Unitat 3

TERMES MITJANS

què treballaràs?

En acabar aquesta unitat has de ser capaç de:

- **Explicar el significat dels diferents tipus de valors numèrics que permeten resumir el comportament d'un conjunt de dades quantitatives.**
- **Calcular els valors numèrics que permeten resumir el comportament d'un conjunt de dades quantitatives.**
- **Comparar les dades estadístiques corresponents a diferents estudis.**
- **Interpretar la informació que ens aporten els termes mitjans.**

Els termes mitjans ens permeten resumir en uns quants valors numèrics el conjunt de dades amb les quals fem l'estudi estadístic.

Hi ha dos tipus de termes mitjans: les mesures de centralització i les mesures de dispersió.

Les mesures de centralització ens indiquen quins són els punts centrals de la distribució de dades, és a dir, cap a quins valors de la variable s'agrupen majoritàriament les dades.

Les mesures de dispersió ens indiquen com d'escampades estan les dades.

Per entendre millor el significat i el càlcul dels diferents termes mitjans, treballarem amb exemples. Exposarem dos exemples per a cada terme mitjà. El primer exemple constarà d'un recull petit de dades. Aquest exemple anirà canviant segons el terme mitjà que s'està estudiant. En el segon exemple hi haurà un recull de dades molt més gran, la qual cosa comportarà treballar amb taules de freqüències per tal d'evitar errades. Aquest segon exemple serà sempre el mateix, per tal de poder comparar els valors dels diferents termes mitjans.

1. Mesures de centralització

Hi ha tres tipus de mesures de centralització: la mitjana, la mediana i el mode.

La mitjana

La mitjana s'obté dividint la suma de les dades entre el nombre total de dades.

El símbol de la mitjana és $\overline{X}$.

Exemple 1

L'estiu passat cinc, amics van recaptar un total de 60 € per fer un regal.

Ens podem fer una idea aproximada del que va aportar cada una de les persones si dividim el total de 60 € entre 5. Així s'obté la mitjana.

Mitjana de diners recollits
60 € : 5 = 12 €

En altres paraules, cada persona va recollir, de mitjana, 12 € .
En realitat, les quantitats aportades per cada un dels amics van ésser:

Jordi 10 € Eva 6 € Joana 16 € Mònica 13 € Pere 15 €

Observa que, en realitat, cap d'ells no va aportar 12 € . De fet, no cal que la mitjana coincideixi amb un dels valors utilitzats per calcular-la.

Exemple 2

En aquest exemple, aprendràs a trobar la mitjana a partir d'una taula de freqüències. Utilitzarem una taula perquè treballarem amb moltes dades, la qual cosa fa que convingui ordenar-les en una taula per tal d'evitar errades.

Demanem als vint-i-cinc estudiants d'una classe quants parells de sabates tenen. Heus aquí els resultats:

2	3	6	10	3	6	3	5	4	4
2	1	6	3	5	11	5	7	10	4
3	5	3	4	10					

Fem una taula de freqüències per tal d'ordenar les dades recollides:

Parells de sabates	n_i
1	1
2	2
3	6
4	4
5	4
6	3
7	1
8	0
9	0
10	3
11	1
	N = 25

Ara, per tal de trobar la mitjana, cal que segueixis els passos següents:

- Multiplica cada un dels valors de la variable per la seva freqüència absoluta. Per estalviar-te feina, no cal que multipliquis els valors de la variable que no recullen cap dada, és a dir, amb $n_i = 0$, ja que el resultat de la multiplicació serà zero, la qual cosa farà que no afectin al resultat final.
- Suma tots els resultats obtinguts.
- Divideix la suma total entre el nombre de dades (N).

Fem-ho:

$$\overline{X} = \frac{(1x1)+(2x2)+(3X6)+(4x4)+(5x4)+(6x3)+(7x1)+(10x3)+(11x1)}{25}$$

$$\overline{X} = \frac{1+4+18+16+20+18+7+30+11}{25} = \frac{125}{25} = 5 \text{ parells}$$

- **Fes les activitats d'aprenentatge 1 i 2**

La mediana

La mediana és la dada que ocupa la posició central quan les dades estan disposades en ordre.

El símbol de la mediana és Me.

Exemple 1

La Sara va fer cinc exàmens de matemàtiques durant el trimestre, i va treure aquestes notes sobre 10 punts:

9 8 1 8 9

La Sara no va fer bé el tercer examen perquè estava malalta.
El seu professor va calcular la nota mitjana dels exàmens fets per la Sara:

$$\overline{X} = \frac{9+8+1+8+9}{5} = \frac{35}{5} = 7 \text{ punts}$$

Aquesta desició no li va semblar justa a la Sara, perquè havia tret més de 7 punts en quatre dels exàmens.

El professor va resoldre el problema utilitzant la mediana de les notes com a terme mitjà.

Per fer-ho, va ordenar les notes des de la més petita fins a la més gran:

1 8 **8** 9 9

La nota que ocupa la posició central és 8 (Me = 8). És el que anomenem mediana. El professor va agafar la mediana per posar la nota final del trimestre i aquest fet li va semblar molt bé a la Sara.

Exemple 2

Ara utilitzarem un altre cop l'exemple del nombre de parells de sabates que tenen els integrants d'un grup d'estudiants, per tal d'aprendre com calcular la mediana a partir d'una taula de freqüències.

Parells de sabates	n_i	N_i
1	1	1
2	2	3
3	6	9
4	4	13
5	4	17
6	3	20
7	1	21
8	0	21
9	0	21
10	3	24
11	1	25
	N = 25	

La taula de freqüències mostra les dades ordenades, des de la més petita fins a la més gran.

En aquest exemple el nombre de dades recollides és de 25 (N=25). Per tant, la posició central està ocupada per la tretzena dada (ja que en té dotze al davant i dotze al darrere), que correspon al valor de la variable 4 parells de sabates.

Així doncs, Me = 4 parells de sabates.

Quan el nombre de dades és parell, per obtenir la mediana hem de calcular la mitjana de les dues dades centrals. Agafem un altre cop l'exemple de les notes de matemàtiques de la Sara, però considerant que va fer sis exàmens durant el trimestre amb aquestes notes:

9 8 1 8 9 9

Primer ordenem les notes des de la més petita fins a la més gran:

1 8 **8** **9** 9 9

Després calculem la mitjana de les dues dades centrals:

$$\text{Me} = \frac{8+9}{5} = 8{,}5 \text{ punts}$$

Si treballem amb una taula i un nombre parell de dades, la situació és similar. Si tornem al segon exemple, però hi afegim una dada més (N = 26):

Parells de sabates	n_i	N_i
1	1	1
2	2	3
3	6	9
4	4	13
5	5	18
6	3	21
7	1	22
8	0	22
9	0	22
10	3	25
11	1	26
	N = 26	

Atès que el nombre de dades és de 26, la posició central de les dades ordenades l'ocupen les dades 13a i 14a. Fixa't que abans de la dada 13a hi ha 12 dades (de la 1a a la 12a) i que després de la dada 14a també hi ha 12 dades (de la 15a a la 26a). La dada 13a correspon al valor de la variable 4 parells i la dada 14a correspon al valor de la variable 5 parells. Això ho podem veure fàcilment a partir de les N_i. Fixa't que la N_i del valor de la variable 4 parells és 13. Això ens indica que fins a 4 parells hi ha 13 dades. L'última d'aquestes dades, la número 13, correspon, doncs, al valor de la variable 4 parells. La N_i del valor de la variable 5 parells és 17, la qual cosa indica que les dades 14a, 15a, 16a i 17a corresponen al valor de la variable 5 parells.

Ara ja sabem els valors de la variable que ocupen la posició central: 4 parells i 5 parells. Calculem ara la mitjana d'aquests dos valors per obtenir la mediana:

$$Me = \frac{4+5}{2} = 4{,}5 \text{ punts}$$

El mode

El mode és la dada que més succeeix.

El símbol del mode és Mo.

Exemple 1

«Quants exàmens us agradaria fer durant aquest trimestre?», va preguntar el professor als seus alumnes.

Els alumnes van respondre i el professor va escriure les respostes a la pissarra:

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

En Joan va calcular ràpidament el valor de la mitjana, 2,5. «Hem de fer 2,5 exàmens durant el trimestre», va dir.

La Maria intervingué aleshores: «Això no té ni solta ni volta. Hem de fer dos exàmens durant el trimestre, perquè és la tria majoritària». La Maria té raó. Hi ha més dosos a la pissarra que qualsevol altre nombre. Així doncs, el 2 és el nombre més comú i l'anomenem mode o valor modal.

Per tant, en aquest cas Mo = 2 exàmens.

De vegades el mode és més útil que no pas la mitjana.

Exemple 2

Ara utilitzarem un altre cop l'exemple del nombre de parells de sabates que tenen els integrants d'un grup d'estudiants, per tal d'aprendre com calcular el mode a partir d'una taula de freqüències.

Parells de sabates	n_i
1	1
2	2
3	6
4	4
5	4
6	3
7	1
8	0
9	0
10	3
11	1
	N = 25

En aquest cas, podem veure fàcilment que el nombre de parells de sabates més habitual, amb sis dades enregistrades, correspon al valor de la variable 3 parells de sabates (en color blau a la taula).

Així doncs, Mo = 3 parells de sabates.

- **Fes les activitats d'aprenentatge 3, 4, 5 i 6**

2. Mesures de dispersió

Si només utilitzem les mesures de centralització, no donem una informació acurada. També calen les mesures de dispersió, que ens indiquen com d'escampades estan les dades. Amb la utilització conjunta de les mesures de centralització i de dispersió podem conèixer amb gran precisió quina és la distribució de les dades.

Les mesures de dispersió que veurem són: el rang, la desviació mitjana, la variància i la desviació típica.

El rang

El rang o recorregut és la mesura de dispersió més senzilla, i és la diferència entre el valor més gran i el més petit que pren la variable.

Per calcular-lo, únicament hem de restar el valor més petit al valor més gran.

El símbol del rang és R.

Passem ara als exemples. Mantindrem els mateixos exemples per a totes les mesures de dispersió. D'aquesta manera, podrem comparar les característiques de cada mesura. El primer exemple contindrà poques dades i el segon exemple constarà d'un nombre més gran de dades, per la qual cosa caldrà treballar amb una taula de freqüències.

Exemple 1

En Sebastià agafa el tren de dilluns a divendres per anar a la feina. La setmana passada va anotar en una llibreta els minuts d'endarreriment que duia el tren cada dia:

2 minuts   5 minuts   3 minuts   7 minuts   8 minuts

El rang d'aquesta col·lecció de dades és:

R = valor més gran - valor més petit

R = 8 minuts - 2 minuts

R = 6 minuts

Exemple 2

Com ja vam comentar, el segon exemple serà el mateix que ja vam veure en l'estudi de les mesures de centralització, el del nombre de parells de sabates que tenen els vint-i-cinc estudiants d'una classe. D'aquesta manera, podrem comparar, a través d'un únic cas, tots els termes mitjans, els de centralització i els de dispersió.

Recorda la taula que ja vam construir quan vam estudiar les mesures de centralització:

Parells de sabates	n_i
1	1
2	2
3	6
4	4
5	4
6	3
7	1
8	0
9	0
10	3
11	1
	N = 25

El rang del nombre de parells de sabates que tenen és:

R = valor més gran - valor més petit

R = 11 parells - 1 parells

R = 10 parells de sabates

La desviació mitjana

La desviació mitjana ens indica de quina manera s'allunyen de la mitjana les dades d'un estudi estadístic.

El símbol de la desviació mitjana és DM.

Per tal de calcular la desviació mitjana, has de seguir aquests passos:

- Compta el nombre de dades per obtenir N.
- Troba la mitjana.
- Compara cada una de les dades amb la mitjana i fes-ne la resta. Resta el nombre petit al nombre gran. Tant se val si el gran és la mitjana o la dada en consideració. Si una dada té el mateix valor que la mitjana, el resultat d'aquesta operació serà 0.
- Suma tots els resultats de les restes.
- Divideix el nombre obtingut entre N.

Tornarem a utilitzar els exemples que vam fer servir per estudiar el rang.

Exemple 1

Recorda els minuts d'endarreriment del tren durant la setmana laboral, enregistrats per en Sebastià:

2 minuts 5 minuts 3 minuts 7 minuts 8 minuts

Primer, hem de calcular la mitjana d'aquesta col·lecció de dades:

$$\overline{X} = \frac{2+5+3+7+8}{5} = \frac{25}{5} = 5 \text{ minuts}$$

Ara, per calcular la desviació mitjana, buscarem les diferències entre cada una de les dades i la mitjana, les sumarem totes i dividirem el resultat de la suma entre N:

$$DM = \frac{(5-2)+(5-5)+(5-3)+(7-5)+(8-5)}{5} = \frac{3+0+2+2+3}{5} = \frac{10}{5} = 2 \text{ minuts}$$

Exemple 2

Continuem amb l'exemple dels parells de sabates per aprendre a trobar la desviació mitjana a partir d'una taula de freqüències.

Recorda la taula de freqüències:

Parells de sabates	n_i
1	1
2	2
3	6
4	4
5	4
6	3
7	1
8	0
9	0
10	3
11	1
	N = 25

Primer cal trobar la mitjana del nombre de parells de sabates. Malgrat que ja ho vam fer quan vam estudiar la mitjana, tornarem a calcular-la:

$$\overline{X} = \frac{(1x1)+(2x2)+(3x6)+(4x4)+(5x4)+(6x3)+(7x1)+(10+3)+(11+1)}{25}$$

$$\overline{X} = \frac{1+4+18+16+20+18+7+30+11}{25} = \frac{125}{25} = 5 \text{ parells}$$

Ara, per calcular la desviació mitjana, hem de calcular les diferències entre cada un dels valors de la variable i la mitjana (excepte pels valors sense cap dada recollida) i multiplicar aquestes diferències pel nombre de dades de cada valor de la variable. Després, s'han de sumar tots els resultats i dividir entre N:

$$DM = \frac{(5-1)x1+(5-2)x2+(5-3)x6+(5-4)x4+(5-5)x4+(6-5)x3+(7-5)x1+(10-5)x3+(11-5)x1}{25}$$

$$DM = \frac{4+6+12+4+0+3+2+15+6}{25} = \frac{52}{25} = 2{,}08 \text{ parells}$$

- **Fes les activitats d'aprenentatge 7, 8, 9 i 10**

La variància i la desviació típica

La variància i la desviació típica (també anomenada desviació estàndard) són unes altres maneres d'indicar en quina mesura, per terme mitjà, les dades s'allunyen de la mitjana.

El símbol de la variància és S^2 i el símbol de la desviació típica és S.

Per calcular la variància cal que segueixis aquests passos:

- Compta el nombre de dades, per obtenir N.
- Calcula la mitjana.
- Compara cada una de les dades amb la mitjana i fes-ne la resta (resta-li al nombre gran el nombre petit, sigui quin sigui el gran o el petit). Si una dada té el mateix valor que la mitjana, el resultat d'aquesta operació serà 0.
- Suma els quadrats de les diferències.
- Divideix el resultat de la suma entre N.

La desviació típica és l'arrel quadrada de la variància.

Tornarem a utilitzar els exemples que vam fer servir per estudiar la desviació mitjana.

Exemple 1

Recorda els minuts d'endarreriment del tren durant la setmana laboral, enregistrats per en Sebastià:

2 minuts 5 minuts 3 minuts 7 minuts 8 minuts

Ja vam calcular la mitjana d'aquestes dades, $\overline{X} = 5$ minuts.

Ara, per calcular la variància, hem de calcular els quadrats de les diferències entre cada una de les dades i la mitjana, sumar-les i dividir el resultat de la suma entre N:

$$S^2 = \frac{(5-2)^2 + (5-5)^2 + (5-3)^2 + (7-5)^2 + (8-5)^2}{5} = \frac{3^2+0^2+2^2+2^2+3^2}{5}$$

$$S^2 = \frac{9+0+4+4+9}{5} = \frac{26}{5} = 5{,}2 \text{ minuts}$$

Finalment, per trobar la desviació típica, cal calcular l'arrel quadrada de la variància:

$$S = \sqrt{5{,}2} = 2{,}28 \text{ minuts}$$

Exemple 2

Reprenem ara l'exemple dels parells de sabates per aprendre a trobar la variància i la desviació típica a partir d'una taula de freqüències.

Tornem a la taula que ja vam fer servir pels altres termes mitjans:

Parells de sabates	n_i
1	1
2	2
3	6
4	4
5	4
6	3
7	1
8	0
9	0
10	3
11	1
	N = 25

Primer, cal trobar la mitjana del nombre de parells de sabates. Recorda que ja ho vam fer a l'apartat anterior, quan vam estudiar el procediment per obtenir la desviació mitjana, i que vam obtenir una mitjana de 5 parells ($\bar{X}$ = 5 parells).

Ara, per calcular la variància, hem de trobar el quadrats de les diferències entre cada un dels valors de la variable i la mitjana (excepte pels valors sense cap dada recollida) i multiplicar-los pel nombre de dades de cada un dels valors de la variable. Després, hem de sumar tots els resultats obtinguts i dividir el nombre obtingut entre N.

$$S^2 = \frac{(5\text{-}1)^2x1+(5\text{-}2)^2x2+(5\text{-}3)^2x6+(5\text{-}4)^2x4+(5\text{-}5)^2x4+(6\text{-}5)^2x3+(7\text{-}5)^2x1+(10\text{-}5)^2x3+(11\text{-}5)^2x1}{25}$$

$$S^2 = \frac{16+18+24+4+0+3+4+75+36}{25} = \frac{180}{25} = 7{,}2 \text{ parells}$$

Finalment, per trobar la desviació típica, cal calcular l'arrel quadrada de la variància:

$$S = \sqrt{7{,}2} = 2{,}68 \text{ parells}$$

- **Fes les activitats d'aprenentatge 11, 12, 13, 14, 15 i 16**

3. Les distribucions estadístiques

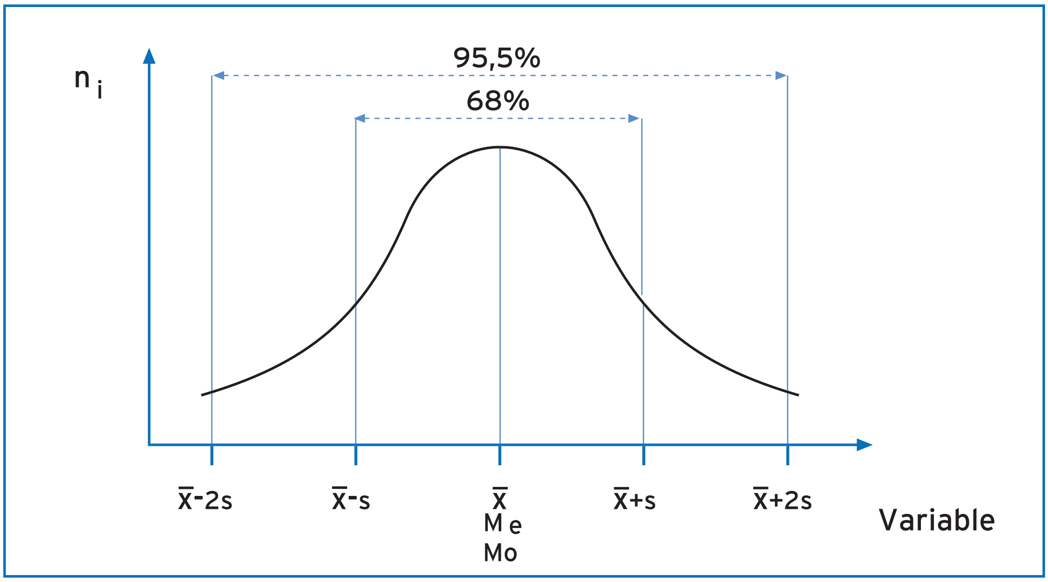
Fins ara, quan hem fet estudis estadístics, hem parlat de grups o col·leccions de dades. La paraula distribució fa referència a les dades que segueixen un patró determinat. Parlem de variables que, quan s'estudien a partir d'un nombre elevat de dades, segueixen un patró que també es repeteix en l'estudi d'altres variables. Com més gran sigui el nombre de dades recollides, més clarament ens n'adonarem de si les dades segueixen una determinada distribució estadística.

Per representar les distribucions, ens basarem en un polígon de freqüències amb els angles suavitzats fins a formar una línia sinuosa. El funcionament és, per tant, idèntic al d'un polígon de freqüències: per determinar la freqüència d'un valor de la variable només cal determinar el punt de la corba que relaciona el valor amb la seva freqüència, de manera semblant a com ho fèiem amb els polígons de freqüències.

Només farem referència a tres tipus de distribucions estadístiques: la normal, les asimètriques i la bimodal. Només veurem les característiques principals de cada una, però no hi entrarem a fons, ja que no és l'objectiu d'aquest mòdul.

La distribució normal

Aquesta distribució també s'anomena campana de Gauss, pels estudis que va fer Carl Friedrich Gauss (1777-1855) sobre aquesta distribució. És una distribució simètrica en forma de campana, que està associada a molts fenòmens naturals i quotidians. Moltes característiques morfològiques, com per exemple l'alçada o el pes, segueixen aquesta distribució.



Les característiques principals d'aquesta distribució són:

- La mitjana, la mediana i el mode coincideixen i marquen el punt de la variable on es divideix el nombre de dades en dues parts iguals (50 % i 50 %).
- La forma de la campana depèn de $\bar{X}$ (la mitjana) i S (la desviació típica). $\bar{X}$ indica la posició de la campana sobre l'eix d'abscisses. Com més gran sigui el seu valor, més «cap a la dreta» se situarà la campana. S determina si la

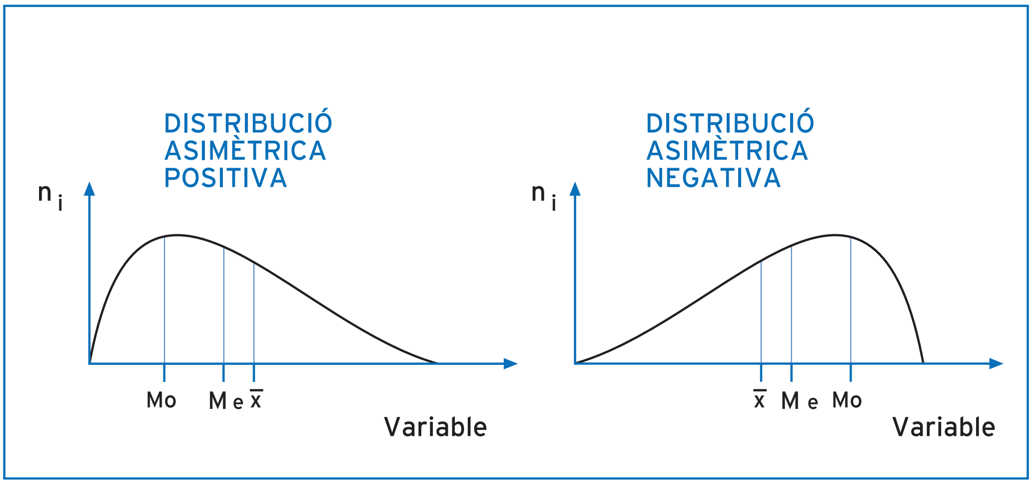
campana és més o menys aplanada. Com més gran sigui el seu valor, més disperses estaran les dades i, per tant, la campana serà més aplanada.

- Un 68 % de les dades se situen entre la mitjana i una desviació típica (S). Un 95,5 % de les dades se situen entre la mitjana i dues desviacions típiques (2S). Així doncs, si en l'estudi del pes d'una població de persones adultes que segueix una distribució normal, $\overline{X}$=72 kg i S = 8 kg, el 68 % de la població pesa entre 64 kg ($\overline{X}$-S) i 80 kg ($\overline{X}$+S) i el 95,5 % de la població pesa entre 56 kg ($\overline{X}$-2S) i 88 kg ($\overline{X}$+2S).

Les distribucions asimètriques

De vegades les dades adopten campanes asimètriques. Aleshores parlem de distribucions asimètriques. En aquests tipus de distribucions, la distància entre la mediana i el mode és aproximadament el doble de la distància entre la mediana i la mitjana.

Quan la campana «s'allarga» cap als valors més alts de la variable, parlem de distribució asimètrica positiva, mentre que, quan la campana «s'allarga» cap als valors més baixos de la variable, parlem de distribució asimètrica negativa.



La distribució bimodal

Hi ha estudis que recullen dades que segueixen dues tendències força diferenciades. En aquests casos, parlem de distribucions bimodals, atès que hi ha dos modes, entre els quals es troben les altres mesures de centralització: la mitjana i la mediana.