Matemàtiques orientades a les ensenyances acadèmiques

# 4tB ESO
# Capítol 13: Estadística



LibrosMareaVerde.tk

www.apuntesmareaverde.org.es


Autora: Raquel Caro

Revisores: María Molero i Nieves Zuasti

Il·lustracions: Banc d'Imatges d'INTEF

Traducció: Pedro Podadera, IES Juan de Garay

L'Estadística s'utilitza en la Ciència. També per a fer **sondejos d'opinió**, com l'acceptació pel públic d'un programa de televisió, o les enquestes sobre la intenció de vot a un partit polític. S'usen tècniques estadístiques als processos de fabricació, és el **control de qualitat**. Per a fer previsions i programes del tràfic, o les necessitats d'energia d'un país. Quan s'analitza un fenomen observable apareixen una sèrie de resultats que han de ser tractats convenientment, de manera que es puguen comprendre millor tant els resultats com la característica objecte d'estudi corresponent al dit fenomen. Per a aquest fi s'utilitza l'Estadística.

En aquest capítol aprendrem a reconèixer i classificar distints tipus de variables estadístiques, construir taules de freqüències i gràfics estadístics per a distints tipus de variables estadístiques i determinar i interpretar mesures de centralització, posició i dispersió.

També ens centrarem a l'estudi de dos variables d'interès corresponents a dues característiques (o variables) distintes. En aquest sentit, pot ser interessant considerar simultàniament els dos caràcters a fi d'estudiar les possibles relacions entre ells.


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

## 1. FASES I TASQUES D'UN ESTUDI ESTADÍSTIC

Ens enfrontem diàriament a la necessitat d'arreplegar, organitzar i interpretar dades i aquesta necessitat augmentarà en el futur, a causa del desenrotllament dels sistemes de comunicació i les bases de dades. És notable l'augment de l'ús de les xarxes socials com ara *Youtube* o *Facebook*, on les persones tenen oportunitat de presentar informació sobre ells mateixos, i de pàgines web on es poden trobar i descarregar gran varietat de dades estadístiques sobre diversos temes d'actualitat: resultats esportius dels seus equips favorits, temperatura màxima i mínima al llarg d'un mes, vendes de torró el passat nadal, etc. Altres vegades les dades són arreplegades per l'investigador mitjançant la realització d'una enquesta o a través d'un experiment. L'enquesta requerirà l'elaboració d'un qüestionari, fixant els objectius del mateix, triant les variables explicatives i redactant les preguntes que permeten obtindre la informació desitjada d'una forma clara i concisa.

En aquest sentit, l'estadística ha jugat un paper primordial en aquest desenrotllament tecnològic que ens està tocant viure, en proporcionar ferramentes metodològiques generals per a analitzar la variabilitat, determinar relacions entre variables, dissenyar de forma òptima experiments, millorar les prediccions i la presa de decisions en situacions d'incertesa.

El tractament estadístic d'un problema comença sempre amb la presentació de la magnitud que es vol analitzar d'una determinada població i la selecció de la mostra pertinent per a passar a l'arreplega de dades. Una vegada obtinguts les dades s'ordenen i presenten en taules o gràfiques, de manera que siga possible observar les particularitats que assenyalen.

D'ací es pot considerar que un estudi estadístic consta d'una sèrie de fases i tasques ben diferenciades :

1. Definició de la població i característica a estudiar.

   *Tasques*: Identificació de les característiques quantitatives i qualitatives; fixació de la població; especificació de la forma d'arreplega de dades (entrevistes, telèfon, correu electrònic, etc.).

2. Selecció de la mostra.

   *Tasques*: Identificació de la grandària de la mostra i pressupost necessari.

3. Arreplega de dades.

   *Tasques*: Disseny del qüestionari; disseny mostral.

4. Organització i representació gràfica.

   *Tasques*: Taules i gràfiques que ajuden a una més fàcil interpretació de les dades; açò consisteix


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

en un estudi de cada variable, la tabulació i representació (ns) gràfica (ques) més apropiada (es).

5. Anàlisi de dades.

*Tasques*: Tractament de les dades. Açò consistirà en una anàlisi descriptiva de les dades i/o una anàlisi multivariant de les dades, depenent del tipus d'estudi a realitzar i costos del mateix.

6. Obtenció de conclusions.

*Tasques*: recomanacions i presa de decisions a partir de les conclusions.

***Exemple:***

- Una llista de punts a tindre en compte en plantejar les preguntes d'investigació és la següent:
  - Què vols provar? Què has de mesurar /observar /preguntar?
  - Quines dades necessites? Com trobaràs les teues dades? Què faràs amb elles?
  - Creus que pots fer-ho? Trobaràs problemes? Quins?
  - ¿Per a què et serviran els resultats?

D'aquesta manera es prepararà una llista de les característiques que volem incloure a l'estudi, analitzant les diferents formes amb què podrien obtindre's les dades. Per simple observació: com el sexe, color de pèl i ulls, si l'alumne usa o no ulleres; Si es requereix un mesurament: com el pes, talla, perímetre de cintura; si caldria preguntar, és a dir, si s'ha de realitzar una enquesta: quant esport practica, nombre del calçat, quantes hores dorm, quantes hores estudia al dia o a la setmana, etc.

Per tant, és important considerar la naturalesa de les escales de mesura i tipus de variable estadística, ja que d'elles depèn el mètode d'anàlisi de dades que es pot aplicar. L'elecció del conjunt de dades és crítica, perquè depenent del tipus de dades la gamma de tècniques estadístiques serà més o menys àmplia, ja que no totes les tècniques són aplicables a qualsevol tipus de dada.

## 2. POBLACIÓ I MOSTRA. VARIABLES ESTADÍSTIQUES

### 2.1. Població

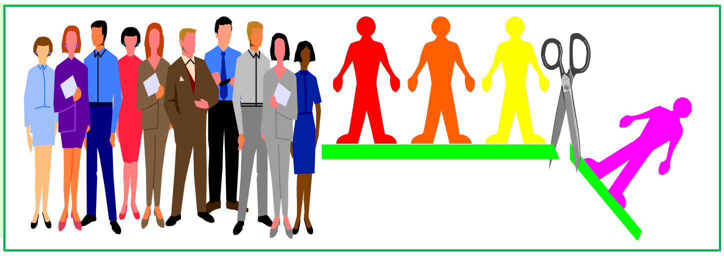
**Població** estadística, col·lectiu o univers és el conjunt de tots els individus (persones, objectes, animals, etc.) que continguen informació sobre el fenomen que s'estudia.

***Exemples:***

- Si estudiem el preu de la vivenda en una ciutat, la població serà el total de les vivendes de la dita ciutat.
- Es va a realitzar un estudi estadístic sobre el percentatge de persones casades a la península. Per a això no és factible estudiar a tots i cada un dels habitants per raons de cost i de rapidesa en l'obtenció de la informació. Per tant, és necessari acudir a examinar només una part d'aquesta **població**. Aqueixa part és la **mostra** triada.

### 1.2. Mostra

**Mostra** és un subconjunt representatiu que se selecciona de la població i sobre el qual es va a realitzar l'anàlisi estadística. La **grandària de la mostra** és el nombre dels seus elements. Quan la mostra comprèn a tots els elements de la població, es denomina cens**.**



***Exemple:***

- Si s'estudia el preu de la vivenda d'una ciutat, el normal serà no arreplegar informació sobre totes les vivendes de la ciutat (ja que seria una labor molt complexa i costosa), sinó que se sol seleccionar un subgrup (mostra) que s'entenga que és prou representatiu.

## Activitats proposades

**1.** Assenyalar en quin cas és més convenient estudiar la població o una mostra:

a) El diàmetre dels caragols que fabrica una màquina diàriament. L'altura d'un grup de sis amics.

**2.** Es pot llegir el següent titular al periòdic que publica el teu institut: *"La nota mitjana dels alumnes de 4t ESO de la Comunitat de Madrid és de* 7'9". Com s'ha arribat a aquesta conclusió? S'ha estudiat a tota la població? Si hagueren seleccionat per al seu càlcul només a les dones, seria representatiu el seu valor?

### 2.3. Individu o unitat estadística

**Individu** o **unitat estadística** és qualsevol element que continga informació sobre el fenomen que s'estudia.

***Exemple:***

- Si estudiem les notes dels alumnes d'una classe, cada alumne és un individu; si estudiem el preu de la vivenda, cada vivenda és una unitat estadística.

## 2.4. Variable estadística

En general, suposarem que s'està analitzant una determinada població, de la que ens interessa certa característica, representada per una **variable** observable o estadística *X*. Les variables que estan baix estudi es poden classificar en dues categories:

**Variables qualitatives o atributs** (dades no mètrics), que no es poden mesurar numèricament. Les escales de mesura no mètriques es classifiquen en nominals (o categòriques) i ordinals.

**Variables quantitatives**, que tenen un valor numèric. Aquest tipus de variables són les que apareixen amb més freqüència i permeten una anàlisi més detallada que les qualitatives. Dins de les variables quantitatives, es poden distingir les **variables discretes** i les **variables contínues**. Les **variables discretes** prenen valors aïllats, mentres que les **variables contínues** poden prendre qualsevol valor dins d'un interval.

***Exemple:***

- Exemples de variables qualitatives són la nacionalitat o la raça d'un conjunt de persones.
- Exemples de variables quantitatives són les notes obtingudes en una assignatura, el pes o altura d'un conjunt de persones.
- Exemples de variables discretes són el nombre d'alumnes que aproven una assignatura, o el nombre de components defectuosos que es produeixen al dia en una fàbrica.
- Exemples de variables contínues són el temps que tardem a arribar a l'institut des de la nostra casa o la velocitat d'un vehicle.

### Activitats resoltes

- Es va a realitzar un estudi estadístic sobre el percentatge de persones amb fills en una localitat madrilenya de 134.678 habitants. Per a això es trien 2.346 habitants i s'estenen les conclusions a tota la població. Identificar: variable estadística, població, mostra, grandària mostral i individu.
    - *Variable estadística*: si una persona té fills o no.
    - *Població*: Els 134.678 habitants de la localitat.
    - *Mostra*: Els 2.346 habitants triats.
    - *Grandària mostral:* 2.346 persones.
    - *Individu*: Cada persona a qui se li pregunte.

### Activitats proposades

**3.** Indicar el tipus de variable estadística que estudiem i raona, en cada cas, si seria millor analitzar una mostra o la població:

a) El sexe dels habitants d'un país.

b) Els diners gastat a la setmana pel teu germà.

c) El color de cabell dels teus Companys de classe.

d) La temperatura de la teua província.

e) La talla de peu dels alumnes de l'institut.

**4.** Per a realitzar un estudi fem una enquesta entre els jóvens d'un barri i els preguntem pel nombre de vegades que van al cine al mes. Indica quines característiques hauria de tindre la mostra triada i si haurien de ser tots els jóvens de la mostra de la mateixa edat.


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

# 3. TAULES DE FREQÜÈNCIES

## 3.1. Freqüència absoluta

Quan s'analitza una *variable discreta*, la informació resultant de la mostra es troba resumida habitualment en una taula o distribució de freqüències. Suposem que s'ha pres una mostra de grandària $N$ en què s'han identificat $k$ valors (o modalitats) distints $x_1, x_2,..., x_k$. Cada un d'ells es produeix amb una **freqüència absoluta** $n_i$, és a dir, el nombre de vegades que apareix a la mostra.

La informació obtinguda es pot resumir en una **taula de freqüències**.

Les taules de freqüència també s'utilitzen per a representar informació d'una *variable contínua* procedent d'una mostra en què s'agrupen les observacions en intervals, que es denominen **intervals de classe** $L_i$ o cel·les.

Encara que aquest procediment suposa, de fet, una pèrdua d'informació, aquesta pèrdua no és de magnitud important i es veu compensada amb l'agrupació de la informació i la facilitat d'interpretació que proporciona una taula de freqüències.

En aquest cas, els valors $x_i$ es corresponen amb el punt mitjà de l'interval i es denominen **marques de classe**.

***Exemple:***

- Quan realitzem un estudi sobre l'oci i enquestem a 40 jóvens d'una localitat sobre el nombre de vegades que van al cine els resultats de la dita enquesta els podem arreplegar en una taula per a resumir la dita informació.

## Activitats resoltes

- S'està realitzant un control del pes d'un grup de xiquets. Per a això, es comptabilitzen el nombre de vegades que mengen al dia una xocolatina 13 xiquets durant un mes, obtenint els nombres següents: 2, 5, 3, 2, 0, 4, 1, 7, 4, 2, 1, 0, 2.

La informació obtinguda es pot resumir en una taula de freqüències absolutes:

| **Valors** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| **Freqüència absoluta** | 2 | 2 | 4 | 1 | 2 | 1 | 0 | 1 |

- En una fàbrica es realitza un estudi sobre la grossària, en *mm*, d'un cert tipus de llandes de refresc. Amb aquest fi, selecciona una mostra de grandària $N$ = 25, obtenint els valors següents: 7'8, 8'2, 7'6, 10'5, 7'4, 8'3, 9'2, 11'3, 7'1, 8'5, 10'2, 9'3, 9'9, 8'7, 8'6, 7'2, 9'9, 8'6, 10'9, 7'9, 11'1, 8'8, 9'2, 8'1, 10'5.

Aquesta informació es pot resumir en la següent taula de freqüències, amb 5 intervals: (7, 8], (8,9], (9, 10], (10, 11], (11, 12], sent les marques de classe els punts mitjans de cada interval: 7'5; 8'5;

9’5; 10’5; 11’5. Comprova que les freqüències absolutes són les indicades en la taula:

| **Intervals de classe** | **(7, 8]** | **(8, 9]** | **(9, 10]** | **(10, 11]** | **(11, 12]** |
|---|---|---|---|---|---|
| **Marques de classe** | 7’5 | 8’5 | 9’5 | 10’5 | 11’5 |
| **Freqüència absoluta** | 6 | 8 | 5 | 4 | 2 |

## Activitats proposades

**5.** Obtindre la taula de freqüències absolutes de les notes en anglès de 24 alumnes:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 6 | 7 | 8 | 4 | 9 | 8 | 7 | 6 | 5 | 3 | 5 |
| 7 | 6 | 6 | 6 | 5 | 4 | 3 | 9 | 8 | 8 | 4 | 5 |

### 3.2. Freqüència relativa

Es denomina **freqüència relativa** ($f_i$) d’un valor de la variable al quocient entre la freqüència absoluta i el nombre total d’observacions $N$. S’escriu:

$$f_i = \frac{n_i}{N} \leq 1$$

***Exemple:***

- De la mateixa manera podem arreplegar la informació obtinguda a partir d’una enquesta a 40 jóvens d’una localitat sobre el nombre de vegades que van al cine mitjançant percentatge del nombre de vegades que es repeteix un valor de la variable sobre el total.

## Activitats resoltes

- S’està realitzant un control del pes d’un grup de xiquets. Per a això, es comptabilitzen el nombre de vegades que mengen al dia una xocolatina 13 xiquets durant un mes, obtenint els nombres següents: 2, 5, 3, 2, 0, 4, 1, 7, 4, 2, 1, 0, 2.

La informació obtinguda es pot resumir en una taula de freqüències relatives:

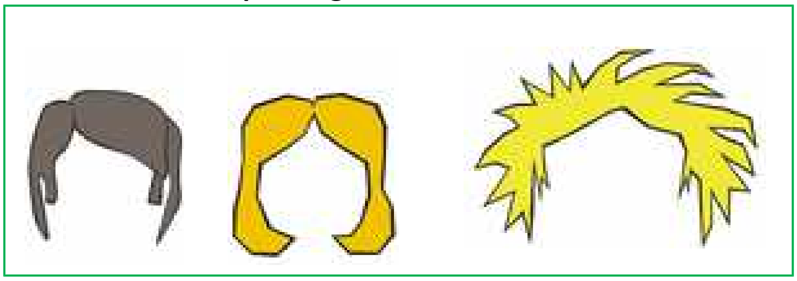
| **Valors** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| **Freqüència relativa** | 0’154 | 0’154 | 0’307 | 0’077 | 0’154 | 0’077 | 0 | 0’077 |

## Activitats proposades

**6.** Construir una taula de freqüències relatives amb el color de cabell de 24 persones triades a l’atzar:

**M=moreno; R=ros; P=pèl-roig**

| | | | | | |
|---|---|---|---|---|---|
| M | R | P | R | R | R |
| R | P | P | M | M | M |
| M | R | R | R | R | R |
| M | M | M | M | M | P |



## 3.3. Freqüència absoluta acumulada

Es denomina **freqüència absoluta acumulada** d'un valor de la variable $N_i$ a la suma de totes les freqüències absolutes dels valors menors o iguals que ell. Es calcula com:

$$N_i = \sum_{j=1}^{i} \left[ n_j \right]$$

Es verifica la següent relació entre els valors de $N_i$:

$$N_1 \leq N_2 \leq \cdots \leq N_k = N$$

***Exemple:***

- De la mateixa manera podem arreplegar la informació obtinguda a partir d'una enquesta a 40 jóvens d'una localitat sobre el nombre de vegades que van al cine mitjançant el nombre acumulat de vegades que es repeteix un valor de la variable sobre el total.

## Activitats resoltes

- S'està realitzant un control del pes d'un grup de xiquets. Per a això, es comptabilitzen el nombre de vegades que mengen al dia una xocolatina 13 xiquets durant un mes, obtenint els nombres següents: 2, 5, 3, 2, 0, 4, 1, 7, 4, 2, 1, 0, 2.

La informació obtinguda es pot resumir en una taula de freqüències absolutes:

| Valors | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Freqüència absoluta | 2 | 2 | 4 | 1 | 2 | 1 | 0 | 1 |
| Freqüència absoluta acumulada | 2 | 4 | 8 | 9 | 11 | 12 | 12 | 13 |

## Activitats proposades

**7.** El nombre d'hores diàries d'estudi de 14 alumnes és el següent:

3 4 2 5 3 4 3 2 3 4 5 4 3 2

a) Efectua un recompte i organitza els resultats obtinguts en una taula de freqüències absolutes acumulades.

b) Què signifiquen les freqüències acumulades que has calculat?

## 3.4. Freqüència relativa acumulada

Es denomina **freqüència relativa acumulada** ($F_i$) d'un valor de la variable a la suma de totes les freqüències relatives dels valors menors o iguals que ell. Es calcula com:

$$F_i = \sum_{j=1}^{i} \left[ f_j \right]$$

Es verifica la següent relació entre els valors de $F_i$:

$$F_1 \leq F_2 \leq \cdots \leq F_k = 1$$

***Exemple:***

- De la mateixa manera podem arreplegar la informació obtinguda a partir d'una enquesta a 40 jóvens d'una localitat sobre el nombre de vegades que van al cine mitjançant el percentatge acumulat del nombre de vegades que es repeteix un valor de la variable sobre el total.

## Activitats resoltes

- S'està realitzant un control del pes d'un grup de xiquets. Per a això, es comptabilitzen el nombre de vegades que mengen al dia una xocolatina 13 xiquets durant un mes, obtenint els nombres següents: 2, 5, 3, 2, 0, 4, 1, 7, 4, 2, 1, 0, 2.

La informació obtinguda es pot resumir en una taula de freqüències relatives:

| Valors | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Freqüència relativa | 0'154 | 0'154 | 0'307 | 0'077 | 0'154 | 0'077 | 0 | 0'077 |
| Freqüència relativa acumulada | 0'154 | 0'308 | 0'615 | 0'692 | 0'846 | 0'923 | 0'923 | 1 |

- En una fàbrica es realitza un estudi sobre la grossària, en *mm*, d'un cert tipus de llandes de refresc. Amb aquest fi, selecciona una mostra de grandària *N* = 25, obtenint els valors següents: 7'8, 8'2, 7'6, 10'5, 7'4, 8'3, 9'2, 11'3, 7'1, 8'5, 10'2, 9'3, 9'9, 8'7, 8'6, 7'2, 9'9, 8'6, 10'9, 7'9, 11'1, 8'8, 9'2, 8'1, 10'5.

Aquesta informació es pot resumir en la següent taula de freqüències, amb 5 intervals:

| Intervals de classe | (7, 8] | (8, 9] | (9, 10] | (10, 11] | (11, 12] |
|---|---|---|---|---|---|
| Marques de classe | 7'5 | 8'5 | 9'5 | 10'5 | 11'5 |
| Freqüència absoluta | 6 | 8 | 5 | 4 | 2 |
| Freqüència relativa | 0'24 | 0'32 | 0'2 | 0'16 | 0'08 |
| Freqüència relativa acumulada | 0'24 | 0'56 | 0'76 | 0'92 | 1 |

- S'organitza en una taula la informació arreplegada de les estatures, en cm, d'un grup de 20 xiquetes:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 130 | 127 | 141 | 139 | 138 | 126 | 135 | 138 | 134 | 131 |
| 143 | 140 | 129 | 128 | 137 | 136 | 142 | 138 | 144 | 136 |

L'estatura és una variable estadística quantitativa contínua. Per tant, podem agrupar els valors de la variable en intervals que anomenem classes o cel·les. L'amplitud de cada interval ve donada per la fórmula: $\frac{Màx - Mín}{\sqrt{N}}$

Al nostre cas concret tenim que: $\frac{144-126}{\sqrt{20}} = 4.02$

Aproximant, l'amplitud de cada interval és de 5 cm.

| Estatura en intervals | [125-130) | [130-135) | [135-140) | [140-145) |
|---|---|---|---|---|
| Freqüència absoluta | 4 | 3 | 8 | 5 |
| Freqüència relativa | 0'2 | 0'15 | 0'4 | 0'25 |
| Freqüència absoluta acumulada | 4 | 7 | 15 | 20 |
| Freqüència relativa acumulada | 0'2 | 0'35 | 0'75 | 1 |

## Activitats proposades

**8.** En una avaluació, dels 30 alumnes d'una classe, el 30 % va aprovar tot, el 10 % va suspendre una assignatura, el 40 % va suspendre dues assignatures i la resta més de dues assignatures.

a) Realitza la taula de freqüències completa corresponent (freqüències absolutes, freqüències relatives, freqüències absolutes acumulades i freqüències relatives acumulades).

b) Hi ha algun tipus de freqüència que corresponga a la pregunta de quants alumnes van suspendre menys de dues assignatures? Raona la resposta.


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

# 4. GRÀFICS ESTADÍSTICS

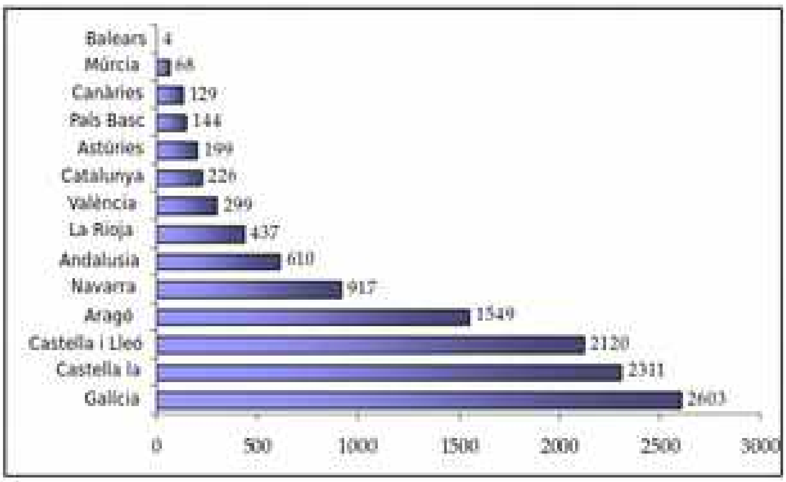
## 4.1. Diagrama de barres

Hi ha nombroses maneres de representar gràficament la informació que s'ha obtingut d'una mostra, depenent del tipus de variable que s'estiga analitzant i del fi que es perseguisca amb la representació.

Quan es vol representar gràficament una variable qualitativa (atribut) o una variable quantitativa discreta es pot utilitzar els ***diagrames de barres o rectangles***. Es col·loquen els valors de la variable (les modalitats de l'atribut o valors de la variable discreta) a l'eix d'abscisses i, a l'eix d'ordenades les freqüències (absolutes o relatives). Sobre cada valor s'alça una barra o rectangle l'altura de la qual és igual a la freqüència. Per comoditat, de vegades també se solen intercanviar els eixos.

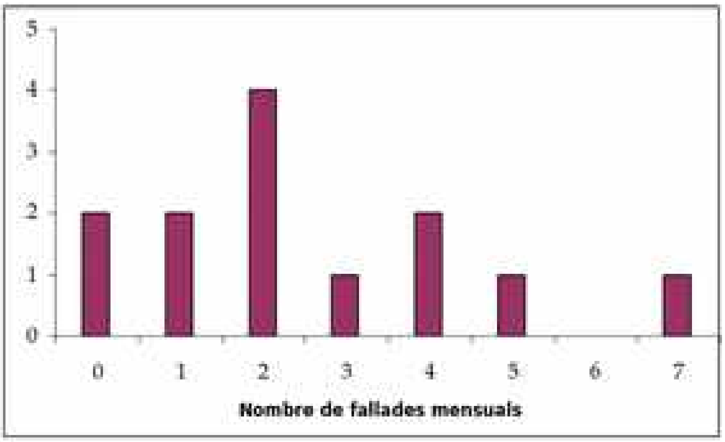
***Exemple:***

- S'ha representat gràficament la potència eòlica (font d'energia elèctrica renovable) instal·lada a Espanya per Comunitat Autònoma Al Gener de 2014 (en Megawats)




***Exemple:***

- S'ha representat gràficament el nombre de fallades mensuals d'una màquina de gelats





Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

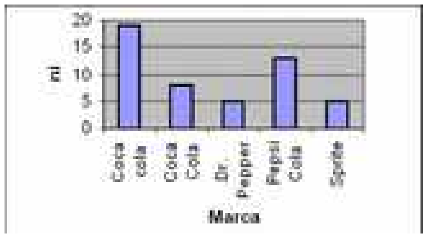
## Activitats resoltes

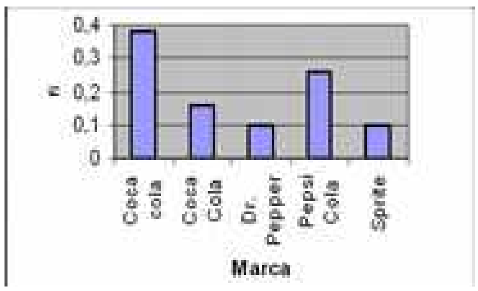
- Donada la següent informació corresponent a les preferències de 50 adolescents americans respecte a la marca de refresc que consumixen, construeix la taula associada a aquestes dades i representa-les gràficament en un diagrama de barres de freqüències absolutes i un altre de freqüències relatives.

**COCA-COLA=CC; COCA-COLA LIGHT=CCL; DR.PEPPER=A; PEPSI-COLA=PC, SPRITE=S**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| CCL | CC | S | A | CC | CC | A | CC | P | CC |
| S | CCL | P | CCL | CC | CC | CCL | P | P | A |
| S | S | CC | CC | CC | A | P | CC | CCL | CC |
| CCL | CC | P | P | P | CCL | P | S | P | CC |
| CC | P | CCL | CC | CC | P | CC | P | CC | A |

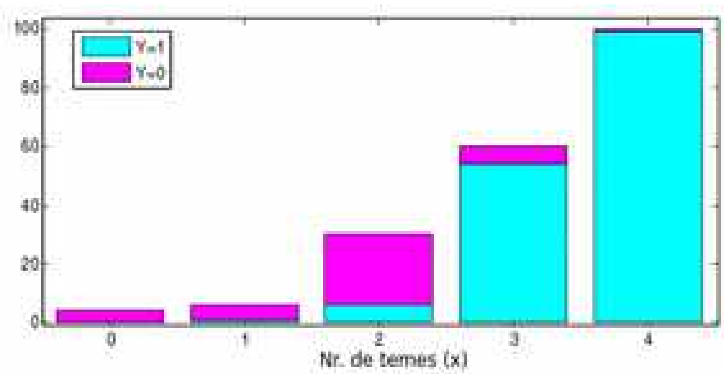
| Marca | $n_i$ | $f_i$ |
|---|---|---|
| Coca Cola | 19 | 0,38 |
| Coca Cola Light | 8 | 0,16 |
| Dr. Pepper | 5 | 0,10 |
| Pepsi Cola | 13 | 0,26 |
| Sprite | 5 | 0,10 |
| | 50 | 1 |




## Activitats proposades

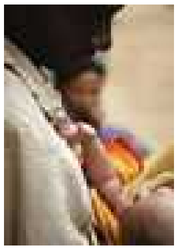
**9.** Si volem representar conjuntament valors de la variable corresponents a diferents períodes de temps, o a distintes qualitats, per a comparar situacions podem construir un diagrama de barres apilades. Podries interpretar aquest gràfic corresponent al nombre de temes que els alumnes d'una assignatura de 4t ESO porten estudiats? Es pren informació en dues classes d'un institut (blau i rosa).





Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

**10.** El sexe de 18 bebès nascuts en un hospital de Madrid ha sigut:



| | | | | | |
|---|---|---|---|---|---|
| H | M | H | H | M | H |
| H | M | M | H | M | H |
| M | M | H | H | M | H |

Construeix la taula associada a aquestes dades i representa-les.

**11.** Representa els valors de la variable de la taula adjunta amb el gràfic adequat corresponents a una enquesta realitzada sobre el sector a què pertanyen un grup de treballadors madrilenys.

| SECTOR | INDUSTRIAL | AGRARI | SERVICIS | ALTRES |
|---|---|---|---|---|
| % TREBALLADORS | 20 | 16 | 45 | 19 |

## 4.2. Histogrames

La representació més utilitzada en variables quantitatives contínues és **l'histograma**.

En l'eix d'abscisses es col·loquen els diferents intervals en què s'agrupen les observacions de la variable. Sobre aquests intervals, s'alcen rectangles **l'àrea** dels quals és proporcional a la freqüència observada en cada un d'ells.

En el cas que tots els intervals tinguen la mateixa amplitud basta amb què l'altura dels rectangles siga proporcional a la freqüència.

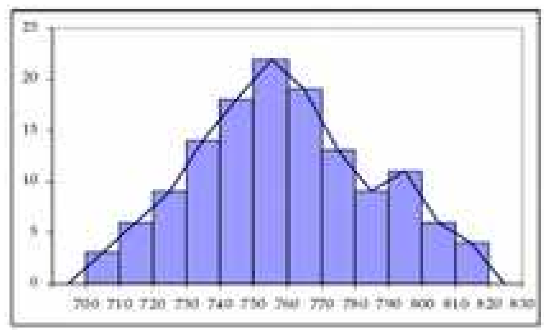
Depenent de les freqüències que s'utilitzen, es tractarà d'un histograma de freqüències relatives, o bé d'un histograma de freqüències absolutes.

De vegades, s'uneixen els punts mitjans dels segments superiors dels rectangles, obtenint-se d'aquesta manera el **polígon de freqüències**, ja siguen absolutes o relatives. Aquests polígons es construeixen utilitzant un interval anterior al primer (de la mateixa longitud que aquest) i un altre posterior a l'últim (de la seua mateixa longitud). D'aquesta manera, els polígons delimiten una àrea tancada.

En ambdós casos, també es poden utilitzar les freqüències acumulades per a construir els respectius histogrames. Aquests histogrames també porten associats els corresponents polígons de freqüències, que en aquest cas es construeixen unint els vèrtexs superiors drets de cada un dels intervals.

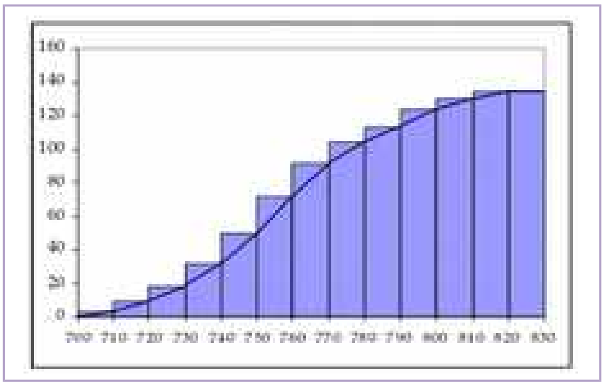
***Exemple***:

- S'ha representat gràficament la informació obtinguda a partir de les emissions específiques de $CO_2$ d'una central de carbó (kg/megawat-hora) a partir d'un histograma i un polígon de freqüències absolutes.



*Exemple:*

- S'ha representat gràficament la informació obtinguda a partir de les emissions específiques de $CO_2$ d'una central de carbó (kg/megawat-hora) a partir d'un histograma i un polígon de freqüències acumulades absolutes.

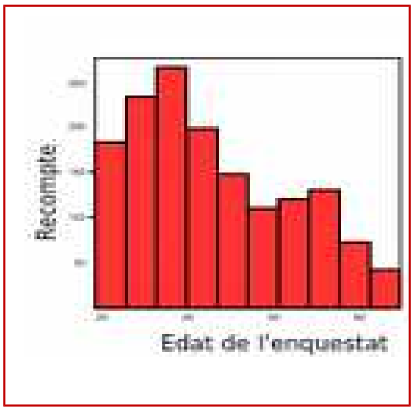



## Activitats proposades

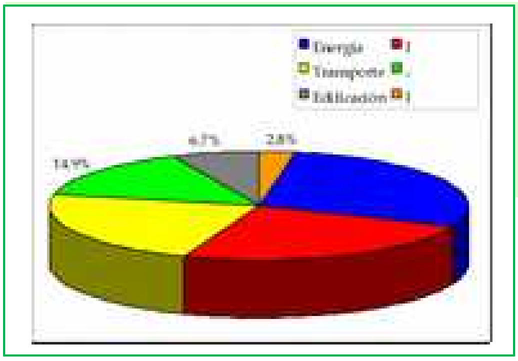
**12.** Completa la taula de freqüències per a poder representar la informació mitjançant l'histograma de freqüències acumulades:

| EDAT | [15, 25) | [25, 35) | [35, 45) | [45, 55) |
|---|---|---|---|---|
| NOMBRE DE PERSONES | 25 | 45 | 55 | 65 |

**13.** A quina representació gràfiques corresponen el següent gràfic corresponent a la informació arreplegada sobre l'edat de 100 persones? Per què creus que s'ha utilitzat aquest i no un altre?




## 4.3. Diagrama de sectors



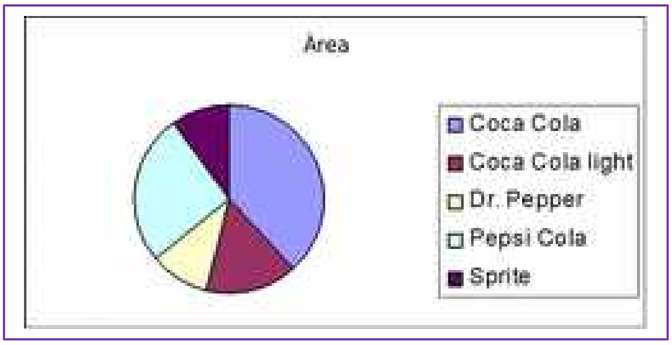

Al **diagrama de sectors** es col·loquen les modalitats de l'atribut (variable qualitativa) o valors d'una variable quantitativa discreta en un cercle, assignant a cada un un sector del cercle **d'angle proporcional a la seua freqüència**. No resolta molt operatiu quan la variable té massa categories.

*Exemple:*

De la mateixa manera podem arreplegar la informació obtinguda d'emissions de gasos d'efecte d'hivernacle a Espanya al període 1999-2012 (%)

## Activitats resoltes

- Donada la informació corresponent a les preferències de 50 adolescents americans respecte a la marca de refresc que consumixen de l'activitat resolta de l'apartat 3.1. realitzar el gràfic de sectors.




## Activitats proposades

**14.** Dels 100 assistents a una boda, el 34 % va menjar vedella de segon plat, 25 % ànec, 24 % corder i la resta peix.

a) Organitza la informació anterior en una taula de freqüències i representa les dades en un gràfic de sectors.

b) Realitza un diagrama de barres i explica com el fas. Quin dels dos gràfics prefereixes? Per què?

**15.** S'ha arreplegat informació sobre el contingut de sals minerals de 24 botelles d'aigua d'un grup d'escolars en una excursió tal que:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 45 | 45 | 65 | 56 | 33 | 65 | 23 | 23 |
| 34 | 23 | 43 | 67 | 22 | 43 | 34 | 23 |
| 12 | 34 | 45 | 34 | 19 | 34 | 23 | 43 |

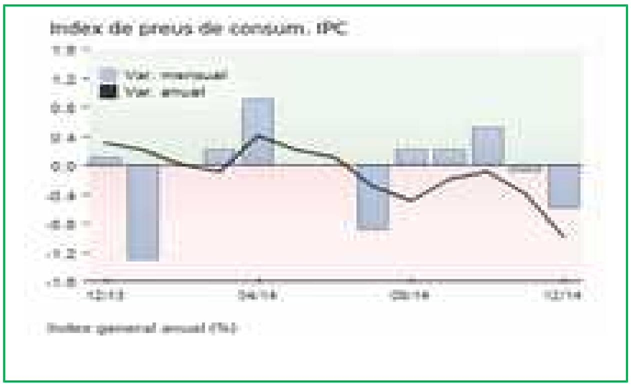
a) Classifica la variable estadística estudiada

b) Seria convenient prendre o no intervals en fer una taula de freqüències?

c) Realitza el gràfic que consideres més oportú.

### 4.3. Anàlisi crítica de taules i gràfiques estadístiques als mitjans de comunicació. Detecció de fal·làcies

Els Mitjans de comunicació recorren ben sovint a taules i gràfiques que ajuden a una més fàcil interpretació de les dades per part del públic en general. Un cas pot ser el següent gràfic que presenta l'Institut Nacional d'Estadística (INE), que representa l'índex dels preus al consum.



No obstant això, no és rar observar com s'utilitzen unes mateixes dades estadístiques per a obtindre conclusions distintes.

- Una pujada de preus o de l'índex de desocupació pot parèixer més o menys accentuada segons qui presente la informació.
- Un índex d'audiència o el colesterol d'un determinat aliment poden parèixer més o menys alts segons amb què se'l compare.

- Les telefonades pareixen ser més barates en una companyia que en una altra.

La llista d'exemples és interminable.

D'aquesta manera, l'Estadística, a més del paper instrumental que hem presentat fins ara, té un important paper en el desenrotllament del pensament crític que ens mantindrà atents a aquests excessos.

Els errors més freqüents, encara que de vegades no es tracta d'errors, sinó de manipulacions tendencioses, són els següents:

- Errors en l'obtenció de dades.
- Limitacions humanes o dels instruments: és impossible, per exemple, mesurar el pes o l'estatura d'una persona amb infinita precisió. Però inclús en estudis exhaustius, com els censos, s'estimen els errors de mostreig.
- Qüestionaris mal plantejats: si no s'arrepleguen totes les possibles respostes, si la pregunta influeix en la resposta, si les preguntes contenen juís de valor o si les diferents opcions de resposta no són equilibrades (per exemple: sí, de vegades, no). El conjunt de respostes possibles pot fer que hi haja duplicacions o omissions. Incórrer en aquest error, deliberadament o no, deixa individus de la població sense representació entre les respostes i, per tant, els resultats que isquen de l'estudi estaran esbiaixats. Les modalitats de la variable han de ser incompatibles i exhaustives (per exemple: si preguntem pel color favorit i oferim com a possibles respostes "Roig", "Blau" o "Groc", deixem sense poder respondre als que volen triar un altre color; si no estem interessats en altres colors, podem incloure un apartat anomenat "Un altre").
- Delimitació imprecisa de la població: Per exemple, si es desitja estudiar si els xiquets madrilenys veuen massa la televisió, caldrà deixar clar quines edats en concret es consideraran, si entenem per madrileny a qualsevol resident o només als nascuts a Madrid, etc.
- Selecció de la mostra inapropiada o no representativa: la mostra no representa a la població. L'elecció dels individus concrets que formen part de la mostra ha de fer-se de forma aleatòria. Per exemple: si estudiem els gustos televisius dels adolescents d'un institut i pensem que aquests gustos poden variar en funció de l'edat, en la selecció de la mostra han de triar-se edats variades, a poder ser, en la mateixa proporció en què es presenten a l'institut.
- Errors en les taules: les dades no estan ordenades, evitar ambigüitats als extrems dels intervals per a variables contínues, etc.
- Errors a les gràfiques: als diagrames de barres falta l'origen, estan truncats o a l'escala als eixos, etc. Cal deixar clares les variables que es mesuren.
- Errors als paràmetres de mesura: per exemple la mitja no és representativa (poblacions heterogènies) o es veu afectada per valors molt grans; confusió entre mitja i mitjana.
- Errors als pictogrames amb superfícies on s'inscriuen proporcionals al quadrat de les freqüènci-es.


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

# 5. MESURES DE TENDÈNCIA CENTRAL

## 5.1. Mesures de grandària

Les mesures de tendència central o de centralització són les que, intuïtivament, apareixen en primer lloc en intentar descriure una població o mostra.

Es poden dividir en tres classes: **mesures de grandària, de freqüència i de posició**.

Al que segueix, suposarem que estem analitzant una població de què es pren una mostra de grandària $N$, és a dir, que està composta per $N$ individus (o observacions), dels quals es desitja estudiar la variable $X$, la qual cosa dóna lloc a l'obtenció de $N$ valors que es representen per $x_1, x_2, \ldots, x_N$. Aquests valors no se suposen ordenats, sinó que el subíndex indica l'orde en què han sigut seleccionats.

Les mesures de grandària es defineixen a partir dels valors de la mostra, així com de la seua freqüència.

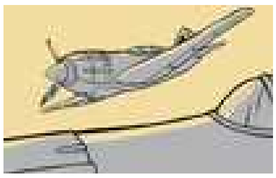
Definim així la **mitja aritmètica** o **promedi** o, simplement, **mitja** com:

$$\overline{x} = \frac{\sum_{i=1}^{N} [x_i]}{N}$$

Es pot interpretar com el centre de masses de les observacions de la mostra. Dins dels seus avantatges es poden destacar que utilitza totes les observacions, que són fàcilment calculables, tenen una interpretació senzilla i bones propietats matemàtiques. El seu inconvenient és que es pot veure afectada pels valors anormalment xicotets o grans que existisquen a la població o mostra (denominats *outliers*).

En el cas que tinguem una variable quantitativa agrupada en intervals el valor de la variable $X$ que representa a l'interval per a poder calcular la mitjana aritmètica és la **marca de classe** i es calcula com la semisuma dels valors extrems de l'interval.

***Exemple:***

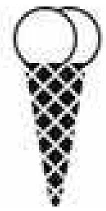


- S'arreplega la informació referida al nombre d'hores de vol diàries de 20 hostesses. Si la mitja és igual a 4'1, açò indica que, generalment, el nombre d'hores de vol és 4'1.

***Exemple:***

- De la mateixa manera si arrepleguem la informació sobre l'edat mitjana de la teua classe obtindrem un valor entre 15 i 16 anys. L'edat mitjana serà per exemple 15'4, valor teòric, que pot no coincidir amb cap dels valors reals.

**Activitats resoltes**



- Un fabricant de gelats està realitzant un control de qualitat sobre certes màquines respecte a la seua capacitat de regular la temperatura de refrigeració. Per a això, selecciona una mostra de $N$ = 16 màquines de la fàbrica i mesura amb precisió el valor de la seua capacitat (en la unitat de mesura μF), obtenint els resultats següents: 20'5, 19'8, 19'6, 19'2, 23'5,


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

28’9, 19’9, 19’2, 20’1, 18’8, 19’5, 20’2, 18’6, 19’7, 22’1, 19’3. Utilitzant aquests valors de capacitat, obtindre la mitja aritmètica.

$$\bar{x} = \frac{\sum_{i=1}^{N}[x_i]}{N} = \frac{20.5 + 19.8 + 19.6 + 19.2 + 23.5 + 28.9 + 19.9 + 19.2 + 20.1 + 18.8 + 19.5 + 20.2 + 18.6 + 19.7 + 22.1 + 19.3}{16} = \mathbf{20.56} \quad \mu\mathrm{F}$$

## Activitats proposades

**16.** Una persona ingressa 10.000 euros en un fons d’inversió l’1 de gener de 2009. Les rendibilitats anuals del fons durant els anys següents van ser les següents:

| Any | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|
| Rendibilitats (%) | 5 | 3 | –1 | 4 |

Si no ha retirat el capital, quin ha sigut la rendibilitat mitja del dit fons durant aquests anys?

**17.** Interpreta els valors de la variable d’aquesta taula que representa el pes de 100.000 bombones de butà d’una fàbrica, en quilograms. Què gràfic utilitzaries? Calcula la mitja i interpreta-la.

| Pes [) | $f_i$ % | $n_i$ | $N_i$ |
|---|---|---|---|
| **14’5-15** | 0’3 | 300 | 300 |
| **15-15’5** | 1’6 | 1600 | 1900 |
| **15’5-16** | 7’4 | 7400 | 9300 |
| **16-16’5** | 21’5 | 21500 | 30800 |
| **16’5-17** | 30’5 | 30500 | 61300 |
| **17-17’5** | 24’5 | 24500 | 85800 |
| **17’5-18** | 10’7 | 10700 | 96500 |
| **18-18’5** | 21’5 | 21500 | 30800 |


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

## 5.2. Mesures de freqüència

Es defineixen tenint en compte únicament la freqüència dels valors de la variable de la mostra.

La **moda** (*Mo*) es defineix com el valor de la variable que s'ha obtingut amb major freqüència. Pot haver-hi més d'una moda.

***Exemple:***

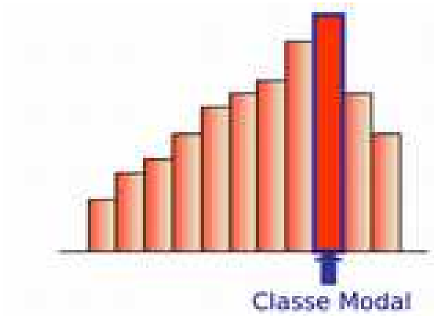
Es realitza un estudi entre 200 espectadors a un musical a Madrid per a determinar el grau de satisfacci-ó, obtenint-se els resultats següents:

| **Opinió** | Molt bo | Bo | Regular | Roí | Molt roí |
|---|---|---|---|---|---|
| **%** | 75 | 25 | 45 | 15 | 40 |

La modalitat que més es repeteix és "molt bo", per la qual cosa la moda és Mo = Molt bo.

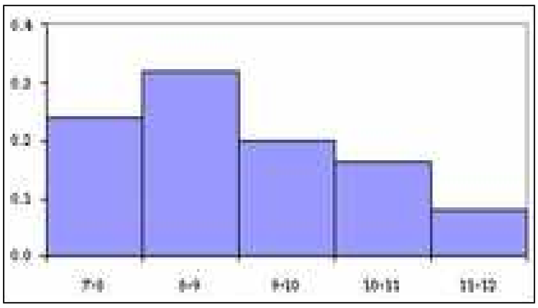
***Exemple:***

Al cas que la distribució estiga agrupada en intervals caldrà identificar la classe modal, és a dir, l'interval on hi ha el nombre més gran de valors de la variable.




## Activitats resoltes

- A partir de la taula de freqüències de la grossària de llandes de refresc, podem dibuixar els seus histogrames de freqüències relatives i determinar on està la seua moda. És a dir en l'interval [8 - 9). La moda assenyala que el més freqüent és tindre una grossària entre 8 i 9 mm.



## Activitats proposades

**18.** Obtindre la mitja i la moda dels següents valors de la variable referits al resultat de llançar un dau 50 vegades.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 2 | 3 | 4 | 3 | 3 | 3 | 5 |
| 5 | 5 | 5 | 6 | 5 | 6 | 5 | 6 | 4 | 4 |
| 3 | 2 | 1 | 2 | 3 | 4 | 5 | 6 | 5 | 4 |
| 3 | 2 | 3 | 4 | 5 | 6 | 5 | 4 | 3 | 2 |
| 3 | 4 | 5 | 5 | 5 | 5 | 6 | 6 | 6 | 3 |

**19.** Realitzar l'activitat anterior però agrupant en intervals d'amplitud 2, començant en 0. Obtens els mateixos resultats? Per què?

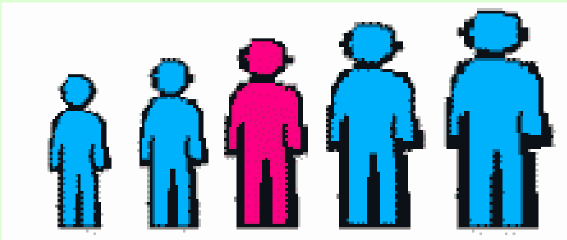
## 5.3. Mesures de posició

Es defineixen a partir de la posició dels valors de la mostra.

En general, es coneixen amb el nom de **centils** o **percentils**.

Si reordenem en orde creixent els valors que hem pres a la mostra i els denotem per $x_{\{1\}}$, $x_{\{2\}}$,…, $x_{\{N\}}$ es poden definir les següents mesures de posició:

- ✓ La ***mitjana*** *Me* és un valor tal que el 50 % de les observacions són inferiors a ell. No té per què ser únic i pot ser un valor no observat.



**Altura mitjana**

- ✓ Els ***quartils*** (o quartiles) $Q_1$, $Q_2$ i $Q_3$ són els valors tals que el 25 %, 50 % i 75 % (respectivament) dels valors de la variable són inferiors a ell.
- ✓ Els ***decils*** $D_1$, $D_2$,…, $D_9$ són els valors tals que el 10 %, 20 %, …, 90 % (respectivament) dels valors de la variable són inferiors a ell.

En general, es defineix el **percentil** o **centil** del *k* % (sent $0 \le k \le 100$) com el valor tal que el *k* % de les observacions són inferiors a ell.

La mitjana i la resta de mesures de posició tenen com principal avantatge la seua fàcil interpretació i la seua robustesa (no es veuen afectades per observacions extremes).

***Exemple:***

- Calcula els quartils i el percentil 65 dels següents valors de la variable referits al nombre de fills de les famílies d'un bloc d'edificis de la localitat de Madrid:

| Nombre de fills | $f_i$ | $F_i$ |
|---|---|---|
| 1 | 11 | 11 |
| 2 | 27 | 38 |

| | | |
|---|---|---|
| 3 | 4 | 42 |
| 4 | 18 | 60 |
| **Total** | 60 | |

Per a trobar el primer quartil calculem el 25 % del total mostral $N$ = 60, és a dir, 60·0'25 = 15. Així, el primer quartil té 15 valors de la variable menors i la resta majors. A la columna de freqüències acumulades, el primer nombre major o igual que 15 és 38, que correspon al valor de la variable 2. Per tant el primer quartil és 2 (o amb millor aproximació un valor entre 1 i 2). De la mateixa manera el 50 % de 60 és 30, és a dir el quartil 2 (Mitjana) seria també 2 (o de nou, un valor entre 1 i 2). El 75% de 60 seria 45 i d'aquesta manera el quartil 3 seria 4 (o un valor entre 3 i 4) ja que el valor major a 45 és 60, que correspon al valor 4 de la variable objecte d'estudi. Finalment, el percentil 65 correspon al valor 3 ja que 65 % de 60 és igual a 39 i el valor major que 39 és 42.

***Resum:***

25 % de 60 = 15 → 38 > 15 > 11 → $Q_1 = 2$

50 % de 60 = 30 → 38 > 30 > 11 → Me = $Q_2 = 2$

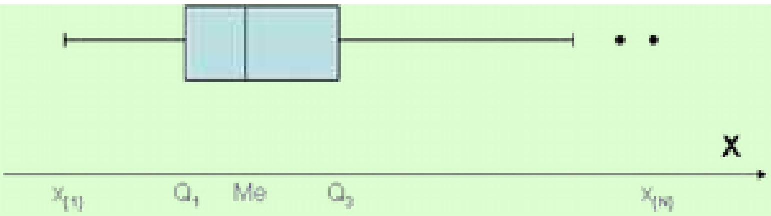
75 % de 60 = 45 → 60 > 45 > 42 → $Q_3 = 4$

65 % de 60 = 39 → 42 > 39 > 38 → $P_{65} = 3$

Les mesures de posició ens permeten realitzar un altre tipus de gràfic estadístic que s'anomena el **gràfic *de caixa***.

Per a realitzar aquest gràfic, es construeix una *caixa* (ja siga horitzontal o vertical), els costats de la qual coincideixen amb el *primer i tercer quartil* $Q_1$ i $Q_3$. Per tant, la caixa comprèn el 50 % de les observacions realitzades. Dins de la dita caixa, s'inclou un segment (o bé un punt) que correspon a la *mitjana*.

De cada costat de la caixa part un segment que s'estén fins als valors corresponents a les observacions *mínima i màxima* $x_{\{1\}}$ i $x_{\{N\}}$.




## Activitats resoltes

- S'està realitzant un control de qualitat sobre les fallades d'unes determinades màquines. Per a això, es comptabilitzen les fallades de $N$ = 13 màquines durant un mes, obtenint els següents nombres de fallades: 2, 5, 3, 2, 0, 4, 1, 7, 4, 2, 1, 0, 2. Utilitzant aquests valors obtindre les mesures de tendència central i resumir en una taula de freqüències la informació obtinguda del nombre de fallades mensuals de les màquines, obtenint la mitja aritmètica d'una altra manera.

$$\bar{x}=\frac{\sum_{i=1}^{N}[x_i]}{N}=\frac{2+5+3+2+0+4+1+7+4+2+1+0+2}{13}=$$ **2,54** fallades/mes

*Mo*=**2 fallades/mes**

$Q_1=x_{[4]}$=**1 fallada/mes**

$Q_3=x_{[10]}$=**4 fallades/mes**

| **Valors** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|

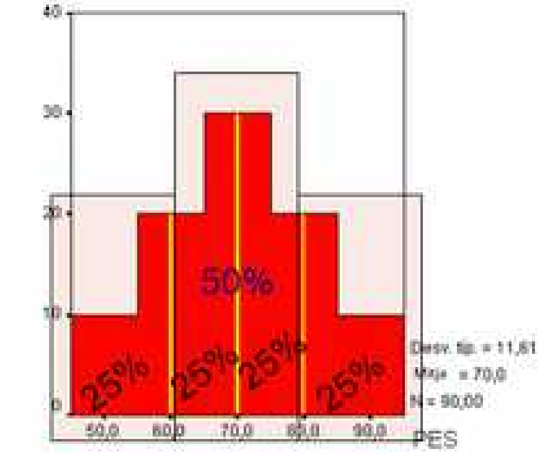

Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

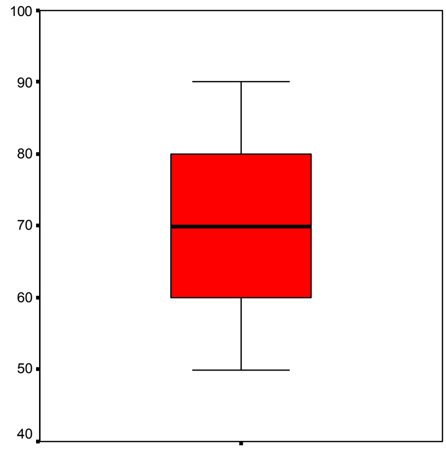
| Freqüència absoluta | 2 | 2 | 4 | 1 | 2 | 1 | 0 | 1 |
|---|---|---|---|---|---|---|---|---|
| Freqüència relativa | 0’154 | 0’154 | 0’307 | 0’077 | 0’154 | 0’077 | 0 | 0’077 |
| Freqüència relativa acumulada | 0’154 | 0’308 | 0’615 | 0’692 | 0’846 | 0’923 | 0’923 | 1 |

$$\bar{x} = \sum_{i=1}^{k} [f_i \times x_i] = 0.154\times 0 + 0.154\times 1 + 0.307\times 2 + 0.077\times 3 + 0.154\times 4 + 0.077\times 5 + 0.077\times 7 = \mathbf{2.54} \quad \textbf{fallos/mes}$$

$$\bar{x} = \sum_{i=1}^{k} [f_i \cdot x_i] = 0{,}154\cdot 0 + 0{,}154\cdot 1 + 0{,}307\cdot 2 + 0{,}077\cdot 3 + 0{,}154\cdot 4 + 0{,}077\cdot 5 + 0{,}077\cdot 7 = \mathbf{2{,}54\ fallades/mes}$$

- S’arreplega informació sobre el pes de 90 xics en una classe de Matemàtiques. Determinar els centils que ens permeten realitzar el gràfic de caixa.



– Primer quartil = percentil 25 = 60 Kg.

– Tercer quartil= percentil 75= 80 kg.

## Activitats proposades

**20.** Dibuixar un diagrama de caixa coneixent les següents dades.
Mínim valor = 2; quartil 1 = 3; mitja = 6; quartil 3 = 7; màxim valor = 12.

**21.** Un corredor de marató entrena, de dilluns a divendres recorrent les distàncies següents: 2, 3, 3, 6 i 4, respectivament. Si el dissabte també entrena:

a) Quants quilòmetres ha de recórrer perquè la mitja siga la mateixa?

b) I perquè la mitja no varie?

c) I perquè la moda no varie?

**22.** El salari mensual en euros dels 6 treballadors d’una empresa tèxtil és el que es presenta. Quin dels tres tipus de mesures de tendència central descriu millor els sous de l’empresa?

| | | | | | |
|---|---|---|---|---|---|
| 1700 | 1400 | 1700 | 1155 | 1340 | 4565 |


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

**23.** Quin valor o valors podríem afegir a aquest conjunt de valors de la variable perquè la mitja continue sent la mateixa?

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 19 | 24 | 23 | 23 | 15 | 21 | 32 | 12 | 6 | 32 | 12 | 12 | 21 |

**24.** Ixen 25 places per a un lloc d'auxiliar d'infermeria i es presenten 200 persones amb les següents notes.

| notes | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|
| $n_i$ | 6 | 34 | 25 | 56 | 29 | 10 | 30 | 10 |


a) Amb quina nota s'obté una de les places mitjançant l'examen?

b) Quin percentil és la nota 5?

# 6. MESURES DE DISPERSIÓ

## 6.1. Mesures de desviacions

Les mesures de tendència central resulten insuficients a l'hora de descriure una mostra. A més de les tendències, és necessari disposar de mesures sobre la variabilitat de les dades. Dins d'aquestes mesures, estudiarem les mesures de desviacions i els rangs.

Les mesures de desviacions arrepleguen les desviacions dels valors de la variable respecte d'una mesura de tendència central.

La ***variança*** es defineix com: $s^2 = \frac{\sum_{i=1}^{N}\left[\left(x_i - \overline{x}\right)^2\right]}{N} = \frac{\sum_{i=1}^{N} x_i^2}{N} - \overline{x}^2$

Els seus principals avantatges són la seua manejabilitat matemàtica i que utilitza totes les observacions. Els seus principals inconvenients són el ser molt sensible a observacions extremes i que la seua unitat és el quadrat de la unitat original de la mostra.

La **desviació típica** és l'arrel quadrada de la variança i té el principal avantatge que utilitza les mateixes unitats que els valors de la variable originals.

Observa que la desviació típica és una distància, la distància dels valors de la variable a la mitja. Recorda que l'arrel quadrada és sempre un nombre positiu.

Associat a la mitjana i la desviació típica, es defineix el **coeficient de variació**, definit en mostres amb mitja diferent de zero com: $g = \frac{s}{\left|\overline{x}\right|}$

Aquest coeficient és adimensional (no té unitats i se sol expressar en percentatge), el que resulta un gran avantatge, ja que permet comparar la variabilitat de distintes mostres, independentment de les seues unitats de mesura. Alguns autors defineixen aquest coeficient utilitzant la mitja al denominador, en compte del seu valor absolut. Valors del coeficient de variació majors del 100 % indiquen que la mitja no es pot considerar representativa del conjunt de valors de la variable.

***Exemple:***

- La nota mitja de 6 alumnes d'una mateixa classe de 4t d'ESO en Matemàtiques és de 5. Si la variança és 0'4, la desviació típica serà de 0'632, per tant la mitjana és prou homogènia en la distribució. Les notes que s'han obtingut estan situades al voltant de la nota mitja 5.

## Activitats resoltes

- El propietari d'una instal·lació mixta solar-eòlica està realitzant un estudi del volum d'energia que és capaç de produir la instal·lació. Per a això, mesura la dita energia al llarg d'un total de *N* = 16 dies que considera prou representatius. L'energia (en quilowatt, KWh) produïda en els dits dies per dues instal·lacions es troba arreplegada a la taula següent:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Generació solar** | 13'1 | 10'5 | 4'1 | 14'8 | 19'5 | 11'9 | 18 | 8'6 |
| **Generació eòlica** | 8'5 | 14'3 | 24'7 | 4 | 2'3 | 6'4 | 3'6 | 9'2 |
| **Generació solar** | 5'7 | 15'9 | 11'2 | 6'8 | 14'2 | 8'2 | 2'6 | 9'7 |
| **Generació eòlica** | 13'5 | 1'4 | 7'6 | 12'8 | 10'3 | 16'5 | 21'4 | 10'9 |

Utilitzant aquests valors de la variable calcula les mesures de dispersió estudiades, comparant els resultats a les dues instal·lacions

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{13'1 + 10'5 + 4'1 + 14'8 + 19'5 + 11'9 + 18 + 8'6 + 5'7 + 15'9 + 11'2 + 6'8 + 14'2 + 8'2 + 2'6 + 9'7}{16} = 10'925 \text{ Kwh}$$

$$\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N} = \frac{8'5 + 14'3 + 24'7 + 4 + 2'3 + 6'4 + 3'6 + 9'2 + 13'5 + 1'4 + 7'6 + 12'8 + 10'3 + 16'5 + 21'4 + 10'9}{16} = 10'463 \text{ Kwh}$$

$$s_x^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - \bar{x}^2 = \frac{13'1^2 + 10'5^2 + 4'1^2 + 14'8^2 + 19'5^2 + 11'9^2 + 18^2 + 8'6^2 + 5'7^2 + 15'9^2 + 11'2^2 + 6'8^2 + 14'2^2 + 8'2^2 + 2'6^2 + 9'7^2}{16} = \frac{141'5}{16} - 10'9^2 = 22'16$$

$$s_y^2 = \frac{\sum_{i=1}^{N} y_i^2}{N} - \bar{y}^2 = \frac{8'5^2 + 14'3^2 + 24'7^2 + 4^2 + 2'3^2 + 6'4^2 + 3'6^2 + 9'2^2 + 13'5^2 + 1'4^2 + 7'6^2 + 12'8^2 + 10'3^2 + 16'5^2 + 21'4^2 + 10'9^2}{16} = \frac{150'48}{16} - 10'5^2 = 41'01$$

$$g_x = \frac{s_x}{|\bar{x}|} = \frac{\sqrt{22'16}}{10'9} = \frac{4'7}{10'9} = 0'43 \qquad g_y = \frac{s_y}{|\bar{y}|} = \frac{\sqrt{41'01}}{10'5} = \frac{6'4}{10'5} = 0'61$$

La mitja de la primera instal·lació és més representativa que la mitja de la segona ja que el coeficient de variació és menor en la primera. Les dades estan menys agrupades a la segona de les instal·lacions. La seua desviació típica és molt major.

- S'està realitzant un control de qualitat sobre les fallades d'unes determinades màquines. Per a això, es comptabilitzen les fallades de $N$ = 13 màquines durant un mes, obtenint els següents nombres de fallades. Utilitzant aquests valors presentats a la taula de freqüències obtindre les mesures de dispersió estudiades.

| **Valors** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
|---|---|---|---|---|---|---|---|---|
| **Freqüència absoluta** | 2 | 2 | 4 | 1 | 2 | 1 | 0 | 1 |
| **Freqüència relativa** | 0'154 | 0'154 | 0'307 | 0'077 | 0'154 | 0'077 | 0 | 0'077 |
| **Freqüència relativa acumulada** | 0'154 | 0'308 | 0'615 | 0'692 | 0'846 | 0'923 | 0'923 | 1 |

$$s^2 = \sum_{i=1}^{k}\left[f_i \times (x_i - \bar{x})^2\right] = 0.154\times(-2.54)^2 + 0.154\times(-1.54)^2 + 0.307\times(-0.54)^2 + 0.077\times 0.46^2 + 0.154\times 1.46^2 +$$
$$+ 0.077\times 2.46^2 + 0.077\times 4.46^2 = \mathbf{3.80} \quad (\textbf{fallos/mes})^2$$

Una altra forma de realitzar aquests mateixos càlculs és:

| **Valors** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **Suma** |
|---|---|---|---|---|---|---|---|---|---|
| **Freqüència absoluta** | 2 | 2 | 4 | 1 | 2 | 1 | 0 | 1 | 13 |

| $x_i^2$ | 0 | 1 | 4 | 9 | 16 | 25 | 36 | 49 | |
|---|---|---|---|---|---|---|---|---|---|
| $x_i^2$·Fr. Abs. | 0 | 2 | 16 | 9 | 32 | 25 | 0 | 49 | 133 |

Apliquem la fórmula: $s^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - \bar{x}^2$ i obtenim que

$s^2$ = 133/13 – 2'54$^2$ = 10'23 – 6'45 = 3'80, por lo que $s$ = 1'95.

## Activitats proposades

**25.** Un grup de gossos pastor alemany té una mitja de 70 kg i desviació típica 2 kg. Un conjunt de gossos canitx té una mitja de 15 kg i desviació típica 2 kg. Compara ambdós grups.

**26.** El temps, en minuts, que un conjunt d'estudiants de 4t ESO dedica a preparar un examen de Matemàtiques és:

| | | | | | | |
|---|---|---|---|---|---|---|
| 234 | 345 | 345 | 123 | 234 | 234 | 556 |
| 234 | 234 | 345 | 223 | 167 | 199 | 490 |

Les qualificacions d'aqueix conjunt d'estudiants són les següents:

| | | | | | | |
|---|---|---|---|---|---|---|
| 4 | 5 | 6 | 7 | 6 | 5 | 8 |
| 9 | 8 | 7 | 8 | 7 | 6 | 8 |

a) Què haurem de fer per a comparar la seua variabilitat? b) En quin conjunt els valors de la variable estan més dispersos? c) És la mitja sempre major que la desviació típica?

### 6.2. Els rangs

Aquestes mesures proporcionen informació sobre l'interval total de valors que pren la mostra analitzada.

El **rang total** o **recorregut** és la diferència entre els valors màxims i mínims que pren la variable en la mostra:

$$R = x_{\{N\}} - x_{\{1\}}$$

El **recorregut interquartílic** és la diferència entre el tercer i el primer quartil:

$$R_I = Q_3 - Q_1$$

***Exemple:***

- S'està realitzant un control de qualitat sobre les fallades d'una determinada màquina. Per a això, es comptabilitzen les fallades de N = 13 màquines durant un mes, obtenint els següents nombres de fallades: 2, 5, 3, 2, 0, 4, 1, 7, 4, 2, 1, 0, 2. Utilitzant aquests valors obtenim el rang total igual a 7 i el recorregut interquartílic igual a 3.

## Activitats resoltes

- Ixen 25 places per a un lloc de caixer en un supermercat i es presenten 200 persones. La següent informació arreplega les notes d'un test de coneixements bàsics.

| notes | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| $n_i$ | 6 | 4 | 30 | 25 | 56 | 29 | 10 | 30 | 10 |

Calcula el rang total de la variable objecte d'estudi.

## Activitats proposades

**27.** S'ha arreplegat una mostra de 20 recipients els diàmetres dels quals són:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0'91 | 1'04 | 1'01 | 1 | 0'77 | 0'78 | 1 | 1'3 | 1'02 | 1 |
| 1 | 0'88 | 1'26 | 0'92 | 0'98 | 0'78 | 0'82 | 1'2 | 1'16 | 1'14 |

a)Calcula totes les mesures de dispersió que conegues.

b) A partir de quin valor de diàmetre dels recipients es consideren el 20 % amb major diàmetre?


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

## 7. DISTRIBUCIONS BIDIMENSIONALS

Aquest apartat se centra en l'anàlisi de dades bidimensionals, en el que són dues les variables d'interès. D'aquesta manera, quan s'està analitzant una població i se selecciona una mostra, per a cada individu es prenen dos valors, corresponents a dues característiques (o variables) distintes. En aquest sentit, pot ser interessant considerar simultàniament els dos caràcters a fi d'estudiar les possibles relacions entre ells.

### 7.1. Taules de freqüència d'una variable bidimensional

Quan es volen resumir els resultats d'una mostra bidimensional utilitzant una taula de freqüències (ja siga per tractar-se d'una variable discreta, o perquè es desitgen agrupar les observacions d'una variable contínua), és necessari utilitzar el que es denomina *taula de doble entrada* (o bidimensional). Siguen $x_1, x_2, \ldots, x_k$ les modalitats de la primera variable e $y_1, y_2, \ldots, y_p$ les de la segona. Aquestes modalitats poden correspondre tant als valors que es donen en la mostra (si la variable és discreta), com a les marques de classe dels intervals utilitzats (si la variable és contínua). Per a construir la taula de freqüències, s'utilitzen les freqüències absolutes $n_{ij}$ corresponents a les observacions que prenen simultàniament valors corresponents a les classes $x_i$ e $y_j$. Òbviament, s'ha de verificar que:

$$\sum_{i=1}^{k}\sum_{j=1}^{p}\left[n_{ij}\right] = N$$

Amb açò, la taula de freqüències absolutes es presenta com:

| | $y_1$ | $y_2$ | …… | $y_p$ | $n_{i\cdot}$ |
|---|---|---|---|---|---|
| $x_1$ | $n_{11}$ | $n_{12}$ | …… | $n_{1p}$ | $n_{1\cdot}$ |
| $x_2$ | $n_{21}$ | $n_{22}$ | …… | $n_{2p}$ | $n_{2\cdot}$ |
| …… | …… | …… | …… | …… | …… |
| $x_k$ | $n_{k1}$ | $n_{k2}$ | …… | $n_{kp}$ | $n_{k\cdot}$ |
| $n_{\cdot j}$ | $n_{\cdot 1}$ | $n_{\cdot 2}$ | …… | $n_{\cdot p}$ | $N$ |

Els valors $n_{i\cdot}$ arrepleguen les freqüències absolutes de la classe $x_i$, mentres que $n_j$ és la suma de freqüències absolutes de la classe $y_{j\cdot}$, amb la qual cosa es verifica:

$$n_{i\cdot} = \sum_{j=1}^{p}\left[n_{ij}\right] \qquad n_{\cdot j} = \sum_{i=1}^{k}\left[n_{ij}\right]$$

$$\sum_{i=1}^{k}\left[n_{i\cdot}\right] = N \qquad \sum_{j=1}^{p}\left[n_{\cdot j}\right] = N$$

De la mateixa manera, es pot realitzar una taula de freqüències relatives $f_{ij}$, utilitzant els quocients entre les freqüències absolutes i el nombre d'observacions:

$$f_{ij} = \frac{n_{ij}}{N} \leq 1$$

## Activitats resoltes

- El propietari d'una instal·lació mixta solar-eòlica està realitzant un estudi del volum d'energia que és capaç de produir la instal·lació. Per a això, mesura la dita energia al llarg d'un total de *N = 16* dies que considera prou representatius. L'energia (en *kWh*) produïda als dits dies per les instal·lacions solar i eòlica es poden resumir a les següents taules de doble entrada de freqüències absolutes i de freqüències relatives:

| | | | | | | |
|---|---|---|---|---|---|---|
| | | [0, 6'5] | (6'5, 13] | (13, 19'5] | (19'5, 26] | |
| | [0, 5] | 0 | 0 | 0 | 2 | |
| | (5, 10] | 0 | 3 | 2 | 0 | |
| | (10, 15] | 2 | 3 | 1 | 0 | |
| | (15, 20] | 3 | 0 | 0 | 0 | |
| | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | [0, 6'5] | (6'5, 13] | (13, 19'5] | (19'5, 26] | |
| | [0,5] | 0 | 0 | 0 | 0'125 | |
| | (5, 10] | 0 | 0'1875 | 0'125 | 0 | |
| | (10, 15] | 0'125 | 0'1875 | 0'0625 | 0 | |
| | (15, 20] | 0'1875 | 0 | 0 | 0 | |
| | | | | | | |

### 7.2. Representació gràfica d'una variable bidimensional

Igual que al cas d'una mostra unidimensional, en nombroses ocasions resulta interessant realitzar una representació gràfica d'una mostra bidimensional.

Un mode senzill de representar una mostra bidimensional és mitjançant el denominat **diagrama de dispersió** o **núvol de punts**. Aquesta tècnica consisteix a representar al pla (*x, y*) els valors obtinguts en la mostra.

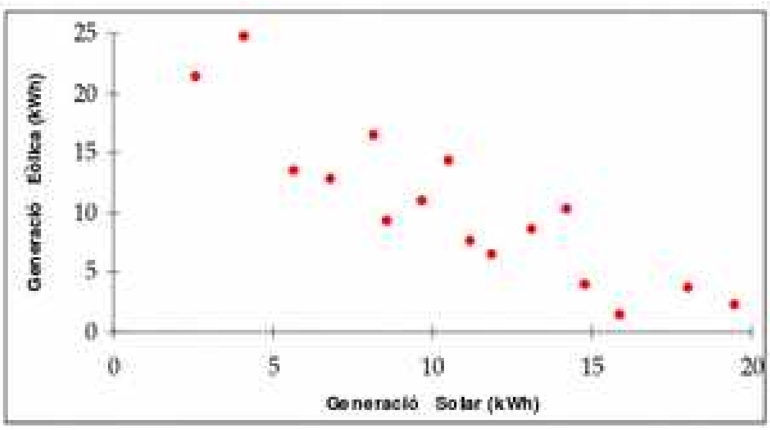



*Diagrama de dispersió de la generació solar i eòlica (en kWh) de l'activitat resolta*

La figura anterior mostra el diagrama de dispersió. Es pot observar l'existència d'una dependència inversa.

### 7.3. Mesures en una variable bidimensional. Coeficient de correlació

Quan s'està analitzant una mostra bidimensional, es poden calcular les mesures que caracteritzen a cada una de les variables de la mostra per separat, tal com s'ha descrit anteriorment. Però en aquest cas es pot fer un pas més i calcular algunes mesures conjuntes, que tenen en compte simultàniament els valors que prenen ambdues variables en cada individu.

Igual que quan s'analitza una única característica, suposarem que es pren una mostra de grandària $N$ de la població, és a dir, que està composta per $N$ individus (o observacions), dels quals es desitja analitzar les característiques (o variables) $X$ e $Y$. Açò dóna lloc a l'obtenció de $N$ valors per a cada una de les dos variables: $(x_1, y_1)$, $(x_2, y_2)$,…, $(x_N, y_N)$. De nou, aquests valors no se suposen ordenats, sinó que el subíndex indica l'orde en què han sigut seleccionats.

Seguint aquesta notació es poden formular els càlculs dels moments respecte a l'origen i respecte a la mitja per a una variable bidimensional. Definim, per tant:

Moments respecte a l'origen d'orde $(r, s)$ com:

$$a_{r,s} = \frac{\sum_{i=1}^{N}\left[x_i^{\,r} \times y_i^{\,s}\right]}{N}$$

Observa que els moments respecte a l'origen d'orde (1, 0) i (0, 1) coincideixen amb les mitges d'ambdues variables:

$$a_{1,0} = \overline{x} \qquad a_{0,1} = \overline{y}$$

També resulta d'interès al moment d'orde (1, 1):

$$a_{1,1} = \frac{\sum_{i=1}^{N}\left[x_i \times y_i\right]}{N}$$

Anàlogament, es poden definir els moments respecte a la mitja d'orde $(r, s)$:

$$m_{r,s} = \frac{\sum_{i=1}^{N}\left[\left(x_i - \overline{x}\right)^r \times \left(y_i - \overline{y}\right)^s\right]}{N}$$

Els moments respecte a la mitja d'orde (2, 0) i (0, 2) coincideixen amb les variàncies d'ambdues variables:

$$m_{2,0} = s_X^{\,2} \qquad m_{0,2} = s_Y^{\,2}$$

El moment respecte a la mitja d'orde (1, 1), que es denomina **covariància** o moment mixt, és de gran importància:


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

$$m_{1,1} = \frac{\sum_{i=1}^{N} \left[ \left( x_i - \bar{x} \right) \times \left( y_i - \bar{y} \right) \right]}{N}$$

Alternativament a la fórmula anterior, la covariància es pot calcular a partir dels moments respecte al origen, segons la fórmula:

$$m_{1,1} = a_{1,1} - a_{1,0} \cdot a_{0,1} = \frac{\sum_{i=1}^{N} (x_i \cdot y_i)}{N} - \bar{x} \cdot \bar{y}$$

La covariància, igual que la variança, té l'inconvenient que depèn de les unitats de la mostra.

Per aquest motiu, s'utilitza el **coeficient de correlació** lineal de Pearson (que es denota, indistintament, com ρ o *r*):

$$\rho = r = \frac{m_{1,1}}{s_x \cdot s_y} = \frac{\frac{\sum_{i=1}^{N} (x_i \cdot y_i)}{N} - \bar{x} \cdot \bar{y}}{s_x \cdot s_y}$$

Aquest coeficient tindrà el signe de la covariància i ens indicarà si la dependència entre les dues variables objecte d'estudi són dependents positiva o negativament. El coeficient de correlació (o simplement correlació) pren un valor comprés entre –1 i 1. Si la correlació és positiva es diu que hi ha dependència directa entre *X* i *Y* (a un augment d'una de les dos variables li correspon una tendència a l'augment a l'altra). En canvi, si la correlació és negativa, es diu que hi ha una dependència inversa (a un augment d'una de les dos variables li correspon una tendència a la disminució a l'altra).

## Activitats resoltes

- El propietari d'una instal·lació mixta solar-eòlica està realitzant un estudi del volum d'energia que és capaç de produir la instal·lació. Per a això, mesura la dita energia al llarg d'un total de *N* = 16 dies que considera prou representatius. L'energia (en kWh) produïda als dits dies per les instal·lacions solar i eòlica es troba arreplegada a la taula següent:

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Generació solar ($x_i$) | 13'1 | 10'5 | 4'1 | 14'8 | 19'5 | 11'9 | 18 | 8'6 | 5'7 | 15'9 | 11'2 | 6'8 | 14'2 | 8'2 | 2'6 | 9'7 |
| Generació eòlica ($y_i$) | 8'5 | 14'3 | 24'7 | 4 | 2'3 | 6'4 | 3'6 | 9'2 | 13'5 | 1'4 | 7'6 | 12'8 | 10'3 | 16'5 | 21'4 | 10'9 |

Utilitzant aquestes produccions, calcularem la covariància i el coeficient de correlació, denotant a la generació solar com a variable ***X*** i la generació eòlica com a variable ***Y***. Afegim noves files a la nostra taula:

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Generació solar ($x_i$) | 13'1 | 10'5 | 4'1 | 14'8 | 19'5 | 11'9 | 18 | 8'6 | 5'7 | 15'9 | 11'2 | 6'8 | 14'2 | 8'2 | 2'6 | 9'7 |
| Generació eòlica ($y_i$) | 8'5 | 14'3 | 24'7 | 4 | 2'3 | 6'4 | 3'6 | 9'2 | 13'5 | 1'4 | 7'6 | 12'8 | 10'3 | 16'5 | 21'4 | 10'9 |
| $x_i^2$ | 171'6 | 110'3 | 16'81 | 219'0 | 380'3 | 141'6 | 324 | 73'96 | 32'49 | 252'8 | 125'4 | 46'24 | 201'6 | 67'24 | 6'76 | 94'09 |
| $y_i^2$ | 72'25 | 204'5 | 610'1 | 16 | 5'29 | 40'96 | 12'96 | 84'64 | 182'3 | 1'96 | 57'76 | 163'8 | 106'1 | 272'3 | 457'9 | 118'8 |
| $x_i * y_i$ | 111'4 | 150'2 | 101'3 | 59'2 | 44'85 | 76'16 | 64'8 | 79'12 | 76'95 | 22'26 | 85'12 | 87'04 | 146'2 | 135'3 | 55'64 | 105'7 |

Prèviament calculem la mitja i la desviació típica de cada variable (que ja coneixem d'una activitat resolta anterior). Sumant la primera fila i dividint per *N* = 16, obtenim la mitja de la Generació Solar en


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

Kwh. Recorda $\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$; per tant

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{13'1 + 10'5 + 4'1 + 14'8 + 19'5 + 11'9 + 18 + 8'6 + 5'7 + 15'9 + 11'2 + 6'8 + 14'2 + 8'2 + 2'6 + 9'7}{16} = 10'925 \text{ Kwh}$$

Sumant la segona fila i dividint per $N$ = 16 obtenim la mitjana de la Generació Eòlica en Kwh:

$$\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N} = \frac{8'5 + 14'3 + 24'7 + 4 + 2'3 + 6'4 + 3'6 + 9'2 + 13'5 + 1'4 + 7'6 + 12'8 + 10'3 + 16'5 + 21'4 + 10'9}{16} = 10'463 \text{ Kwh}$$

A la tercera fila hem calculat els quadrats dels valors de la primera variable i els utilitzem per a calcular la variança: Recorda $s_x^{\,2} = \frac{\sum_{i=1}^{N} x_i^{\,2}}{N} - \bar{x}^2$; per tant

$$s_x^{\,2} = \frac{\sum_{i=1}^{N} x_i^{\,2}}{N} - \bar{x}^2 = \frac{13'1^2 + 10'5^2 + 4'1^2 + 14'8^2 + 19'5^2 + 11'9^2 + 18^2 + 8'6^2 + 5'7^2 + 15'9^2 + 11'2^2 + 6'8^2 + 14'2^2 + 8'2^2 + 2'6^2 + 9'7^2}{16} = \frac{141'5}{16} - 10'9^2 = 22'16$$

A la quarta fila els quadrats dels valors de la segona variable i calculem la seua variança tal que

$$s_y^{\,2} = \frac{\sum_{i=1}^{N} y_i^{\,2}}{N} - \bar{y}^2 = \frac{8'5^2 + 14'3^2 + 24'7^2 + 4^2 + 2'3^2 + 6'4^2 + 3'6^2 + 9'2^2 + 13'5^2 + 1'4^2 + 7'6^2 + 12'8^2 + 10'3^2 + 16'5^2 + 21'4^2 + 10'9^2}{16} = \frac{150'48}{16} - 10'5^2 = 41'01$$

La desviació típica és l'arrel quadrada de la variança, per tant:

$$s_x = \sqrt{22'16} = 4'71 \text{ y } s_y = \sqrt{41'01} = 6'4$$

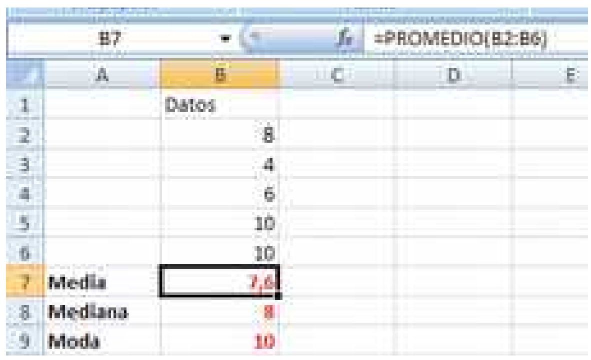
Per a calcular el coeficient de correlació calculem a la cinquena fila els productes de la variable $x$ per la variable $y$. Així, 13'1*8'5 = 111'4. Volem calcular el terme: $\frac{\sum_{i=1}^{N} (x_i \cdot y_i)}{N}$. En sumar obtenim 1401'2, que dividim entre 16, li restem el producte de les mitges i dividim pel producte de les desviacions típiques:

$$\rho = \frac{\frac{\sum_{i=1}^{N} (x_i \cdot y_i)}{N} - \bar{x} \cdot \bar{y}}{s_x \cdot s_y} = \frac{1401'2 - (10'9 \cdot 10'5)}{4'71 \cdot 6'4} = \frac{-26'728}{4'71 \cdot 6'4} = -0'887$$

Aquest coeficient de correlació negatiu i pròxim a –1 ens indica que la relació entre les dos variables és negativa i prou important.

## Utilitza l'ordinador

- *Nieves ha tingut en Matemàtiques les notes següents: 8, 4, 6, 10 i 10. Calcula la seua mitja, la seua moda i la seua mitjana.*




Per a calcular la mitja, la mitjana i la moda amb el full de càlcul, copiem a la casella B2, B3… les dades: 8, 4, 6, 10 i 10. Escrivim a la casella A7, Mitja, i per a calcular la mitja escrivim un signe igual en B7. Busquem, desplegant les possibles funcions, la funció MITJA, i


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

escrivim

=PROMEDI(B2:B6),

que significa que calcule la mitja dels valors que hi ha a les caselles des de B2 fins a B6.
De la mateixa manera calculem la mitjana buscant a les funcions o escrivint =MITJANA(B2:B6) i la moda buscant a les funcions o escrivint =MODA(B2,B6).
*(Nota del traductor: el nom de les funcions depèn de l'idioma en què estiga configurat el full de càlcul)*


Igual que hem calculat la mitja, la mitjana i la moda, el full de càlcul es pot utilitzar per a obtindre:

- El recorregut calculant MAX – MIN → 6.
- La variança utilitzant VARP → 5'44.
- La desviació típica usant DESVESTP →2'33
- Els quartils, (QUARTIL), sent el quartil 0 el mínim; el quartil 1, Q1; el quartil 2, la mitjana; el quartil 3, Q3; i el quartil 4, el màxim.
- Q1 = 6.
- Q3 = 10.
- Interval interquartílic = 10 – 6 = 4.

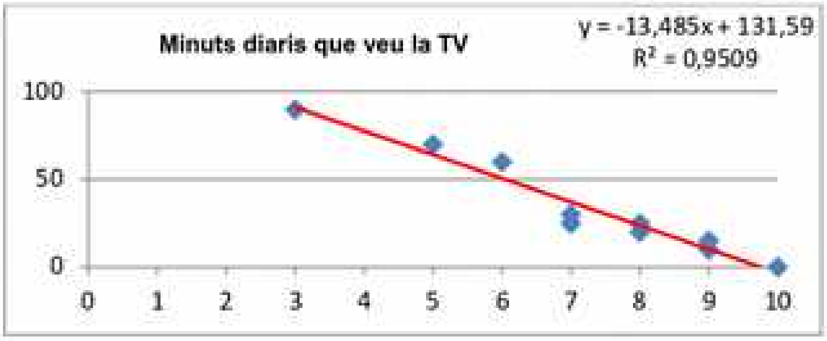
## Utilitza l'ordinador

- Preguntem a 10 alumnes de 4t d'ESO per les seues qualificacions en Matemàtiques, pel nombre de minuts diaris que veuen la televisió, pel nombre d'hores setmanals que dediquen a l'estudi, i per la seua estatura en centímetres. Les dades s'arrepleguen a la taula adjunta. Volem dibuixar els núvols de punts que els relacionen amb les qualificacions de Matemàtiques, el coeficient de correlació i la recta de regressió.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Qualificacions de Matemàtiques | 10 | 3 | 7 | 8 | 5 | 9 | 9 | 8 | 6 | 7 |
| Minuts diaris que veu la TV | 0 | 90 | 30 | 20 | 70 | 10 | 15 | 25 | 60 | 25 |
| Hores setmanals d'estudi | 15 | 2 | 9 | 12 | 7 | 14 | 13 | 11 | 7 | 8 |
| Estatura (en cm) | 177 | 168 | 157 | 159 | 163 | 179 | 180 | 175 | 169 | 170 |

Per a fer-ho, entrem en Excel, i copiem les dades. Seleccionem la primera i la segona fila, després la primera i la tercera i finalment la primera fila i la quarta.

Amb la primera i segona files seleccionades, inserirem, *Dispersió* i triem el núvol *de punts*. Podem aconseguir que l'eix d'abscisses vaja de 0 a 10 en "Donar *format a l'eix*". Punxem sobre un punt del núvol, i triem "*Agregar línia de tendència*". Perquè dibuixe l'ordinador la recta de regressió la línia de tendència ha de ser *Lineal*. En la pantalla que apareix marquem la casella que diu: "*Presentar equació al gràfic*" i la casella que diu "*Presentar el valor de R quadrat al gràfic*".



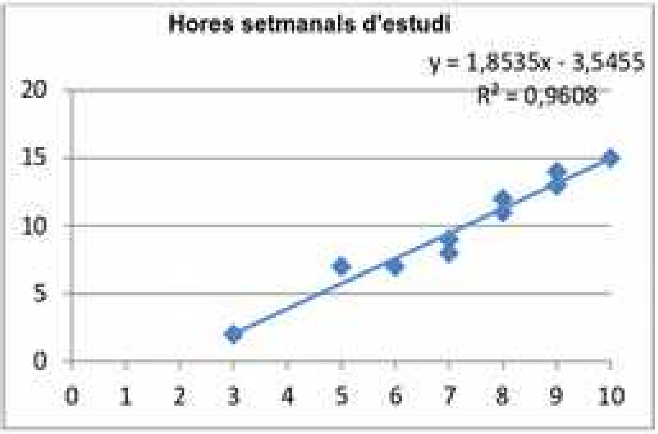

Observa, la recta de regressió, en color roig, és decreixent i la seua equació és aproximadament:

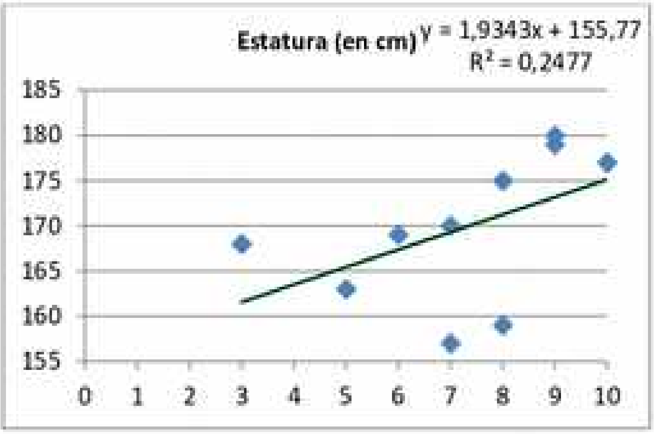
$$y = -13'5\,x + 132.$$

El quadrat del coeficient de correlació és $\rho^2 = 0'95$. La correlació és negativa i alta:

$$\rho = \sqrt{0'95} = -0{,}975$$

Fem el mateix amb la primera i tercera fila i amb la primera i quarta fila. Obtenim els gràfics:




Observa que en ambdós casos el pendent de la recta de regressió és positiva però al primer el coeficient de correlació, positiu, és pròxim a 1, $\rho = \sqrt{0{,}96} = 0{,}98$ . La correlació és alta i positiva.

Al segon $\rho = \sqrt{0'25} = 0{,}5$.

## Activitats proposades

**28.** S'han mesurat els pesos i altures de 6 persones, com a mostra de les persones que estan en una fila o cua d'espera, obtenint-se els resultats següents:

| | | | | | | |
|---|---|---|---|---|---|---|
| Pesos (kg) | 65 | 60 | 65 | 63 | 68 | 68 |
| Altures (cm) | 170 | 150 | 168 | 170 | 175 | 180 |

Es demana:

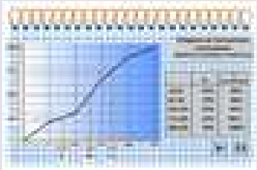
a) Calcular les mitges i les variàncies d'aqueixos dos conjunts de dades unidimensionals.
b) Quines mesures estan més disperses, els pesos o les altures?
c) Representar gràficament aqueix conjunt de dades bidimensionals. Calcular la covariància i interpretar el seu valor.

d) Donar una mesura de la correlació entre ambdues variables. Interpretar el seu valor.

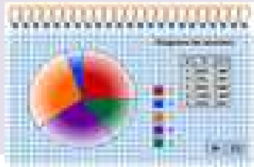

Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

## CURIOSITATS. REVISTA

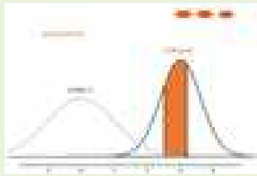
### UTILITZEM L'ESTADÍSTICA PER DAMUNT DE LES NOSTRES POSSIBILITATS?



A les últimes dècades l'ús de dades estadístiques és una de les principals maneres amb què es presenta informació de qualsevol tipus, provinga la seua font dels Mitjans de comunicació, a través de missatges publicitaris o relacionada amb treballs d'investigació. Actualment consumir informació es converteix, moltes vegades, entrar en un món de nombres, percentatges, gràfics, probabilitats, mapes i altres conceptes bàsics d'aquesta disciplina que costa entendre.



### "TINC ELS MEUS RESULTATS FA TEMPS, PERÒ NO SÉ COM ARRIBAR A ELLS"

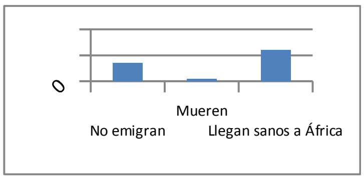


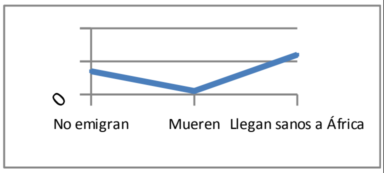
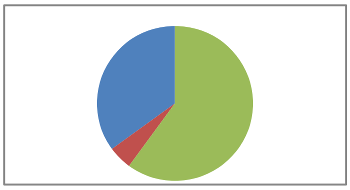
Aquesta expressiva frase de Gauss -descobridor de la campana que porta el seu nom, i que al·ludeix a la distribució normal quan la quantitat de dades és prou gran-, és aplicable a moltes de les informacions errònies que veiem diàriament. Tenen les dades però no saben com arribar al nucli de la seua interpretació. Moltes vegades quan un mitjà de comunicació vol impressionar mitjançant un titular sobre la gravetat d'una situació que afecta tota la població, fa ús de nombres absoluts en compte de percentatges.


Per exemple: Quan llegim el titular sens dubte que tots pensem que 40 morts són molts morts siguen per accident de tràfic o per una altra causa. L'argúcia està ben pensada per a cridar l'atenció del lector, però informativament parlant aquesta presentació dels fets utilitzant nombres sense comparar-los amb altres nombres es mereix *"un suspens"*. Les dades estadístiques no "parlen per si mateixes". Una dada sempre cal relacionar-la amb altres dades per a comprendre la variabilitat que ha experimentat el cas que estem analitzant. Si la notícia s'haguera acompanyat amb les estadístiques de morts per accident de tràfic dels últims anys en períodes vacacionals de quatre dies, ràpidament el lector es donaria compte que no és per a alarmar-se més que altres vegades ja que el nombre de morts ni ha pujat ni ha baixat, és més o menys el mateix que en qualsevol altre pont semblant en dies. O siga, aquest "impactant" titular recolzat en dades numèriques, en realitat *ni tan sols és notícia*…


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

## RESUM

| | | *Exemples* |
|---|---|---|
| **Població estadística, col·lectiu o univers** | El conjunt de tots els individus (persones, objectes, animals, etc.) que continguen informació sobre el fenomen que s'estudia. | Nombre de persones a Espanya entre 16-65 anys |
| **Mostra** | És un subconjunt representatiu que se selecciona de la població i sobre el qual es va a realitzar l'anàlisi descriptiva. La grandària de la mostra és el nombre dels seus elements. Quan la mostra comprèn a tots els elements de la població, es denomina cens. | Nombre de persones en un barri de Madrid entre 16 i 65 anys. |
| **Variable observable o estadística *X*** | En general, suposarem que s'està analitzant una determinada població, de la que ens interessa certa característica que ve donada per la variable *X*. | Les variables que estan baix estudi es poden classificar en dues categories:<br>Variables qualitatives o atributs (dades no mètrics)<br>Variables quantitatives, que tenen un valor numèric. |
| **Freqüència absoluta** | Nombre de vegades que es repeteix un valor de la variable | Si en tirar un dau hem obtingut 2 vegades el 3, 2 és la freqüència absoluta de 3. |
| **Freqüència relativa** | Freqüència absoluta dividit pel nombre d'experiments | Si es realitza un experiment 500 vegades i la freqüència absoluta d'un succés és 107, la freqüència relativa és 107/500. |
| **Freqüència acumulada** | Se sumen les freqüències anteriors | |
| **Diagrama de rectangles o barres** | Els valors de la variable es representen mitjançant rectangles de la mateixa base i d'altura proporcional a la freqüència. S'indica a l'eix horitzontal la variable i al vertical les freqüències. |  |

| | | |
|---|---|---|
| **Polígon de freqüències** | S'uneixen els punts mitjans superiors d'un diagrama de barres |  |
| **Diagrama de sectors** | En un cercle es dibuixen sectors d'angles proporcionals a les freqüències |  |
| **Mitja aritmètica** | És el quocient entre la suma de tots els valors de la variable i el nombre total de dades. | En les dades 3, 5, 5, 7, 8, la mitja és: (3 + 5 + 5 + 7 + 8)/5 = 28/5 = 5'6. |
| **Mitjana** | Deixa per davall la mitat dels valors i per damunt l'altra mitat | La mitjana és 5 |
| **Moda** | El valor que més es repeteix. | La moda és 5. |
| **Variança** | Mesura de desviació que arreplega les desviacions dels valors de la variable respecte de la mitja aritmètica. | $s^2 = \frac{\sum_{i=1}^{N}\left[\left(x_i - \bar{x}\right)^2\right]}{N}$ |
| **Desviació típica** | La desviació típica és l'arrel quadrada de la variança | |
| **Coeficient de variació** | Permet comparar la variabilitat de distintes mostres, independentment de les seues unitats de mesura. | $g = \frac{s}{\lvert\bar{x}\rvert}$ |
| **Rang total o recorregut** | Diferència entre els valors màxims i mínims que pren la variable a la mostra | $R = x_{\{N\}} - x_{\{1\}}$ |
| **Recorregut interquartílic** | Diferència entre el tercer i el primer quartil | $R_I = Q_3 - Q_1$ |
| **Núvol de punts** | Un mode senzill de representar una mostra bidimensional. Aquesta tècnica consisteix a representar al pla (*x*, *y*) els valors obtinguts a la mostra. |  |
| **Coeficient de correlació** | Ens indica si la dependència entre dos variables objecte d'estudi són dependents positiva o negativament. | $\rho = r = \frac{m_{1,1}}{s_x \times s_y}$ |

# EXERCICIS I PROBLEMES.

## Població i mostra. Variables estadístiques. Taules de freqüències

**1.** Es llança una moneda 700 vegades i s'obté cara 355 vegades. Expressa en una taula les freqüències absolutes, relatives i calcula també les freqüències acumulades absolutes i acumulades relatives de cares i creus en aquest experiment.

**2.** Es llança un dau 500 vegades i s'obtenen els resultats següents:

| **Resultat** | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| **Nombre de vegades** | 70 | 81 | 92 | 85 | | 81 |

a) Quantes vegades ha eixit el 5?

b) Construir taula amb les freqüències absolutes i les freqüències absolutes acumulades

c) Construir una taula amb les freqüències relatives i les freqüències relatives acumulades

**3.** Una urna que conté 10 boles numerades del 0 al 9, traiem una bola, anotem el nombre i tornem la bola a l'urna. Repetim l'experiment 1000 vegades i s'han obtingut els resultats indicats a la taula:

| **Resultat** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Freqüència absoluta** | 79 | 102 | | | 93 | 98 | 104 | 77 | | |
| **Freqüència relativa** | | | 0'12 | 0'13 | | | | | 0'1 | |
| **Freqüència absoluta acumulada** | 79 | 181 | | | | | | | | |
| **Freqüència relativa acumulada** | | | | | | | | | | 1 |

a) Quina és la freqüència absoluta de 9?

b) Quina és la freqüència absoluta acumulada de 2?

c) Quina és la freqüència relativa acumulada d'1?

d) Copia la taula al teu quadern i completa-la.

**4.** Pepa ha tirat un dau 25 vegades i ha obtingut els resultats següents:

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

a) Construir una taula de freqüències absolutes.

b) Construir una taula de freqüències relatives.

c) Dibuixa un diagrama de barres.

d) Dibuixa un polígon de freqüències i una representació per sectors.


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

**5.** En una classe s'ha mesurat la grandària de les mans de cada un dels alumnes, i el resultat en centímetres ha sigut el següent:

19, 18, 20, 19, 18, 21, 19, 17, 16, 20,

16, 19, 20, 21, 18, 17, 20, 19, 22, 21,

23, 21, 17, 18, 17, 19, 21, 20, 16, 19

a) Quina grandària ha sigut el valor mínim? I el màxim? Quin és el rang total de la variable?

b) Construir una taula de freqüències absolutes i una altra de freqüències relatives.

c) Construir una taula de freqüències absolutes acumulades i una altra de freqüències relatives acumulades.

**6.** Calcula la freqüència absoluta de les dades d'una enquesta en què s'ha triat entre veure la televisió, t, o llegir un llibre, l:

t, l, t, t, t, l, t, t, l, t, l, t, l, t, t, t, l, l, t, l, t, l, t, l, t.

**7.** La duració en minuts d'unes telefonades ha sigut:

7, 3, 6, 3, 7, 5, 4, 3, 5, 7, 10, 1, 9, 12, 2

Construir una taula de freqüències absolutes i una taula de freqüències relatives.

## Gràfics estadístics

**8.** S'ha preguntat en un poble de la província de Madrid el nombre de germans que tenien i s'ha obtingut la següent taula de freqüències absolutes sobre el nombre de fills de cada família:

| **Nombre de fills** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 o més |
|---|---|---|---|---|---|---|---|---|
| **Nombre de famílies** | 46 | 249 | 205 | 106 | 46 | 21 | 15 | 6 |

a) Escriu al teu quadern una taula de freqüències relatives.

b) Fes un diagrama de barres de freqüències absolutes i un altre de freqüències relatives.

c) Fes un polígon de freqüències absolutes i un altre de freqüències absolutes acumulades.

**9.** Fes una enquesta semblant amb els teus companys i companyes de curs preguntant el nombre de germans i confeccionant una taula sobre el nombre de fills i el nombre de famílies.

a) Construeix una taula de freqüències relatives

b) Fes un diagrama de barres de freqüències absolutes i relatives. Completa amb un polígon de freqüències

c) Compara la taula de freqüències relatives i el diagrama de barres de freqüències relatives que obtingues amb l'obtingut en l'exercici anterior.

**10.** Un batut de fruites conté 25 % de taronja, 15 % de plàtan; 50 % de poma i, la resta de llet. Representa en un diagrama de sectors la composició del batut.

**11.** En un campament d'estiu s'han gastat deu mil euros. El gràfic mostra la distribució del gasto:


1. Menjar: 40 %
2. Neteja i manteniment: 30 %
3. Aigua, gas, electricitat i telèfon: 25 %
4. Vestuari: ..............

a) Quin percentatge es va gastar en vestuari?

b) Quants euros es van gastar en menjar?

c) Quant mesura l'angle del sector corresponent a activitats?

**12.** Busca en revistes o periòdics dues gràfiques estadístiques, retalla-les i apega-les al teu quadern. Moltes vegades aquestes gràfiques tenen errors. Observa-les detingudament i comenta les qüestions següents:

a) Està clara la variable a què es referix? I les freqüències?

b) Són correctes les unitats? Poden millorar-se?

c) Comenta les gràfiques.

**13.** Es fa una enquesta sobre el nombre de vegades que van al cine uns jóvens al mes. Els valors de la variable estan a la taula:

| **Vegades que van al cine** | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| **Freqüència absoluta** | 1 | 7 | 9 | 5 | 2 | 1 |

a) Representa un diagrama de barres de freqüències absolutes.

b) Representa un polígon de freqüències relatives.

c) Representa els valors de la variable en un diagrama de sectors.

**14.** Es fa un estudi sobre el que es recicla en una ciutat i es fa una taula amb el pes en percentatge dels distints tipus de residus:

| **Tipus de residu** | **Percentatge** |
|---|---|
| Orgànic | 15 |
| Paper i cartó | 1 |
| Vidre | 15 |
| Plàstic | 1 |
| Piles | 15 |

a) Construeix un diagrama de barres

b) Representa un polígon de freqüències.

c) Representa els valors de la variable en un diagrama de sectors.

**15.** En un exercici anterior s'ha tingut el resultat de mesurar en una classe la grandària de les mans de cada un dels alumnes, i el resultat en centímetres ha sigut el següent:

19, 18, 20, 19, 18, 21, 19, 17, 16, 20,

16, 19, 20, 21, 18, 17, 20, 19, 22, 21,

23, 21, 17, 18, 17, 19, 21, 20, 16, 19

Representa els valors de la variable en un diagrama de barres i en un polígon de freqüències.

**16.** El 35 % de les cigonyes no ha emigrat enguany a Africa i el 6 % va morir pel camí. Dibuixa un diagrama de sectors que descriga aquesta situació.

**17.** En una classe s'ha preguntat per les preferències esportives i s'ha obtingut:

| Futbol | Bàsquet | Natació | Karate | Ciclisme |
|---|---|---|---|---|
| 8 | 9 | 7 | 6 | 10 |

a) Copia la taula al teu quadern i construeix una taula de freqüències relatives.

b) Representa aquests valors de la variable en un diagrama de sectors.

## Mesures de centralització i dispersió

**18.** Pepa ha tirat un dau 25 vegades d'un exercici anterior i ha obtingut els resultats següents:

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

a) Calcula la mitja aritmètica
b) Calcula la mitjana
c) Quina és la moda? És única?
d) Calcula la variança i desviació típica interpretant el seu resultat

**19.** Sara ha tingut les següents notes als seus exàmens de Matemàtiques: 9, 7, 8, 6, 9, 10, 9

a) Calcula la mitja aritmètica

b) Calcula la mitjana

c) Quina és la moda? És única?

d) Calcula el percentil 45 interpretant el seu resultat

e) Calcula el percentil 75 interpretant el seu resultat. quin altre nom rep?


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

f) Calcula la variança i desviació típica interpretant el seu resultat

g) Calcula el coeficient de variació interpretant el seu resultat

**20.** En un exercici anterior s'ha tingut el resultat de mesurar en una classe la grandària de les mans de cada un dels alumnes, i el resultat en centímetres ha sigut el següent:

19, 18, 20, 19, 18, 21, 19, 17, 16, 20,

16, 19, 20, 21, 18, 17, 20, 19, 22, 21,

23, 21, 17, 18, 17, 19, 21, 20, 16, 19

a) Calcula la mitja aritmètica

b) Calcula la mitjana

c) Quina és la moda? És única?

d) Calcula el percentil 45 interpretant el seu resultat

e) Calcula el percentil 75 interpretant el seu resultat. Quin altre nom rep?

f) Calcula la variança i desviació típica interpretant el seu resultat

g) Calcula el coeficient de variació interpretant el seu resultat

**21.** Ens interessa conèixer la distribució de notes obtingudes per 40 estudiants. Les notes són:

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

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

a) Escriu al teu quadern una taula de freqüències absolutes.

b) Fes un polígon de freqüències absolutes.

c) Calcula la mitja

d) Calcula la mitjana

e) Calcula la moda

f) Calcula el percentil 45 interpretant el seu resultat

g) Calcula el percentil 75 interpretant el seu resultat. quin altre nom rep?

h) Calcula la variança i desviació típica interpretant el seu resultat

i) Calcula el coeficient de variació interpretant el seu resultat

j) Si les notes dels mateixos alumnes respecte a una altra assignatura tenen una mitjana de 5,3 i desviació típica de 2, quina de les dues assignatures té una mitja més homogènia?


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

**22.** Els jugadors d'un equip d'handbol tenen les edats següents:

12, 14, 13, 12, 15, 11, 12, 12, 13, 14, 11, 12, 12.

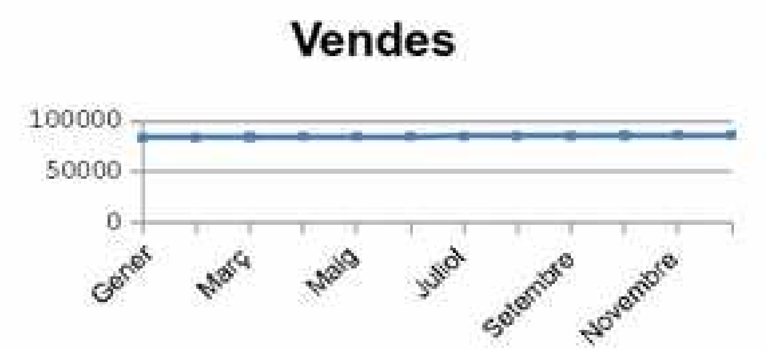
a. Calcula la mitja
b. Calcula la mitjana
c. Calcula la moda
d. Calcula el percentil 45 interpretant el seu resultat
e. Calcula el percentil 75 interpretant el seu resultat. quin altre nom rep?
f. Calcula la variança i desviació típica interpretant el seu resultat
g. Calcula el coeficient de variació interpretant el seu resultat

## Problemes

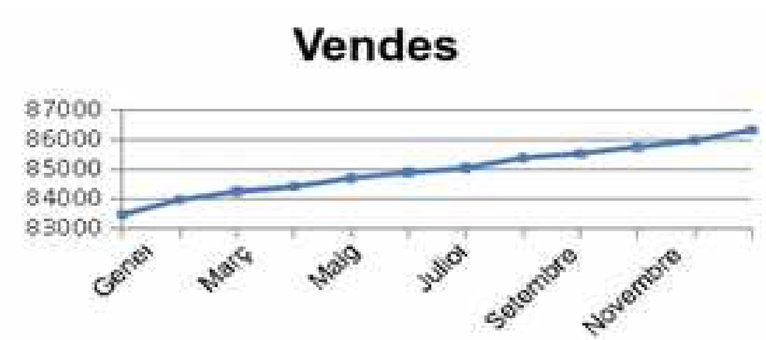
**23.** El Director Comercial d'una empresa serà avaluat. Per a això ha de donar compte dels resultats obtinguts. Vol quedar bé, perquè això li pot suposar un augment de sou. S'han venut les quantitats següents:

| **Mesos** | Gener | Febrer | Març | Abril | Maig | Juny | Juliol | Agost | Setembre | Octubre | Novembre | Desembre |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Vendes** | 83451 | 83962 | 84238 | 84401 | 84693 | 84889 | 85032 | 85378 | 85524 | 85751 | 859967 | 86316 |

L'estadístic de l'empresa li ha entregat la següent gràfica:




No li ha agradat gens, i per a la presentació ell s'ha confeccionat el següent gràfic:




Ambdós gràfics són correctes. Escriu un inform sobre com poden els distints gràfics donar impressions tan diferents.


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

**24.** Llança una moneda 15 vegades i anota les vegades que cau cara i les que no. Construeix després dues taules: una de freqüències absolutes i una altra de freqüències relatives. Representa el resultat en un diagrama de freqüències i en un polígon de freqüències.

**25.** La mitja de sis nombres és 5. S'afigen dos nombres més però la mitja continua sent 5. Quant sumen aquests dos nombres?

**26.** La següent taula expressa les estatures, en metres, de 1000 soldats:

| Talla | 1'50 - 156 | 1'56 – 1'62 | 1'62 - 168 | 1'68 - 1'74 | 1'74 - 1'80 | 1'80-1'92 |
|---|---|---|---|---|---|---|
| Nr de soldats | 20 | 150 | 200 | 330 | 200 | 100 |

Calcula:

a) La mitja i la desviació típica.

b) Els intervals on es troben la mitjana i els quartils.

c) L'interval ($\bar{x}$ - σ, $\bar{x}$ + σ) i el percentatge d'individus en el dit interval.

d) Representa les dades en un histograma.

**27.** Una companyia aèria sospita que hi ha una relació entre les variables *X*, temps d'un vol, en hores; i *Y*, consum de combustible (gasoil) per al dit vol, en litres. Per aquesta raó, s'han obtingut les següents dades, dins del rang de nivells d'interès per a X en aquesta companyia.

| $X_i$ | 0´4 | 0´5 | 0´6 | 0´65 | 0´7 | 0´8 | 1 | 1´15 | 1´2 |
|---|---|---|---|---|---|---|---|---|---|
| $Y_i$ | 1.350 | 2.220 | 2.900 | 3.150 | 3.350 | 3.550 | 3.900 | 4.330 | 4.500 |

| $X_i$ | 1´4 | 1´5 | 1´6 | 1´8 | 2´2 | 3 |
|---|---|---|---|---|---|---|
| $Y_i$ | 5.050 | 5.320 | 5.650 | 6.400 | 7.500 | 10.250 |

Es demana:

a) Mitjançant la representació del diagrama de dispersió raonar l'interès de relacionar les dites variables.

b) Obtindre la covariància i el coeficient de correlació entre ambdues variables. Interpretar els resultats.


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF

## AUTOAVALUACIÓ

**1.** Un diagrama de caixa informa sobre:

a) Els quartils i qurtosis. b) Asimetria i variança. c) Dades atípiques i simetria.

**2.** Siga la variable aleatòria nombre de persones que és capaç d'alçar un ascensor. Per a calcular el nombre de persones a partir del qual s'arreplega el 30 % dels valors de la variable necessitem obtindre

a) El percentil 30 b) El percentil 3 c) El percentil 70

**3.** El 25 % dels madrilenys gasten en la factura del mòbil per damunt de 100 euros, mentres que el 25 % gasten per davall de 20 euros. Llavors coneixem:

a) 100 i 20 són valors que corresponen al quartil 1 i 3, respectivament.

b) 100 i 20 són valors que corresponen al quartil 3 i 1, respectivament.

c) 100 i 20 són valors que no corresponen a cap quartil.

**4.** En un diagrama de barres de freqüències absolutes, la suma de les seues altures és proporcional a:

a) 100 b) 1 c) Total de valors de la variable d) Suma de les seues bases

**5.** La mitja dels següents valors de la variable 3, 4, 6, 7, 5, 8, és:

a) 6 b) 7 c) 4,8 d) 5,5

**6.** La mitjana dels següents valors de la variable 3, 4, 6, 7, 8, és:

a) 6 b) 7 c) 4 d) 5

**7.** La moda dels següents valors de la variable 3, 4, 6, 7, 5, 8, 7, 7, és:

a) 6 b) 7 c) 4 d) 5

**8.** La mitja de 7 nombres és 8. S'afigen dos nombres més però la mitja continua sent 8. Quant sumen aquests dos nombres?

a) 10 b) 16 c) 20 d) 14

**9.** Dues revistes especialitzades en ocupació, A i B, han publicat una mitja d'ofertes de treball, de $m_A$ = 10 i $m_B$ = 20 amb variàncies, respectivament de $s^2_A = 4$ i $s^2_B = 9$.

a) La revista B presenta major dispersió absoluta que la revista A, mentres que la revista A presenta major dispersió relativa que la B

b) La revista A presenta major dispersió absoluta que la revista B, mentres que la revista B presenta major dispersió relativa que la A

c) La revista B presenta major dispersió absoluta i relativa que la A

**10.** El 70 % dels madrilenys gasten en regals nadalencs per damunt de 100 euros, mentres que el 5 % gasten per damunt de 500 euros. Llavors coneixem:

a) El valor corresponent al percentil 30.

b) El valor corresponent al percentil 70.

c) El valor corresponent al percentil 5.


Autora: Raquel Caro
Traducció: Pedro Podadera, IES Juan de Garay
Il·lustracions: Banc de Imatges d'INTEF