

Objectius

En aquesta quinzena aprendràs a:

- Distingir els diferents tipus de variables estadístiques.
- Agrupar en intervals les dades d'un estudi estadístic.
- Fer la taula estadística associada a un conjunt de dades.
- Representar i interpretar gràfics estadístics, i saber quan és convenient utilitzar cada tipus.
- Calcular la mitjana, la moda, la mediana i els quartils d'un conjunt de dades.
- Què són i com es calculen els paràmetres de dispersió: el rang o recorregut, la variància i la desviació típica, el coeficient de variació.

1.Fer estadística	pàg. 4
Necessitat	
Població i mostra	
Variables	
2.Recompte i gràfics	pàg. 5
Recompte de dades	
Gràfics	
Agrupació de dades en intervals	
3.Mesures de centralització	pàg.9
i posició	
Mitjana	
Moda	
Quartils i mediana	
Diagrames de caixa i bogotis	
4.Mesures de dispersió	pàg. 12
Rang i desviació mitjana	
Desviació típica	
Coeficient de variació	

Exercicis per practicar

Per saber-ne més

Resum

Autoavaluació

Activitats per enviar al tutor

Annex: Utilitzar el full de càlcul

Abans de començar

Estadística

L'estadística, a nivell primari, és una activitat que tothom fa des de molt petit. El mer fet de comptar i/o classificar els teus joguines (els teus cotxes, nines/ots, bales, videojocs,...) ja és una activitat estadística.

Classificar objectes

El tornavis a la caixa d'eines, els coberts al calaix de la cuina, els llibres a la prestatgeria, els videojocs al costat de la consola,...

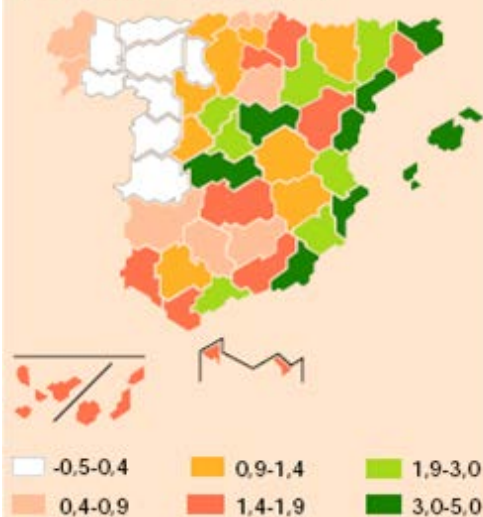


Competicions escolars



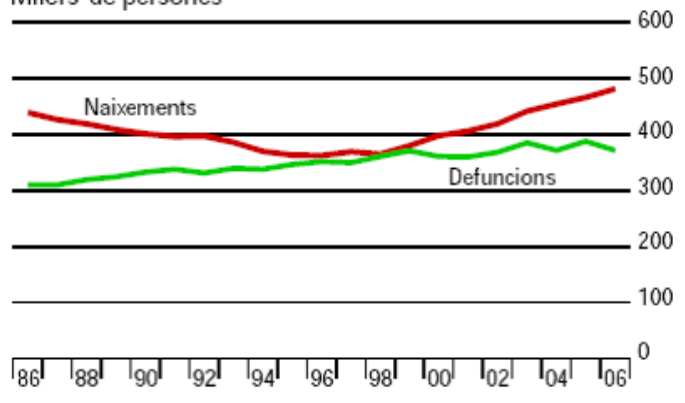
Al teu institut hi haurà equips i competicions, i s'haurà de portar un còmput de resultats i una ordenació d'equips segons una puntuació. Fins i tot és possible que hi hagi un registre de diversos anys.

Creixement relatiu de la població total
1 de gener de 2008 (%)



Creixement natural de la població

Milers de persones



1. Fer estadística

Necessitat

En posar en pràctica una mesura social per saber la seva acceptació. A quantes persones pot anar dirigida?, quins són els diferents nivells? Davant una iniciativa com aquesta, preguntar a tota la població pot esgotar els recursos destinats a ella, una enquesta prèvia pot estalviar-nos algun que un altre equivoc.

Població i mostra

Quan es fa un estudi estadístic l'investigador decideix si analitzarà tota la població o una mostra escollida prèviament.

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

La **mostra** és un subconjunt de la població. Ha d'escollir-se de forma que sigui representativa de tota la població de la característica estudiada.

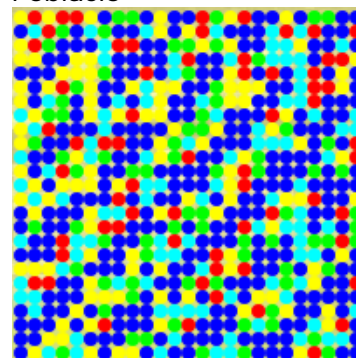


Atributs i Variables.

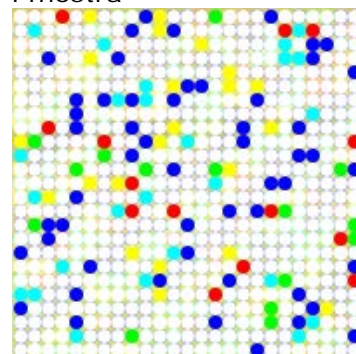
Cada una de les propietats o característiques que podem estudiar és una **variable estadística**. Depenent dels possibles valors que poden prendre es classifiquen en:

- **Variables qualitatives** o atributs. Els valors de la variable no són nombres sino qualitats, s'expressen amb paraules. El color, la forma, el sexe,...són exemples de variables qualitatives.
- **Variables quantitatives**. Les dades s'expressen numèricament i poden ser:
 - Discretes. Cada una de las variables només pot prendre valors enters (1, 2, 3...). El nre. de germans, el nre. de finestres de casa, el nre. De col·legis de la teva població,...
 - Contínues. Poden prendre qualsevol valor d'un interval donat. El nostre pes, altura, força, no es poden mesurar amb nombres enters, ni la densitat de l'aire, ni la velocitat mitjana dels fórmula 1 en una carrera,...

Població



i mostra



El color dels llapis, és una variable qualitativa



L'altura, edat i pes, són variables quantitatives.



2. Recompte de dades

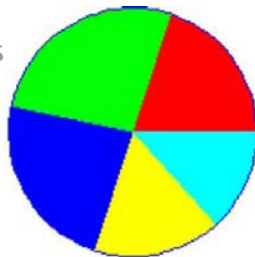


X_i	f_i	$f_i/30$	F_i
Vermell	6	0,20	15
Verd	8	0,21	25
Blau	7	0,15	32
Groc	5	0,33	42
Turquesa	4	0,11	47
Total	30	1	

És part del procés, després de recopilar les dades es procedeix al seu recompte per a expressar-les de forma ordenada i perquè sigui més fàcil treballar amb elles. Generalment s'elabora una taula com en la simulació de l'esquerra on pots practicar.

- Freqüència **absoluta**, és el nombre de vegades que apareix una dada. A la de x_i l'anomenarem f_i .
- Freqüència **relativa**, és el quocient entre la freqüència absoluta i el nombre total de dades.
- Freqüència **acumulada** d'una dada, és la suma de les freqüències absolutes dels valors que són menors o iguals que ell, la indicarem amb F_i . També es poden calcular les freqüències relatives acumulades.

Diagrama de sectors



Per calcular els graus de cada sector, fixa't:

$\frac{360 \cdot 4}{30} = 48^\circ$	
$\frac{360 \cdot 5}{30} = 60^\circ$	
$\frac{360 \cdot 7}{30} = 84^\circ$	
$\frac{360 \cdot 8}{30} = 96^\circ$	
$\frac{360 \cdot 6}{30} = 72^\circ$	

Diagrames de barres i de sectors

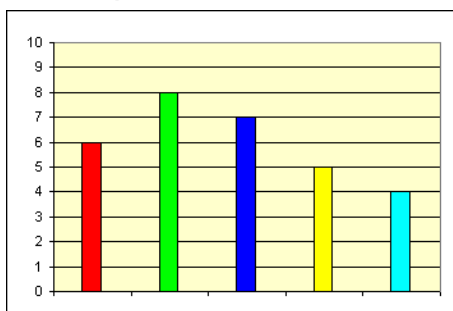
Les dades estadístiques solen representar-se de forma gràfica, ja que d'aquesta forma podem fer-nos una idea de la seva distribució només amb una ullada. En funció del tipus de variable convé més utilitzar un tipus de gràfic o un altre.

- **Diagrama de sectors**, pot aplicar-se a qualsevol tipus de variable, encara que és el més adequat en variables qualitatives i per una primera presa de contacte amb els valors d'una població. És un cercle dividit en sectors d'angle proporcional a la freqüència de cada valor.

L'amplitud de cada sector s'obté multiplicant la freqüència relativa per 360° .

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

Diagrama de barres



- **Diagrama de barres**. També pot aplicar-se a qualsevol tipus de variable, encara que es considera l'adequat per a variables discretes. Cada valor es correspon amb una barra de longitud proporcional a la seva freqüència.

EXERCICIS resolts

1. Quantes persones suposen una mostra del 10% d'una població de 10.000 habitants? I d'una de 6000 habitants?.

Solució: a) $10.000 \cdot 10 / 100 = 1000$, b) $6000 \cdot 10 / 100 = 600$

2. Una empresa de sondejos estadístics té capacitat per entrevistar a 1000 persones per setmana. Si disposa de 4 setmanes a quin percentatge d'una població de 100.000 habitants pot entrevistar per obtenir una mostra.

Solució: En 4 setmanes pot entrevistar a 4000 persones. 4000 de 100.000 equival a 4 de 100. Així doncs el 4%.

3. Amb la finalitat de conèixer millor la forma de viatjar d'una població han preparat una enquesta. Algunes de les preguntes tractaren sobre: Nre. de dies de viatge, diners emprats, nombre d'embalums, zones geogràfiques, mitjà de transport, naturalesa del viatge (negocis, turisme, familiar, salut...) i nre. de persones. Classifica aquestes variables estadístiques.

Solució:

V qualitativa: Zones geogràfiques, mitjà de transport i naturalesa del viatge.

V. quantitativa discreta: Nre. de dies, nombre d'embalums i nre. de persones.

V. quantitativa contínua: Diners emprats.

4. Fes un recompte de les següents dades

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

Solució:

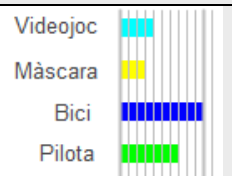
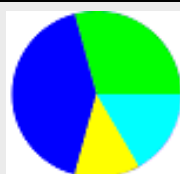
Xi	fi
1	3
2	8
3	4
4	9

5. Fes un recompte de les següents dades, un gràfic de sectors i un altre de barres. Indica l'angle de cada sector.

Pilota, màscara, pilota, màscara, màscara, bici, màscara, bici, bici, màscara, màscara, màscara, màscara, videojoc, màscara, pilota, videojoc, pilota, videojoc, pilota, pilota, videojoc, pilota, màscara.

Solució:

Xi	fi	graus
Videojoc	4	60
Màscara	3	45
Bici	10	150
Pilota	7	105



Les dades

55	491	42	465	653
829	798	254	155	427
153	533	945	878	230
690	652	476	110	87
14	751	47	211	341
737	473	452	352	246
499	109	694	308	933
326	209	729	651	397
161	329	975	848	823
240	640	319	526	

Agrupades en 5 intervals

Observa las marques de cada classe com es corresponen amb la mitjana dels seus extrems.

Interval	Marca	Freqüència
[0 , 200)	100	10
[200 , 400)	300	13
[400 , 600)	500	9
[600 , 800)	700	10
[800 , 1000)	900	7

Agrupats ara en 8 intervals

Interval	Marca	Fr.
[0 , 125)	62,5	7
[125 , 250)	187,5	8
[250 , 375)	312,5	7
[375 , 500)	437,5	8
[500 , 625)	562,5	2
[625 , 750)	687,5	8
[750 , 875)	812,5	5
[875 , 1000)	937,5	4

Agrupació de dades en intervals

En variables contínues, o en discretes quan el nombre de dades diferents es fa quasi tan gran com el nombre de dades, per a poder estudiar-les, es fa necessari agrupar-les en **intervals** o **classes**, habitualment de la mateixa amplitud i com a mínim quatre.

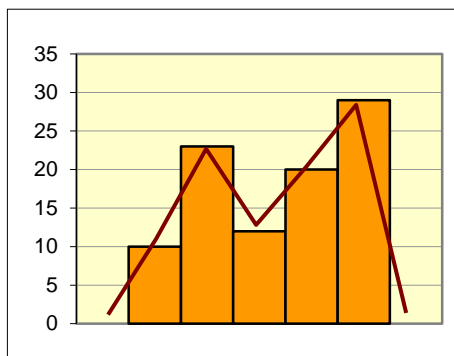
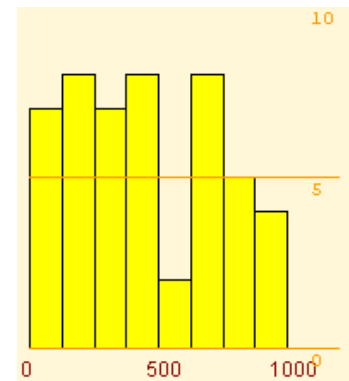
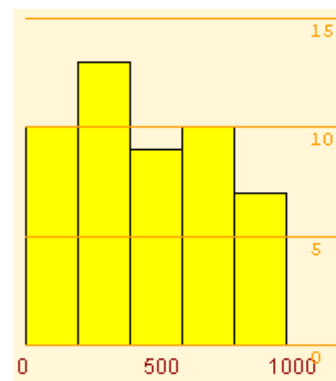
Per exemple, en una població hi ha quasi tantes altures com individus, però podem agrupar-los en baixos, mitjans i alts; també podríem fer baixos, mitjans-baixos, mitjans-alts i alts, o classificar-los de 10 en 10 cm, o de 20 en 20...

- Par a representar totes les dades d'un interval triem un valor, el punt mitjà de l'interval, que l'anomenem **marca de classe**.

Histograma

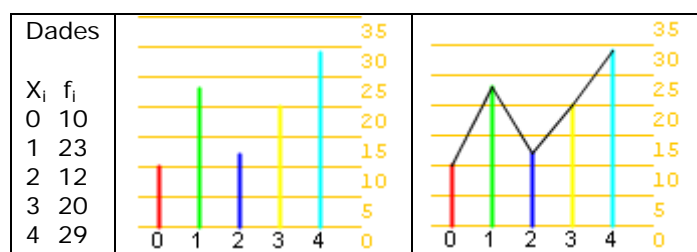
Quan les dades vénen agrupades en intervals s'utilitza per a representar-les gràficament l'**histograma**. Cada valor es representa amb un rectangle d'amplada l'interval corresponent i amb l'altura proporcional a la seva freqüència.

Els histogrames per a les dades del marge agrupades en cinc i vuit intervals:



Polígon de freqüències.

El creem en unir els extrems superiors de les barres dels histogrames o dels diagrames de barres.

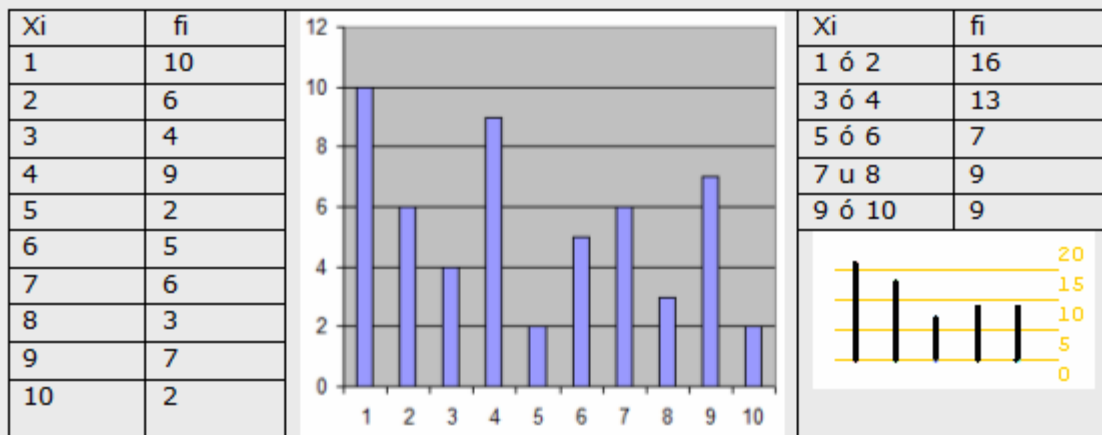


EXERCICIS resolts

6. Agrupa les següents dades en 10 grups. Agrupa les mateixes dades, ara, en 5 grups i fes un gràfic per a cada agrupació.

2	9	9	8	2	9	5	4	1	7	7	1
2	8	4	1	6	1	9	1	4	7	4	9
4	1	3	2	3	4	3	1	1	1	4	5
10	6	6	2	1	4	3	7	6	6	10	2
9	8	9	7	7	4						

Solució:

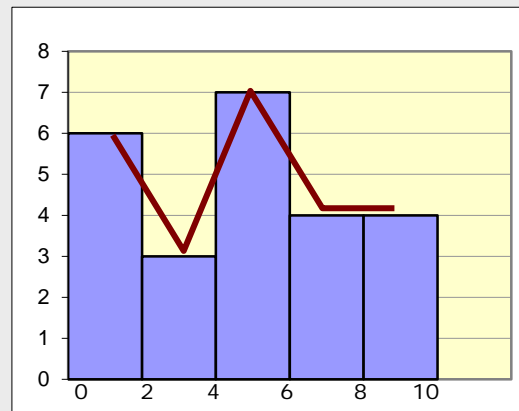


7. Agrupa les dades següents en 5 intervals d'igual amplitud, fes un gràfic i un polígon de freqüències.

7,2	6	6,3	9,8	9,1	9,3
5,7	6,7	8,4	5,7	3,1	1,4
5,4	1,1	4,8	2,5	0,1	4
5,3	1,3	3,6	1,9	5,2	1,7

Solució:

	f_i
[0,2)	6
[2,4)	3
[4,6)	7
[6,8)	4
[8,10)	4



Mitjana

Exemple 1

10, 12, 10, 14 i 13

$$\bar{x} = \frac{10+12+10+14+13}{5} = \frac{59}{5} = 11.8$$

Exemple 2

Xi	fi	Xi · fi
5	4	20
10	6	60
15	7	105
20	9	180
25	4	100
30	6	180
36	645	

La mitjana.
 $\bar{x} = \frac{645}{36} = 17,91$

Moda

Exemple 1

Xi	fi	Moda
5	8	5
10	5	
15	1	
20	8	20
25	0	
30	3	

Exemple 2

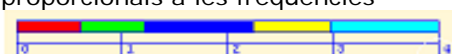
Xi	fi	Moda
100	5	
200	5	
300	1	
400	3	
500	2	
600	9	600

Mesures de posició

Ampliem la taula amb la columna de freqüències acumulades i percentatges d'aquestes

xi	fi	Fi	%
1	10	10	15,152
2	10	20	30,303
3	17	37	56,061
4	12	49	74,242
5	17	66	100

i/o una barra amb longituds proporcionals a les freqüències



i podrem saber la mediana i quartils
 $Q_1=2, Q_2=Me=3$ i $Q_3=5$

3. Mesures de centralització i posició

La mitjana

Tots els alumnes saben que amb un 6 i un 4 tenen de mitjana 5. Doncs la mitjana en estadística no és una altra cosa que això, només que, habitualment, amb més dades.

Per a calcular la mitjana si són poques les dades, se sumen totes i es divideix entre el nombre total. Si en són moltes, les tindrem agrupades, llavors se sumen els productes de cada dada per la seva freqüència absoluta i es divideix aquesta suma pel nombre total de dades. Se indica amb \bar{x} .

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum x_i f_i}{N}$$

La moda

Qui no ha escoltat alguna vegada: "Està de moda anar a...", "Es porten aquest tipus de pantalons, estan de moda", o "S'ha posat de moda el grup"... i tothom entén que hi ha una bona quantitat de persones en aquestes opcions.

Així doncs, el valor que més freqüència tingui serà "el de moda", encara que pot passar que n'hi hagi més d'un.

- La **moda**, **Mo**, d'una distribució estadística és el valor de la variable que més es repeteix, el de major freqüència absoluta.

La mediana i els quartils

La mediana i els quartils, com la mitjana aritmètica, només es poden calcular quan la variable és quantitativa.

- La **mediana**, **Me**, és el valor que ocupa la posició central una vegada ordenades les dades en ordre creixent, és a dir, el valor que és major que el 50% i menor que l'altre 50%.

La mediana divideix la distribució en dues parts amb igual nombre de dades, si la dividim en quatre parts obtenim els **quartils**, 1r, 2n i 3r, que s'indiquen respectivament **Q₁**, **Q₂** i **Q₃**.

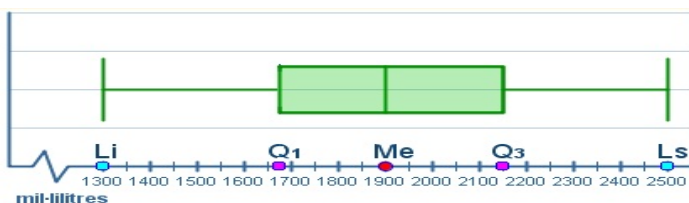
Ordenades les dades, **el primer quartil**, és major que el 25% d'aquestes; **el tercer quartil**, major que el 75%, i el segon coincideix amb la mediana.

Diagrames de caixa i bogotis

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

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.

$$\text{mín}=1300 \quad Q_1=1675 \quad \text{Me}=1900 \quad Q_3=2150 \quad \text{màx}=2500$$



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

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	Q ₁
Rosa	2100	Juan	1650	
Lupe	1700	Lupe	1700	
Cesc	1900	David	1750	
Tere	1500	Pep	1850	
Iris	1900	Alex	1900	Me
Pep	1850	Iris	1900	
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 resolts

8. Calcula la mitjana en cada cas:

a) 4, 6, 8

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

b) 4, 6, 8, 6

b) $(4+6+8+6)/4 = 24/4 = 6$

c) 100, 120, 180, 200

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

9. Calcula la mitjana dels següents dades

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

Solució:

Ho sumem tot i dividim entre 24, el nombre de dades.

$$\bar{X} = 2,29$$

10. Calcula la mitjana de les següents dades

2,4	3	1,1	4	3,5	0,7	0	2,8	3,8	0,2	2,8	1,9
0,6	3,8	3,1	4	2,8	0,2	0,4	3,1	1,5	1,9	1,8	3,1

Solució: $\bar{X} = 2,19$

11. Determina la moda per a les dades

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

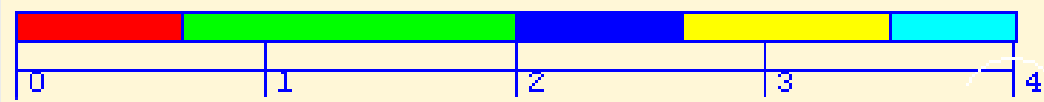
Solució:

Mo = 1, és el valor que més vegades apareix, 8 en total.

12. Calcula la mediana, el primer i el segon quartil de les dades de l'exercici anterior.

Solució:

Fem el recompte, $0 \rightarrow 4$, $1 \rightarrow 8$, $2 \rightarrow 4$, $3 \rightarrow 5$ y $4 \rightarrow 3$, dibuixem barres de colors de longituds proporcionals a les freqüències, valdrien per exemple de 4mm, 8mm, 4mm, 5mm y 3mm. Dividim tota la barra en 4 parts i ens fixem en el color en que queda la divisió.

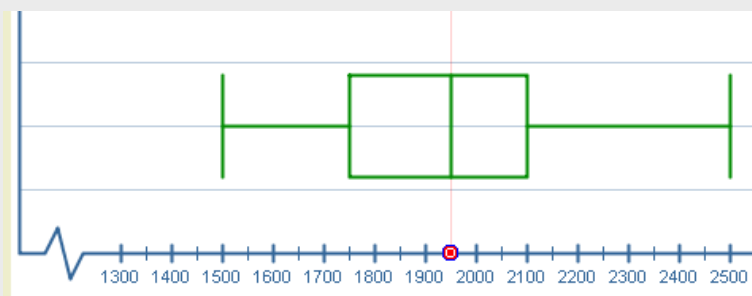


Mediana = 1,5,
Q1=1 y Q3=3

Ó bé construïm la taula i observem on queden en l'última columna els valors 25%, 50% i 75%.

X_i	f_i	F_i	%
0	4	4	16,667
1	8	12	50
2	4	16	66,667
3	5	21	87,5
4	3	24	100

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



Mínim = 1500

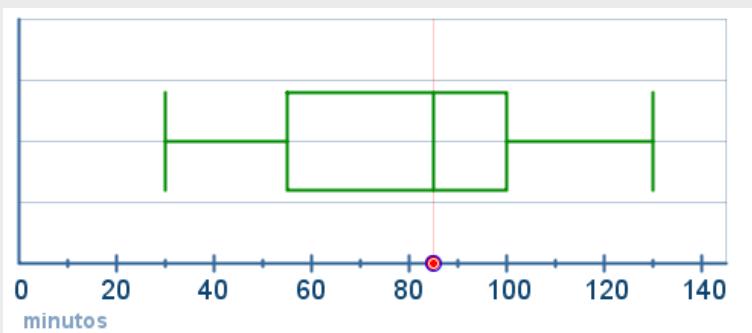
$Q_1 = 1750$

Me = 1950

$Q_3 = 2100$

Màxim = 2500

14. 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 i Desviació mitjana

Les mesures de **dispersió** indiquen si les dades estan més o menys agrupades respecte de les mesures de centralització.

- **Rang** o recorregut, és la diferència entre el major i el menor valor de la variable, indica la longitud de l'interval en el que se troben totes les dades.

Encara que el rang dona una informació important, resulta més interessant calcular quant es desvien en terme mig les dades de la mitjana.

- **Desviació mitjana**, és la mitjana dels valors absoluts de les diferències entre la mitjana i les diferents dades.

Variància i desviació típica

És una altra forma de mesurar si les dades estan o no pròximes a la mitjana i és la més utilitzada.

- La **variància** és la mitjana dels quadrats de les desviacions.
- La **desviació típica** és l'arrel quadrada positiva de la variància. Per a designar-la fem la lletra grega "sigma" σ .

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

És important que entenguis el significat d'aquestes mesures, com més grans siguin, més disperses estaran les dades.

Els intervals al voltant de la mitjana d'amplitud 2 o 4 vegades la desviació típica, tenen molta importància en estadística pel percentatge de dades que hi ha en elles.

Coeficient de variació

És el quocient entre la desviació típica i la mitjana, s'utilitza per a comparar les dispersions de dades de diferent mitjana.

Per exemple, per les dades 4 i 6, el $CV=1/5=0,2$ i per 101 i 99 és $CV=1/100=0,01$. En ambdós casos la desviació típica és la mateixa, però en relació a la mitjana és més important en el primer.

Desviació mitjana

Exemple

X_i	f_i	$X_i \cdot f_i$	$ \bar{X} - X_i \cdot f_i$
5	2	10	29,44
10	1	10	9,72
15	3	45	14,16
20	4	80	1,11
25	6	150	31,66
30	2	60	20,55
	18	355	106,66

$$DM = \frac{106,66}{18} = 5,92$$

Desviació típica

Exemple

X_i	f_i	$X_i \cdot f_i$	$(X_i - \bar{X})^2 \cdot f_i$	$f_i \cdot X_i^2$
5	5	25	1164,81	125
10	1	10	105,33	100
15	1	15	27,7	225
20	2	40	0,13	800
25	1	25	22,43	625
30	9	270	853,25	8100
	19	385	2173,68	9975

$$\sigma = \sqrt{\frac{2173,68}{19}} = 10,69$$

O bé, com $\bar{x} = 20,26$

$$\sigma = \sqrt{\frac{9975}{19} - 20,26^2} = 10,69$$

Coeficient de variació

$$\bar{x} = 12 \quad \sigma = 7 \quad CV = \boxed{58,33} \%$$

$$\bar{x} = 110 \quad \sigma = 7 \quad CV = \boxed{6,36} \%$$

$$\bar{x} = 1148,25 \quad \sigma = 7 \quad CV = \boxed{0,60} \%$$

Utilitzar la Calculadora científica

MODEL 1

Mode estadístic

Primer s'ha d'escollir el mode estadístic. En moltes calculadores es fa prement:

[MODE][.]

Dades desordenades

A continuació s'han d'introduir les dades, per exemple per a 2, 3, 4, 3 teclejarem:

[2][M+] [3][M+] [4][M+] [3][M+]

I per a fer els càlculs:

- Per a la mitjana **[SHIFT][x̄]**
- Per a la desviació típica **[SHIFT][σ_n]**

També es poden sumar totes, o els quadrats, o comptar el nre. de dades introduïdes, prement respectivament:

[SHIFT][Σx] [SHIFT][Σx²] [SHIFT][n]

Dades en una taula

x_i	f_i	S'introdueixen les dades segons la
2	4	seqüència: [2][x][4][M+]
3	3	[3][x][3][M+]
4	5	[4][x][5][M+]

I ara ja es poden realitzar els càlculs com abans.

MODEL 2

Mode estadístic i introducció de dades

Escollim el mode estadístic (mode stat 1-VAR) i ens apareix una columna on introduir les dades, una darrere l'altra, no importa que vagin desordenades. Si tinguéssim una taula amb freqüències hauríem d'activar les freqüències (Setup frequency on) i emplenar les columnes. Després de l'última dada premer AC.

Càlculs

Prement SHIFT STAT ens apareix un menú, **1:type, 2:Data, 3>Edit, 4:Sum, 5:Var 6:MinMax**. Amb l'opció **5:Var** accedirem a calcular la mitjana, desviació típica i quantitat de dades. Amb l'opció **4:Sum** les sumes que habitualment necessitem. Amb l'opció **6:MinMax** el mínim i el màxim. I amb l'opció **2:Data** podrem modificar les dades introduïdes.



Nota: Hi ha molts models de calculadores, però afortunadament, totes són bastant semblants, tracta d'esbrinar el funcionament de la teva si no coincideixen amb aquestes, consulta el manual o pregunta al teu professor.

EXERCICIS resoltos

13. Calcula el rang i la desviació mitjana de les dades:

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

x_i	f_i	x_i·f_i	 x_i - x̄ · f_i
5	2	10	4,8
6	4	24	5,6
7	6	42	2,4
8	3	24	1,8
9	3	27	4,8
10	2	20	5,2
20	147	24,6	

Solució:
El rang oscil·la entre 8 i 12 amb una amplitud de 4.
Fem el recompte.
La mitjana: $\bar{x} = \frac{147}{20} = 7,4$
Calculem la desviació de cada dada respecte a la mitjana, en valor absolut. La mitjana de les desviacions:
 $DM = \frac{24,6}{20} = 1,23$

14. Calcula la desviació mitjana de les dades tabulades següents:

x_i	f_i	x_i·f_i	 x_i - x̄ · f_i	
[0,200)	100	7	700	2831,82
[200,400)	300	8	2400	1636,36
[400,600)	500	13	6500	59,09
[600,800)	700	9	6300	1759,09
[800,1000)	900	7	6300	2768,18
Total:	44	22200	9054,55	

Solució:
Calculem la mitjana:
 $\bar{x} = \frac{22200}{44} = 504,55$
Completem l'última columna:
 $DM = \frac{9054,55}{44} = 205,59$

EXERCICIS resolts

15. Calcula la mitjana i la desviació típica en

a) 200, 250

b) 175, 275

Solució:

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

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

16. Calcula la mitjana i la desviació típica en:

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

b) 20, 25, 20, 22, 21

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

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

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

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

17. Quina de les dues distribucions anteriors presenta major dispersió?

Solució:

Las desviacions típiques són molt semblants 1,59 i 1,85

Per a comparar calculem el coeficient de variació:

$$a) CV = (1,59/4,33) \cdot 100 = 36,72\% \quad b) CV = (1,85/21,6) \cdot 100 = 8,56\%$$

Ara es pot apreciar que la dispersió és molt major en la distribució a).

18. Calcula la mitjana i la desviació típica de les dades agrupades següents:

X_i	5	10	15	20	25	30
f_i	9	2	3	5	9	4

Solució:

X_i	f_i	$X_i \cdot f_i$	$(X_i - \bar{X})^2 \cdot f_i$	$f_i \cdot X_i^2$
5	9	45	1371,31	225
10	2	20	107,86	200
15	3	45	16,47	675
20	5	100	35,27	2000
25	9	225	527,56	5625
30	4	120	640,72	3600
32	555	2699,21	12325	

Calculem prèviament la mitjana:

$$\bar{x} = \frac{555}{32} = 17,34$$

$$1a \text{ Forma de calcular-la } \sigma = \sqrt{\frac{2699,21}{32}} = 9,18$$

$$2a \text{ Forma } \bar{X} = 17,34 \quad \sigma = \sqrt{\frac{12325}{32} - 17,34^2} = 9,18$$

19. Quin és el coeficient de variació de la distribució anterior?

$$\text{Solució: } CV = \frac{9,18}{17,34} \cdot 100 = 52,94\%$$



Per practicar

- Quantes persones suposen una mostra del 5% d'una població de 20.000 habitants? I d'una de 1000 habitants?
- D'una població de 30000 individus s'han estudiat diverses característiques en 150 individus. Quin percentatge del total ha estat estudiat?
- Un veterinari estudia les següents característiques en una mostra d'animals d'una granja tipus d'animal, pes, color dels ulls, temperatura corporal, nombre de companys i metres quadrats per animal.
- Fes un recompte de les següents dades, un gràfic de sectors i un altre de barres. Indica l'angle de cada sector.

a	b	c	a	c	c
d	c	d	b	d	a
d	a	b	b	c	c
a	a	b	a	b	d
- Fes un recompte de les següents dades i un diagrama de barres amb polígon de freqüències

3	3	1	1	3	2
3	3	2	1	3	2
2	3	1	1	4	3
2	2	4	4	3	3
- Agrupa les següents dades en 10 grups. Agrupa les mateixes dades, ara, en 5 grups.

3	6	5	9	2	6
2	2	7	9	4	6
2	5	9	9	1	0
2	5	3	6	7	8
6	4	3	6	7	9
10	10	9	1	6	8
6	2	3	9	6	5
6	6	5	7	6	6
10	1	3	4	4	4
- Calcula la mitjana en cada cas:
 - 14, 16, 18
 - 24, 26, 28, 26
 - 1000, 1200, 1800, 2000
- Calcula la mitjana de les següents dades

3	3	1	1	3	2
3	3	2	1	3	2
2	3	1	1	4	3
2	2	4	4	3	3
- Calcula la mitjana de les següents dades

10	1,5	18	20	16	1
9,5	5,50	15,5	6,5	4,5	4
8,5	7,5	1,5	15	13	0
20	12,5	7,5	4,5	14,5	9
- Determina la moda per a les dades

3	3	1	1	3	2
3	3	2	1	3	2
2	3	1	1	4	3
2	2	4	4	3	3
- Calcula la mediana, el primer i el segon quartil de les dades de l'exercici anterior.
- Calcula la desviació mitjana en cada cas:
 - 14, 16, 18
 - 34, 36, 38, 36
 - 1000, 1200, 1800, 2000
- Calcula el rang i la desviació mitjana de les dades:

23	8	21	24	20	9
33	20	11	36	13	1
40	25	30	12	18	5
40	27	16	26	9	7
- Calcula la desviació mitjana de les dades tabulades següents:

Interval	Marca = X_i	Fr	$F_i \cdot \bar{X} - X_i $
[0 , 200)	100	1	450
[200 , 400)	300	3	750
[400 , 600)	500	3	150
[600 , 800)	700	2	300
[800 , 1000)	900	3	1050

15. Calcula la mitjana i la desviació típica a

- a) 2000, 2500
- b) 1750, 2750
- c) 2500, 2500

16. Calcula la mitjana i la desviació típica de les dades:

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

17. Calcula el coeficient de variació de les dades de l'exercici anterior.

18. Calcula la mitjana i la desviació típica de les dades:

25	29	40	9	32	4
15	35	26	24	16	2
11	16	37	10	30	2
35	17	8	40	38	5

19. Calcula el coeficient de variació de les dades de l'exercici anterior.

20. Calcula la mitjana i la desviació típica de les dades agrupades següents:

X_i	f_i
5	7
10	0
15	2
20	2
25	4
30	2

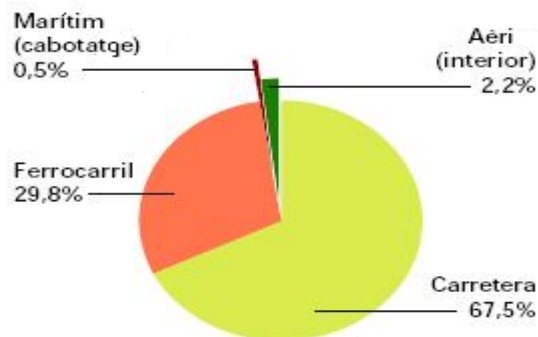
21. Fes els càlculs per a un milió d'habitants en cada comunitat.

Taxa de criminalitat. 2006
Infraccions penals per 1.000 hab.

Taxes més altes

Illes Balears	78,8
Comunitat de Madrid	70,8
Comunitat Valenciana	67,5
Ceuta	67,4
Catalunya	65,3

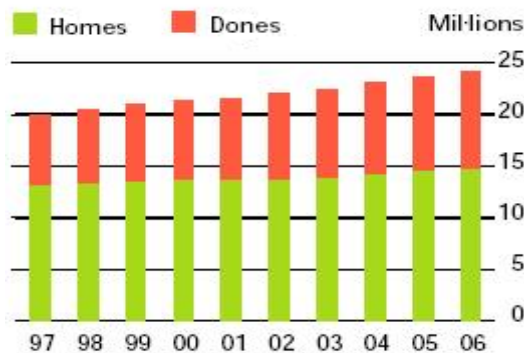
22. De cada milió de viatgers, quants corresponen a cada sector?
Viatgers que utilitzen transport interurbà. 2007



Fonts: INE, RENFE, FEVE, D. Gral. d'Aviació Civil i D. Gral. de Ports i Costes

23. Quants conductors hi havia l'any 2002? Quants eren homes i quantes dones?

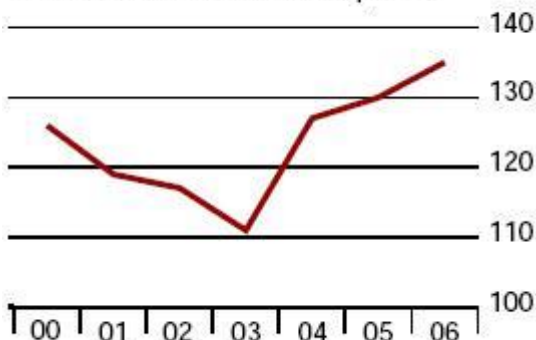
Cens de conductors



Font: Direcció General de Trànsit

24. Entre quins anys augmentaren més els detinguts per infraccions penals?

Taxa de detinguts
Per cada 1.000 infraccions penals



Font: Ministeri de l'Interior

España en cifras

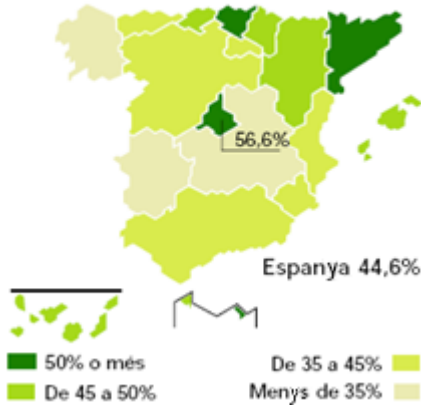
Per saber-ne més



L'Institut Nacional d'Estadística publica periòdicament dades com aquestes. Consulta la web. (www.ine.es)



Habitatges que disposen d'accés a internet. 2007



Ús de les TIC a les llars. 2007

	Llars amb accés a internet (%)	Persones que van comprar per internet (%)
Països Baixos	83	55
Suècia	79	53
Dinamarca	78	56
Luxemburg	75	47
Alemanya	71	52
Finlàndia	69	48
Regne Unit	67	53
Bèlgica	60	21
Àustria	60	36
Eslovènia	58	16
Irlanda	57	33
UE-27	54	29
Estònia	53	9
Letònia	51	11
França	49	:
Eslovàquia	46	16
Espanya	45	18
Lituània	44	6
Itàlia	43	10
Polònia	41	16
Portugal	40	9
Xipre	39	10
Hungria	38	11
República Txeca	35	17
Grècia	25	8
Rumània	22	3
Bulgària	19	3
Malta	:	:

: Dada no disponible

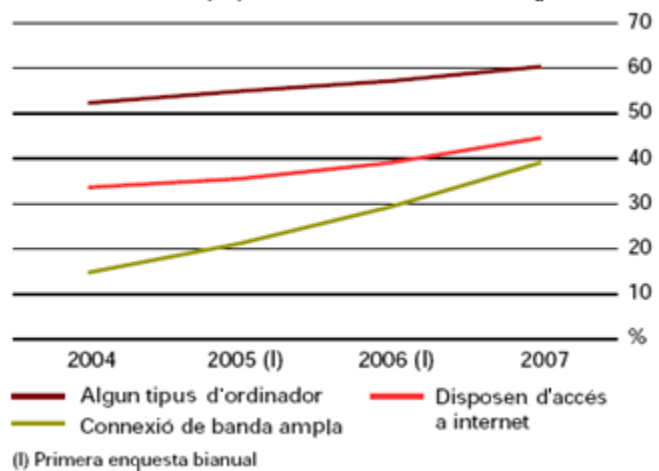
Font: Eurostat

Equipament tecnològic de les llars

Segons dades de l'Enquesta sobre Equipament i Ús de les Tecnologies d'Informació i Comunicació en les Llars, el 2007 el 99% de les llars disposa de telèfon (ja sigui fix o mòbil), el 99,5% disposa de televisió i el 22,8% té recepció de televisió digital terrestre (TDT). En una de cada cinc habitatges amb televisió, alguna d'elles és de pantalla plana (plasma, LCD). El DVD es troba present en tres de cada quatre llars, en detriment del vídeo.

Un 60,4% disposa d'algun tipus d'ordinador.

Evolució de l'equipament TIC en habitatges



La banda ampla guanya adeptes

6,5 milions d'habitatges familiars tenen accés a internet (un 44,6% del total). El 39% de les llars disposes de connexió mitjançant banda ampla (ADSL, xarxa de cable,...), el que suposa un increment de 10 punts respecte l'any 2006.

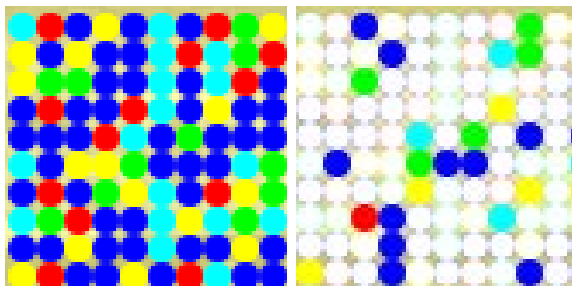
El major percentatge d'habitatges amb accés a internet es dona a Comunitat de Madrid (56,6%) i Catalunya (51,3%). La diferència en punts percentuals entre sexes respecte als principals indicadors d'ús de TIC es redueix el 2007, tant pel que fa a l'ús d'ordinador com en els usuaris freqüents; no obstant això, augmenta en l'accés a internet. El 60,5% dels homes han fet ús d'internet alguna vegada davant d'un 54,1% de les dones. En la Unió Europea, els països que disposen d'indicadors TIC superiors a la mitjana es troben, principalment, al nord d'Europa. Per sota de la mitjana, se situen els països de l'àrea mediterrània o els que són de recent ingrés.



Recorda el més important

Població

Mostra

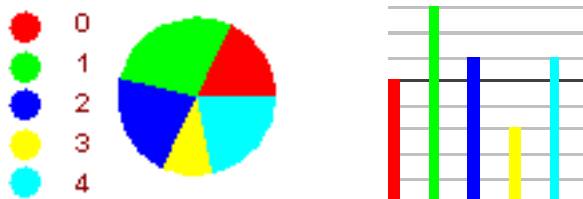


Nre. de germans: 4 3 2 3 1 2 0 2 0 1 2 3 1 2 4
0 1 1 4 1 1 4 0 4 2 0 4 1

Recompte de dades:

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

Gràfics de sectors i barres



Mitjana i moda

$$\text{Mitjana} = \bar{X} = \frac{53}{28} = 1,89$$

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

Quartils i mediana



$$Me=2, Q1=1, Q3=3$$

Rang

De 0 a 4, de amplitud 4

Desviació típica

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

Coefficient de variació

$$CV = 1,39 / 1,89 = 0,73 = 73\%$$

Variabls estadístiques:

- Qualitativa, color preferit;
- Quantitativa discreta, nre. de germans
- Quantitativa contínua, altura.

Altura: 172 162 147 184 140 156
153 186 157 189 162 175 162 158 163
150 152 163 151 182 146 154 163 170
183 162 176 167 168 165

Interval	x_i	f_i	$x_i \cdot f_i$	$(x_i - \bar{x})^2 \cdot f_i$
[140,150)	145	3	435	1200
[150,160)	155	8	1240	800
[160,170)	165	10	1650	0
[170,180)	175	4	700	400
[180,190)	185	5	925	2000
Total		30	4950	4400

Mitjana:

$$\bar{x} = \frac{4950}{30} = 165$$

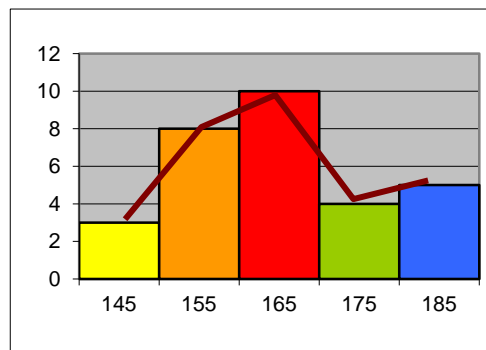
Variància i Desviació típica:

$$\sigma^2 = \frac{4400}{30} = 146,67 \quad \sigma = \sqrt{146,67} = 12,11$$

Coefficient de variació:

$$CV = 12,11 / 165 = 0,073 = 7,3\%$$

Histograma i polígon de freqüències:



Autoavaluació



1. Fes un recompte de les dades següents

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

2. Fes un gràfic de barres per a les dades anteriors.

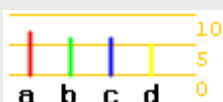

3. Calcula la mitjana de les dades donades per la taula

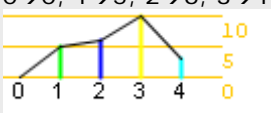
X_i	f_i
1	11
2	3
3	5
4	5

4. Calcula la mediana de les dades anteriors.
5. Calcula el primer quartil de les dades de l'exercici 3
6. Calcula el tercer quartil de les dades de l'exercici 3
7. Calcula en rang de les dades de l'exercici 3
8. Calcula la desviació mitjana de les dades anteriors.
9. Calcula la desviació típica de les dades de l'exercici 3
10. Calcula el coeficient de variació per a les dades del exercici 3.

Solucions dels exercicis per practicar

1. a) 400 b) 50
2. 0,5%
3. V- qualitativa: tipus animal, color d'ulls.
V. quantitativa discreta: nombre de companys.
V. quantitativa contínua: pes, temperatura i metres quadrats.
4. a→7, b→6, c→6 i d→5

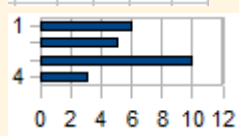


5. 0→0, 1→5, 2→6, 3→10, 4→3


6. En 10: 1→3, 2→6, 3→5, 4→5, 5→5, 6→13, 7→4, 8→2, 9→7 i 10→4 .
En 5: (1 ó 2)→9, (3 ó 4)→10, (5 ó 6)→18, (7 u 8)→6 i (9 ó 10)→11
7. a) $(14+16+18)/3 = 16$
b) $(24+26+28+26)/4 = 104/4 = 26$
c) $(1000+1200+1800+2000)/4 = 1500$
8. $\bar{X} = 2,46$
9. $\bar{X} = 9,77$
10. $M_o = 3$, és el valor que més vegades apareix, 10 en total.
11. 1→5, 2→6, 3→10 i 4→3.
Mediana = 3, $Q_1=2$ i $Q_3=3$
12. a) $(2+0+2)/3 = 1,3333\dots$
b) $(2+0+2+0)/4 = 1$
c) $(500+300+300+500)/4 = 1600/4 = 400$
13. El rang oscil·la entre 1 i 40 amb una amplitud de 39. $\bar{X} = 20,58$, $DM=8,92$

$$DM = \frac{2700}{12} = 225$$
- 14.
15. a) $\bar{X} = 2250$ $\sigma = 250$
b) $\bar{X} = 2250$ $\sigma = 500$
c) $\bar{X} = 2500$ $\sigma = 0$
16. Mitjana = 2,58 D. típica = 1,21
17. $CV = 1,21/2,58 = 0,46$
18. $\bar{X} = 21,08$ i $\sigma = 12,98$
19. $CV = 0,62$
20. $\bar{X} = 15,58$ i $\sigma = 9,68$
21. 78800, 70800, 67500, 67400 i 65300
22. Carretera 67500, aeri 22000, marítim 5000 i ferrocarril 298000
23. Aproximadament 22.000.000, dels quals homes són 14.000.000 i dones 8.000.000

Solucions AUTOAVALUACIÓ

- | | | | | |
|-------|---|---|----|---|
| x_i | 1 | 2 | 3 | 4 |
| f_i | 6 | 5 | 10 | 3 |
1.


 - 2.
 3. $\bar{x} = 2,17$
 4. $Me = 2$
 5. $Q_1 = 1$
 6. $Q_3 = 3$
 7. rang=3
 8. $DM = 1,11$
 9. $\sigma = 1,24$
 10. $Cv = 0,571$

Utilitzar el Full de Càlcul

	A
1	3
2	4
3	5
4	3
5	4
6	6
7	3
8	3
9	4

Introducció de dades i primers càlculs

En general, les següents opcions estan disponibles en tots els fulls de càlcul, en particular en la d'Openoffice i en Excel.

✓ **Obre** un full de càlcul i introdueix les dades 3, 4, 5, 3, 4, 6, 3, 3, 4 en la columna A, en A1, A2,... aquests valors seran les dades sobre les que farem els primers càlculs.

Les dades estan en l'àrea "A1:A9" del full de càlcul.

Si les dades estiguessin en 3 files x 3 columnes, començant per A1, seria "A1:C3".

I per exemple, si amb el ratolí selecciono les cel·les de la imatge adjunta, l'àrea que cobreixo és "A2:C3".

	A	B	C
1	3		
2	4		
3	5		

Hi ha funcions per a tots i cada un dels càlculs estadístics que hem estudiat i moltes més que pots investigar pel teu compte. Les funcions fan càlculs amb els nombres de l'àrea que se'ls proporciona.

Vegem com es farien els càlculs per a les dades que hem introduït. Farem els càlculs en la columna C, la B la usarem per a indicar el nom del càlcul que hi ha a continuació.

B26		f _x	
	A	B	C
1	3	Moda	3
2	4	Màxim	6
3	5	Mínim	3
4	3	Rang	3
5	4	Suma	35
6	6	Mitjana	3,888888889
7	3	Mediana	4
8	3	Quartil 1r	3
9	4	Quartil 3r	4
10		Desviació típica	1,054092553
11		Desviació mitjana	0,790123457
12			

Moda: En la cel·la C1 introdueix la fórmula =**moda(A1:A9)**.

Màxim: En C2 escriu =**max(A1:A9)**

Mínim: Introdueix =**min(A1:A9)**

Rang: =**C2-C3**

Suma: =**suma(A1:A9)**

Mitjana: =**promedio(àrea)** Recorda el que és l'àrea de les dades, **A1:A9**

Mediana: =**mediana(àrea)**

Quartil 1r. =**cuartil(àrea;1)**

Quartil 3r. =**cuartil(àrea;3)**

Desviació mitjana: =**desprom(àrea)**

Desviació típica: =**desvest(àrea)**

Exercici

Introdueix dades en un àrea més gran que l'anterior, al menys de 20 cel·les i fes tots els càlculs de dues formes, amb el full de càlcul i com ho has estudiat. Si tots els resultats són correctes aquest full et val per a qualsevol conjunt de dades d'aquest grandària.

Valors aleatoris.

Si en 20 cel·les escrivim **=aleatorio()** (podem escriure-ho en un i copiar-ho en els altres) tindrem 20 dades entre 0 i 1 aleatoris. Si volem que siguin nombres compresos entre 0 i 20 escrivem **=20*aleatorio()** I si volem que no portin decimals escrivem **=entero(20*aleatorio())** D'aquesta forma ens inventem 20 dades per fer càlculs estadístics.

Comptar

Si tenim introduïts una quantitat gran de dades. Amb la funció **=contar(àrea)** ens dona la quantitat de nombres de l'àrea (imagina que hi ha cel·les buides i no és fàcil comptar les que estan plenes de nombres).

Si el que tenim són dades qualitatives (textos) usarem **=contara(àrea)** per a què ens compti les cel·les ocupades de l'àrea indicada.

Agrupar dades.

- En **variable discreta**. Si en un àrea del full de càlcul tenim 20 nombres enters entre 0 i 4, per exemple. Com podem comptar quants d'ells són un 4? Escrivem en una cel·la en blanc **=contar.si(àrea;4)**. Per comptar la resta usarem en la cel·la que ens convingui **=contar.si(àrea;0)**, **=contar.si(àrea;1)**, **=contar.si(àrea;2)**,...

14	Dades	Marca	Freqüència
15	1	X1=0	5
16	0	X2=1	2
17	3	X3=2	5
18	2	X4=3	3
19	2	X5=4	5
20	4		Suma
21	2		20
22	4		
23	0		
24	2		

Per exemple. En la imatge de l'esquerra hi ha 20 dades generades aleatòriament entre 0 i 4. En la columna Marca indico quins valors comptaré en la següent columna, i en la columna freqüència uso la fórmula **'=contar.si(àrea;valor)'** escrivint:
=contar.si(àrea;0), **=contar.si(àrea;1)**,
=contar.si(àrea;2), ...

Y finalment comprovo sumant les freqüències que s'han comptat totes les dades (en la cel·la D21 escric **=suma(d15:d19)**).

Exercici. Genera 20 dades aleatòries entre 0 i 4 i fes una taula de freqüències usant la funció **contar.si** Calcula tots els paràmetres estadístics estudiats (rang, màxim, mínim, mitjana,...).

- En **variable contínua**. Si tenim 100 altures entre 150 i 200 centímetres en una àrea d'un full de càlcul i volem saber quantes estan en l'interval [150,160) usarem **=contar.si(àrea;">=150")-contar.si(àrea;">=160")**; per l'interval [160,170) usarem **=contar.si(àrea;">=160")-contar.si(àrea;">=170")** i així per a cadascun dels intervals.

- En **variable qualitativa**. Si el que tenim és 100 colors i volem saber quants vermells hi ha s'escriu: **=contar.si(àrea;"vermell")**

Classificar

Segons el tipus de variable que emprem, usarem la fórmula 'contar.si' sobre l'àrea de dades tantes vegades com sigui necessari per classificar totes les dades i ens pugui ser d'utilitat. Així, per a colors usarem **=contar.si(àrea;"vermell")**, **=contar.si(àrea;"verd")**, ... tants cops com colors tinguem, de forma ordenada, en una columna per exemple.

Exemple

Anem a inventar-nos el pes de 50 persones amb valors entre 50 i 110. Usarem la fórmula **=aleatorio()*60+50**, si volem que siguin valors enters es posa **=entero(aleatorio()*50+50)** i la copiarem en l'àrea **A1:E10**.

Ara s'han de classificar les dades. Com totes estan entre 50 i 110 els intervals ho tindran que reflectir. Realitzarem dues taules diferents, una amb valors de 20 en 20 i una altra amb valors de 10 en 10.

Per la primera, copiarem la següent taula en un lloc buit del full de càlcul

Interval, marca	Freqüència
[50,70] 60	=contar.si(A1:E10;">=50")-contar.si(A1:E10;">=70")
[70,90] 80	=contar.si(A1:E10;">=70")-contar.si(A1:E10;">=90")
[90,110] 100	=contar.si(A1:E10;">=90")-contar.si(A1:E10;">=110")

Per la segona, copiarem aquesta taula.

Interval, marca	Freqüència
[50,60] 55	=contar.si(A1:E10;">=50")-contar.si(A1:E10;">=60")
[60,70] 65	=contar.si(A1:E10;">=60")-contar.si(A1:E10;">=70")
[70,80] 75	=contar.si(A1:E10;">=70")-contar.si(A1:E10;">=80")
[80,90] 85	=contar.si(A1:E10;">=80")-contar.si(A1:E10;">=90")
[90,100] 95	=contar.si(A1:E10;">=90")-contar.si(A1:E10;">=100")
[100,110] 105	=contar.si(A1:E10;">=100")-contar.si(A1:E10;">=110")

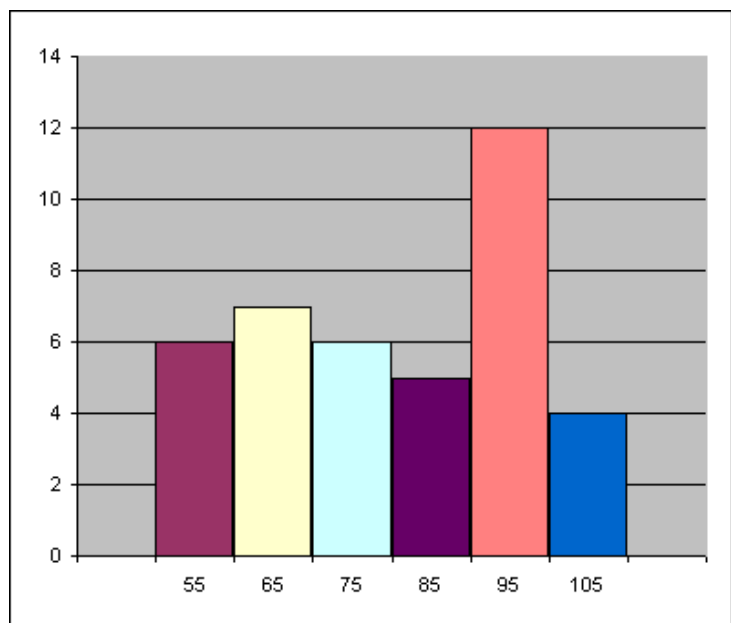
Exercici

Genera 100 nombres de 0 a 80 i fes dues agrupacions en intervals diferents. Per exemple, una de 10 en 10 i una altra de 20 en 20.

Gràfics

Senzill i sense etiquetes. És prou seleccionar l'àrea de les freqüències de les dades i crear el gràfic prement en gràfics del menú inserir. Després podem escollir el tipus de gràfic que volem, i l'aspecte d'aquest entre una gran varietat d'opcions que incorporen tots els fulls de càlcul.

Amb etiquetes: El més senzill és utilitzar la primera fila o primera columna per les etiquetes. En les opcions del gràfic s'haurà d'activar i/o desactivar la que correspongui amb la forma de presentar les dades en el full.



Exercici

Fes dos gràfics diferents, un per a cada una de les distribucions de l'exercici anterior.