

# ESTADÍSTICA.

Recordeu:

**Població:** És el conjunt de tots els elements que compleixen una determinada propietat, que anomenem **caràcter estadístic**. Els elements de la població s'anomenen individus.

**Mostra:** És qualsevol subconjunt de la població. El nombre d'elements de la mostra s'anomena grandària de la mostra.

El **caràcter estadístic** és una propietat que permet classificar els individus d'una població. Classifiquem els caràcters estadístics en **Qualitatius**, els que no podem mesurar, i en **Quantitatius**, els que podem mesurar.

Els valors que prenen els caràcters estadístics quantitatius els anomenem **variables estadístiques**.

Les variables estadístiques poden ser **discretes** i **contínues**.

Una variable estadística és discreta quan només pot prendre un nombre finit de valors.

Representem els seus valors mitjançant:  $x_1, x_2, \dots, x_n$ .

Una variable estadística és contínua quan pot prendre tots els valors possibles dins d'un interval de la recta real. Les representem mitjançant el interval o classe, i agafem com a representant del interval la marca de classe del interval (punt mig de l'interval). Cal que les classes tinguin la mateixa amplitud de l'interval.

Anomenem **distribució estadística** al conjunt de totes les dades estadístiques.

Podem presentar les dades estadístiques mitjançant taules amb les variables estadístiques i les diferents freqüències d'aquestes variables.

**Freqüència absoluta d'un valor**  $x_i$  és el nombre de vegades que apareix aquest valor  $x_i$ . La representarem per  $f_i$

**Freqüència absoluta acumulada d'un valor**  $x_i$  és la suma de les freqüències absolutes dels valors menors o iguals que  $x_i$ . La representem per  $F_i$ . Sabem que:  $F_i = f_1 + f_2 + \dots + f_i$

**Freqüència relativa d'un valor**  $x_i$  és el quocient entre la freqüència absoluta d'aquest valor i el nombre total de dades que tenim a la distribució estadística. La representarem per  $h_i = \frac{f_i}{N}$ .

**Freqüència relativa acumulada d'un valor**  $x_i$  és el quocient entre la freqüència absoluta acumulada del valor  $x_i$  i el nombre total de dades de la distribució estadística. La representem

per  $H_i$ . Sabem que:  $H_i = \frac{F_i}{N} = \frac{f_1 + f_2 + \dots + f_i}{N} = \frac{f_1}{N} + \frac{f_2}{N} + \dots + \frac{f_i}{N} = h_1 + h_2 + \dots + h_i$

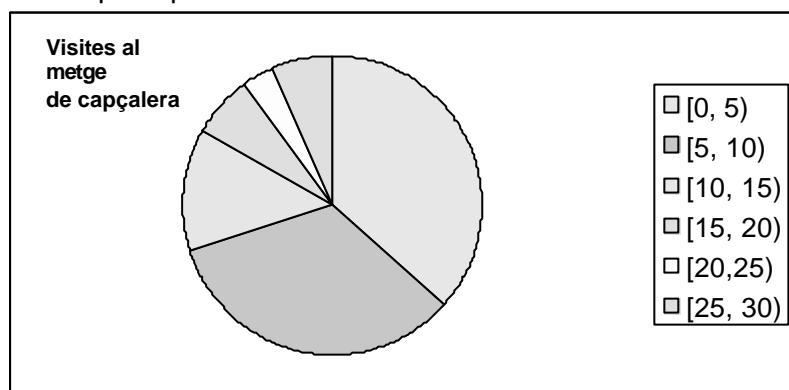
Podem agrupar les dades fent representacions gràfiques. D'aquestes representacions les més importats són:

**Diagrama de barres:** representem les dades mitjançant barres d'alçada proporcional a la seua magnitud. En el diagrama de barres les dades poden ser qualitatives o quantitatives, però sempre sense agrupar en classes.

**Histogrames:** És un diagrama de barres però les dades són sempre quantitatives agrupades en classes o intervals.

El **polígon de freqüències** és la línia trencada que uneix els punts mitjos dels costats superiors dels rectangles que formen les barres. Tant el diagrama de barres com el histograma tenen un polígon de freqüències associat.

**Diagrama de sectors:** Dividim un cercle en sectors d'amplària proporcional a la magnitud de les dades que representen.



Aquest és un exemple de diagrama de sectors, que es correspon amb les dades del segon problema d'autoaprenentatge, el del metge de capçalera.

Què cal fer per analitzar una mostra?

Recollir les dades

Ordenar les dades (en ordre creixent o decreixent)

Recòmpte de freqüències.

Agrupació de les dades: segons la variable aleatòria siga discreta o contínua.

Construcció de la taula estadística.

Els **paràmetres** són uns nombres que ens proporcionen una idea, el més aproximada possible, del comportament de tots els elements d'una població en relació al caràcter que estudiem. Els dividim en dos grups: els paràmetres de centralització i els paràmetres de dispersió.

**Paràmetres de centralització** pretenen agrupar o centralitzar les dades corresponents a tota una població en només un valor numèric, representant del conjunt total. Els més importants són: mitjana aritmètica, moda i mediana.

**La mitjana aritmètica:**  $\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{i=1}^N x_i$ , on  $x_i$  és cadascun dels valors de la

variables estadística i N és el nombre total de dades. El càlcul de la mitjana aritmètica és molt senzill i intervenen totes les dades. Presenta un inconvenient, els efectes, que de vegades són greus, que produeixen els valors extrems. Aquestos valors, normalment no són significatius.

**La moda** és el valor de la variable amb freqüència major. Si les dades les tenim agrupades en classes l'anomenem classe modal.

**La mediana** és el valor de la variable que ocupa el lloc central si tenim un nombre imparell de dades, en cas contrari la mediana coincideix amb la mitjana aritmètica dels valors centrals. La mediana té la propietat de que el 50% de les dades són menors o iguals que ella i l'altre 50% són majors o iguals. Aleshores la mediana divideix el conjunt de dades en dos subconjunts iguals.

**Paràmetres de dispersió** tenen per objectiu presentar-nos una idea de la proximitat o llunyania de les dades de la població respecte al valor que hem pres com a valor central. Els més importants són: el rang, la variança i la desviació típica. Una desviació típica molt gran indica que hi ha molta dispersió de les dades, i una desviació típica molt menuda indica que els dades estan totes al voltant de la mitjana.

**El rang** és la diferència entre el valor màxim i el mínim d'una variable estadística. Si el rang és gran existeix la possibilitat que les dades estiguen allunyades unes d'altres, i per tant, que els valors centrals no siguin massa representatius. Per altra banda, si el rang és menut, les dades no poden trobar-se molt distanciades i els valors centrals poden ser representatius del conjunt.

**La variança** és la mitjana aritmètica dels quadrats de les desviacions de les dades respecte a la

$$\text{mitjana: } \sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N} = \frac{1}{N} \sum_{i=1}^N x_i^2 - \bar{x}^2.$$

**La desviació típica** és l'arrel quadrada de la variança:  $\sigma = \sqrt{\sigma^2}$

Quan en les N dades només apareixen k valors diferents (lògicament  $k < N$ ),  $x_1, x_2, \dots, x_k$ , podem utilitzar la taula de freqüències per obtenir la mitjana i la variança:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k x_i f(x_i) = \sum_{i=1}^k x_i h(x_i) \quad \sigma^2 = \frac{1}{N} \sum_{i=1}^k x_i^2 f(x_i) - \bar{x}^2 = \sum_{i=1}^k x_i^2 h(x_i) - \bar{x}^2$$

(Recordem que f és la freqüència absoluta i h la freqüència relativa)

Exercicis d'autoaprenentatge:

1.

Un professor de 3r ESO de anglés, d'una classe de 25 alumnes, té les següents notes al seu quadern: 5, 3, 4, 1, 2, 8, 9, 7, 6, 8, 6, 7, 9, 8, 7, 7, 1, 0, 1, 5, 8, 0, 8, 8, 4

Es tracta d'una variable aleatòria discreta.

Construïm la taula estadística:

Variable $x_i$	Fre. Abs. $f_i$	Fre. abs. acu. $F_i$	Fre. relat. $h_i$	Fre. rel. acu $H_i$
0	2	2	$\frac{2}{25} = 0'08$	0'08
1	3	5	$\frac{3}{25} = 0'12$	0'2
2	1	6	$\frac{1}{25} = 0'04$	0'24
3	1	7	$\frac{1}{25} = 0'04$	0'28
4	2	9	$\frac{2}{25} = 0'08$	0'36
5	2	11	$\frac{2}{25} = 0'08$	0'44
6	2	13	$\frac{2}{25} = 0'08$	0'52
7	4	17	$\frac{4}{25} = 0'16$	0'68
8	6	23	$\frac{6}{25} = 0'24$	0'92
9	2	25	$\frac{2}{25} = 0'08$	1

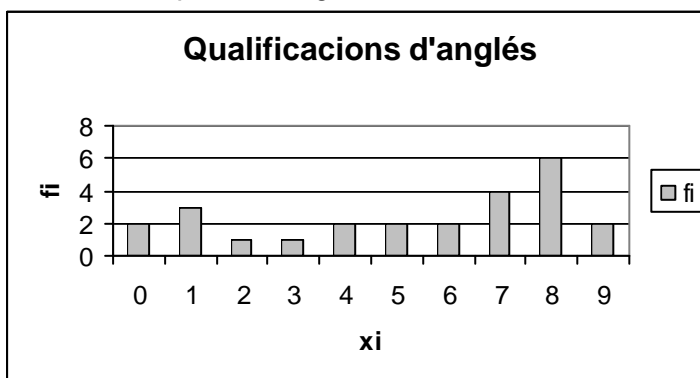
25

1

La suma de totes les freqüències absolutes és el nombre total de dades, en aquest cas concret aquesta suma es correspon amb el nombre d'alumnes: 25.

La suma de totes les freqüències relatives és sempre la unitat.

Fem el corresponent diagrama de barres:



Cal fer un estudi estadístic, calculant les mesures de centralització i les de dispersió i traient les conclusions d'aquest estudi.

Les mesures de centralització són:

La mitjana aritmètica és  $\bar{x} = 5'04$ ,  
la moda és el 8 i la mediana és 6.

Les mesures de dispersió són:

El recorregut és 9, la varianza és 9'1584 i la desviació típica és  $\sigma = 3'02628$

$x_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$
0	2	0	0
1	3	3	3
2	1	2	4
3	1	3	9
4	2	8	32
5	2	10	50
6	2	12	72
7	4	28	196
8	6	42	336
9	2	18	162
	25	126	864

Les nostres conclusions són les següents:

sabem que la mitjana és de 5'04, però amb la dispersió que hi ha de al voltant de 3, podem dir que les notes d'anglès són o molt bones o més dolentes. Sobretot podem dir que, com que la moda és 8 i la mediana és 6, hi ha més gent en bones notes, però també hi ha gent suspesa.

$$\bar{x} = \frac{126}{25} = 5'04$$

$$\sigma^2 = \frac{864}{25} - \left(\frac{126}{25}\right)^2 = 9'1584$$

$$\sigma = \sqrt{9'1584} = 3'02628$$

2.

Hem recollit el nombre de persones que han visitat el metge de capçalera al llarg del mes de novembre: 3, 2, 13, 4, 2, 4, 5, 6, 7, 3, 4, 5, 3, 2, 5, 6, 15, 21, 4, 3, 6, 29, 13, 6, 17, 13, 6, 5, 12, 26. Es demana construir una taula de freqüències amb el seu diagrama corresponent i també un estudi estadístic calculant les mesures de centralització i de dispersió.

Es tracta d'una variable aleatòria contínua. Agrupem les dades en classes que són intervals d'amplària 5. La marca de classe és el punt mig de cada interval.

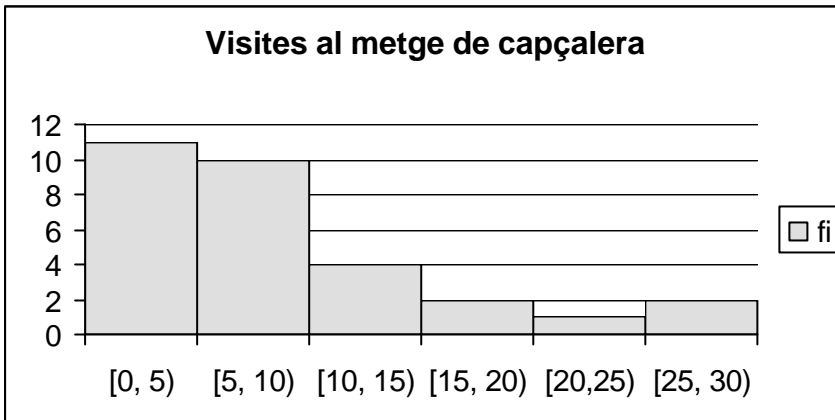
Construïm la taula estadística:

Classes	Marca de classe $x_i$	$f_i$	$F_i$	$h_i$	$H_i$
[0, 5)	2'5	11	11	$\frac{11}{30} = 0'36$	0'36
[5, 10)	7'5	10	21	$\frac{10}{30} = 0'3$	0'7
[10, 15)	12'5	4	25	$\frac{4}{30} = 0'13$	0'83
[15, 20)	17'5	2	27	$\frac{2}{30} = 0'06$	0'9
[20,25)	22'5	1	28	$\frac{1}{30} = 0'03$	0'93
[25, 30)	27.5	2	30	$\frac{2}{30} = 0'06$	1

30

1

Construïm el histograma:



Calcula la mitjana aritmètica, la variança i la desviació típica.

Classes	Marca de classe: $x_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$
[0, 5)	2'5	11	27'5	68'75
[5, 10)	7'5	10	75	562'5
[10, 15)	12'5	4	200	2500
[15, 20)	17'5	2	35	612'5
[20, 25)	22'5	1	22'5	506'25
[25, 30)	27'5	2	54	1485
		30	412	5735

$$\bar{x} = \frac{412}{30} = 13'7\hat{3} \quad \sigma^2 = \frac{5735}{30} - \left(\frac{412}{30}\right)^2 = 2'56\hat{2}$$

$$\sigma = \sqrt{\sigma^2} = 1'6006$$

La mitjana és  $\bar{x} = 13'7\hat{3}$ , la variança és  $\sigma^2 = 2'56\hat{2}$  i la desviació típica és:  $\sigma = 1'6006$ . Podem dir que, més o menys, els nombre de pacients que ha rebut el metge ha sigut, més o menys al voltant de 12-13-14.

Ús de la calculadora:

Cal posar la calculadora en mode STAT polsant  $\boxed{2ndF}$   $\boxed{STAT ON/F}$

Introduir les dades. Per exemple: 3, 4, 7, 9, 12  $\rightarrow$  3  $\boxed{DATA}$   $\boxed{M+}$ ; 4  $\boxed{DATA}$   $\boxed{M+}$ ; 7  $\boxed{DATA}$   $\boxed{M+}$ ; 9  $\boxed{DATA}$   $\boxed{M+}$ ; 12  $\boxed{DATA}$   $\boxed{M+}$

Ara podem saber les mesures diferents:

Per a saber:

$\bar{x}$   $\rightarrow$  cal polsar  $\boxed{x \rightarrow M}$   $\rightarrow$  a la pantalla apareix  $\bar{x} = 7$

$\sigma$   $\rightarrow$  cal polsar  $\boxed{2ndF}$   $\boxed{RM}$   $\rightarrow$  a la pantalla apareix  $\sigma = 3.29$

$\sum x$   $\rightarrow$  cal polsar  $\boxed{2ndF}$   $\boxed{\Sigma}$   $\rightarrow$  a la pantalla apareix  $\sum x = 35$

$\sum n$   $\rightarrow$  cal polsar  $\boxed{n}$   $\rightarrow$  a la pantalla apareix  $n=5$



Problemes:

1. Hem fet un estudi a 100 famílies per a estudiar el nombre de persones que formen la unitat familiar i hem obtingut la taula següent:

Persones per família	Nombre de famílies
1	2
2	8
3	19
4	27
5	25
6	14
7	4
9	1

Construïu la taula de freqüències

Feu la representació gràfica de la distribució.

2. Els pesos, en kg, dels alumnes d'un grup de 4t ESO són:

54, 73, 52, 58, 66, 52, 53, 67, 53, 54 59, 68, 43, 60, 46, 56, 48, 62, 49, 62, 56, 50, 57, 64, 57, 51.
--

Agrupeu les dades en intervals d'amplària 5 kg, sent el primer [40, 45), i construir una taula de freqüències.

Representeu aquesta distribució de pesos mitjançant un histograma.

3. Les altures de tots els alumnes del institut de 4t ESO estan replegades en la taula següent:

Altura en cm	Alumnes
155 a 160	4
160 a 165	7
165 a 170	18
170 a 175	19
175 a 180	12
180 a 185	11
185 a 190	6
190 a 195	3

Construïu la taula de freqüències, agrupant les dades en classes i elegint la marca de classe adient.

Feu el corresponent histograma.

4. Les qualificacions finals de Matemàtiques dels alumnes de 3èr ESO venen donades en la taula següent:

Qualificacions	EX	NT	B	SF	IN
Nº Alumnes	2	5	13	2	3

Construïu una taula de freqüències.

Calculeu el tant per cent de aprovats i el de suspensos.

Representeu gràficament aquesta distribució mitjançant un diagrama de barres i un diagrama de sectors.



5. En una prova de Biologia, de tipus test, de 60 qüestions, els alumnes han contestat correctament:

23; 41; 19; 35; 25; 49; 55; 8; 38; 46; 59; 35; 25; 28; 32; 37; 45; 41; 25; 35; 42; 51; 21; 29; 43; 39; 33; 47; 30.

Construiu una taula de freqüències agrupant les dades en 6 intervals.

Representeu gràficament aquesta distribució.

6. A 100 alumnes els hem fet un test amb 60 preguntes. En la taula següent hem replegat el nombre de preguntes que han contestat correctament:

Nombre de preguntes	Nombre d'alumnes
0-9	5
10-19	10
20-29	24
30-39	33
40-49	21
50-60	7

Representeu-la gràficament mitjançant un histograma. Quin percentatge d'alumnes contesten bé a més de la meitat de les preguntes?

7. Ompliu la taula següent, amb els pesos de tots els alumnes de la vostra classe quan vàreu nàixer:

Peso en Kg	Nº de Xiquets/Xiquetes
2,5 a 2,9	
2,9 a 3,3	
3,3 a 3,7	
3,7 a 4,1	
4,1 a 4,5	

Construiu la taula de freqüències. Representar gràficament aquesta distribució.

Calculeu el recorregut

Calculeu el pes mitjà

Calculeu la desviació típica.

Escriviu les conclusions al voltant d'aquest estudi.

8. Compareu els resultats acadèmics de dues classes que han obtingut les qualificacions finals següents:

CLASSE A	
Insuficients	7
Suficients	20
Notables	8
Excel·lents	2

CLASSE B	
Insuficients	7
Suficients	15
Notables	4
Excel·lents	4

a) Per fer aquest estudi calculeu el tant per cent dels alumnes que han aconseguit cadascuna de les qualificacions i representeu-la gràficament mitjançant un diagrama de sectors.

b) Feu l'estudi amb les mesures de centralització i de dispersió

9. La quantitat de diners que porten els 25 alumnes d'una classe del institut en un moment determinat és:

Euros	0 – 0'5	0'5 -1	1 – 1'5	1'5 - 2	2 – 2'5	2'5 - 3
Nombre d'alumnes	2	4	2	10	4	3

Calculeu la mitjana i la desviació típica dels diners que porten a la butxaca aquestos alumnes. Escriviu conclusions a propòsit d'aquest estudi.

10. El pes mitjà dels alumnes d'una classe del institut és de 65 Kg.

Incorporem una persona al grup que pesa 65 Kg, quin serà ara el nou pes mitjà?

Quin és el nou pes mitjà si la persona que s'incorpora pesa 70kg?

Coneixem que la persona que s'incorpora és un jutge, i baixa la mitjana a 64 kg. Pot ser veritat?. I si a la persona que s'incorpora li agraden molt les llepolies?

11. Una família consta d'un matrimoni i quatre fills. Tots els membres de la família són treballadors a sou. Quin paràmetre estadístic, dels seus sis sous, informa millor de la riquesa familiar: la moda, la mitjana o la variança? Quin paràmetre estadístic informa millor sobre la diversitat de sous: la mediana, la mitjana o la desviació típica?

12. Els augments dels preus de cinc productes alimentaris han sigut, respectivament, del 21, 31, 34, 48, 3 per cent. Quina mesura de centralització reflexa millor el terme mitjà del augments, la mediana o la mitjana aritmètica? Calculeu ambdós valors.

13. En una classe de 4t ESO de 25 alumnes, hi ha 10 alumnes de francès i 15 d'anglès. La nota mitjana dels 25 alumnes en Llengua Estrangera és 6.5. Coneixem que la nota mitjana dels alumnes de francès és de 6. Quina és la mitjana dels alumnes d'anglès?

14. Calculeu totes les mesures de centralització i de dispersió de les qualificacions obtingudes pels alumnes de la classe en Matemàtiques:

3, 7, 8, 5, 4, 0, 6, 6, 7, 5, 2, 10, 9, 3, 4, 6, 6, 5, 4, 10, 1, 8, 7, 6, 3, 8, 6, 7, 5, 4.

15.

Considerem les notes d'un examen de matemàtiques en dos grups A, B. La taula de freqüències és:

Grup A	
Notes	Fre.abs.
$x_i$	$f_i$
2	2
3	2
4	4
5	6
6	2
8	4
Total	10

Grup B	
Notes	Fre.abs.
$x_i$	$f_i$
0	4
1	4
3	2
4	2
7	4
8	2
9	2
10	4
Total	

Representeu els resultats dels dos grups en diagrames de barres. Estudieu les mesures de centralització i dispersió dels dos grups.

Comenta els resultats obtinguts.

16.

En un grup de 40 alumnes, comptem el nombre de suspensos per alumne i s'obté la següent taula de freqüències.

Calculeu les mesures de centralització i les mesures de dispersió.

Nombre de suspensos $x_i$	freqüència absoluta $f_i$
0	8
1	7
2	1
3	15
4	5
5	3
6	1
Total	40

17.

Les edats dels professors d'un centre escolar són:

35, 26, 40, 27, 37, 36, 37, 50, 55, 35, 60, 26, 30, 36, 40, 41, 43, 36, 38, 42, 46, 52, 62, 57, 36, 41, 46, 36, 46, 37, 37, 41, 55, 32, 50, 52, 37, 42, 62, 55, 56, 47, 30, 42, 44, 44, 37, 48, 50, 55.

a) Ompliu la taula següent:

Classes	Marca de classe $x_i$	Fre. Abs. $f_i$
[25,30[		
[30,35[		
[35,40[		
[40,45[		
[45,50[		
[50,55[		
[55,60[		
[60,65[		

Representeu els resultats en un histograma.

Estudieu les mesures de centralització i dispersió.