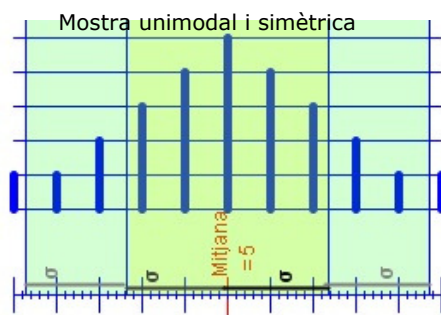
$$\sigma = \sqrt{\frac{\sum f_i \cdot (X_i - \bar{X})^2}{n}} \quad \text{o} \quad \sigma = \sqrt{\frac{\sum f_i \cdot X_i^2}{n} - \bar{X}^2}$$

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. Però pots comprovar que en cap de les tres mesures de dispersió definides a dalt coincideixen.

Mitjana i desviació típica.

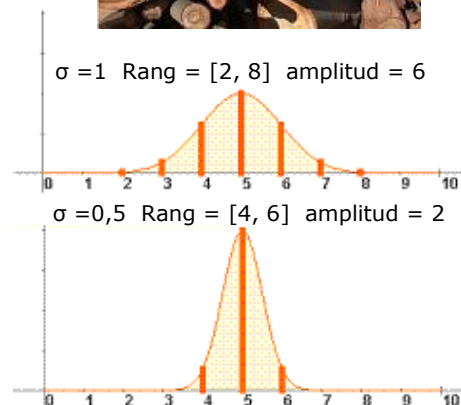
Per a mostres unimodals, amb una sola moda, i quasi simètriques, al voltant de la mitjana podem considerar un interval que contingui la majoria de les dades. Per exemple, per a una mostra amb mitjana 100 i desviació típica 10, aproximadament el 68% de les dades estaran entre 90 i 110; entre 80 i 120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



L'estatura dels tributaris present poca dispersió



El gruix dels trunks presenta força dispersió



En els dos gràfics la mitjana, mediana i moda valen 5

En la pràctica es sol utilitzar la fórmula reduïda per al càlcul de la desviació típica.

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

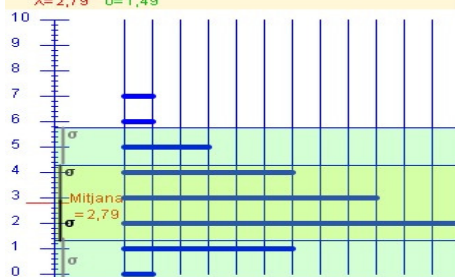
Així, per a

Marca	Fr
4	3
5	3
6	2

Tenim que la mitjana $\bar{X} = 4,85$ i

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

Dades= 39
Franja interior [2,79 - 1,49 , 2,79 + 1,49] = [1,31 , 4,28]
amb un total de 27 dades, el que suposa un 69,23 %.
L'ampla té un total de 37 dades, el 94,87 %
 $\bar{X} = 2,79$ $\sigma = 1,49$



EXERCICIS resoltos

13. Calcula la mitjana i la desviació típica en

- a) 200, 250
- b) 175, 275
- c) 250, 250

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

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

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

14. Calcula la mitjana i la desviació típica en:

- a) 7, 5, 3, 2, 4, 5
- b) 20, 25, 20, 22, 21

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

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

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

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

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

15. Organitza les dades següents en intervals de 10 cm des de 150 a 200. Amplia 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. Calcula 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	

	x_i	f_i	$x_i \cdot f_i$	$f_i \cdot (x_i - \bar{X})^2$
[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 s'obté:

Mitjana i Desviació típica

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

5. Representativitat

Mostreig aleatori

La característica més important d'una mostra és la seva **representativitat** respecte a l'estudi estadístic que s'estigui fent. Si la mostra no és representativa direm que està **esbiaixada**.

El procés mitjançant el qual es tria una mostra s'anomena **mostreig**, i per a què ens proporcioni una mostra representativa ha de ser aleatori. Un mostreig és **aleatori** quan els individus de la mostra es trien a l'atzar, de forma que tots tenen la mateixa probabilitat de ser escollits.

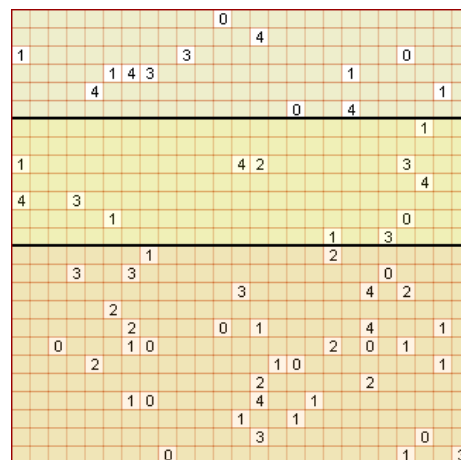
Exemple: Trucades telefòniques voluntàries. Aquestes enquestes tenen diverses fonts d'esbiaixada. Hi ha famílies que no tenen telèfon, el cost de la trucada no tot el món està disposat a assumir-ho. Però sobre tot, el factor de resposta voluntària, els enquestats s'autoseleccionen. Solen respondre aquells amb una forta opinió negativa sobre el tema. L'enuig els anima a participar.

Exemple

En la imatge tens 625 quadradets que representen els alumnes d'un institut fictici, es vol estudiar el "nombre de germans" i per això s'ha triat una mostra aleatòria com pots veure a la dreta.



Fes-ho així: Decideix primer la mida de la mostra, per exemple 62 alumnes, ordenats els alumnes en triem un d'ells a l'atzar (ho pots simular triant un quadradet amb els ulls tancats), a partir d'aquest compta i senyala cada 10 quadradets ($625/62 \approx 10$), quan arribis al final de la llista (quadradet) segueix des del principi. Aquest tipus de mostreig aleatori s'anomena **sistemàtic**.



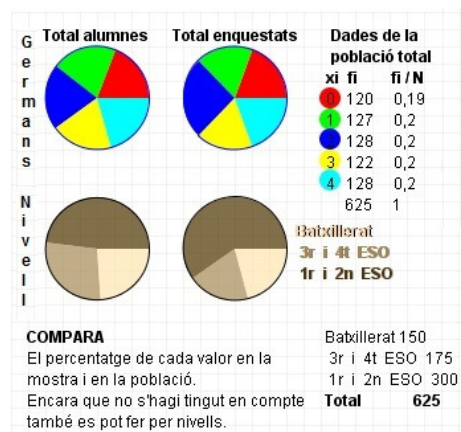
xi	fi	fi / N	DADES DE LA MOSTRA
0	12	0,19	
1	11	0,18	
2	16	0,26	
3	11	0,18	
4	12	0,19	
	62	1	

EN AQUEST MOSTREIG

No has de tenir en compte els nivells, sols que cada alumne sigui triat d'entre tots aleatòriament.

Encara així hi haurà correlació en els nivells entre mostra i població.

Batxillerat	13
2n cicle ESO	12
1r cicle ESO	37
Total	62
Percentatge	9,92 %



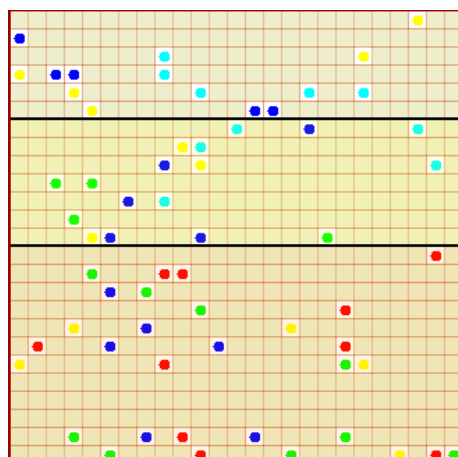
Mostreig estratificat

En ocasions quan la població objecte d'estudi pertany a diferents grups o estrats convé triar la mostra de forma que tots ells quedin representats.

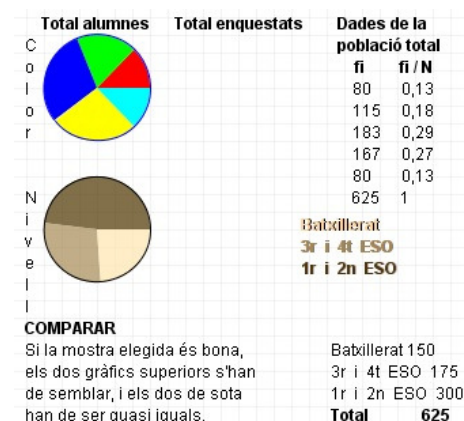
Aquest tipus de mostreig, escollint un repartiment proporcional als estrats, s'anomena **estratificat**.

Per exemple, si volem estudiar el poder adquisitiu d'una població, i només triem 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 agafant 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.

Tot seguit sobre la població de l'institut fictici anterior s'ha fet una enquesta sobre el color preferit i en aquest cas s'ha decidit fer estratificada. De cada nivell s'ha seleccionat aleatòriament un nombre d'individus proporcional al nombre de components.



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



EXERCICIS resoltos

16. Una gran empresa té treballadors en quatre àrees. Operaris, Representants, Administració i Direcció. Les condicions de treball són força diferents en cada àrea, per tant el grau de satisfacció no és igual en cadascuna d'elles. Per esbrinar-ho, si hi ha 1000, 500, 300 i 200 treballadors en les àrees d'operaris, representants, administratius i directius, quants se n'han de seleccionar de cada àrea per a una mostra de mida?

a) 200 b) 100 c) 300

- a) D'un total de 2000 empleats, els percentatges per a operaris, repartidors, administratius i directius són del 50%, 25%, 15% i 10%. Això fa que la mostra agafi 100 operaris, 50 repartidors, 30 administratius i 20 directius.
b) 50, 25, 15 i 10.
c) 150, 75, 45 i 30

Alguns dels exercicis proposats tot seguit estan elaborats a partir d'aquesta publicació de INE. Pots veure articles semblants a

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

4/2007

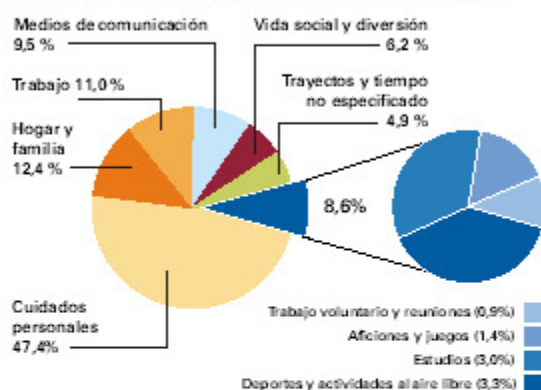


Encuesta de Empleo del Tiempo

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.



Más información en:

www.ine.es

DEPÓSITO LEGAL: M-12643-2001

ISSN: 1579-3207

NºP: 005-01-005-1

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.

Fuentes estadísticas utilizadas:

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



Per practicar

1. 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.

2. Escribe tres variables qualitatives que estiguin relacionades amb embarcacions.

3. Escribe tres variables quantitatives discretes que estiguin relacionades amb avions.

4. Escribe tres variables quantitatives contínues que estiguin relacionades amb trens.

5. 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?

6. Fes una taula i un gràfic de sectors de les dades: R R A A R A R V N V R N

7. 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

8. Agrupa les dades següents en intervals

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

9. Fes un histograma de les dades de l'exercici anterior

10. Calcula la mitjana en cada cas:

- a) 4, 6, 8
- b) 4, 6, 8, 6
- c) 100, 120, 180, 200

11. 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

12. Determina la moda i la mediana

- a) 50,60,60
- b) 12,12,22,32
- c) 10,20,30,40,20
- d) 35,25,35,25,25,25

13. 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

14. Quina o quines de les dades següents es pot considerar una observació atípica en cadascuna de les dues sèries?

- a) 4 5 6 5 7 8 4 5 8 7 5 12 6 7 6 5 4
- b) 8 9 1 9 8 9 7 9 6 7 8

15. 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

16. 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

17. Calcula la mitjana i la desviació típica en cadascun dels següents casos:

- a) 100 i 100
- b) 99 i 101
- c) 110 i 90
- d) 120 i 80

18. 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			

Estadística

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

20. 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

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

a)

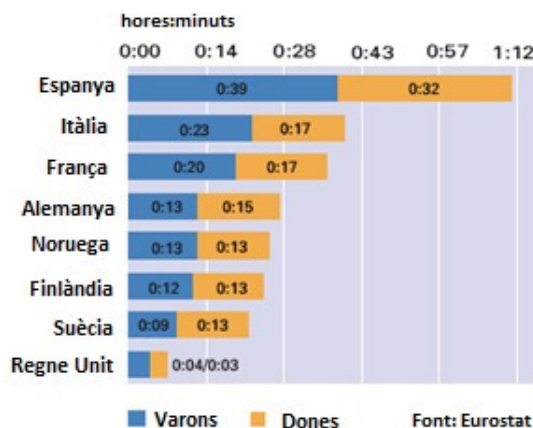


a1. Quina és la variable estudiada? i la freqüència?

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)

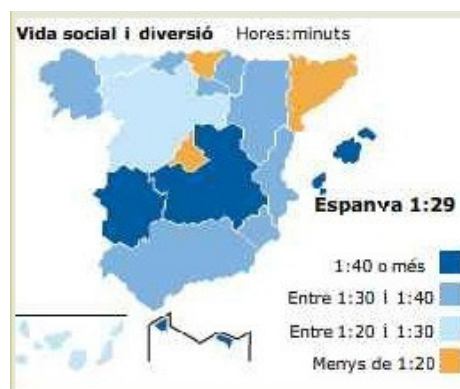


c1. Creus que el dormir s'ha considerat com activitat de cura personal?

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?

d)



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 pare de les comunitats?

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



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 se'n va adonar que la causa principal de les morts de ferits en combat era la falta de mesures sanitàries. A l'aplicar-les, la taxa de mortalitat passà d'un 42,7% a un 2,2%. Gràcies a un ús eficaç de les dades, aconseguí modificar el sistema d'atenció sanitària a la seva tornada a Gran Bretanya. Canvià el sistema de registre de dades i fou una de les primeres persones en utilitzar els gràfics estadístics per a representar les dades d'una forma senzilla de forma que fins i tot els parlamentaris i generals poguessin entendre.

Per Florence, les dades no eren quelcom abstracte, era 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 ajudaren a organitzar l'estadística com un camp d'estudi precis els mètodes de la qual s'apliquen a problemes pràctics de moltes disciplines. Com quasi tots els pioners en l'estadística, els seus treballs nasqueren de la necessitat de resoldre problemes pràctics.

Inferència estadística

L'estadística desenvolupada en aquest tema és el que es coneix com estadística descriptiva, en ella es 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 està propera a la de la mostra? Si considero un interval al voltant de la mitjana mostral, la veritable, amb quina probabilitat hi estarà inclosa? 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 en **Indústria i empreses**. Per al control de qualitat en la producció en cadena, per a l'anàlisi de mercats, per a l'estudi de preu de venda al públic dels articles fabricats, en gestió financera,...

A la part dreta es citen algunes altres de les seves aplicacions.

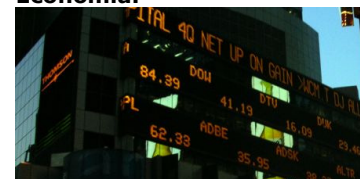
Alguns camps d'aplicació de l'estadística

Administració pública



Mitjançant les Delegacions territorials i provincials, es recullen dades per analitzar-les i sotmetre-les a processos estadístics. D'aquesta forma es coneixen dades referides a naixements, defuncions, matrimonis, preus, salaris, treball, ensenyament, sanitat,... Totes aquestes dades es solen publicar per l'INE.

Economia.



En aquest camp és imprescindible, sobre tot en macro-magnituds.

Psicologia.



La major part dels treballs científics en psicologia experimental tenen com a principal eina de treball l'estadística.

Medicina.



En qualsevol estudi experimental d'aquestes àrees Existeix una assignatura específica anomenada Bioestadística per aquests estudis experimentals. En Genètica i antropometria trobem dos dels camps de major aplicació.



Recorda 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

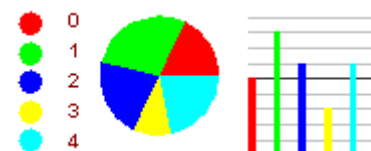
Altura: 182 172 157 194 150 166 163
196 167 199 172 185 172 168 173
160 162 173 161 192 156 164 173
180 193 172

Recompte de dades:

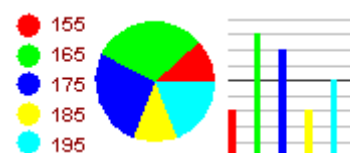
x_i	f_i	Intervalo	x_i	f_i
0	5	[150,160)	155	3
1	8	[160,170)	165	8
2	6	[170,180)	175	7
3	3	[180,190)	185	3
4	6	[190,200)	195	5
	28	Total		26

Gràfics de **sectors i barres**

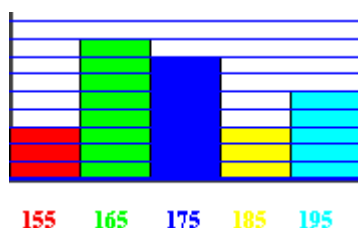
Nre. de germans



Altura.



Histograma



Mitjana i moda i desviació típica

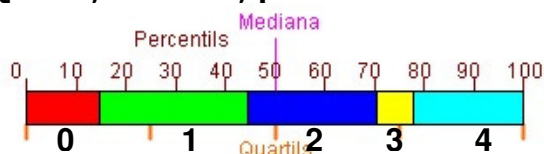
x_i	f_i	$x_i \cdot f_i$	$f_i \cdot (x_i - \bar{x})^2$
0	5	0	0
1	8	8	6,37
2	6	12	0,06
3	3	9	3,67
4	6	24	26,64
Total	28	53	54,67

$$\text{Mitjana} = \bar{X} = \frac{53}{28} = 1.89$$

$$\text{Moda} = Mo = 1$$

$$\sigma = \sqrt{\frac{54.67}{28}} = 1.39$$

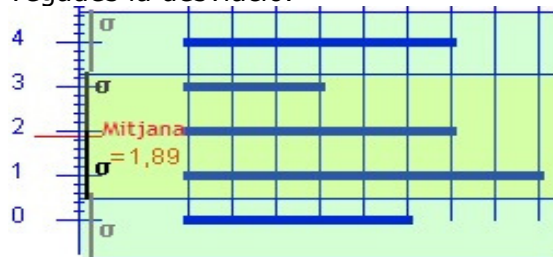
Quartil, mediana, percentil



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

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.

Autoavaluació



1 Quants graus corresponen en un diagrama de sectors a la marca 2?

xi	fi
1	4
2	4
3	7
4	5

2 La mediana de la distribució anterior és?

3 Quina és la moda ?

xi	fi
15	40
25	45
35	37
45	51

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

xi	fi
100	4
200	4
300	7
400	5

5 Quin és el percentil 30 ?

xi	fi
1	4
2	4
3	7
4	5

6 Quina és la mitjana de les dades anteriors?

7 Quina és la desviació típica de les dades del nre.5?

8 Quina és la mitjana?

xi	fi
180	40
200	25
220	27
240	50

9 Quina és la desviació típica de les dades anteriors?

10 Quin és el percentil 70?

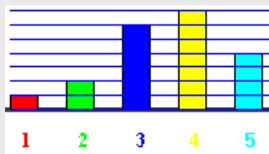
Solucions dels exercicis per practicar

- 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 $T \rightarrow 36^\circ$

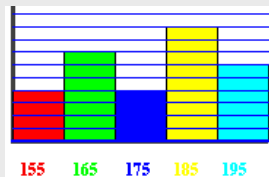
- $R \rightarrow 5$,
 $A \rightarrow 3$,
 $V \rightarrow 2$,
 $N \rightarrow 2$



- 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



- \rightarrow
- 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

15. $Me=2$, $Q1=2$, $Q3=3$, $P90=4$

16. $Me=3$, $Q1=1$, $Q3=4$ y $P20=1$

17. La mitjana és 100 en els 4, i la desviació 0, 1, 10 i 20.

18.

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

19. $\bar{x} = 171,3$ $\sigma \approx 14.02$

20. A (0.42, 2.9) hi ha 11,
i a (-0.88, 4.14) tots

21. 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

Solucions AUTOAVALUACIÓ

- Sol 72°
- Sol 3
- Sol 51
- Sol 40%
- Sol 2
- Sol 2.65
- Sol 1.06
- Sol 212.25
- Sol 24.53
- Sol 240

No t'oblidis d'enviar les activitats al tutor

