

## Repàs:

Una **població** és el conjunt de tots els individus u objectes d'interés. (Totes les observacions possibles).

Per exemple, si ens preguntem per la intenció de vots dels habitants dels EEUU en les properes eleccions, la població estarà formada per tots els ciutadans censats als EEUU.

Per exemple, si ens preguntem per l'alçada dels eucaliptus, la població estarà per tots els eucaliptus vius del planeta.

Una **mostra** és un subconjunt de la població.

Per exemple, si realitzem una enquesta a 1000 habitants dels EEUU per saber la seva intenció de vot, el conjunt d'individus que han contestat l'enquesta és una mostra.

## **Estadística descriptiva:**

Mètode: recollir dades, presentar dades, caracteritzar les dades numèricament.

Objectiu: descriure un conjunt de dades. (fer afirmacions sobre la mostra).

## **Estadística inferencial:**

Mètode: mostreig aleatori, estimacions, test d'hipòtesis.

Objectiu: prendre decisions sobre una població basades en les característiques d'una mostra.

## **Variables i dades:**

Una variable és una mesura que pot prendre valors diferents en individus i circumstàncies diferents. Quan el seu gran d'incertesa és molt gran direm que és una variable aleatòria.

## **Tipus de variables:**

- Numèriques o quantitatives
  - Discretes
  - Contínues
  
- No-numèriques o qualitatives
  - Nominal
  - Ordinal

**Tipus d'escala de mesura:** (depèn de la natura de les observacions)

- Escala nominal: només permet una classificació dels individus.
- Escala Ordinal: classifica els individus i permet establir ordres.
- Escala numèrica: a més de classificar els individus, ens permet operar amb els valors.

El tipus de variable i escala determinarà els mètodes descriptius que podrem utilitzar en cada cas. Té sentit calcular la mitja d'una variable qualitativa nominal?

## 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 ( $n$ ).

**Exemple:** D'una població A hem agafat una mostra i preguntat a la gent el seu nivell d'estudis. Hem obtingut que 9 persones tenen un nivell alt, 54 mitjà i 27 baix.

| NIVELL | $n_i$      |
|--------|------------|
| ALT    | $n_1 = 9$  |
| MITJÀ  | $n_2 = 54$ |
| BAIX   | $n_3 = 27$ |
| Total  | $n = 90$   |

La suma de totes les freqüències absolutes és número total de casos (en aquest cas, enquestats).

**Exemple:** Ara enquestem a tota la població i no només a una mostra, obtenim

| NIVELL | $n_i$         |
|--------|---------------|
| ALT    | $n_1 = 11233$ |
| MITJÀ  | $n_2 = 81234$ |
| BAIX   | $n_3 = 30144$ |
| Total  | $n = 122611$  |

Com podem comparar els resultats de la mostra amb els de la població? Imagineu que tenim una altra població, com podem comparar les freqüències?

És difícil fer-ho ja que el número total d'enquestats és diferent.

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%  |



Ara comparem les dues taules

| NIVELL | $n_i$ | $f_i$ | $p_i$ |
|--------|-------|-------|-------|
| ALT    | 9     | 0.1   | 10%   |
| MITJÀ  | 54    | 0.6   | 60%   |
| BAIX   | 27    | 0.3   | 30%   |
|        | 90    | 1     | 100%  |

| 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%  |        |       |       |

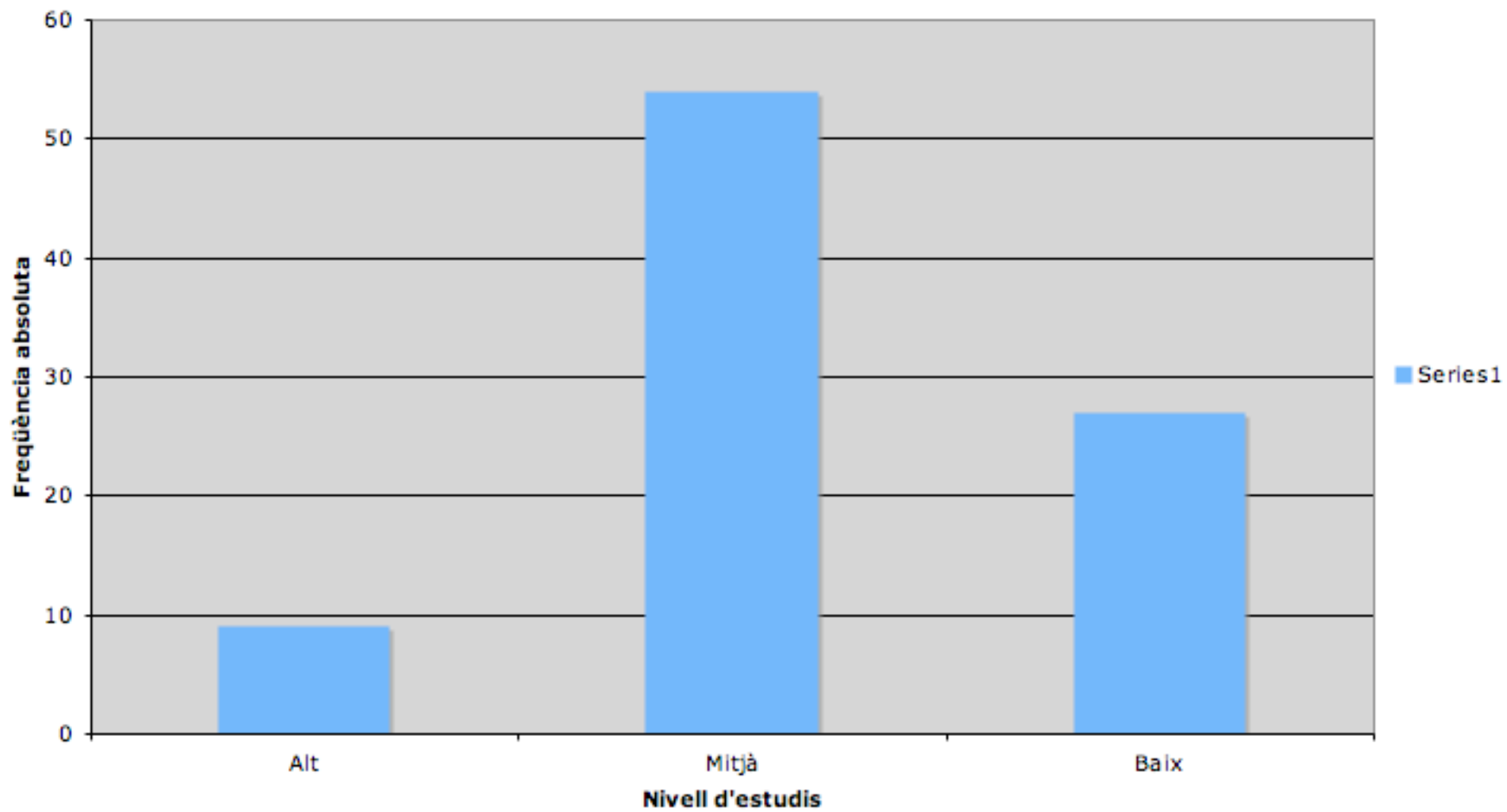
Observeu que la última línia ha de contenir els totals.

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

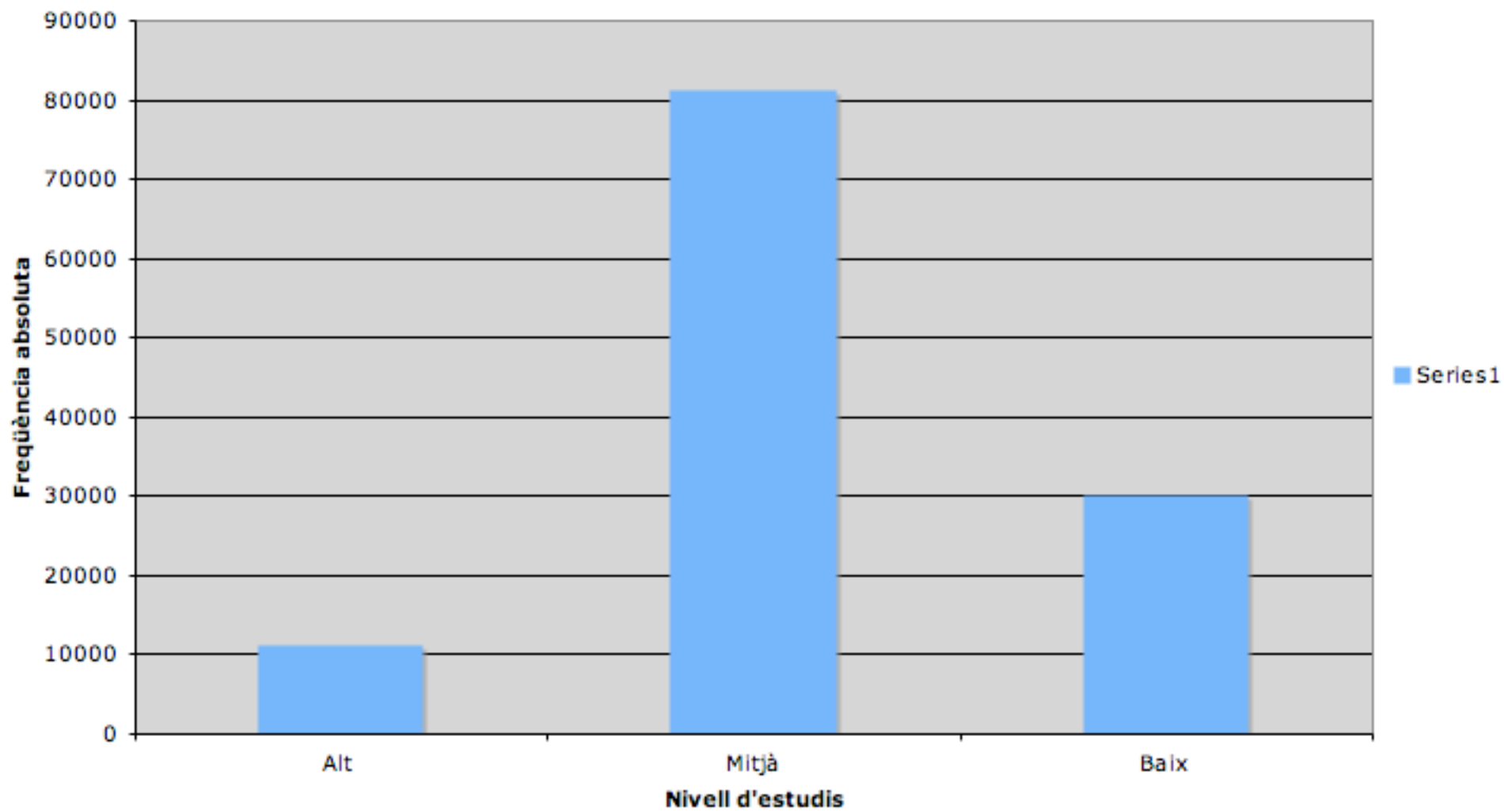
## Representació gràfica d'una variable qualitativa

- **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).

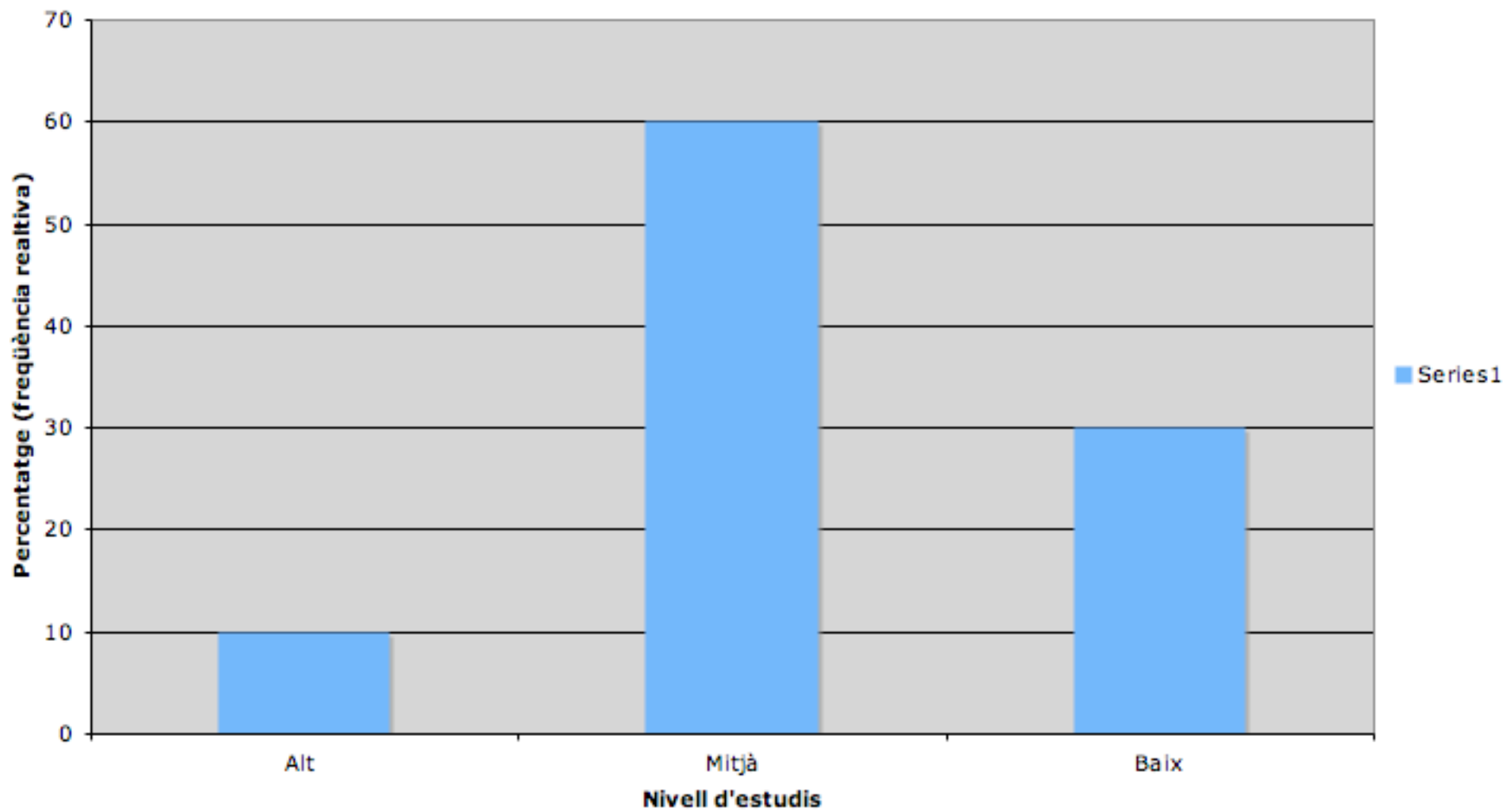
**Diagrama de barres del nivell d'estudis**



**Diagrama de barres de Nivell d'estudis**



**Diagrama de barres de Nivell d'estudis**



## Representació gràfica d'una variable qualitativa

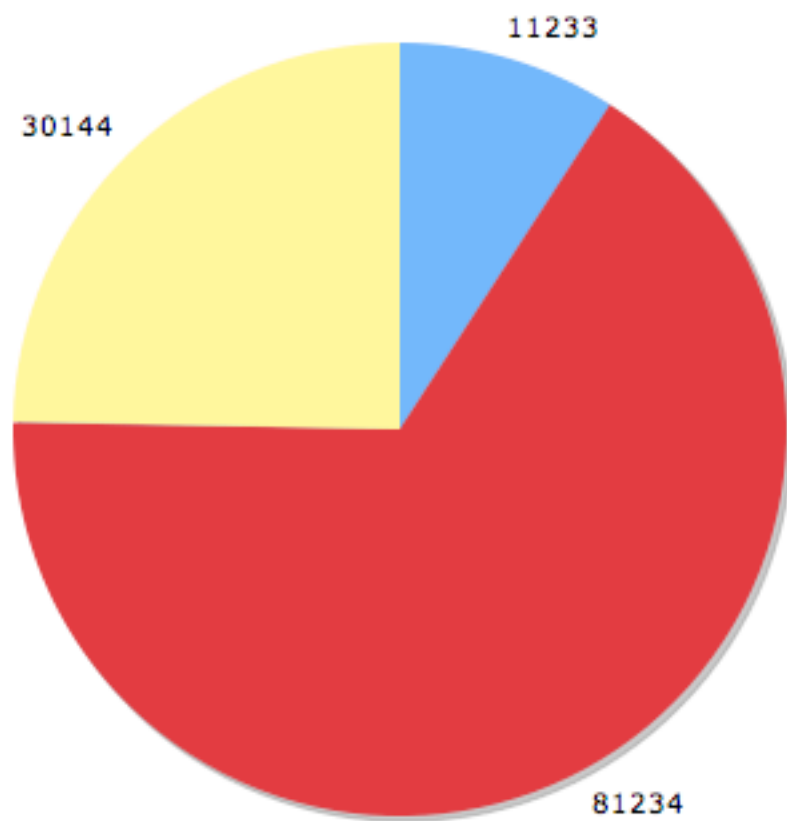
- **El diagrama de sectors** Consisteix en repartir l'àrea d'un cercle (o els  $360^\circ$ ) de manera proporcional (percentatge) entre els diferents valors que pren la variable.

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

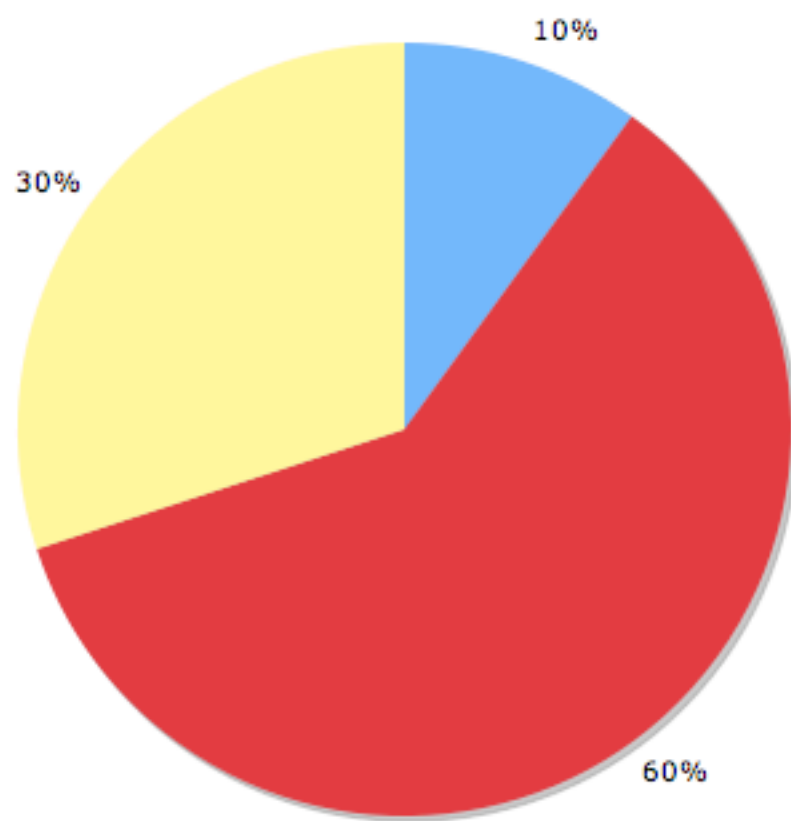
**Diagrama de sectors de nivell d'estudis**



Alt  
Mitjà  
Baix



**Diagrama de sectors de Nivell d'estudis**

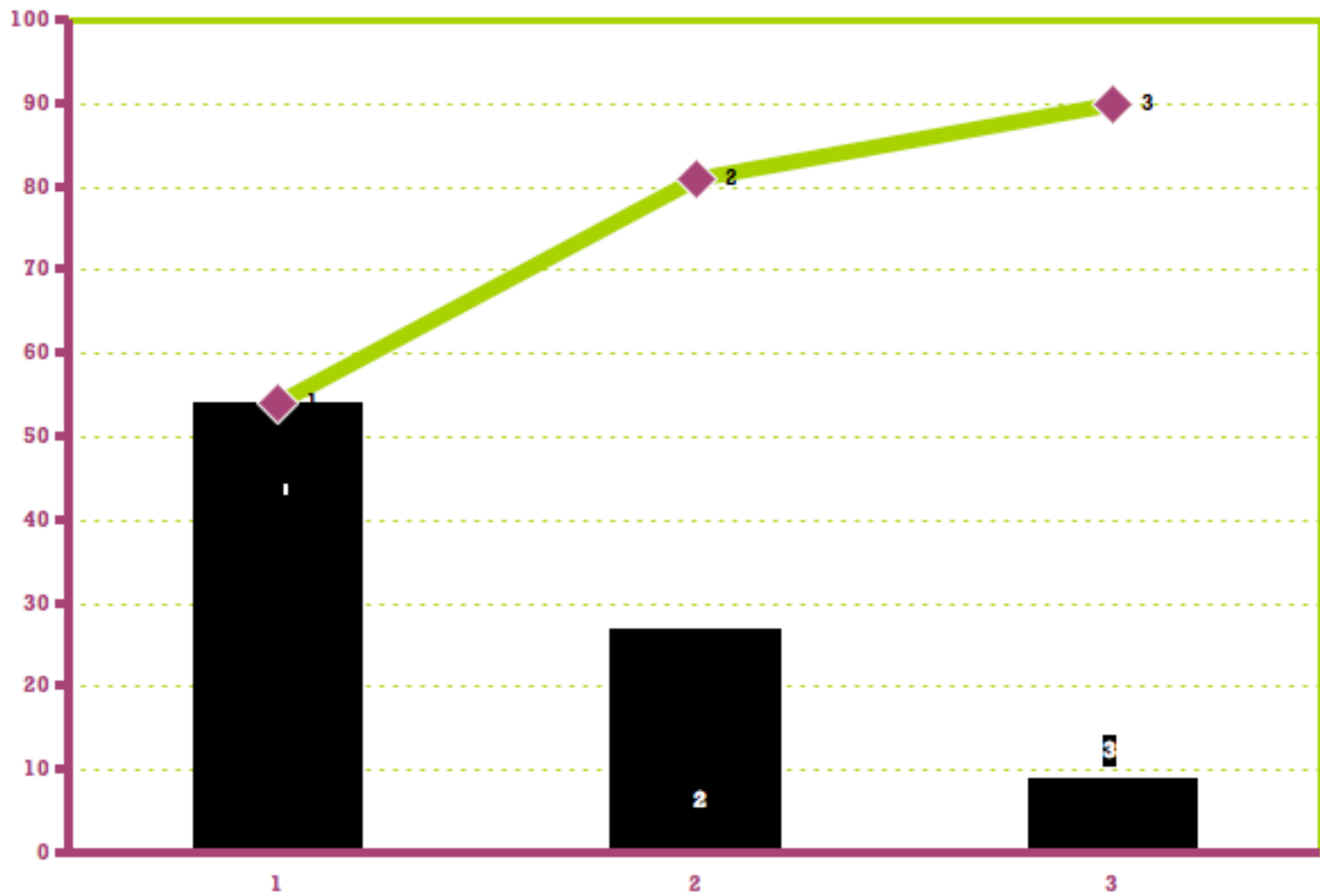


■ Alt  
■ Mitjà  
■ Baix

## 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.



De quin tipus (tipus de variable i escala) són les variables següents:

1. Edats de la població de Catalunya (nens, joves, adults, gent gran).
2. Nombre d'anuncis en una pausa publicitària d'una cadena.
3. Temps necessari per arribar a la feina (mesurat en segons).
4. Temperatura del lloc de treball.
5. Nombre de menors d'edat per llar a Catalunya.

Considereu la taula següent (la informació correspon a Catalunya l'any 2001):

| Persones que hi habiten | Milers d'habitatges |
|-------------------------|---------------------|
| 1                       | 381.2               |
| 2                       | 693.6               |
| 3                       | 491.9               |
| 4                       | 454.8               |
| 5 o més                 | 159.2               |

1. Quina és la variable estudiada i de quin tipus és.
2. Quins són els objectes estudiats?
3. Feu una taula de freqüència (amb les freqüències absolutes, relatives, percentatges i les acumulades). Quin sentit tenen els percentatges acumulats?
4. Feu un diagrama de barres, de sectors i de pareto.