

## Distribucions (taules) de freqüències

És el primer pas per resumir un conjunt d'observacions i consisteix en enumerar els diferents valors observats en una variable i el número de repeticions.

La **freqüència absoluta** d'un valor ( $n_i$ ) és el número de vegades que es repeteix aquest valor. La llista de freqüències constitueix la taula de freqüències. La suma de totes les freqüències absolutes és el número total de casos ( $n$ ).

La **freqüència relativa** ( $f_i$ ) és  $\frac{n_i}{n}$  on  $n$  és el número total de casos. La suma de totes les freqüències relatives és 1.

El **percentatge** ( $p_i$ ) és la freqüència relativa multiplicat per 100. La suma de tots els percentatges és 100.

### Exemple:

NIVELL	$n_i$	$f_i$	$p_i$
ALT	11233	0.09	9%
MITJÀ	81234	0.66	66%
BAIX	30144	0.25	25%
	122611	1	100%

Els valors acumulats s'obtenen sumant els valors anteriors a cada freqüència o percentatge. Així s'obté la **freqüència absoluta acumulada** ( $N_i$ ), la **freqüència relativa acumulada** ( $F_i$ ), i el **percentatge acumulat** ( $P_i$ ).

Té més sentit en escales ordinals. Normalment s'ordenen les dades en sentit ascendent i ens donen el número o percentage de casos amb valor menor o igual que el donat.

### Exemple:

	$n_i$	$f_i$	$p_i$	$N_i$	$F_i$	$P_i$
Baix	30144	0.25	25%	30144	0.25	25%
Mig	81234	0.66	66%	111378	0.91	91%
Alt	11233	0.09	9%	122611	1	100%
	122611	1	100%			

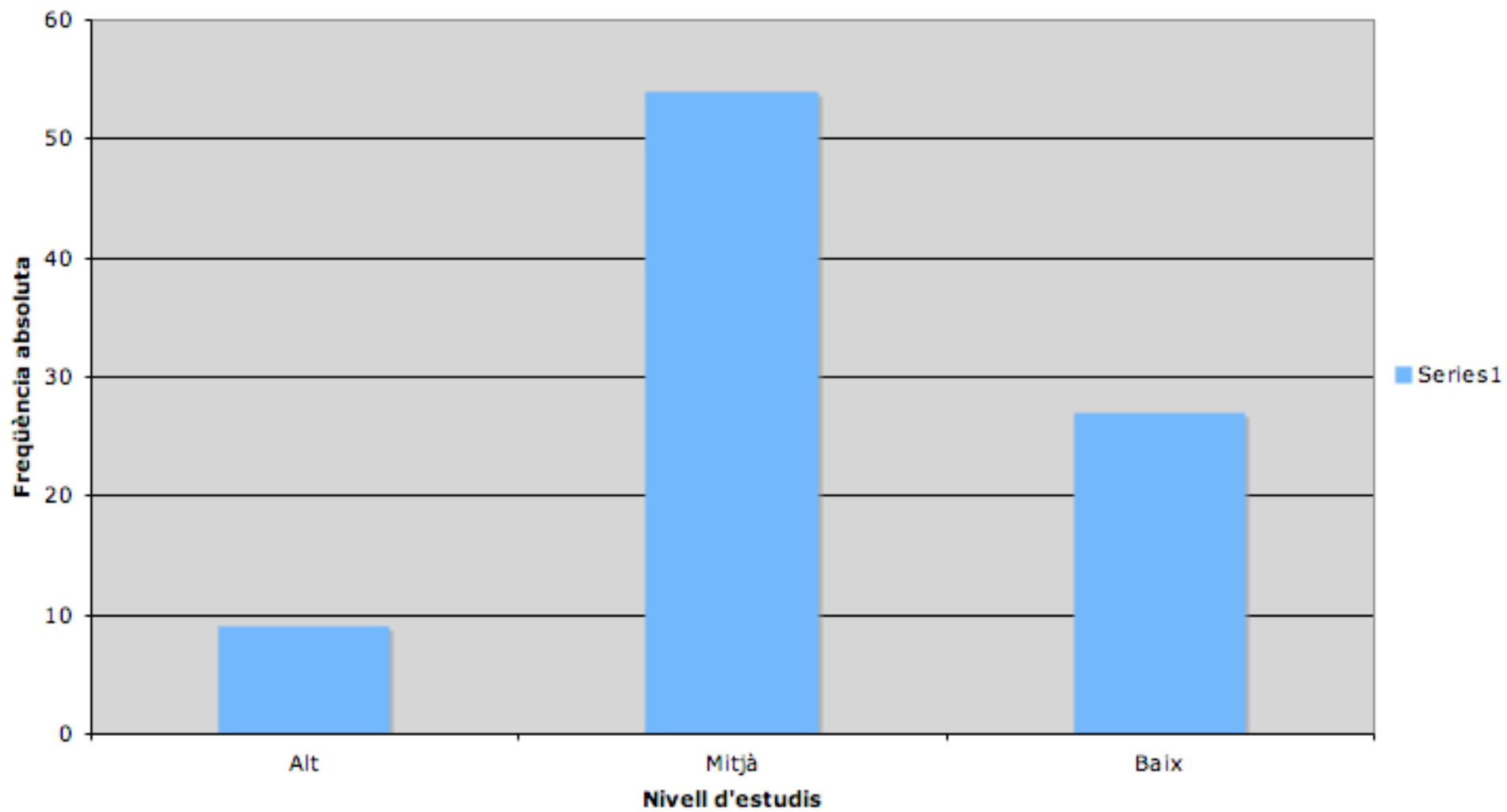
Observeu que la última línia ha de contenir els totals.

Podem afirmar que més d'un 90% de la població té un nivell d'estudis inferior o igual a mig.

## Representació gràfica d'una variable qualitativa

- **El diagrama de barres** Consisteix en aixecar, a sobre de l'eix horitzontal on es representen els valors de la variable, una barra d'altura igual a la freqüència (absoluta) o proporcional (freqüència relativa).

**Diagrama de barres del nivell d'estudis**



## Representació gràfica d'una variable qualitativa

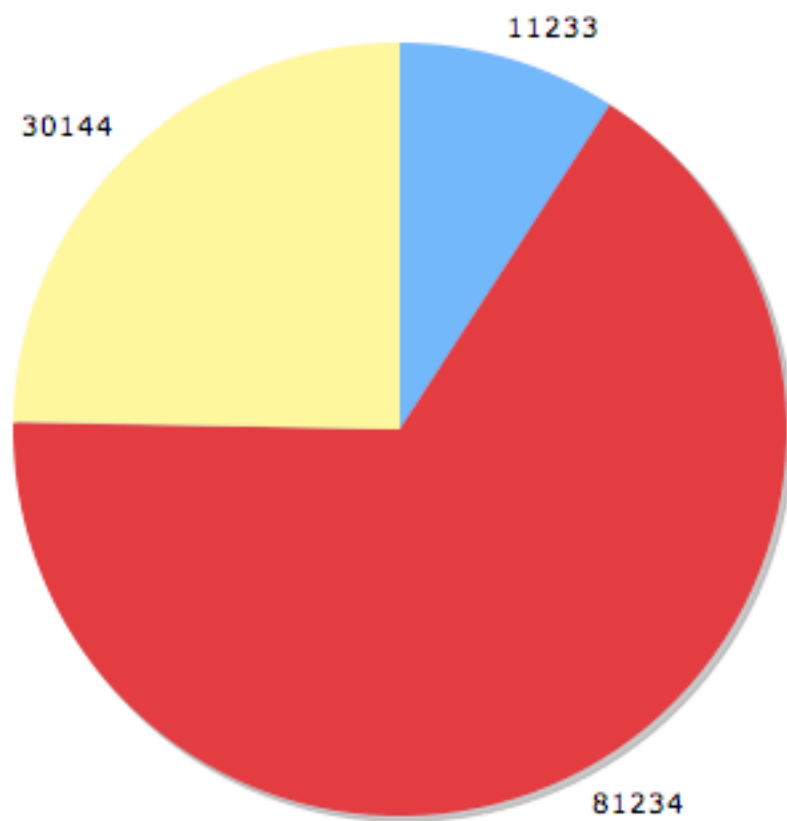
- **El diagrama de sectors** Consisteix en repartir l'àrea d'un cercle (o els  $360^\circ$ ) de manera proporcional (percentatge) entre els diferents valors que pren la variable.

A cada valor li correspon l'angle

$$360^\circ \cdot f_i = 360^\circ \cdot \frac{n_i}{n}$$

Recordeu que  $p_i$  és la freqüència relativa

**Diagrama de sectors de nivell d'estudis**



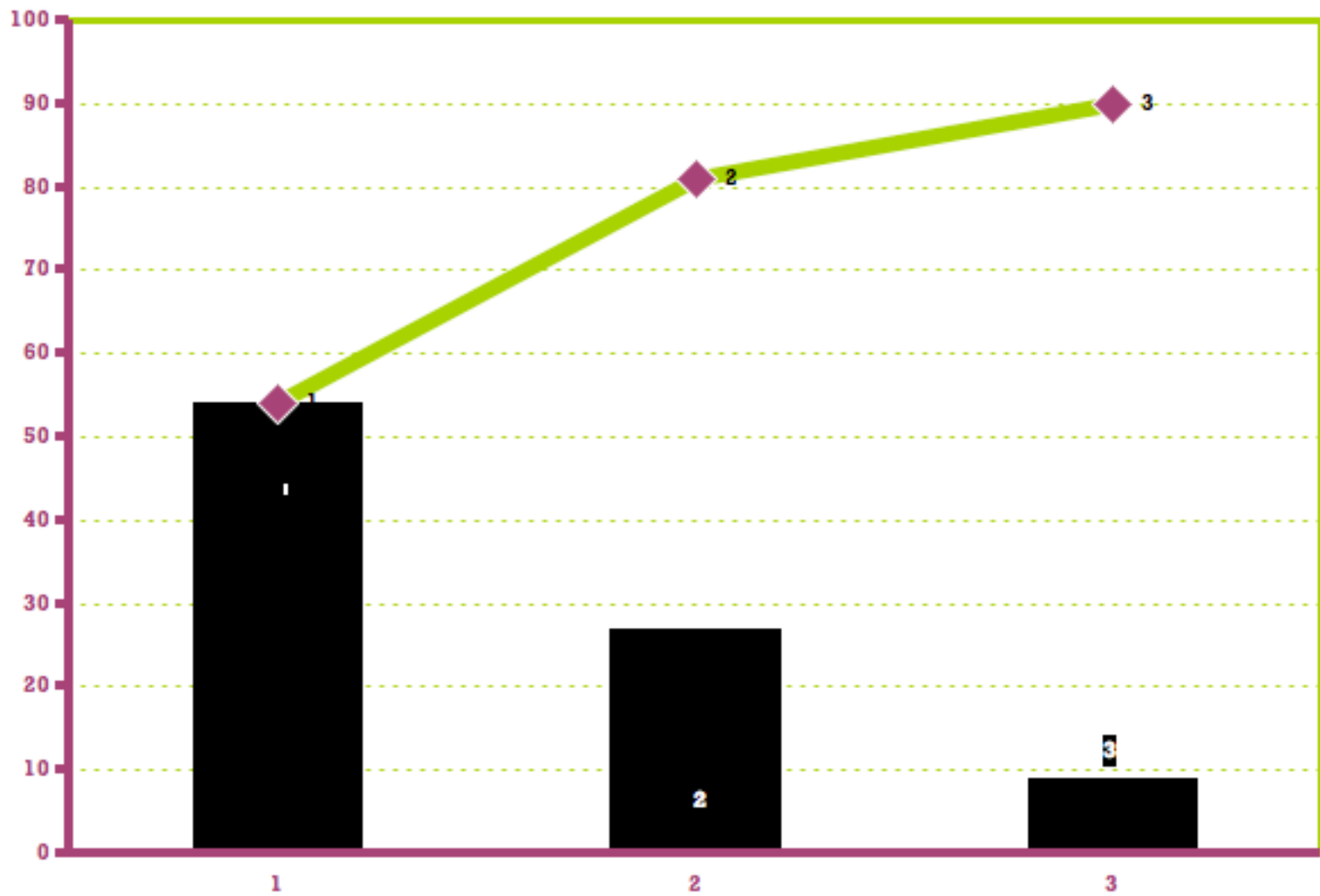
Alt  
Mitjà  
Baix

## Representació gràfica d'una variable qualitativa

- **El diagrama de Pareto** És una combinació de diagrama de barres i diagrama poligonal o de línia (s'uneixen els valors amb una línia).

És un diagrama de barres en què els valors s'ordenen de més freqüent (freqüència més alta) a menys freqüent (freqüència més baixa). S'acompanya d'una línia que uneix les freqüències acumulades.





## Mesures de tendència central : variable qualitativa

- **La Moda:** la moda és el valor més freqüent de la variable. És el corresponent a la freqüència absoluta, relativa o percentatge més alt.

Per exemple, s'ha fet una enquesta on es pregunta per la qualitat de l'atenció sanitària pública: 3 persones han dit que és excel·lent, 6 que és bona, 7 que és correcte i 4 que és dolenta.

El valor més freqüent és el 'correcte'. La moda és el valor 'correcte'.

## Propietats:

- La moda també és calculable per variables numèriques un cop han estat agrupades en intervals. Aleshores podem parlar de l'interval modal com aquell que té una freqüència més alta o bé podem dir que la moda és la marca de classe de l'interval amb freqüència més alta.

Recordeu el següent exemple. Les següents dades són les notes (puntuació sobre 100) obtingudes en un examen,

55.8	60.9	37.0	91.3	65.8
42.3	33.8	60.6	76.0	69.0
45.9	39.1	35.5	56.0	44.6
71.7	61.2	61.5	47.2	74.5
83.2	40.0	31.7	36.7	62.3
47.3	94.6	56.3	30.0	68.2
75.3	71.4	65.2	52.6	58.2
48.0	61.8	78.8	39.8	65.0
60.7	77.1	59.1	49.5	69.3
69.8	64.9	27.1	87.1	66.3

En aquest cas, l'agrupació que té més sentit és la de considerar els intervals  $[0, 50)$ ,  $[50, 70)$ ,  $[70, 90)$  i  $[90, 100]$  (que correspon a suspens, aprovat, notable i excel·lent).

Nota	$n_i$	$f_i$	$p_i$	$n_i$	$f_i$	$p_i$
[0, 50)	17	0.34	34%	17	0.34	34%
[50, 70)	22	0.44	44%	39	0.78	78%
[70, 90)	9	0.18	18%	48	0.96	96%
[90, 100]	2	0.04	4%	50	1	100
Total	50	1	100			

L'interval modal seria [50, 70), o bé, la moda és 60.

## Propietats:

- La moda pot no ser única (quan a dos o més valors els correspon la màxima freqüència) i aleshores tenim dues o més modes.

Per exemple, s'ha preguntat a un grup de persones pel seu gust preferit de gelat: 10 han triat vainilla, 40 preferiexen xocolata, 40 han triat maduixa i 10 han triat cafè.

Aleshores hi ha dues modes: xocolata i maduixa.

## Propietats:

- La moda és més significativa quan la seva freqüència és sensiblement superior a les altres.

Per exemple, a la pregunta de quantes hores mires la televisió a la setmana, hem obtingut que 10 persones han dit que no tenen televisió, 3 persones només 1 hora, 3 persones 2 hores, 2 persones 3 hores, 9 persones la miren 4 hores, 8 persones 5 hores, 8 persones 6 hores.

La moda és 0 hores però no ens dóna informació de com s'acumulen les dades. Ens calen altres mesures de tendència central.

- **La Mediana:** en una variable de tipus ordinal (o numèrica) la mediana és el valor tal que la meitat dels casos prenen valors menors o igual que ell (i l'altra meitat pren valors més grans o iguals).

És a dir, si ordenem els valors de més petit a més gran, la mediana és el valor que trobem a al mig.

Per exemple, recordeu l'últim exemple de les hores que es mira la televisió a la setmana. Els ordenem

0,0,0,0,0,0,0,0,0,0,1,1,1,2,2,2,3,3,4,4,4,4,4,  
4,4,4,4,5,5,5,5,5,5,5,5,6,6,6,6,6,6,6

En total tenim 43 persones. La posició 22 es troba just a la meitat. Hi ha un 4. Per tant, la mediana és 4.



1. **Si el número de casos es senar**, hi ha exactament un valot situat al mig i aquest és la mediana. Per les dades 3, 27, 30, 51, 100, la mediana és 30.
2. **Si el número de casos es parell**, la mediana és la mitjana dels dos valors que ocupen les posicions centrals. Per les dades 3, 3, 5, 7, els dos valors que ocupen les posicions centrals són 3 i 5, aleshores la mediana és  $\frac{3+5}{2} = 4$ .

## **Propietats:**

- Només té sentit en escales ordinals (qualitatives o numèriques).
- Representa un punt central en la distribució de dades (el punt que es troba al mig de tots els valors ordenats).
- No es veu afectada per la introducció d'uns pocs valors extrems.

## Com calcular la mediana amb la taula de freqüències

Sigui  $n$  el número total de casos. Les taules de freqüència es calculen amb els valors de menor a més gran.

1. Si  $n$  és senar, la mediana és el primer valor amb un percentatge acumulat superior a 50%.

Per exemple,

	$n_i$	$f_i$	$p_i$	$N_i$	$F_i$	$P_i$
1	30144	0.25	25%	30144	0.25	25%
2	81234	0.66	66%	111378	0.91	91%
3	11233	0.09	9%	122611	1	100%
	122611	1	100%			

La mediana és el valor 2.

2. Si  $n$  és parell i cap valor té un percentatge acumulat igual al 50%, la mediana és el primer valor amb un percentatge acumulat superior a 50%.

	$n_i$	$f_i$	$p_i$	$N_i$	$F_i$	$P_i$
0	10	0.24	24%	10	0.24	24%
1	3	0.07	7%	13	0.31	31%
2	3	0.07	7%	16	0.38	38%
3	2	0.05	5%	18	0.43	43%
4	9	0.21	21%	27	0.64	64%
5	8	0.19	19%	35	0.83	83%
6	7	0.17	17%	42	1	100%
	42	1	100%			

La mediana és 4.

3. Si  $n$  és parell i un valor té un percentatge acumulat igual al 50%, llavors la mediana és la mitjana entre aquest valor i el següent.

	$n_i$	$f_i$	$p_i$	$N_i$	$F_i$	$P_i$
1	30	0.5	50%	30	0.50	50%
2	20	0.33	33%	50	0.83	83%
3	10	0.17	17%	60	1	100%
	60	1	100%			

La mediana és  $(1 + 2)/2 = 1.5$ .

Enlloc del 50% també podem considerar altres percentatges de les dades que quedin acumulats a sota d'un valor. Aleshores obtenim els quartils (25%, 50% i 75%).

El **primer quartil**  $Q_1$  és el valor que satisfà que el 25% de les dades prenen valors més petits o iguals.

El **segon quartil**  $Q_2$  és el valor que satisfà que el 50% de les dades prenen valors més petits o iguals. (És la mediana)

El **tercer quartil**  $Q_3$  és el valor que satisfà que el 75% de les dades prenen valors més petits o iguals.

Important: quan un dels valors té un percentatge acumulat exactament igual a 25%, 50% o 75%, aleshores el quartil corresponent és la mitjana entre aquest valor i el següent.

El rang interquartílic és la diferència entre el tercer quartil i el primer quartil,  $Q_3 - Q_1$ .

En general podem parlar de percentils.

El **percentil  $r$**  o  **$r$ -éssim centil** és el valor que satisfà que el  $r\%$  de les dades prenen valors més petits o iguals.

Quan un dels valors té un percentatge acumulat exactament igual a  $r\%$ , aleshores el quartil corresponent és la mitjana entre aquest valor i el següent.

	$n_i$	$f_i$	$p_i$	$N_i$	$F_i$	$P_i$
0	10	0.24	24%	10	0.24	24%
1	3	0.07	7%	13	0.31	31%
2	3	0.07	7%	16	0.38	38%
3	2	0.05	5%	18	0.43	43%
4	9	0.21	21%	27	0.64	64%
5	8	0.19	19%	35	0.83	83%
6	7	0.17	17%	42	1	100%
	42	1	100%			

$Q_1 = 1, Q_2 = 4, Q_3 = 5.$

El percentil 30 és 1. El percentil 90 és 6.