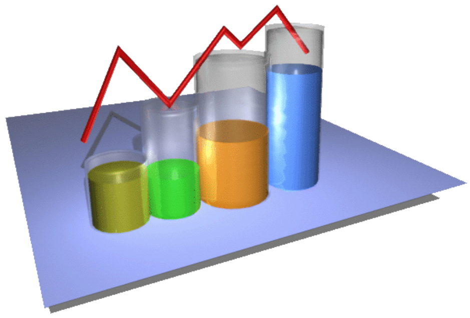
Introducció a l'estadística aplicada per a educadors

Autores:
Judit Fullana Noell
Montse Tesouro Cid

Departament de Pedagogia



UdG

Crèdits

Introducció a l'estadística aplicada per a educadors

1a edició electrònica, octubre 2006

Autores: Judit Fullana Noell i Montse Tesouro Cid
Departament de Pedagogia

ISBN: 84-8458-229-9
Dipòsit legal: GI.1398-2006

Edita:
Universitat de Girona
La Factoria de Recursos Docents
Biblioteca de Montilivi
Campus de Montilivi s/n
Tel. (972)41 89 06 / 639 40 89 73
lafactoria@udg.es

1. Estadística descriptiva

L'estadística descriptiva ajuda a resumir i descriure quantitats grans de dades. Segons Etxeberria i Tejedor (2005:32), l'"estadística descriptiva té com a objectiu recollir, organitzar, resumir, descriure i presentar les dades corresponents a un conjunt d'elements". Dins d'aquest tema parlarem de:

1. Les distribucions de freqüències i percentatges.
2. La representació gràfica.
3. El càlcul dels índexs o mesures de tendència central: mitjana aritmètica, mediana i moda.
4. El càlcul dels índexs o mesures de dispersió: variància i desviació típica.

.

1.1. Distribucions de freqüències i percentatges

Una manera de presentar un conjunt de dades de forma resumida i comprensible és escriure ordenadament tots els valors de la variable que s'analitza posant al costat de cada valor el nombre de vegades que ha ocorregut. El nombre d'ocurrències de cada valor és el que es coneix com a *freqüència.*

Per exemple, imaginem-nos que tenim les notes del primer quadrimestre de 79 estudiants de l'assignatura Bases metodològiques de la investigació educativa:

Cas	**Nota 1r quadrimestre**
1	NP
2	5
3	6
4	3
5	5
6	NP
7	7
8	7
9	NP
10	8
11	NP
12	6
13	3
14	8
15	3
16	8
17	NP
18	7
19	NP
20	7

21	7
22	NP
23	9
24	7
25	NP
26	4
27	6
28	7
29	6
30	4
31	6
32	5
33	6
34	NP
35	7
36	5
37	5
38	4
39	6
40	6
41	3
42	5
43	NP
44	8
45	8
46	8
47	5
48	NP
49	7
50	6
51	6
52	NP
53	NP
54	1
55	8
56	6
57	7
58	7
59	8
60	NP
61	5
62	6
63	7
64	6
65	NP
66	4
67	6
68	7
69	6
70	4

71	7
72	7
73	7
74	4
75	7
76	NP
77	8
78	6
79	NP

Freqüències simples o absolutes: *f*

Consisteix a escriure ordenadament tots els valors possibles, registrant al costat de cadascun el nombre de vegades que ha passat. Serveixen per resumir les dades i són especialment útils quan no hi ha un nombre gaire elevat de valors diferents.

En el cas de la taula anterior, la distribució de freqüències serà:

(Valors de la variable) ***x***	***f***
1	1
3	4
4	6
5	8
6	16
7	17
8	9
9	1
No presentats	17
	$\Sigma f = 79$

La suma de les freqüències absolutes (Σf) és igual al nombre de casos de la distribució:

$$\Sigma f = N$$

Freqüències acumulades: *fa*

Serveixen per veure quants casos hi ha fins a un determinat valor. Només es poden calcular en cas que estiguem treballant amb una variable mesurada en escala d'interval.

En l'exemple anterior, podem buscar les freqüències acumulades, però hem d'eliminar els casos corresponents als "no presentats".

(Valors de la variable) ***x***	***f***	***fa***
1	1	1
3	4	5
4	6	11
5	8	19
6	16	35
7	17	52
8	9	61
9	1	62
	Σf = 62	

<u>Freqüències agrupades en intervals</u>

Quan hi ha massa dades amb molts valors diferents, el que es pot fer és agrupar els valors de la variable en intervals. Un cop decidits els intervals, es fa el recompte de tots els casos que se situen dins d'aquell interval. Són útils quan volem resumir un conjunt de dades, però en general no s'han d'utilitzar per calcular estadístics ja que no proporcionen els valors exactes corresponents a cada cas de la distribució.

Podem agrupar les notes en intervals de dos valors:

Intervals	***f***	***fa***
1-2	1	1
3-4	10	11
5-6	24	35
7-8	26	61
9-10	1	62
	Σf = 62	

<u>Freqüències relatives: *fr*</u>
S'obtenen dividint la freqüència corresponent a cada valor pel nombre total de casos de la distribució:

$$fr_i = f_i / N$$

La $_i$ ens indica que es calcula un valor de freqüència relativa per cada valor de la distribució.

La suma de les freqüències relatives és 1.

Així, en l'exemple, la freqüència relativa corresponent al primer valor de la taula és 1/62; pel valor 3 és 4/62, etc.

(Valors de la variable) *x*	***f***	***fa***	$fr_i = f_i / N$
1	1	1	0,016
3	4	5	0,061
4	6	11	0,097
5	8	19	0,129
6	16	35	0,258
7	17	52	0,274
8	9	61	0,145
9	1	62	0,016
	$\Sigma f = 62$		$\Sigma fr_i = 0,996$

<u>Freqüències relatives acumulades:</u> f_{ra}

$f_{ra} = fa / N$

La freqüència relativa acumulada corresponent al darrer valor de la distribució és sempre 1.

En l'exemple anterior, la freqüència relativa acumulada corresponent al valor 3 és 5/62; la corresponent al valor 6 és 35/62, etc.

(Valors de la variable) *x*	***f***	***fa***	$fr_i = f_i / N$	$f_{ra} = fa / N$
1	1	1	0,016	0,016
3	4	5	0,061	0,080
4	6	11	0,097	0,177
5	8	19	0,129	0,306
6	16	35	0,258	0,565
7	17	52	0,274	0,839
8	9	61	0,145	0,984
9	1	62	0,016	1
	$\Sigma f = 62$		$\Sigma fr_i = 0,996$	

<u>Percentatge: *%* o *P*</u>

Ens indica quina proporció de la distribució representen les freqüències de cada valor, tenint en compte que el nombre total de casos, *N*, és el 100 % de la distribució.
Es calcula multiplicant la freqüència relativa per 100.

$P = fr_i * 100$

(Valors de la variable) *x*	***f***	***fa***	$fr_i = f_i$ / N	$f_{ra} = fa / N$	$P = fr_i * 100$
1	1	1	0,016	0,016	1,6 %
3	4	5	0,061	0,080	6,1 %
4	6	11	0,097	0,177	9,7 %
5	8	19	0,129	0,306	12,9 %
6	16	35	0,258	0,565	25,8 %
7	17	52	0,274	0,839	27,4 %
8	9	61	0,145	0,984	14,5 %
9	1	62	0,016	1	1,6 %
	$\Sigma f = 62$		Σ fr_i = 0,996		

<u>Percentatges acumulats</u>

Pa = $f_{ra} * 100$

(Valors de la variable) *x*	***f***	***fa***	$fr_i = f_i$ / N	$f_{ra} = fa / N$	$P = fr_i * 100$	$Pa = f_{ra} * 100$
1	1	1	0,016	0,016	1,6 %	1,6 %
3	4	5	0,061	0,080	6,1 %	8 %
4	6	11	0,097	0,177	9,7 %	17,6 %
5	8	19	0,129	0,306	12,9 %	30,6 %
6	16	35	0,258	0,565	25,8 %	65,5 %
7	17	52	0,274	0,839	27,4 %	83,9 %
8	9	61	0,145	0,984	14,5 %	98,4 %
9	1	62	0,016	1	1,6 %	100 %
	$\Sigma f = 62$		$\Sigma fr_i = 0,996$			

1.2. Els índexs o mesures de tendència central

Les mesures de tendència central són nombres que descriuen la localització general d'un conjunt de dades. Es tracta de buscar un punt mitjà que serveixi per representar la distribució de valors.

Per exemple, imaginem-nos que tenim les edats de les persones següents. Quin valor representa millor totes aquestes edats?

20, 20, 20, 22, 23, 25, 25, 28, 30, 30, 30, 30, 31

Una possibilitat seria considerar que el valor que més bé representa aquesta distribució és la seva mitjana, és a dir, 25,7 anys, que es troba sumant totes les edats i dividint el resultat pel nombre de persones (13).

Però observem que no hi ha ningú que tingui aquesta edat; per això es pot considerar que 30 anys, que és el valor que més es repeteix, és el que representa millor aquesta distribució. Una altra persona pot pensar que l'edat que expressa millor aquest conjunt de persones és 25 perquè, un cop ordenades totes les edats, és la que queda al mig de la distribució, de manera que la meitat de les persones tenen menys de 25 anys i l'altra meitat en tenen més.

Darrere d'aquests tres números hi ha els conceptes de *mitjana*, *moda* i *mediana*, que tot seguit s'explicaran breument.

1.2.1. Mitjana aritmètica

És l'índex més utilitzat. Es calcula sumant els valors observats i dividint aquesta suma pel nombre d'observacions.

$$\boxed{\overline{X} = \frac{\Sigma x}{N}} \quad (1)$$

En el cas anterior, aplicant la fórmula (1) sumarem tots els valors de les edats i les dividirem per 13:

$$\frac{20 + 20 + 20 + 22 + 23 + 25 + 25 + 28 + 30 + 30 + 30 + 30 + 31}{13}$$

$$\overline{X} = \frac{\Sigma x}{N} = 25{,}7 \text{ anys}$$

Si les dades estan agrupades en freqüències, la fórmula que utilitzarem serà:

$$\boxed{\overline{X} = \frac{\Sigma fx}{N}} \quad (2)$$

En l'exemple anterior, podem agrupar les dades en una taula de freqüències com la següent:

Edat (*x*)	*f*	*fx*
20	3	3 * 20 = 60
22	1	1 * 22 = 22
23	1	23
25	2	50
28	1	28
30	4	120
31	1	31
	$\Sigma f = 13 = N$	$\Sigma fx = 334$

$$\overline{X} = \frac{\Sigma fx}{N} = \frac{334}{13} = 25{,}7$$

1.2.2. Mediana

Ens indica quin és el punt mitjà d'una distribució. Per sota de la mediana se situa exactament la meitat de la distribució, i per sobre, l'altra meitat.

En l'exemple, següent, on tenim un número senar de valors, el valor central de la distribució és 25.

20, 20, 20, 22, 23, 25, **25**, 28, 30, 30, 30, 30, 31

En 6 casos tenen edats inferiors a 25, i en 6 més tenen edats superiors a 25. Si la distribució tingués un número parell de valors, la mediana se situaria entre els valors 25 i 28. En aquest cas, es calcula la mitjana d'aquests dos valors 25 + 28 = 53

Mdn = 53/2 = 26,5

$$\underbrace{20, 20, 20, 22, 23, 25, \mathbf{25},}_{\text{7 persones}}\ \underbrace{\mathbf{28}, 30, 30, 30, 30, 31, 32}_{\text{7 persones}}$$

Cal calcular el lloc de la mediana = (N + 1) / 2. En aquest cas: (14 + 1) / 27,5.

El valor 7,5 no és la Mdn. Només ens indica a quina posició de la distribució ordenada la trobarem. Per tant, en aquest cas, hem de buscar quin valor ocupa la posició 7,5. La Mdn és, per tant, i com ja hem dit, 26,5.

Quan tenim les dades agrupades en freqüències, el que fem és buscar el lloc de la mediana.

Edat (x)	f	fa
20	3	3
22	1	4
23	1	5
25	2	7
28	1	8
30	4	12
31	1	13
	$\Sigma f = 13 = N$	

Lloc de la *Mdn* = (13 + 1) / 2 = 7

Quin valor ocupa la posició número 7? Si mirem les freqüències acumulades veiem que la posició 7 està ocupada pel valor 25. Per tant, *Mdn* = 25.

1.2.3. Moda

La moda és el valor que es presenta amb més freqüència en una distribució. Per exemple, en la distribució anterior, la moda és el valor 30, perquè hi ha 4 casos que tenen 30 anys i és, per tant, el valor més repetit en aquesta distribució.

Tal com diuen Welkowitz, Ewen i Cohen (1981), la moda és una mesura descriptiva que indica *grosso modo* la localització d'una distribució, ja que ignora una part substancial de les dades, per això no s'utilitza normalment en investigació.

Com més gran sigui la mostra que estudiem i com menys esbiaixada sigui, més properes entre si es troben les mesures de tendència central.

Exercici 1

S'han de buscar les mesures de tendència central de les notes del primer quadrimestre de l'assignatura Bases metodològiques de la investigació educativa. Tenim la taula de freqüències següent:

(Valors de la variable) ***x***	***f***
1	1
3	4
4	6
5	8
6	16
7	17
8	9
9	1
	Σ*f* = 62

Mitjana aritmètica:

Hem de fer servir la fórmula (2) perquè tenim les dades agrupades en freqüències:

(Valors de la variable) ***x***	***f***	***fx***
1	1	1
3	4	12
4	6	24
5	8	40
6	16	96
7	17	119
8	9	72
9	1	9
	Σ*f* = 62	Σ*fx* = 373

$$\overline{X} = \frac{\Sigma fx}{N} = \frac{373}{62} = 6{,}02$$

Per tant, la mitjana de puntuació obtinguda és 6,02. Ens indica la localització de la distribució i ens ajuda a conèixer la tendència central de les notes del primer quadrimestre, deixant de banda, en aquest cas, les persones no presentades.

Mediana

(Valors de la variable) x	***f***	***fa***
1	1	1
3	4	5
4	6	11
5	8	19
6	16	35
7	17	52
8	9	61
9	1	62
	Σf=62	

El lloc de la mediana el buscarem calculant (*N* + 1) / 2. En aquest cas serà (62+1) / 2 = 31,5.

Localitzarem la mediana, doncs, a la posició 31,5. Si mirem les freqüències acumulades, veiem que de la posició 20 a la posició 35 el valor que trobem és 6. Si despleguem part de la distribució anterior a la taula següent, podem observar que la posició 31 està ocupada pel valor 6 i la posició 32 també està ocupada pel valor 6. Per tant, el valor de la mediana és 6.

Posició	18	19	20	21	22	23	24	25	26	27	28	29	30	**31**	**32**	33	34	35	36
Valor (*X*)	5	5	6	6	6	6	6	6	6	6	6	6	6	**6**	**6**	6	6	6	7

Mdn = 6

Per tant, sabem que la meitat, és a dir, el 50 % dels casos d'aquesta distribució, ha obtingut una puntuació de 6 o menys i que l'altra meitat ha obtingut puntuacions de 6 o més.

Moda

La moda és el valor més repetit. En aquest exemple, la moda és 7.

Moda = 7

1.3. Els índexs o mesures de dispersió

La dispersió o variabilitat, tal com expliquen Welkowitz, Ewen i Cohen (1981:76), “es refereix al grau de dispersió-separació entre les dades d’una distribució, és a dir, fins a quin punt s’assemblen entre si els valors de la distribució.”

Imaginem-nos els quatre conjunts de dades següents:

Distribució A:	10	10	10	10	10	10	10
Distribució B:	9,8	9,8	9,9	10	10,1	10,2	10,2
Distribució C:	7	8	9	10	11	12	13
Distribució D:	4	6	8	10	12	14	16

En totes la mitjana és 10. No obstant això, no en fem prou amb la mitjana per descriure aquestes distribucions. Tot i que la mitjana és 10, observem que en les distribucions A i B tots els valors es troben molt agrupats entorn de la mitjana, mentre que en les distribucions C i D els valors no estan tan agrupats, i és la distribució D la que presenta més variabilitat o dispersió.

Les mesures de dispersió permeten analitzar fins a quin punt les característiques que es mesuren varien d’unes persones a les altres, permeten analitzar i quantificar les diferències individuals.

Les mesures de dispersió que estudiarem són l’amplitud, la variància, la desviació típica i el coeficient de variació.

1.3.1. Amplitud o rang de la distribució

És la diferència entre la puntuació més gran i la puntuació més petita de la distribució, és a dir, entre el valor mínim i el valor màxim. En el cas de les quatre distribucions anteriors, els rangs són:

Distribució A: mitjana = 10, rang = 0
Distribució B: mitjana = 10, rang = 0,4
Distribució C: mitjana = 10, rang = 6
Distribució D: mitjana = 10, rang = 12

Cal tenir en compte que l’amplitud és la mesura més pobra de la dispersió, ja que en algunes distribucions ens podem trobar que els valors extrems no són típics de la variació total d’una distribució. Així, per exemple, observem les dues distribucions següents:

Distribució E:1, 1, 3, 4, 5, 5, 6, 7, 8, 10, 11, 11, 11
Distribució F: 1, 5, 5, 6, 6, 6, 6, 6, 6, 6, 7, 7, 11

L'amplitud és 10 en ambdós casos, però, en canvi, en la distribució F la dispersió és molt menor que en la distribució E. És per això que l'amplitud és poc utilitzada en investigació.

1.3.2. Variància i desviació típica

Igual com la mitjana aritmètica, és un valor mitjà que pren en consideració totes les observacions; es tracta de buscar una mesura de dispersió que també tingui en compte totes les observacions o puntuacions de la distribució. Una de les formes més utilitzades per mesurar la variabilitat, doncs, és mesurar les distàncies de cada puntuació a un punt central de la distribució, que normalment és la mitjana aritmètica. És a dir, es tracta de calcular una dispersió mitjana basada en tots els valors observats (Welkowitz, Ewen i Cohen, 1981).

$$\text{Desviació} = X - \overline{X}$$

Els valors molt allunyats de la mitjana tindran desviacions numèricament grans, mentre que els valors pròxims a la mitjana tindran desviacions numèricament petites.
Per obtenir la mitjana de desviacions, la forma més usual seria calcular

$$\frac{\Sigma\,(X - \overline{X})}{N}$$

Recordem, però, que la suma de totes les desviacions respecte de la mitjana és igual a 0. Per resoldre aquest problema, el que es fa és elevar al quadrat cada desviació abans de sumar-les. Així, la suma de quadrats de totes les desviacions respecte de la mitjana, dividit pel nombre total d'observacions fetes, serà una mesura de dispersió que s'anomena *variància* i que se simbolitza amb una S^2:

$$S^2 = \frac{\Sigma(X - \overline{X})^2}{N}$$

Cal tenir en compte, però, que, en elevar les puntuacions al quadrat, la variància és un índex de dispersió que també està elevat al quadrat. És a dir, si calculem la variància d'una distribució constituïda per valors corresponents a la variable "alçada", la mesura que s'obté és un valor d'alçada al quadrat. Si l'hem mesurat en centímetres, la mesura serà en centímetres quadrats. Per resoldre aquesta qüestió, es busca l'arrel quadrada de la variància. Amb això obtenim una altra mesura de dispersió molt més útil i utilitzada, que és la desviació típica i que se simbolitza amb una S.

$$S = \sqrt{S^2}$$

$$S = \sqrt{\frac{\Sigma(X - \overline{X})^2}{N}}$$

La fórmula de càlcul que s'utilitza, però, és la que surt de desenvolupar aquesta. Per tant, és la següent.

Quan les dades no estan agrupades:

Variància

$$S^2 = \frac{1}{N}\left[\Sigma x^2 - \frac{(\Sigma x)^2}{N}\right] \quad (3)$$

Desviació típica

$$\boxed{S = \sqrt{S^2}} \quad (4)$$

Quan les dades estan agrupades en taules de freqüències:

$$S^2 = \frac{1}{N}\left[\Sigma fx^2 - \frac{(\Sigma fx)^2}{N}\right] \quad (5)$$

1.3.3. Coeficient de variació

A vegades ens interessa comparar la variabilitat de dos conjunts de dades o de dues variables. Quan el conjunt de puntuacions o les dues variables estan mesurades amb la mateixa escala, es comparen les desviacions típiques o variàncies, sabent que, com més gran sigui la desviació típica, més gran serà la variabilitat del conjunt de dades. Ara bé, quan les variables estan mesurades amb escales diferents, no es pot fer aquesta comparació directa. Imaginem-nos que un professor puntua una prova del 0 al 10 i un altre la puntua del 0 al 100.

El grup d'estudiants del primer professor té les puntuacions següents:

5, 5, 4, 6, 7, 4, 8, 3 Mitjana = 5,25 S = 1,56

Les puntuacions obtingudes pels estudiants del segon professor són:

30, 30, 24, 36, 42, 24, 48, 18 Mitjana = 31,5 S = 9,36

Sembla, doncs, que les puntuacions del grup del segon professor estan més disperses. Però no podem fer la comparació directament perquè les puntuacions s'han obtingut amb escales de mesura diferents. Per resoldre aquesta qüestió s'utilitza el coeficient de variació:

$$CV = \frac{S}{\overline{X}} * 100 \qquad (6)$$

Per cada grup de puntuacions el coeficient de variació serà la desviació típica dividida per la mitjana i multiplicada per 100:

CVa = (1,56/5,259) *100 = 30

CVb = (9,36/31,5)*100 = 30

Observem, doncs, que el valor del CV és el mateix. Per tant, la variabilitat d'ambdós grups de puntuacions és exactament la mateixa.

Exercici 2

S'han de buscar les mesures de dispersió de les notes del primer quadrimestre de Bases metodològiques de la investigació educativa.

<u>Amplitud</u>

En aquest cas l'amplitud és 9-1 = 8. Aquest és el rang de la distribució.

<u>Variància</u>

Hem d'aplicar la fórmula (5), perquè tenim les dades agrupades en freqüències:

(Valors de la variable) ***x***	***f***	***fx***	***x²***	***fx²***
1	1	1	1	1
3	4	12	9	36
4	6	24	16	96
5	8	40	25	200
6	16	96	36	576
7	17	119	49	833
8	9	72	64	576
9	1	9	81	81
	$\Sigma f = 62$	$\Sigma fx = 373$		$\Sigma fx^2 = 2399$

$$S^2 = \frac{1}{N}\left[\Sigma fx^2 - \frac{(\Sigma fx)^2}{N}\right] =$$

$$= \frac{1}{62}\left[2399 - \frac{(373)^2}{62}\right] =$$

$$= \frac{1}{62}\left[2399 - 2244{,}02\right] = 2{,}5$$

<u>Desviació típica</u>

$S = \sqrt{S^2} = \sqrt{2{,}5} = 1{,}6$

<u>Coeficient de variació</u>

La professora de l'assignatura, per calcular la nota final del quadrimestre, té en compte els resultats de diferents exercicis. Un dels exercicis que fa és una prova individual que puntua del 0 al 6. Ha buscat la mitjana de les puntuacions d'aquest exercici per als 62 estudiants de la mostra i surt un resultat de 3,5. També ha buscat la desviació típica, i ha obtingut una desviació d'1,1. Es pregunta si hi ha més variabilitat en la distribució global de les notes del quadrimestre (mitjana = 6,02 i S = 1,6) o en les puntuacions de la prova individual.

Com que les distribucions corresponen a escales de mesures diferents, cal que busquem el coeficient de variació.

Aplicant la fórmula (6) tenim que:

CV de les notes del primer quadrimestre = (1,6 / 6,02) * 100 = 26,6

CV de les puntuacions de la prova individual = (1,1 / 3,5) * 100 = 32,4

Per tant, observem que, tot i que aparentment hi havia més dispersió en les notes finals del quadrimestre perquè la desviació típica era més alta, *un cop aplicat el CV veiem que realment hi ha més dispersió en les puntuacions de la prova individual que en les puntuacions finals del quadrimestre.*

1.4. Exercicis d'estadística descriptiva

A continuació hi ha un resum d'una recerca de la qual podeu trobar un article a la referència següent:

SÁNCHEZ, I.; PONT, A.; MAGRE, J.; PARÈS, A. (1994): "Condicions de vida i valors dels joves de Barcelona". *Revista d'Informació i Estudis Socials. Barcelona Societat*, 1, pàg. 24-51.

Llegeix aquest resum, que et permetrà situar-te en una recerca en concret. Imagina't que tenim les dades que han donat 30 persones a través d'unes quantes preguntes d'un qüestionari. A continuació es presenta el resum de la recerca i la matriu de dades. Sobre la base d'aquestes dades respon a les qüestions següents:

1.1. Resumeix les variables gènere, nombre de germans i estudis acabats, en taules de freqüències en les quals apareguin les freqüències absolutes, les freqüències absolutes acumulades, els percentatges i els percentatges acumulats. Quants casos hi ha que tinguin dos germans o menys?

1.2. Calcula els índexs de tendència central i els índexs de dispersió de les variables nombre de germans, ingressos i nombre de persones actives a la llar.

1.3. Explica breument a quines conclusions podem arribar una vegada feta aquesta petita anàlisi amb aquests 30 casos.

Resum de la recerca

SÁNCHEZ, I.; PONT, A.; MAGRE, J.; PARÈS, A. (1994): "Condicions de vida i valors dels joves de Barcelona". *Revista d'Informació i Estudis Socials. Barcelona Societat*, 1, pàg. 24-51.

◊ Origen de la recerca

L'Ajuntament de Barcelona fa l'encàrrec a l'equip de l'Àrea d'Estudis d'Opinió d'ICB, SA (Informació Cartogràfica i de Base).

◊ Objectiu de l'estudi

"Descriure els aspectes que conformen la vida quotidiana dels joves i el seu entorn més immediat i esbrinar les actituds que posen de manifest davant de la societat en general i de la política en particular" (pàg. 24-25)

◊ Població:

Joves de Barcelona d'edats compreses entre els 15 i els 25 anys.

◊ Mostra

Es va seleccionar una mostra de 1.200 joves de Barcelona d'entre 15 i 25 anys. La mostra va ser dissenyada segons quotes de sexe i edat.

◊ Mètode

Enquesta

◊ Instrument

Es va elaborar un qüestionari amb preguntes sobre diversos temes:

- Origen dels joves (edat, lloc de naixement, lloc de naixement dels pares, llengua habitual...)
- Lloc de residència (domicili familiar, sols, en parella...) i satisfacció pel que fa al lloc de residència

- Els estudis
- Inserció al mercat de treball
- Recursos econòmics
- L'ocupació del temps lliure
- La parella i les relacions sexuals
- El consum de tabac, alcohol i altres drogues
- Els valors i les actituds davant del món (sentiment religiós, problemàtica social, intolerància i permissivitat, distància social i discriminació, confiança en les institucions, compromís social, actituds polítiques...)

◊ Recollida de dades

Es va passar un qüestionari a 1.200 joves durant els mesos de novembre i desembre del 1992. L'enquesta es va fer per domicilis seleccionats aleatòriament.

◊ Anàlisi de les dades:

Anàlisi estadística.

◊ Conclusions:

"Els joves entre 15 i 25 anys es troben encara a mig camí en el seu procés de transició cap a la vida adulta: viuen amb els pares, en depenen econòmicament i els estudis representen la seva activitat principal. Encara no s'han plantejat un canvi en la seva situació, tot i que aquest va apareixent a mesura que es fan grans, que acaben els estudis i s'integren al món laboral. Mostren una gran confiança en la institució familiar, estan satisfets de la seva família i aposten per ella. En les seves relacions externes els amics ocupen el primer lloc i amb ells els agrada compartir el temps d'oci. A mesura que es fan grans la parella i l'amor van adquirint importància.

"Es mostren més tolerants amb les qüestions relatives a l'àmbit privat de l'individu que amb les de caràcter públic. Són amants de la llibertat, però sense supeditar-la a la renúncia de les normes socials que garanteixin la convivència. Estan preocupats pels problemes socials, però se'ls fa difícil materialitzar les seves inquietuds en alguna activitat o organització concreta. Els partits polítics no apareixen, entre els joves, com a instruments aglutinadors d'interessos. Estan d'acord que la democràcia és el millor sistema polític, però en canvi mostren desconfiança en algunes institucions bàsiques que la sustenten.

"Sembla, doncs, que sota una actitud d'aparent indiferència hi ha una capacitat de reflexió i adhesió a la realitat que els toca viure de prop. Ni l'aclaparament ni l'apatia formen part del seu tarannà sinó que intenten fer-se el seu propi espai en la societat. Els joves d'avui estableixen un ordre de prioritats i abans que res volen i estan preparant-se per ser adults. La seva eina principal és l'estudi. D'aquesta manera demostren un gran coneixement del joc que tenen entre les mans i una serietat soterrada i allunyada, en la majoria dels casos, dels exabruptes que hom atribueix normalment a l'etapa juvenil" (p. 50-51).

Matriu de dades

1. A continuació apareix una possible matriu de dades de 30 casos corresponent a aquest estudi. (Les dades no són les que s'han obtingut realment sinó que ens basem en aquest plantejament de recerca per treballar amb una matriu fictícia.)

cas	Any de naixement	Gènere	Nombre de germans	Lloc de residència habitual	Ingressos familiars	Estudis acabats	Activitat principal	Nombre de persones actives a la llar	T'has plantejat marxar de casa?	Et preocupa independitzar-te?	Com viuries si poguessis escollir ?
1	68	Dona	3	Amb els pares	1.500	BUP	Estudio i treballo	3	Sí	Poc	Amb la parella
2	69	Dona	1	Amb la parella	1.350	Universitaris grau superior	Només estudio	2	No	Gens	Amb la parella
3	70	Dona	2	Amb la parella	1.200	BUP	Estudio i treballo	1	No	Gens	Amb la parella
4	68	Dona	1	Amb amics	840	Universitaris grau mitjà	Treballo i estudio	1	No	Gens	Amb els amics
5	75	Dona	1	Amb els pares	1.620	Primaris	Ni estudio ni treballo	2	Sí	Poc	Amb la parella
6	74	Dona	2	Amb els pares	900	BUP	Ni estudio ni treballo	2	Sí	Gens	Sol/a
7	74	Dona	1	Amb els pares	1.200	Universitaris grau mitjà	Estudio i treballo	2	Sí	Bastant	Sol/a
8	73	Dona	1	Amb els pares	1.800	Primaris	Treballo i estudio	2	Sí	Poc	Amb els pares
9	70	Dona	3	Amb la parella	1.320	Formació professional	Treballo i estudio	2	No	Gens	Amb la parella
10	69	Dona	1	Amb amics	1.680	BUP	Només estudio	3	No	Gens	Amb la parella
11	68	Dona	1	Amb la parella	2.400	Universitaris grau superior	Només treballo	2	No	Gens	Amb els pares
12	70	Dona	2	Amb els pares	1.620	Formació professional	Només treballo	3	Sí	Gens	Amb la parella
13	75	Dona	2	Amb els pares	1.020	BUP	Només treballo	1	No	Poc	Amb els amics
14	75	Dona	2	Amb els pares	1.920	Formació professional	Estudio i treballo	2	No	Bastant	Sol/a
15	69	Dona	0	Amb els pares	960	Universitaris grau mitjà	Només estudio	1	Sí	Poc	Amb la parella
16	68	Home	0	Sol/a	720	Formació professional	Només estudio	1	Sí	Poc	Amb els pares
17	69	Home	0	Sol/a	810	BUP	Treballo i estudio	1	No	Gens	Amb la parella
18	70	Home	0	Amb la parella	1.410	Universitaris grau superior	Només treballo	2	No	Gens	Amb els amics
19	71	Home	1	Amb amics	750	Universitaris grau mitjà	Treballo i estudio	1	Sí	Bastant	Amb la parella
20	72	Home	1	Amb els pares	1.080	Formació professional	Estudio i treballo	2	Sí	Bastant	Amb els pares
21	74	Home	1	Amb els pares	1.140	Formació professional	Estudio i treballo	1	No	Molt	Amb els pares
22	74	Home	2	Amb els pares	1.800	BUP	No contesta	3	No	Molt	Amb la parella
23	71	Home	2	Amb els pares	2.100	BUP	No contesta	3	Sí	Molt	Amb els pares
24	71	Home	1	Amb els pares	810	BUP	Treballo i estudio	1	Sí	Bastant	Amb els pares
25	72	Home	1	Amb amics	2.520	Universitaris grau mitjà	No contesta	3	No	Bastant	Amb la parella
26	75	Home	0	Amb els pares	1.020	Primaris	No contesta	1	No	Molt	Amb la parella
27	69	Home	0	Amb els pares	1.500	Primaris	Ni estudio ni treballo	2	Sí	Poc	Amb els amics
28	70	Home	0	Sol/a	900	BUP	Ni estudio ni treballo	1	Sí	Gens	Amb els pares
29	73	Home	0	Sol/a	1.020	Primaris	Ni estudio ni treballo	2	No	Gens	Amb la parella
30	73	Home	0	Amb la parella	1.200	Primaris	No contesta	2	No	Gens	Amb la parella

2. A continuació tens una taula extreta de l'*Estadística dels joves*[1] *de Catalunya 2002*. Hi consta el nombre de joves estudiants segons la previsió d'acabament dels estudis. Respon a les preguntes següents:

   2.1. Busca els percentatges acumulats per al total de la població.
   2.2. Què vol dir que el percentatge acumulat corresponent al valor "en 3 anys" sigui del 60,5 %?
   2.3. Quina és la mitjana d'anys en què els joves del 2002 tenien previst finalitzar els seus estudis? (Fixa't que la variable "nombre d'anys previstos" es pot considerar una variable mesurada en escala d'interval, si substituïm el valor "1 o menys" per 1 i el "4 anys o més" per 4.)
   2.4. Amb tot, seria millor utilitzar un altre índex de tendència central, ja que no tenim els valors exactes. Quin seria aquest índex? Calcula'l per al cas dels homes i per a les dones i digues què significa.

Nombre de joves estudiants segons la previsió d'acabament dels seus estudis. 2002

	Homes		Dones		Total	
	Població	%	Població	%	Població	%
En 1 any o menys	92.338	31,4	105.206	31,8	197.544	31,6
En 2 anys	50.902	17,3	44.050	13,3	94.953	15,2
En 3 anys	40.374	13,7	45.092	13,6	85.467	13,7
En 4 anys o més	110.578	37,6	136.203	41,2	246.781	39,5
Total	294.192	100,0	330.551	100,0	624.744	100,0

Font: Observatori Català de la Joventut i IDESCAT (2002): *Estadística dels joves de Catalunya 2002*. Generalitat de Catalunya. Departament de Presidència. Secretaria General de Joventut.

3. A partir de la taula que tens a continuació, extreta del Departament de Benestar i Família, respon a les preguntes següents:

   3.1. Quin és el percentatge de persones amb discapacitat de les comarques de Girona respecte del total de Catalunya?
   3.2. Quina és la mitjana del nombre de persones amb discapacitat psíquica per comarca?
   3.3. I quina és la mitjana del nombre de persones amb discapacitat per comarca?

[1] Per fer aquest estudi s'ha treballat amb una mostra de joves d'entre 15 i 29 anys.

Persones amb discapacitat per tipologia. Comarques. Xifres absolutes. Juny 2005

	Físics						Malalts	No	Total
Comarca	Motòrics	no motòrics	Total	Visuals	Auditius	Psíquics	mentals	consta	discapacitats
Alt Camp	613	448	1.061	107	55	244	260	6	1.733
Alt Empordà	1.410	802	2.212	231	128	765	470	15	3.821
Alt Penedès	1.350	849	2.199	290	159	480	600	8	3.736
Alt Urgell	302	215	517	40	27	95	128	4	811
Alta Ribagorça	66	51	117	10	9	17	32	0	185
Anoia	1.567	1.440	3.007	364	176	592	776	16	4.931
Bages	3.221	2.245	5.466	784	363	821	1.730	11	9.175
Baix Camp	2.353	1.843	4.196	520	314	1.750	1.304	34	8.118
Baix Ebre	1.123	1.030	2.153	242	120	442	546	8	3.511
Baix Empordà	1.404	782	2.186	212	161	754	488	12	3.813
Baix Llobregat	12.276	11.464	23.740	2.658	1.531	3.662	6.250	28	37.869
Baix Penedès	1.022	711	1.733	173	130	376	412	7	2.831
Barcelonès	43.014	39.471	82.485	12.400	6.102	11.668	24.050	278	136.983
Berguedà	1.528	1.256	2.784	375	188	195	629	2	4.173
Cerdanya	130	65	195	28	15	81	49	0	368
Conca de Barberà	279	211	490	56	21	136	128	3	834
Garraf	1.611	1.240	2.851	392	183	494	690	7	4.617
Garrigues	390	260	650	90	51	108	146	2	1.047
Garrotxa	612	276	888	118	57	401	202	2	1.668
Gironès	2.561	1.278	3.839	386	259	1.207	1.041	12	6.744
Maresme	4.178	3.357	7.535	955	598	1.684	2.172	38	12.982
Montsià	976	746	1.722	137	109	362	390	7	2.727
Noguera	768	457	1.225	128	104	270	256	3	1.986
Osona	1.673	1.255	2.928	459	220	741	985	6	5.339
Pallars Jussà	260	133	393	65	32	72	117	2	681
Pallars Sobirà	114	52	166	16	15	27	24	2	250
Pla d'Urgell	652	440	1.092	110	72	211	234	5	1.724
Pla de l'Estany	336	159	495	46	23	227	126	3	920
Priorat	151	82	233	22	15	79	53	3	405
Ribera d'Ebre	295	276	571	55	34	129	113	3	905
Ripollès	372	199	571	70	60	206	151	3	1.061
Segarra	322	183	505	63	29	99	140	0	836
Segrià	4.619	3.244	7.863	887	632	1.481	1.982	23	12.868
Selva	1.954	994	2.948	287	178	844	577	16	4.850
Solsonès	198	99	297	40	33	87	92	0	549
Tarragonès	3.366	2.545	5.911	775	407	1.440	1.520	34	10.087
Terra Alta	158	116	274	31	11	61	62	1	440
Urgell	575	324	899	118	69	208	228	2	1.524
Val d'Aran	99	45	144	14	9	51	27	0	245
Vallès Occidental	11.453	9.009	20.462	2.694	1.853	4.493	5.867	120	35.489
Vallès Oriental	4.827	3.352	8.179	1.015	644	1.676	2.152	21	13.687
No consta comarca	4	1	5	0	0	0	0	0	5
Total Catalunya	**114.182**	**93.005**	**207.187**	**27.463**	**15.196**	**38.736**	**57.199**	**747**	**346.528**

Font: Departament de Benestar i Família. Secretaria General. Elaboració a partir de la Base de dades de persones amb discapacitat.

4. A partir de les dades de la taula següent, respon a les qüestions que es plantegen a continuació. Pensa que per contestar cada pregunta només has d'utilitzar les dades que siguin necessàries d'aquesta taula.

   4.1. Quantes persones tenen menys del 74 % de discapacitat a Catalunya?
   4.2. Quin percentatge representa aquest nombre respecte del total de discapacitats?
   4.3. Quin percentatge de persones tenen un grau de discapacitat del 75 % o més?
   4.4. Quin tant per cent de persones amb discapacitat representen les dones?

Persones amb discapacitat per grau i sexe. Comarques. Xifres absolutes. Juny 2005

	Del 33% al 64%			Del 65% al 74%			75% i més			Total discapacitats		
Comarca	Homes	Dones	Total	Homes	Dones	Total	Homes	Dones	Total	Homes	Dones	Total
Alt Camp	437	383	820	246	293	539	179	195	374	862	871	1.733
Alt Empordà	1.177	757	1.934	498	605	1.103	367	417	784	2.042	1.779	3.821
Alt Penedès	964	928	1.892	462	516	978	424	442	866	1.850	1.886	3.736
Alt Urgell	207	187	394	90	148	238	97	82	179	394	417	811
Alta Ribagorça	64	38	102	18	35	53	15	15	30	97	88	185
Anoia	1.329	1.303	2.632	570	719	1.289	485	525	1.010	2.384	2.547	4.931
Bages	2.526	2.686	5.212	1.092	1.193	2.285	735	943	1.678	4.353	4.822	9.175
Baix Camp	2.029	1.526	3.555	1.255	1.378	2.633	955	975	1.930	4.239	3.879	8.118
Baix Ebre	817	676	1.493	537	575	1.112	435	471	906	1.789	1.722	3.511
Baix Empordà	1.131	914	2.045	468	579	1.047	348	373	721	1.947	1.866	3.813
Baix Llobregat	11.445	9.598	21.043	4.446	5.613	10.059	3.171	3.596	6.767	19.062	18.807	37.869
Baix Penedès	792	570	1.362	421	532	953	251	265	516	1.464	1.367	2.831
Barcelonès	36.172	34.243	70.415	18.002	18.773	36.775	13.422	16.371	29.793	67.596	69.387	136.983
Berguedà	1.627	1.416	3.043	350	337	687	204	239	443	2.181	1.992	4.173
Cerdanya	101	82	183	50	54	104	41	40	81	192	176	368
Conca de Barberà	181	141	322	114	173	287	102	123	225	397	437	834
Garraf	1.172	1.160	2.332	594	700	1.294	453	538	991	2.219	2.398	4.617
Garrigues	264	291	555	70	172	242	115	135	250	449	598	1.047
Garrotxa	554	355	909	206	208	414	185	160	345	945	723	1.668
Gironès	2.048	1.597	3.645	864	898	1.762	624	713	1.337	3.536	3.208	6.744
Maresme	3.391	3.285	6.676	1.663	1.770	3.433	1.381	1.492	2.873	6.435	6.547	12.982
Montsià	529	597	1.126	348	599	947	309	345	654	1.186	1.541	2.727
Noguera	592	464	1.056	172	325	497	206	227	433	970	1.016	1.986
Osona	1.561	1.369	2.930	610	634	1.244	571	594	1.165	2.742	2.597	5.339
Pallars Jussà	201	182	383	63	86	149	74	75	149	338	343	681
Pallars Sobirà	71	67	138	24	37	61	20	31	51	115	135	250
Pla d'Urgell	479	500	979	133	294	427	139	179	318	751	973	1.724
Pla de l'Estany	261	196	457	119	114	233	103	127	230	483	437	920
Priorat	101	71	172	58	81	139	30	64	94	189	216	405
Ribera d'Ebre	198	182	380	133	151	284	109	132	241	440	465	905
Ripollès	295	274	569	125	161	286	99	107	206	519	542	1.061
Segarra	208	231	439	80	126	206	78	113	191	366	470	836
Segrià	3.837	3.441	7.278	1.300	1.855	3.155	1.147	1.288	2.435	6.284	6.584	12.868
Selva	1.437	1.205	2.642	549	747	1.296	430	482	912	2.416	2.434	4.850
Solsonès	132	153	285	59	88	147	57	60	117	248	301	549
Tarragonès	2.621	2.095	4.716	1.408	1.815	3.223	986	1.162	2.148	5.015	5.072	10.087
Terra Alta	113	84	197	47	76	123	55	65	120	215	225	440
Urgell	408	394	802	145	243	388	155	179	334	708	816	1.524
Val d'Aran	92	54	146	25	27	52	22	25	47	139	106	245
Vallès Occidental	10.522	9.497	20.019	3.875	5.169	9.044	3.062	3.364	6.426	17.459	18.030	35.489
Vallès Oriental	3.647	3.377	7.024	1.683	2.298	3.981	1.315	1.367	2.682	6.645	7.042	13.687
No consta comarca	2	1	3	0	2	2	0	0	0	2	3	5
Total Catalunya	**95.735**	**86.570**	**182.305**	**42.972**	**50.199**	**93.171**	**32.956**	**38.096**	**71.052**	**171.663**	**174.865**	**346.528**

Font: Departament de Benestar i Família. Secretaria General. Elaboració a partir de la Base de dades de persones amb discapacitat.

5. A la taula següent es presenta la distribució per grups d'edat de persones amb discapacitat que constaven a la comarca del Gironès el juny del 2005. Respon a les preguntes següents:

   5.1. Quin percentatge de persones amb discapacitat de la comarca del Gironès representen les persones que tenen entre 6 i 15 anys?
   5.2. Quin percentatge de persones amb discapacitat representen les dones que tenen més de 65 anys?
   5.3. Quantes persones de menys de 19 anys de la comarca del Gironès tenen una discapacitat? Quin percentatge representa aquest nombre respecte del total de les persones amb discapacitat de la comarca el juny del 2005?

5.4. Quin interval d'edat constitueix la moda d'aquesta distribució?
5.5. Quin interval d'edat constitueix la moda per als homes?
5.6. Quin interval d'edat constitueix la moda per a les dones?

Grup d'edat	Nombre de persones amb discapacitat. Comarca del Gironès. Juny 2005		
	Homes	**Dones**	**Total**
O-3 anys	19	16	35
4-5 anys	23	13	36
6-15 anys	234	137	371
16-17	61	34	95
18-19	58	46	104
20-34	552	327	879
35-44	536	372	908
45-54	588	482	1070
55-64	561	523	1084
65-74	436	511	947
75 anys o més	468	747	1215
Total	**3536**	**3208**	**6744**

Font: Departament de Benestar i Família. Secretaria General. Elaboració a partir de la base de dades de persones amb discapacitat.

6. Es tenen les dades de dos centres d'acollida de la Direcció General d'Atenció a la Infància sobre el nombre de nens i nenes que han atès cada any durant el període del 1995 al 2004. Les dades són les següents:

	1995	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**
Centre A	30	50	66	100	150	80	85	90	85	102
Centre B	73	70	84	105	98	80	60	71	64	135

(Dades fictícies)

6.1. Quants nens i nenes ha atès de mitjana cada centre durant aquests 10 anys?

6.2. A quin centre hi ha hagut més variabilitat quant al nombre d'infants atesos?

7. S'ha escollit una mostra d'universitaris als quals s'ha aplicat una escala d'ansietat, i s'han obtingut les puntuacions que apareixen a la taula següent. Busca els índexs de tendència central i els de dispersió d'aquesta distribució.

68	45	48	32	51	52	70	47	50	34	53
54	62	50	53	37	56	57	67	55	58	42
61	64	68	44	45	46	55	39	41	47	49
43	52	66	47	54	56	32	36	39	34	65
61	59	67	66							

8. Un tribunal qualifica dos educadors socials d'una oposició que consta de quatre exercicis amb les puntuacions següents:

Exercicis	**Opositor A**	**Opositor B**
Coneixements teòrics	9	7
Coneixements pràctics	6	5
Prova psicotècnica	7	10
Ordinador	10	4

A la convocatòria es diu que la plaça serà adjudicada a l'opositor que obtingui una mitjana superior, ponderada segons els barems següents: coneixements teòrics 1, coneixements pràctics 3, prova psicotècnica 6, ordinador 2.

8.1. A quin dels dos opositors correspon la plaça?

8.2. Què hauria passat si l'ordre de convocatòria hagués legislat que la mitjana fos simple en lloc de ponderada?

9. A continuació apareix una taula que exposa el perfil dels nois i noies internats en centres educatius per a menors de la Direcció General de Mesures Penals Alternatives i Justícia Juvenil (Departament de Justícia). Són dades del 1998.

9.1. Completa la taula amb els percentatges corresponents a cada variable de la taula 1.

9.2. Quina va ser la mitjana de l'edat dels menors que van ser internats en centres l'any 1998? A quin valor li correspon la mediana? (Per fer aquest exercici, assumim que n'hi ha 4 que tenien 18 anys tot i que no és ben bé així, ja que en tenien 18 o més.)

9.3. En l'exercici anterior hem fet una mica de trampa, perquè hem igualat tots els casos de 18 anys o més a 18. Això no es pot fer i, per tant, ens cal buscar una mesura de tendència central que sigui més adequada a la informació que tenim. En aquest cas, la mesura que podem buscar és la mediana, ja que per calcular-la no ens cal conèixer tots els valors de la distribució. Calcula, doncs, la mediana de l'edat dels menors que van ser internats en centres l'any 1998.

9.4. La taula 1.2 presenta el mateix tipus de dades, però corresponen a l'any 1999. Observant els percentatges, què podríem comentar?

9.5. Per a l'any 1999, quina és la mitjana de l'edat dels menors que van ser internats en centres en la jurisdicció de menors?

Taula 1.1: Perfil dels menors internats en la jurisdicció de menors:[2] casos finalitzats (1998)

Total casos finalitzats: 219

Programes	ICS[3]	ICC[4]	IC[5]	Total	%
Casos finalitzats	8	59	152	219	
Sexe					
Nois	8	54	143	205	
Noies	0	5	9	14	
Edat (1)					
12 anys	0	2	0	2	
13 anys	0	5	10	15	
14 anys	0	11	18	29	
15 anys	2	19	29	50	
16 anys	5	15	75	95	
17 anys	1	5	18	24	
18 anys o més	0	2	2	4	
Delictes (2)					
Contra la vida	0	0	5	5	
Lesions	2	18	36	56	
Contra la llibertat	1	6	12	19	
Contra la llibertat sexual	0	4	3	7	
Contra la intimitat	0	2	7	9	
Contra l'honor	0	0	0	0	
Contra el patrimoni i l'ordre socioeconòmics.	2	63	169	234	
Contra la seguretat col·lectiva	0	4	3	7	
Falsedat documental	0	0	0	0	
Contra l'Administració de justícia	0	0	1	1	
Contra l'ordre públic	0	6	7	13	
Altres delictes i faltes	0	2	7	9	
Total	5	105	250	360	

1. Es tracta de l'edat en el moment de l'inici del programa.
2. El nombre de delictes pot ser superior al de casos finalitzats, perquè hi ha casos en què s'intervé per més d'un delicte.

Font: Generalitat de Catalunya. Departament de Justícia.

[2] Cal tenir present que, a més de la jurisdicció de menors, la jurisdicció penal de justícia juvenil també pot imposar penes en centres de menors. En general són joves d'entre 16 i 22 anys.

[3] Internament de cap de setmana.

[4] Internament cautelar en centres.

[5] Internament en centres.

Taula 1.2: Perfil dels menors internats en la jurisdicció de menors: casos finalitzats (1999)

Programes	ICS	ICC	IC	ICT	Total	%
Casos finalitzats	7	79	112	1	199	
Sexe						
Nois	6	75	106	1	188	94,47
Noies	1	4	6	-	11	5,53
Edat (1)						
12 anys	-	5	2	-	7	3,52
13 anys	-	6	5	-	11	5,53
14 anys	2	17	10	-	29	14,57
15 anys	-	39	50	-	89	44,72
16 anys	4	8	37	1	50	25,13
17 anys	1	4	8	-	13	6,53
18 anys o més	-	-	-	-	-	-
Delictes (2)						
Contra la vida	-	2	1	-	3	0,76
Lesions	1	58	38	-	97	24,62
Contra la llibertat	-	17	13	-	30	7,61
Contra la integritat moral	-	-	-	-	-	-
Contra la llibertat sexual	-	16	8	-	24	6,09
Contra la intimitat	-	1	2	-	3	0,76
Contra l'honor	-	1	-	-	1	0,25
Contra les relacions familiars	-	-	-	-	-	-
Contra el patrimoni i l'ordre socioeconòmic	6	84	116	1	207	52,54
Contra patrimoni històric i medi ambient	-	-	-	-	-	-
Contra la seguretat col·lectiva	-	6	5	-	11	2,79
Falsedat documental	-	-	-	-	-	0,00
Contra l'Administració de justícia	-	-	3	-	3	0,76
Contra els drets fonamentals	-	-	-	-	-	-
Contra l'ordre públic	-	4	9	-	13	3,30
Altres delictes i faltes	-	-	2	-	2	0,51
No consta el delicte o la falta	-	-	-	-	-	-
Total	7	189	197	1	394	

1. Es tracta de l'edat en el moment de l'inici del programa.
2. El nombre de delictes pot ser superior al de casos finalitzats perquè hi ha casos en què s'intervé per més d'un delicte.
3. No consta en l'ofici del jutjat i no s'ha pogut determinar durant el procés d'intervenció.

Font: Generalitat de Catalunya. Departament de Justícia.

10. A la taula 1.3 tens el nombre de casos finalitzats dins del programa de Mediació i reparació penal (Departament de Justícia), distribuïts per edats i per tipus de delictes. Respon a les qüestions següents:

    10.1. Quin percentatge de casos finalitzats han estat homes i quin han estat dones?
    10.2. Quin percentatge dels casos finalitzats dins d'aquest programa correspon a delictes comesos contra el patrimoni i l'ordre socioeconòmic?
    10.3. Quin tant per cent de persones que han participat en aquest programa ho han fet per haver comès delictes de lesions?
    10.4. Quin percentatge de casos finalitzats correspon a persones de més de 20 anys?

Taula 1.3: Mediació i reparació penal: perfil dels casos finalitzats. Any 2000

Casos finalitzats	**68**
Sexe	**Nombre**
Homes	52
Dones	16
Edat (1)	**Nombre**
16-20 anys	28
21-25 anys	4
26-30 anys	4
31-35 anys	3
36-40 anys	9
4145 anys	2
46-50 anys	9
51-55 anys	1
56-60 anys	2
61 anys o més	5
No consta	1
Delictes (2)	**Nombre**
Contra la vida	-
Lesions	15
Contra la llibertat	16
Contra la integritat moral	-
Contra la llibertat sexual	1
Contra la intimitat	-
Contra l'honor	5
Contra les relacions familiars	-
Contra el patrimoni i l'ordre socioeconòmic	23
Contra el patrimoni històric i el medi ambient	-
Contra la seguretat col·lectiva	-
Contra l'ordre públic	2
Altres delictes i faltes	-
No consta el delicte o la falta	6
Total	69

1. Es tracta de l'edat en el moment de l'inici del programa.
2. El nombre de delictes pot ser superior al de casos finalitzats perquè hi ha casos en què s'intervé per més d'un delicte.

Font: Generalitat de Catalunya. Departament de Justícia. http://www.gencat.net/justicia/ciutadans/orientacio/mpenal/dades/mrp00.pdf. Consultat el 14 de febrer de 2006.

11. A continuació tens la taula 1.4 amb les dades corresponents al nombre d'alumnes matriculats a cada etapa educativa des del curs 1990-1991 fins al curs 1999-2000, per sectors públic i privat.

    11.1. Quin tipus de gràfic podríem fer per observar de manera clara l'evolució que han seguit les dades de matriculació, comparant els sectors públic i privat?

    11.2. Després de la taula 1.4 trobaràs tres gràfics referits a les dades de la taula. Mira-te'ls detingudament i fes-ne un breu comentari.

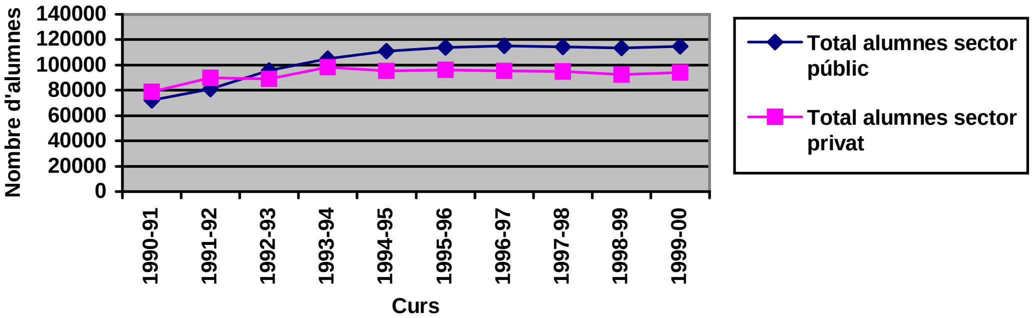
    11.3. Quina diferència hi ha entre els tres primers gràfics i el gràfic 1.4?

Taula 1.4: Dades de matriculació a les diferents etapes educatives (1990-2000). Sectors públic i privat

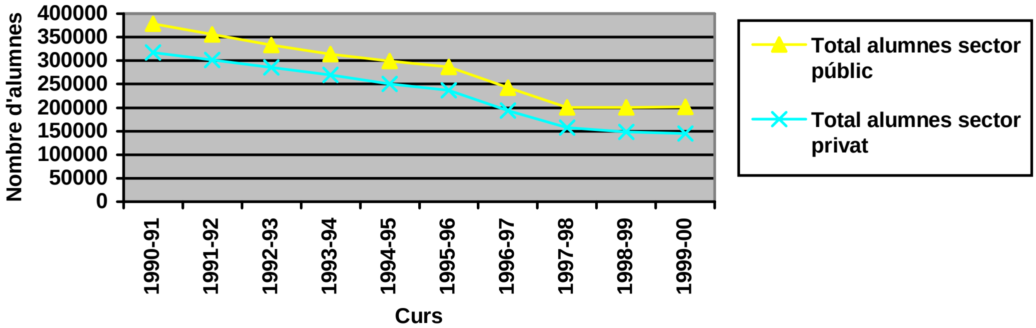
	Educació infantil		Educació primària		Educació secundària obligatòria	
Curs	Total alumnes sector públic	Total alumnes sector privat	Total alumnes sector públic	Total alumnes sector privat	Total alumnes sector públic	Total alumnes sector privat
1990-1991	72.239	78.714	378.484	317.269	7.166	3.279
1991-1992	80.911	89.967	356.024	301.448	8.688	3.454
1992-1993	95.639	88.811	332.942	285.551	11.937	3.663
1993-1994	104.920	98.383	313.907	269.300	15.980	5.421
1994-1995	110.707	95.120	298.758	250.884	26.169	8.841
1995-1996	114.039	96.152	286.848	236.452	41.030	16.251
1996-1997	115.110	95.221	242.671	193.731	95.609	53.844
1997-1998	114.438	94.804	201.126	157.450	140.260	89.195
1998-1999	113.554	92.437	200.772	148.758	146.538	105.625
1999-2000	114.753	94.067	202.020	145.437	147.950	118.517

Font: Departament d'Ensenyament.

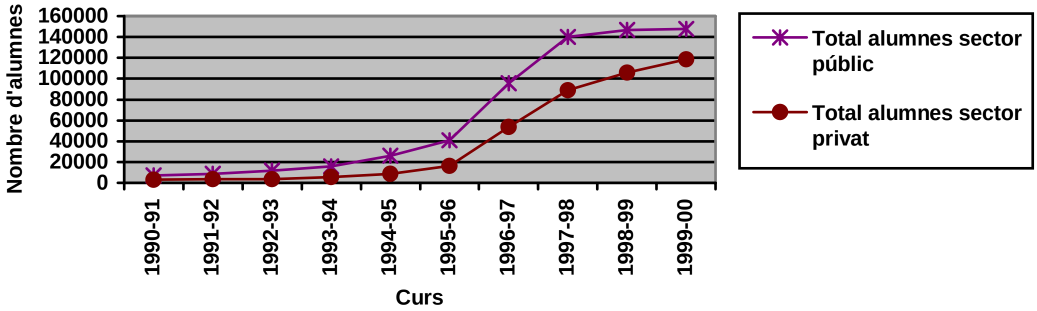
Gràfic 1.1: Total alumnat d'educació infantil (1990-2000) Sectors públic i privat



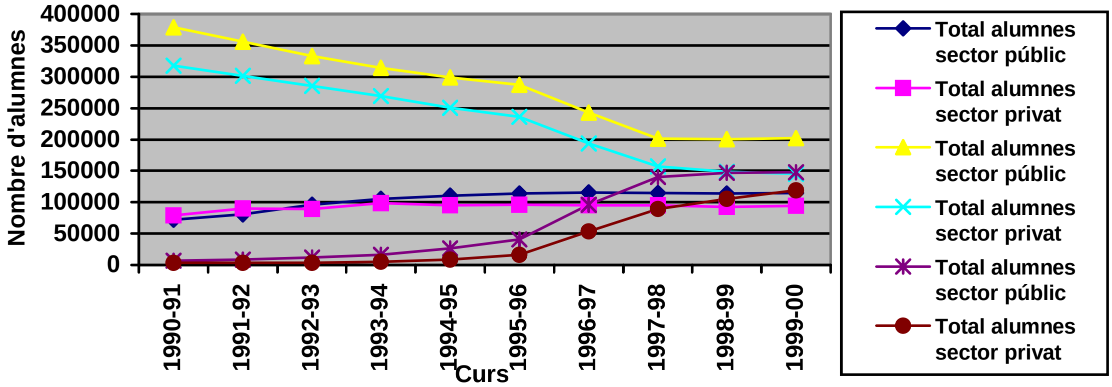
Gràfic 1.2: Total alumnat d'educació primària (1990-2000) Sectors públic i privat



Gràfic 1.3: Total alumnat d'educació secundària obligatòria (1990-2000) Sectors públic i privat



Gràfic 1.4: Evolució de la matrícula de les diferents etapes educatives (1990-2000) Sectors públic i privat



12. En una investigació per conèixer el procés d'inserció laboral dels diplomats en Educació Social per la UdG, es va passar un qüestionari a tots els diplomats des de l'any 1996 fins a l'any 2005. En una de les preguntes del qüestionari se'ls demanava que valoressin de l'1 al 7 la feina que feien en el moment de respondre. Concretament, la pregunta era:

Valora d'1 a 7 cadascuna de les qüestions presentades, sent 1 la mínima valoració i 7 la màxima (només en cas que estiguis treballant actualment)

	Gens d'acord					Totalment d'acord	
	1	2	3	4	5	6	7
El contingut de la feina és satisfactori.							
Les perspectives de millora i promoció en la feina actual són bones.							
El nivell de retribució et sembla satisfactori.							
Els coneixements derivats de la formació universitària t'han estat útils per a la feina.							
Les perspectives d'estabilitat són bones.							
La jornada laboral (horaris, torns de treball, etc.) és satisfactòria.							
Et sents realitzat com a educador social en aquesta feina.							

Les dades es van introduir en una matriu de dades del programa SPSSx i són les que apareixen a la taula següent (el "cas" és el número del qüestionari). A partir d'aquestes dades, respon a les preguntes següents:

12.1. Busca la taula de freqüències i els percentatges corresponents a les puntuacions de l'ítem "El contingut de la feina és satisfactori" (Valoració contingut).
12.2. Quina és la mediana d'aquesta variable?
12.3. Quin valor constitueix la moda?
12.4. Quin percentatge de persones valoren amb una puntuació de 4 o més que el contingut de la feina és satisfactori?
12.5. Busca les mitjanes de cada un dels ítems. Quin dels aspectes és, globalment, valorat d'una manera més positiva?
12.6. Quina de les set valoracions presenta més dispersió?

(Veure taula a la pàgina següent)

Cas	Valoració contingut (1-7)	Valoració perspectives millora (1-7)	Valoració retribució (1-7)	Valoració coneixements de formació (1-7)	Valoració perspectives estabilitat (1-7)	Valoració jornada laboral	Valoració realització personal (1-7)
1	7	5	6	4	6	7	7
2	6	3	6	5	7	6	7
3	5	6	5	3	5	5	.
4	6	4	3	6	6	6	6
5	7	4	6	4	6	7	1
6	7	6	6	3	7	6	7
7	7	4	6	5	6	7	7
8	4	1	1	3	7	5	2
9	5	2	3	6	6	6	4
10	6	4	4	5	6	6	6
11	4	4	1	3	6	1	2
12	4	2	4	2	6	6	5
13	3	4	6	4	7	6	3
14	5	3	3	3	6	4	.
15	6	6	6	5	7	7	7
16	5	2	3	2	7	6	5
17	6	4	4	6	3	2	5
18	1	2	5	1	7	7	.
19	7	4	6	1	1	2	4
20	6	3	6	5	6	7	6
21	5	5	5	1	7	6	1
22	6	5	6	4	4	6	7
23	7	6	3	7	6	7	7
24	7	5	.	6	4	7	7
25	4	2	6	5	2	4	3
26	6	1	5	2	7	5	7
27	.	.	.	.	.	.	.
28	6	4	5	2	5	4	6
29	6	5	5	5	6	5	6
30	5	7	5	4	7	3	6
31	6	3	1	5	2	2	5
32	6	7	5	7	7	7	3
33	6	3	6	1	6	7	1
34	6	6	3	3	7	5	6
35	6	6	7	4	5	4	6
36	6	5	6	5	5	5	5
37	4	3	4	6	6	6	5
38	5	3	5	4	4	2	4
39	7	4	1	7	4	1	.
40	7	6	2	2	7	6	1
41	2	3	6	5	7	5	1
42	7	1	3	7	1	7	7

43	7	4	4	4	4	6	7
44	6	2	4	7	7	5	6
45	7	3	7	5	7	5	7
46	6	4	3	2	6	4	5
47	6	4	6	3	6	6	6
48	.	.	.	.	.	.	.
49	6	5	6	5	6	6	6
50	5	6	6	4	4	7	6
51	7	7	7	7	7	7	7
52	7	2	5	7	2	5	6
53	7	6	7	5	5	7	7
54	6	5	7	4	2	5	6
55	6	6	4	6	7	6	6
56	6	3	7	5	7	7	7
57	6	5	5	5	6	6	5
58	6	5	7	6	6	5	7
59	5	6	6	2	5	3	1
60	5	7	6	3	7	4	5
61	7	5	3	6	6	2	6
62	7	5	6	5	7	7	6
63	.	.	.	.	.	.	.
64	7	7	7	4	5	6	7
65	4	4	6	4	6	4	6
66	7	6	5	4	6	7	6
67	3	1	2	3	4	3	1
68	6	4	1	5	5	5	2
69	2	2	3	3	5	4	1
70	6	5	3	2	5	5	5
71	5	4	4	5	5	5	6
72	5	5	4	5	6	7	7
73	3	2	4	2	4	2	1
74	6	6	7	5	3	4	5
75	5	5	4	4	1	6	6
76	3	1	1	3	7	1	2
77	3	6	6	6	5	2	7
78	1	1	.	2	1	1	1
79	6	3	5	5	4	5	6
80	3	1	6	7	6	6	3
81	6	6	5	6	5	7	6
82	3	4	5	4	7	7	3
83	7	3	5	7	7	7	4
84	6	5	5	4	6	6	6
85	6	5	2	5	6	4	6
86	5	6	4	5	6	6	6
87	.	.	.	.	.	.	.

88	.	.	.	.	.	.	.
89	7	1	7	2	7	7	7
90	6	6	7	2	4	6	1
91	6	6	6	6	3	7	6
92	2	2	6	5	5	2	2
93	.	.	.	.	.	.	.
94	5	3	2	5	4	2	4
95	6	6	1	7	1	6	6
96	3	3	7	4	3	5	4
97	.	.	.	.	.	.	.
98	6	5	4	4	6	5	6
99	7	2	6	2	2	4	7
100	5	6	6	4	6	6	6
101	.	.	.	.	.	.	.
102	6	5	3	5	6	4	6
103	6	6	5	3	4	6	6
104	6	2	4	6	2	6	6
105	6	4	7	6	1	6	5
106	7	4	5	2	5	5	7
107	7	5	6	3	2	6	7
108	6	2	2	4	7	7	7
109	7	5	3	2	5	3	6
110	7	6	6	2	6	4	7
111	.	6	6	4	7	7	2
112	2	1	1	5	1	1	1
113	7	6	6	5	7	5	4
Σx	573	437	485	447	543	536	506
Σx^2	3385	2125	2597	2179	3151	3054	2942

1.5. Respostes als exercicis d'estadística descriptiva

Exercici 1.

1.1.

GÈNERE	Freqüències absolutes	Freqüències acumulades	Percentatges	Percentatges acumulats
dona	15	15	50	50
home	15	30	50	100
Total	30		100	

NOMBRE DE GERMANS	Freqüències absolutes	Freqüències acumulades	Percentatges	Percentatges acumulats
0	9	9	30,0	30,0
1	12	21	40,0	70,0
2	7	28	23,3	93,3
3	2	30	6,7	100,0
Total	30		100,0	

Dels 30 casos n'hi ha 28 que tenen dos germans o menys.

ESTUDIS ACABATS	Freqüències absolutes	Freqüències acumulades	Percentatges	Percentatges acumulats
BUP	10	10	33,3	33,3
Formació professional	6	16	20,0	53,3
Primaris	6	22	20,0	73,3
Universitaris grau mitjà	5	27	16,7	90,0
Universitaris grau superior	3	30	10,0	100,0
Total	30		100,0	

1.2.

	Nombre de germans	Ingressos	Persones actives a la llar
mitjana	1,07	1.337,00	1,83
mediana	1,00	1.200,00	2,00
moda	1,00	1.200,00	2,00
desviació típica	0,892	473,689	0,734
mínim	0,00	720,00	1,00
màxim	3,00	2.520,00	3,00
rang o amplitud	3,00	1.800,00	2,00

Exercici 2

2.1.

Anys	% total	Percentatge acumulat
En 1 any o menys	31,6	31,6
En 2 anys	15,2	46,8
En 3 anys	13,7	60,5
En 4 anys o més	39,5	100

2.2. En aquest cas vol dir que hi ha un 60,5 % de la població de joves (d'entre 15 i 29 anys) de Catalunya que tenen previst acabar els estudis en 3 anys o menys.

2.3.

Anys (x)	**Població**	***fx***
En 1 any	197.544	197.544
En 2 anys	94.953	189.906
En 3 anys	85.467	256.401
En 4 anys	246.781	987.124
Total població	624.745	Σfx =1.630.975

$$\overline{X} = \frac{\Sigma fx}{N} = \frac{1.630.975}{624.745} = 2,6$$

La mitjana és de 2,6 anys.

2.4.

Anys (x)	**Població (homes)**	***fa***
En 1 any o menys	92.338	92.338
En 2 anys	50.902	143.240
En 3 anys	40.374	183.614
En 4 anys	110.578	294.192
Total població		

Lloc de la mediana: (294.192 +1) / 2 = 147.096,5

Mdn = 3 anys, per al cas dels homes

Anys (x)	**Població (dones)**	***fa***
En 1 any o menys	105.206	105.206
En 2 anys	44.050	149.256
En 3 anys	45.092	194.348
En 4 anys	136.203	330.551
Total població		

Lloc de la mediana: (330.551 +1) / 2 = 165.276

Mdn = 3 anys

Per tant, tant en el cas dels homes com en el de les dones, la meitat de la població té previst acabar els estudis en 3 anys o menys i l'altra meitat, en 3 anys o més.

Exercici 3

3.1.

Comarca	Persones amb discapacitat
Alt Empordà	3.821
Baix Empordà	3.813
Cerdanya	368
Garrotxa	1.668
Gironès	6.744
Pla de l'Estany	920
Ripollès	836
Selva	4.850
TOTAL	23.020

Les persones amb discapacitat de les comarques de Girona representen el 6,64 % del total de persones amb discapacitat de Catalunya.

3.2. Si agafem tot Catalunya (38.736 persones amb discapacitat psíquica), la mitjana per comarca (41 comarques) és de 944,78 persones amb discapacitat psíquica.

Si ens fixem només en el nombre de persones amb discapacitat psíquica de les comarques de Girona (4.485), la mitjana per comarca (8 comarques) és, en aquest cas, de 560,6 persones amb discapacitat psíquica.

3.3. Si agafem el total de persones amb discapacitat de Catalunya (346.528), la mitjana de persones amb discapacitat per comarca (41 comarques) és de 8.451 persones amb discapacitat.

Si ens fixem només en les comarques de Girona (23.020), la mitjana per comarca (8) és de 2.878 persones amb discapacitat.

Exercici 4

4.1. Del 33 % al 64 % de disc.: 182.305
Del 65 % al 74 % de disc.: 93.171
Per tant, hi ha 275.476 persones amb un 74 % de discapacitat o menys.

4.2. Total de discapacitats = 346.528 persones

Per tant, representa el 79,5 % de les persones amb discapacitat.

4.3. El 20,5 % tenen un grau de discapacitat de 75 % o més.

4.4. Hi ha 174.865 dones amb discapacitat, que representen el 50,5 % de les persones amb discapacitat.

Exercici 5

5.1. Les persones d'entre 6 i 15 anys representen el 5,5 % de les persones amb discapacitat de la comarca del Gironès.

5.2. Les dones que tenen més de 65 anys representen el 18,7 % del total de persones amb discapacitat de la comarca del Gironès.

5.3. Hi ha 641 persones de menys de 19 anys que tenen una discapacitat. Això representa el 9,5 % del total de les persones amb discapacitat de la comarca.

5.4. L'interval "75 anys o més"

5.5. L'interval "45-54 anys"

5.6. L'interval "75 anys o més"

Exercici 6

6.1. Centre A Mitjana = 83,8 Desviació típica = 30,64

Centre B Mitjana = 84 Desviació típica = 21,71

6.2. Com que les mitjanes són molt semblants, podem comparar les dues desviacions típiques. Veiem que la desviació típica del centre B és menor que la del centre A i, per tant, podem dir que al centre A hi ha hagut més variabilitat durant aquests 10 anys.

Exercici 7

Mitjana = 51,6 Mediana = 52 Moda = 47 Desviació típica = 10,5

Amplitud = 38

Exercici 8

8.1. Mitjana ponderada opositor A = 7,42

Mitjana ponderada opositor B = 7,5

Per poca diferència, la plaça correspon a l'opositor B.

8.2. Si la mitjana fos simple la plaça hauria correspost a l'opositor A.

Exercici 9

9.1. Buscar els percentatges per a cada variable (any 1998).

Programes	ICS	ICC	IC	Total	%
Casos finalitzats	8	59	152	219	
Sexe					
Nois	8	54	143	205	93,61
Noies	0	5	9	14	6,39
Edat (1)					
12 anys	0	2	0	2	0,91
13 anys	0	5	10	15	6,85
14 anys	0	11	18	29	13,24
15 anys	2	19	29	50	22,83
16 anys	5	15	75	95	43,38
17 anys	1	5	18	24	10,96
18 anys o més	0	2	2	4	1,83
Delictes (2)					
Contra la vida	0	0	5	5	1,39
Lesions	2	18	36	56	15,56
Contra la llibertat	1	6	12	19	5,28
Contra la llibertat sexual	0	4	3	7	1,94
Contra la intimitat	0	2	7	9	2,50
Contra l'honor	0	0	0	0	0
Contra el patrimoni i l'ordre socioeconòmics.	2	63	169	234	65
Contra la seguretat col·lectiva	0	4	3	7	1,94
Falsedat documental	0	0	0	0	0
Contra l'Administració de justícia	0	0	1	1	0,28
Contra l'ordre públic	0	6	7	13	3,61
Altres delictes i faltes	0	2	7	9	2,50
Total	5	105	250	360	

9.2.

Edat (1)	Nombre total de menors	*fx*	*fa*
12 anys	2	24	2
13 anys	15	195	17
14 anys	29	406	46
15 anys	50	750	96
16 anys	95	1.520	191
17 anys	24	408	215
18 anys	4	72	219
	N= 219	Σfx= 3375	

Mitjana = Σfx / N = 3375 / 219 = 15,41 anys

9.3.
Mediana = (N + 1) / 2 = 220 / 2 = 110 lloc **=** 16 anys
Moda = 16 anys

9.4. **Pel que fa al gènere,** l'any 1999 s'observa un petit increment en el nombre total de nois i una disminució en el de noies.
Respecte a l'edat, l'any 1998 teníem més menors de 16 anys i l'any 1999 teníem més menors de 15 anys. L'any 1999, de 17 i 18 anys n'hi ha molts menys; en canvi, n'hi ha algun més de 12 anys que l'any 1998.
Si mirem els delictes, l'any 1999 veiem que la majoria (52,54 %) són contra el patrimoni i l'ordre socioeconòmic, tot i que l'any 1998 eren d'un 65 %; per tant, han disminuït un 13 %.
L'any 1999 els delictes per lesions eren d'un 24,62 %; en canvi, l'any 1998 eren d'un 15,56 %, és a dir, s'han incrementat un 9 %.
L'any 1999 també s'han incrementat una mica els delictes contra la llibertat sexual (més d'un 4 % d'increment).

9.5.

Edat (1)	**Nombre de joves**	***fx***	***fa***
12 anys	7	84	7
13 anys	11	143	18
14 anys	29	406	47
15 anys	89	1335	136
16 anys	50	800	186
17 anys	13	221	199
	$N =$ 199	fx= 2989	

Mitjana = $\Sigma fx / N$ = 2989 / 199 **=** 15,02 anys
Mediana = (N + 1) / 2 = 200 / 2 = 100 lloc **=** 15 anys
Moda = 15 anys

Exercici 10

10.1. El 76,5 % han estat homes i el 23,5 % han estat dones.

10.2. Un 33,8 % dels casos finalitzats corresponen a delictes contra el patrimoni i l'ordre socioeconòmic.

10.3. El 22,1 % de les persones que han participat en aquest programa ho han fet per haver comès delictes de lesions.

10.4. El 58,8 % dels casos finalitzats eren persones de més de 20 anys.

Exercici 12

12.1.

x	*f*	%
1	2	1,9
2	4	3,8
3	8	7,7
4	6	5,8
5	16	15,4
6	41	39,4
7	27	26,0

12.2.

x	*f*	*fa*
1	2	2
2	4	6
3	8	14
4	6	20
5	16	36
6	41	**77**
7	27	104

Lloc de la mediana = (104 + 1) / 2 = 52,5   Mediana = 6

12.3. Moda = 6

12.4. El 86,5 % de les persones valoren amb un 4 o més la satisfacció que els suposa el contingut de la feina.

12.5 i 12.6.

	Contingut de la feina	**Perspectives de millora**	**Retribució**	**Coneixe-ments de formació**	**Perspectives d'estabilitat**	**Jornada laboral**	**Realitza-ció personal**
Mitjana	5,51	4,16	4,71	4,26	5,17	5,10	5,01
Desviació típica	1,49	1,72	1,75	1,63	1,82	1,75	2,02

L'aspecte que de manera global es valora més positivament és el contingut de la feina, no només perquè té una mitjana més alta sinó també perquè és l'ítem en què la valoració presenta menor dispersió.

Presenta més dispersió la valoració del nivell de realització personal que implica la feina. Tot i que la valoració és també alta en conjunt, veiem que la dispersió és la més gran.

2. Puntuacions típiques i corba normal

Hem vist que les característiques importants d'una distribució de valors són la seva localització, representada normalment per la mitjana, i la seva dispersió, per a la qual se sol utilitzar la desviació típica. Amb aquesta informació podem deduir el valor relatiu d'una dada concreta respecte al grup al qual pertany. Però ens podem trobar en situacions com la següent.

Imaginem-nos que un estudiant ha fet quatre exàmens en quatre assignatures diferents i ha obtingut les puntuacions següents:

Matemàtiques	6
Llengua catalana	8
Anglès	7,5
Ciències socials	6,5

Aparentment, sembla que la millor qualificació de l'estudiant és la de llengua. El que passa és que les puntuacions directes (s'anomenen així les puntuacions obtingudes directament a través de l'aplicació d'algun tipus de prova) difícilment són comparables entre si. Pot passar, per exemple, que l'examen de llengua catalana hagi estat molt més fàcil que els altres, de manera que tot el grup ha obtingut puntuacions elevades. O bé que l'examen d'anglès hagi puntuat sobre 20 punts i el de matemàtiques sobre 6. Aquestes qualificacions proporcionen informació sobre el valor directe obtingut, però no ens permeten saber fins a quin punt la puntuació és bona, ni si és bona en comparació de les puntuacions dels altres.

Ara suposem que, a més de la puntuació directa, coneixement la mitjana:

	X	**Mitjana**
Matemàtiques	6	4
Llengua catalana	8	8,5
Anglès	7,5	6,5
Ciències socials	5,5	7

Mirant les mitjanes ja podem veure que hi ha unes notes que estan per sota de la mitjana i unes altres que estan per sobre. En fer les diferències de cada valor respecte a la mitjana ens surt que:

	X	**Mitjana**	**X - Mitjana**
Matemàtiques	6	4	2
Llengua catalana	8	8,5	-0,5
Anglès	7,5	6,5	1
Ciències socials	5,5	7	-1,5

Per tant, les notes de llengua catalana i les de ciències socials no seran gaire bones ja que se situen per sota de la mitjana obtinguda pel grup. El resultat més dolent de l'estudiant ha estat en ciències socials, ja que la puntuació de 5,5 es troba a una distància d'1,5 punts per sota de la mitjana, i la puntuació de llengua catalana es troba mig punt per sota de la mitjana.

Tant les qualificacions de matemàtiques com les d'anglès estan per sobre de la mitjana. Podríem pensar que la millor qualificació és la de matemàtiques, que està dos punts per sobre de la mitjana, mentre que la d'anglès està només un punt per sobre. No obstant això, també hem de tenir en compte la dispersió de la distribució. Per tant, hem de mirar les desviacions típiques. Suposem que són les següents:

	X	**Mitjana**	**X - Mitjana**	**S**
Matemàtiques	**6**	**4**	**2**	**0,5**
Llengua catalana	8	8,5	-0,5	1
Anglès	**7,5**	**6,5**	**1**	**1**
Ciències socials	5,5	7	-1,5	1,5

Si ens fixem en les desviacions típiques, veiem que la dispersió és més elevada en anglès (1) que en matemàtiques. La desviació típica ens mostra que la dispersió en les qualificacions d'anglès és d'un punt respecte de la mitjana, la qual cosa vol dir que hi va haver puntuacions que van superar la mitjana en més d'un punt i altres puntuacions que estaven més d'un punt per sota de la mitjana. La puntuació de 7,5 es troba un punt exacte per sobre de la mitjana, la qual cosa indica que es troba a una distància d'una desviació típica respecte de la mitjana. La desviació típica de la prova de matemàtiques és 0,5. Igualment, això vol dir que la puntuació de 6 es troba a una distància de quatre desviacions típiques per sobre de la mitjana. Dit d'una altra manera, en la prova de matemàtiques, les puntuacions obtingudes estan molt més concentrades al voltant de la mitjana que en la prova d'anglès. Mirant els valors de les desviacions típiques respectives veiem que per a l'estudiant ha estat molt més difícil tenir una puntuació de 6 en matemàtiques que una puntuació de 7,5 en anglès.

En conclusió, malgrat que una primera mirada de les puntuacions ens feia pensar que la millor qualificació era la de llengua catalana, després d'aquesta anàlisi resulta que la millor puntuació en termes relatius ha estat la de matemàtiques, que era la més baixa en valor absolut.

El que s'ha fet, doncs, és transformar les puntuacions directes en unes altres puntuacions mesurades amb el mateix tipus d'escala, de manera que puguem comparar les puntuacions (puntuacions típiques).

2.1. Les puntuacions típiques o puntuacions Z

El procediment habitual per transformar puntuacions directes en típiques és convertir-les en unes puntuacions en una escala de valors que tingui de mitjana 0 i de desviació típica 1. D'aquesta manera, si la puntuació resultant és positiva, podrem dir que està per sobre de la mitjana (ja que estarà per sobre de 0), i si la puntuació és negativa, podrem afirmar que el valor està per sota de la mitjana. A més, el valor resultant ens indicarà a quantes desviacions típiques per sobre o per sota de la mitjana es troba el valor que estem transformant. La fórmula que s'utilitza és la següent:

$$Z = \frac{X - \overline{X}}{S} \qquad (7)$$

Exercici

Imaginem-nos que tenim les notes del primer quadrimestre de l'assignatura de Bases de dos cursos diferents, les del curs 2003-2004 i les del curs 2004-2005. En el curs 2003-2004, la mitjana de les qualificacions va ser de 5,5 i la desviació típica, de 2,2. En el curs 2004-2005 la mitjana aritmètica va ser de 6 i la desviació típica, d'1,6.
Un estudiant que obtingués una nota de 7,5, a quin dels dos cursos representaria que ha aconseguit una millor qualificació?

$$Z_1 = \frac{X_1 - \overline{X_1}}{S_1} = \frac{7,5 - 5,5}{2,2} = 0,91 \qquad \text{(Curs 2003-2004)}$$

$$Z_2 = \frac{X_2 - \overline{X_2}}{S_2} = \frac{7,5 - 6}{1,6} = 0,93 \qquad \text{(Curs 2004-2005)}$$

Observem, doncs, que en termes relatius l'estudiant que hagi obtingut una qualificació de 7,5 l'any 2004-2005 ha tret una millor nota que el que ha aconseguit la mateixa puntuació l'any 2003-2004, tot i que la diferència és mínima.

2.2. Exercicis sobre puntuacions típiques

1. Un mateix alumne obté en els exàmens d'Estadística i d'Avaluació de programes un 6,3 i un 7,1 respectivament. Si la mitjana d'Estadística de tota la classe és de 4,6 amb desviació típica d'1,2 i la mitjana d'Avaluació de programes és de 6,9 amb desviació típica de 0,8, calcula en quina assignatura aquest alumne ha tret comparativament una puntuació més alta.
2. En una prova de domini de la lectura aplicada a 800 alumnes de 3r curs d'educació primària d'un centre escolar, la puntuació mitjana obtinguda és de 85 i la S = 11. En un altre centre, els alumnes de 3r obtenen una puntuació mitjana de domini de la lectura de 86 i una S = 14. Una alumna obté una puntuació de 88. En quin dels dos centres, comparativament, aquest valor representa una puntuació més alta de domini de la lectura?
3. La marca de cotxes X té un consum mitjà de 5 litres de gasolina cada cent quilòmetres, amb una desviació típica d'1,3 litres. La marca de cotxes Y té un consum mitjà de 5,5 litres, amb una desviació típica d'1,2 litres. Un cotxe de cada marca que hagi consumit durant el mateix trajecte 8 litres de gasolina, a quina de les dues marques representa que ha tingut un consum més elevat?
4. En un estudi sobre el procés d'inserció laboral dels diplomats en Educació Social de la UdG es va passar un qüestionari a tots els diplomats des de l'any 1996 fins a l'any 2005. En una de les preguntes del qüestionari se'ls demanava que valoressin de l'1 al 7 la feina que feien en el moment de respondre al qüestionari. Concretament, es demanava que valoressin els ítems següents:

	Gens d'acord					Totalment d'acord	
	1	2	3	4	5	6	7
1. El contingut de la feina és satisfactori.							
2. Les perspectives de millora i promoció en la feina actual són bones.							
3. El nivell de retribució et sembla satisfactori.							
4. Els coneixements derivats de la formació universitària t'han estat útils per a la feina.							
5. Les perspectives d'estabilitat són bones.							
6. La jornada laboral (horaris, torns de treball, etc.) és satisfactòria.							
7. Et sents realitzat com a educador social en aquesta feina.							

A continuació tens una taula amb els estadístics descriptius una vegada analitzat cada un dels ítems.

	N	Rang	Suma	Mitjana	Desv. típ.	Variància
Valoració contingut (1-7)	104	6	573	5,51	1,488	2,213
Valoració perspectives millora (1-7)	105	6	437	4,16	1,716	2,945
Valoració retribució (1-7)	103	6	485	4,71	1,752	3,071
Valoració coneixements de formació (1-7)	105	6	447	4,26	1,629	2,654
Valoració perspectives estabilitat (1-7)	105	6	543	5,17	1,816	3,297
Valoració jornada laboral	105	6	536	5,10	1,748	3,056
Valoració realització personal (1-7)	101	6	506	5,01	2,017	4,070

4.1. Una persona ha valorat amb un 6 tant les perspectives de millora com la retribució que rep. Aquesta puntuació, en quina de les dues variables representa una millor valoració tenint en compte les característiques de cada distribució?

4.2. Una persona ha valorat amb un 4 tant la utilitat dels coneixements derivats de la formació universitària per a la seva feina com el nivell de realització personal a la feina. En quina de les dues variables representa una puntuació pitjor?

4.3. Hi ha persones que han valorat amb un 3 la jornada laboral i persones que han valorat amb un 2 les perspectives de millora. En quina de les dues variables aquestes puntuacions representen una millor valoració?

4.4. Una puntuació de 3, en quina de les tres variables representa una millor puntuació, en la valoració de les perspectives de millora, en la valoració de les perspectives d'estabilitat o en la valoració del nivell de realització personal que suposa la feina?

4.5. Una persona ha valorat amb un 5 les perspectives de millora i amb un 6 el contingut de la feina. En quin dels dos ítems la seva valoració ha estat més positiva, tenint en compte les característiques de la distribució de valors de cada variable?

2.3. Respostes als exercicis sobre puntuacions típiques

Exercici 1

$Z_1 = 1{,}42$

$Z_2 = 0{,}25$

La puntuació més alta, comparativament, és la d'Estadística.

Exercici 2

Una alumna que aconsegueix 88 punts té una puntuació més alta de domini de la lectura al centre 1 perquè la puntuació típica del centre 1 (0,27) és més gran que la puntuació típica del centre 2 (0,14).

Exercici 3

$$Z_1 = \frac{X_1 - \overline{X_1}}{S_1} = \frac{8-5}{1{,}3} = 2{,}3$$

$$Z_2 = \frac{X_2 - \overline{X_2}}{S_2} = \frac{8-5{,}5}{1{,}2} = 2{,}1$$

Comparant les dues marques, veiem que el consum de 8 litres representa un consum més elevat en el cas de la marca X que en la marca Y.

Exercici 4

4.1.

$$Z_1 = \frac{X_1 - \overline{X_1}}{S_1} = \frac{6 - 4{,}16}{1{,}716} = 1{,}07$$

$$Z_2 = \frac{X_2 - \overline{X_2}}{S_2} = \frac{6 - 4{,}71}{1{,}752} = 0{,}74$$

Mirant les dues puntuacions tipificades veiem que 1,07 > 0,74, la qual cosa vol dir que aquesta puntuació de 6 representa una valoració més bona de les perspectives de millora que no pas de la retribució que percep, tenint en compte les característiques de la distribució obtinguda a partir de les respostes a aquesta pregunta.

4.2.

$Z_1 = -0{,}16$
$Z_2 = -0{,}5$ $\qquad -0{,}5 < -0{,}16$

Per tant, tenint en compte les característiques de les dues mostres, la puntuació de 4 atorgada a la variable "valoració del nivell de realització com a educador social" representa una pitjor valoració que la puntuació de 4 atorgada a la "valoració de la utilitat per a la feina dels coneixements universitaris".

4.3.

$Z_1 = -1{,}2$ (puntuació típica per al valor 3 de "valoració de la jornada laboral")
$Z_2 = -1{,}26$ (puntuació típica per al valor 2 de "valoració de les perspectives de millora")

$-1{,}26 < -1{,}2$
El valor 3 representa una millor puntuació, si bé hi ha poca diferència. Vol dir que gairebé són equivalents les dues valoracions (el 2 i el 3) per a les dues variables diferents.

4.4.

$Z_1 = -0{,}68$ (puntuació típica per al valor 3 de "valoració de les perspectives de millora")
$Z_2 = -1{,}20$ (puntuació típica per al valor 3 de "valoració de les perspectives d'estabilitat")
$Z_3 = -0{,}99$ (puntuació típica per al valor 3 de "valoració del nivell de realització personal")

$-1{,}20 < -0{,}99 < -0{,}68$

Per tant, una puntuació de 3 és una millor puntuació en la valoració de les perspectives de millora.

4.5.

$Z_1 = 0{,}49$ (puntuació típica per al valor 5 de "valoració de les perspectives de millora)
$Z_2 = 0{,}33$ (puntuació típica per al valor 6 de "valoració del contingut de la feina")

$0{,}33 < 0{,}49$. Per tant, haver puntuat amb un 5 les perspectives de millora representa una millor valoració que haver puntuat amb un 6 el contingut de la feina.

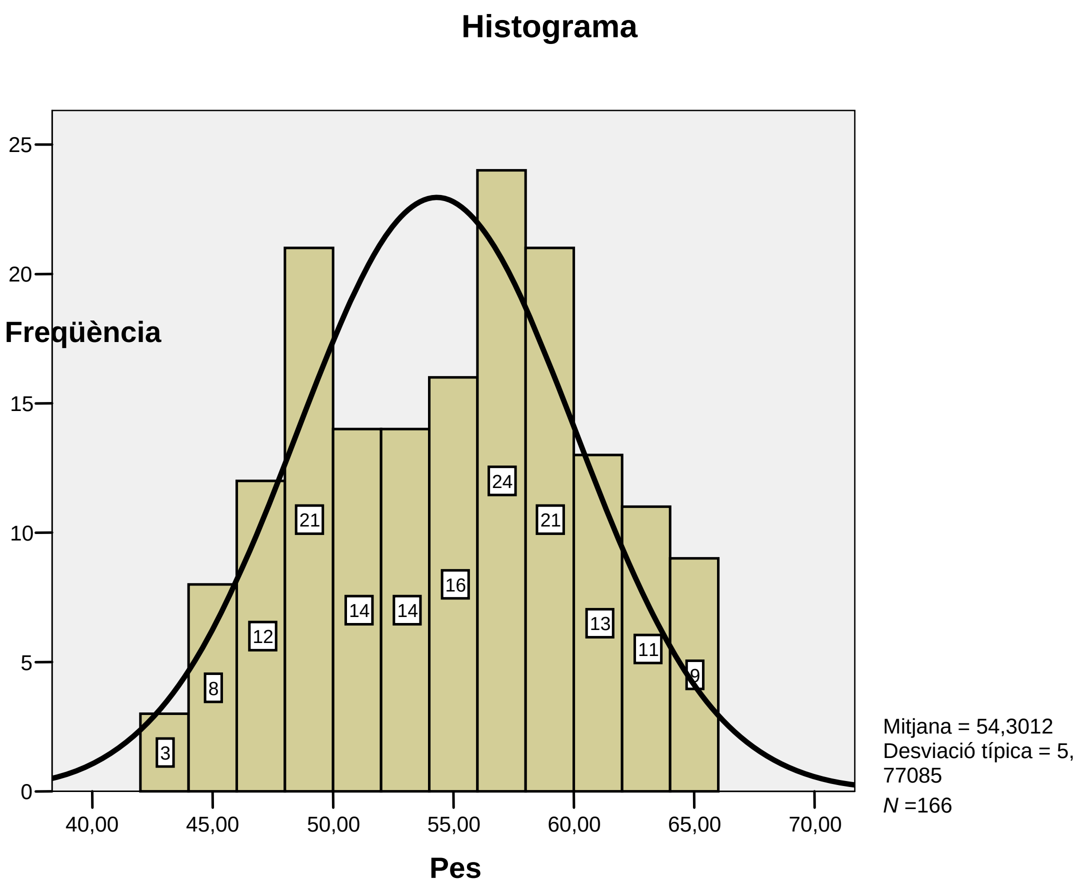
2.4. El model de la corba normal

En aquest document fem només una petita introducció al model de la corba normal, que ampliarem més endavant. De moment, el que ens cal saber és que el model de la corba normal és una distribució teòrica, és a dir, és una distribució que mai no coincidirà exactament amb la distribució d'unes dades empíriques. Les distribucions de valors es poden representar gràficament. Quan l'escala de mesura d'una variable és una escala d'interval, el tipus de gràfic que podem utilitzar és l'histograma. Així, imaginem-nos que tenim la següent distribució de valors de la variable "pes" de les dones d'entre 18 i 30 anys:

Valors		*f*
	43,00	3
	44,00	3
	45,00	5
	46,00	2
	47,00	10
	48,00	11
	49,00	10
	50,00	10
	51,00	4
	52,00	10
	53,00	4
	54,00	6
	55,00	10
	56,00	13
	57,00	11
	58,00	9
	59,00	12
	60,00	8
	61,00	5
	62,00	7
	63,00	4
	64,00	4
	65,00	4
	66,00	1
	Total	166

Al gràfic 2.1 podem veure la distribució d'aquests valors. Cada columna representa un interval de 2 valors referits al pes: [42-44), [44-46); [46-48); [48-59), etc.

Gràfic 2.1: Distribució de la mostra segons el pes



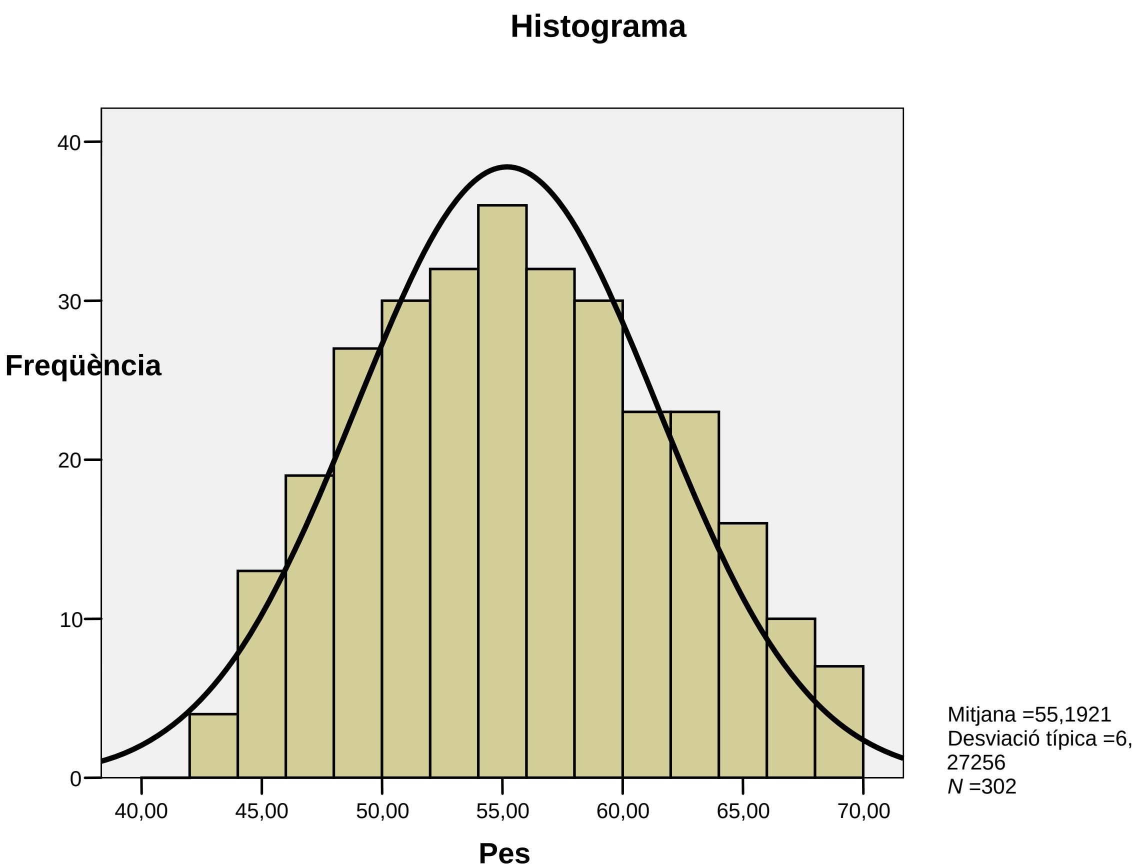
Ara imaginem-nos que hem augmentat la mida de la mostra. Els valors són els següents (i a continuació veiem l'histograma):.

pes

Valors		*f*	%
	42,00	1	,3
	43,00	3	1,0
	44,00	6	2,0
	45,00	7	2,3
	46,00	6	2,0
	47,00	13	4,3
	48,00	15	5,0
	49,00	12	4,0
	50,00	18	6,0
	51,00	12	4,0
	52,00	16	5,3
	53,00	16	5,3
	54,00	16	5,3

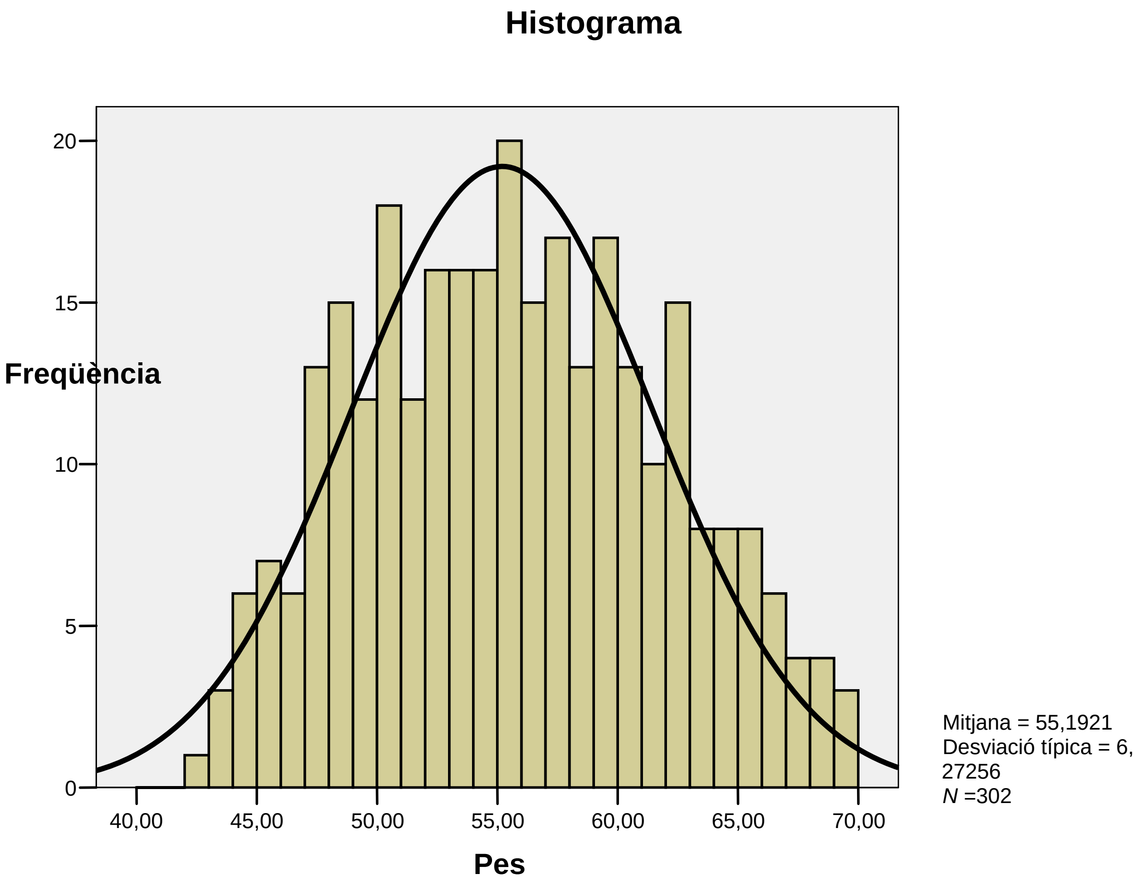
	55,00	20	6,6
	56,00	15	5,0
	57,00	17	5,6
	58,00	13	4,3
	59,00	17	5,6
	60,00	13	4,3
	61,00	10	3,3
	62,00	15	5,0
	63,00	8	2,6
	64,00	8	2,6
	65,00	8	2,6
	66,00	6	2,0
	67,00	4	1,3
	68,00	4	1,3
	69,00	1	,3
	70,00	2	,7
	Total	302	100,0

Gràfic 2.2: Distribució de la mostra segons el pes



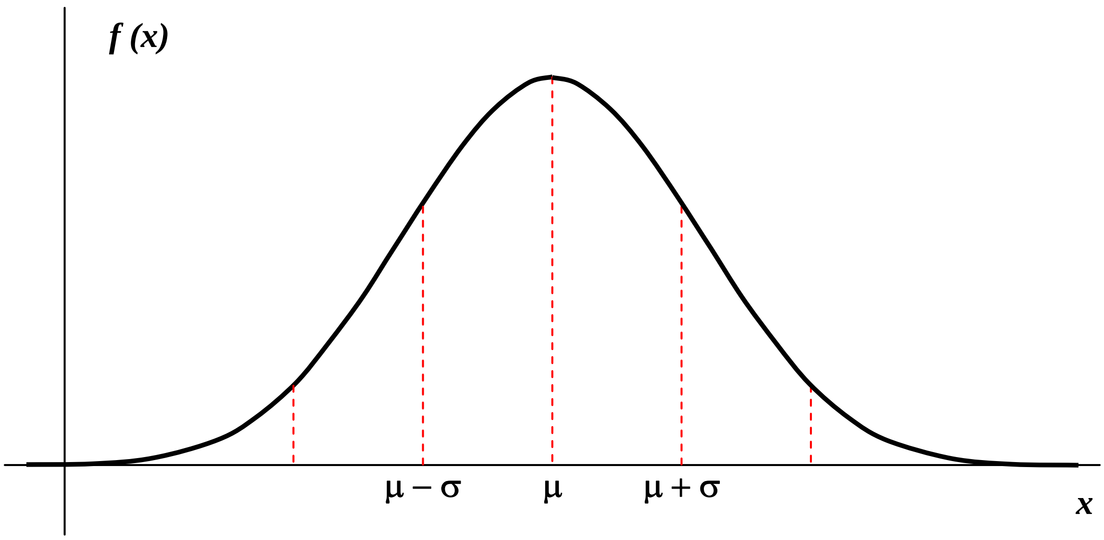
En ambdós histogrames veiem que els intervals són de 2, és a dir, cada columna representa les freqüències sumades de dos valors. Si anéssim fent intervals més petits, fins a tenir la distribució per a cada valor de pes, ens trobaríem que la forma de la gràfica cada vegada s'acosta més a la que hi ha dibuixada en el gràfic 2.3.

Gràfic 2.3: Distribució de la mostra segons el pes. Intervals d'1.



Si féssim la distribució de tots els valors de la població de dones d'entre 18 i 30 anys tindríem que, teòricament, la distribució seria molt similar a la que es representa en el gràfic 2.4. Al punt central trobaríem la mitjana de la població (μ). Entre la mitjana i una desviació típica per sobre i per sota seu, trobaríem la majoria de casos. A mesura que ens allunyem de la mitjana cada vegada hi ha menys casos, és a dir, menys freqüències. Així, als extrems és on trobem menys casos.

Gràfic 2.4: Corba normal



L'equació que descriu aquesta funció és la següent, que es coneix també com a *funció de densitat*:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

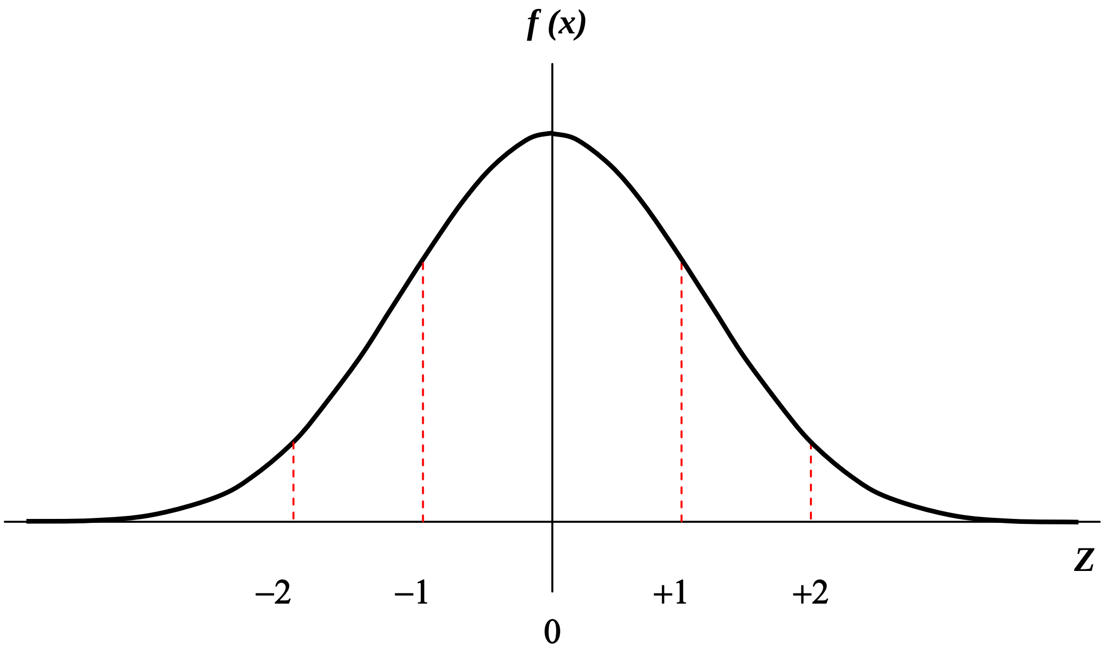
μ mitjana	π = 3,1416
σ desviació típica	e = 2,7182
σ^2 variància	x abscissa

Característiques de la corba normal:

1. La corba assoleix l'alçada màxima al punt central, que correspon al valor de la mitjana.
2. La corba és simètrica respecte a l'eix vertical, que passa per la mitjana.
3. En el model de la corba normal, la mitjana, la mediana i la moda coincideixen al mateix punt.

La distribució amb què es treballa normalment és la corba normal típica, és a dir, una distribució normal que té de mitjana 0 i de desviació típica 1.

Gràfic 2.5: Corba normal tipificada

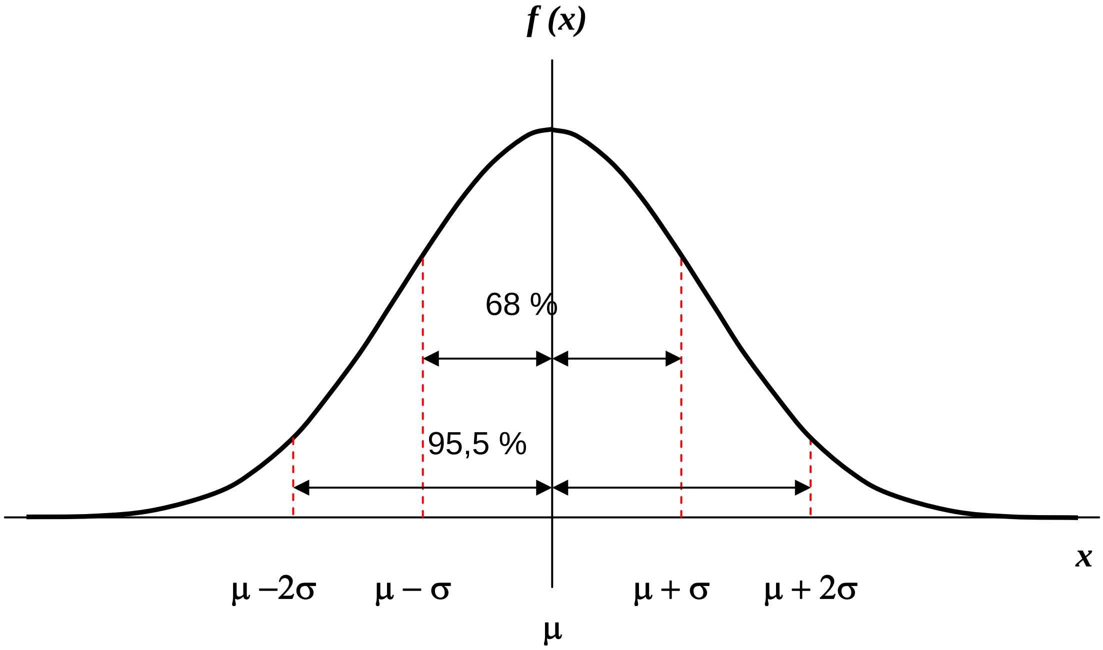


Tal com expliquen Etxeberria i Tejedor (2005), la corba normal representa un histograma en el qual la mida de l'interval és molt i molt petita.

Quan la distribució és normal, la proporció de casos que es troben entre la mitjana menys una desviació típica i la mitjana més una desviació típica sempre és del 68 %. I la proporció de casos que estan entre la mitjana menys dues desviacions típiques i la mitjana més dues desviacions típiques sempre és del 95 %.

En el cas de la corba normal tipificada, l'àrea delimitada per l'eix d'abscisses (x) i la corba val 1. En general, l'àrea sota la corba normal representa el 100 % dels casos.

Gràfic 2.6: Àrees sota la corba normal



3. Correlació lineal i predicció

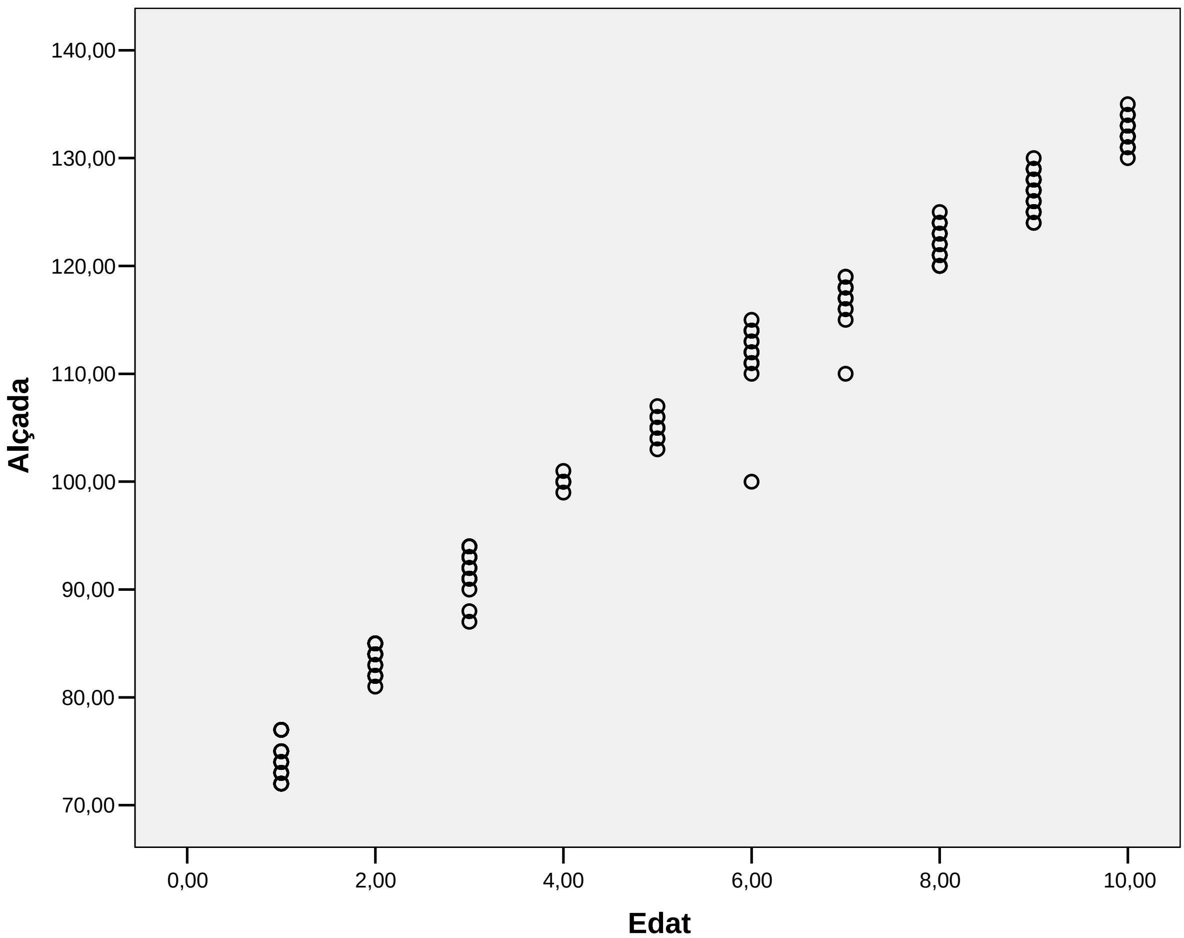
3.1. Introducció

Fins ara hem estat treballant amb una sola variable. Quan buscàvem els índexs de tendència central i de dispersió els buscàvem per a una variable cada vegada.

Però ens pot interessar estudiar conjuntament dues variables. Una de les anàlisis que podem fer és el de la *correlació lineal* entre dues variables.

Per entendre què és això de la correlació, vegem com es representaria gràficament la relació entre dues variables, "edat" i "alçada" d'una mostra de 150 nens i nenes d'entre 1 i 10 anys.

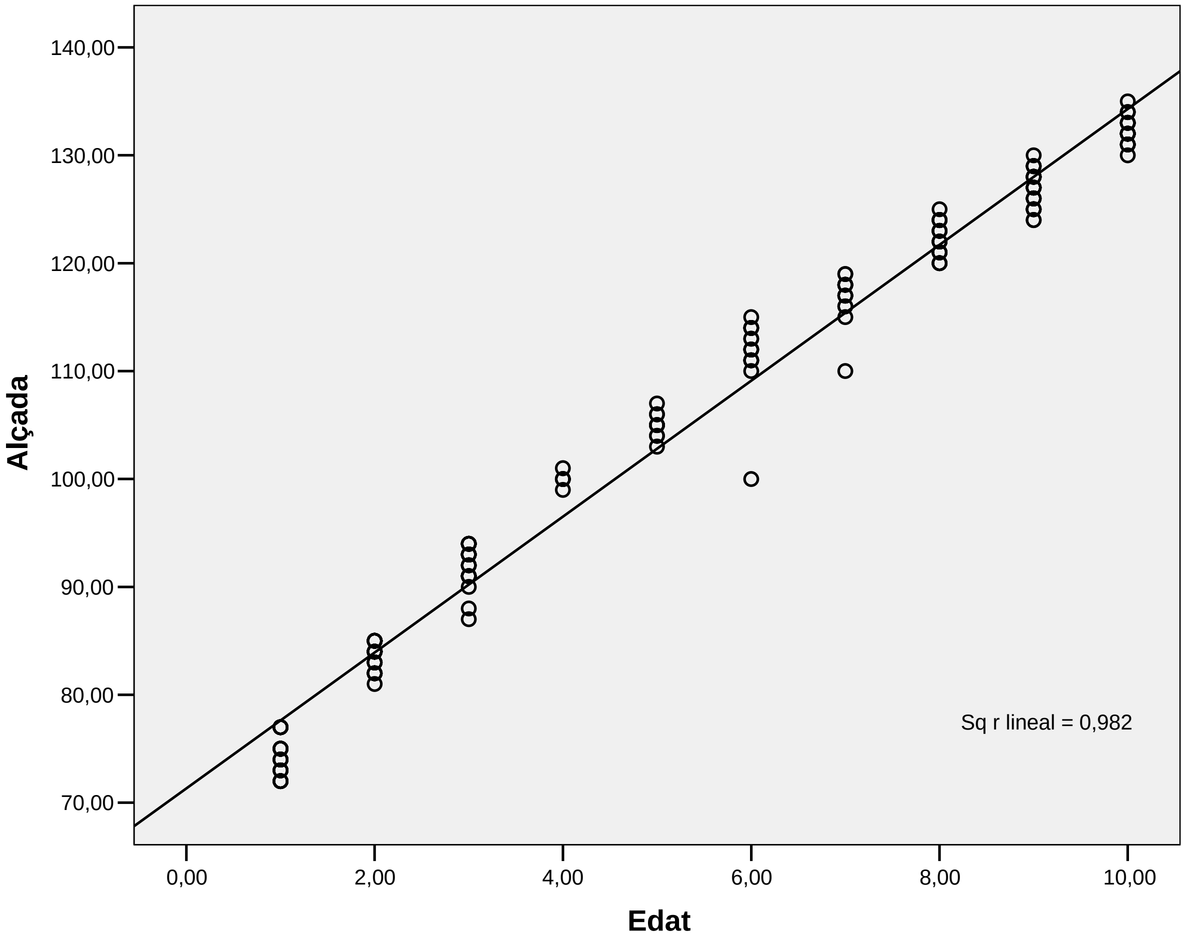
Gràfic 3.1: Diagrama de dispersió de les variables "edat" i "alçada" d'una mostra de nens i nenes d'entre 1 i 10 anys



Aquest diagrama ens indica on se situa cada un dels 150 casos, creuant l'edat amb l'alçada. Una vegada col·locats tots els punts, podem veure que tendeixen a situar-se en uns llocs concrets del diagrama i que segueixen una tendència determinada, de manera que els nens més grans tenen més alçada i els nens més petits en tenen menys. Aquest cas és força obvi, però altres vegades voldrem saber si entre dues variables hi ha alguna relació d'aquest tipus, i no resultarà tan obvi com entre l'edat i l'alçada.

Observeu que cada punt del gràfic representa dos valors per a un mateix nen: l'edat i l'alçada. La tendència que segueixen aquests valors es pot representar amb una línia recta que sintetitza la inclinació del diagrama de dispersió i la seva direcció, d'esquerra a dreta, en aquest cas. Aquesta línia recta s'anomena *recta de regressió*.

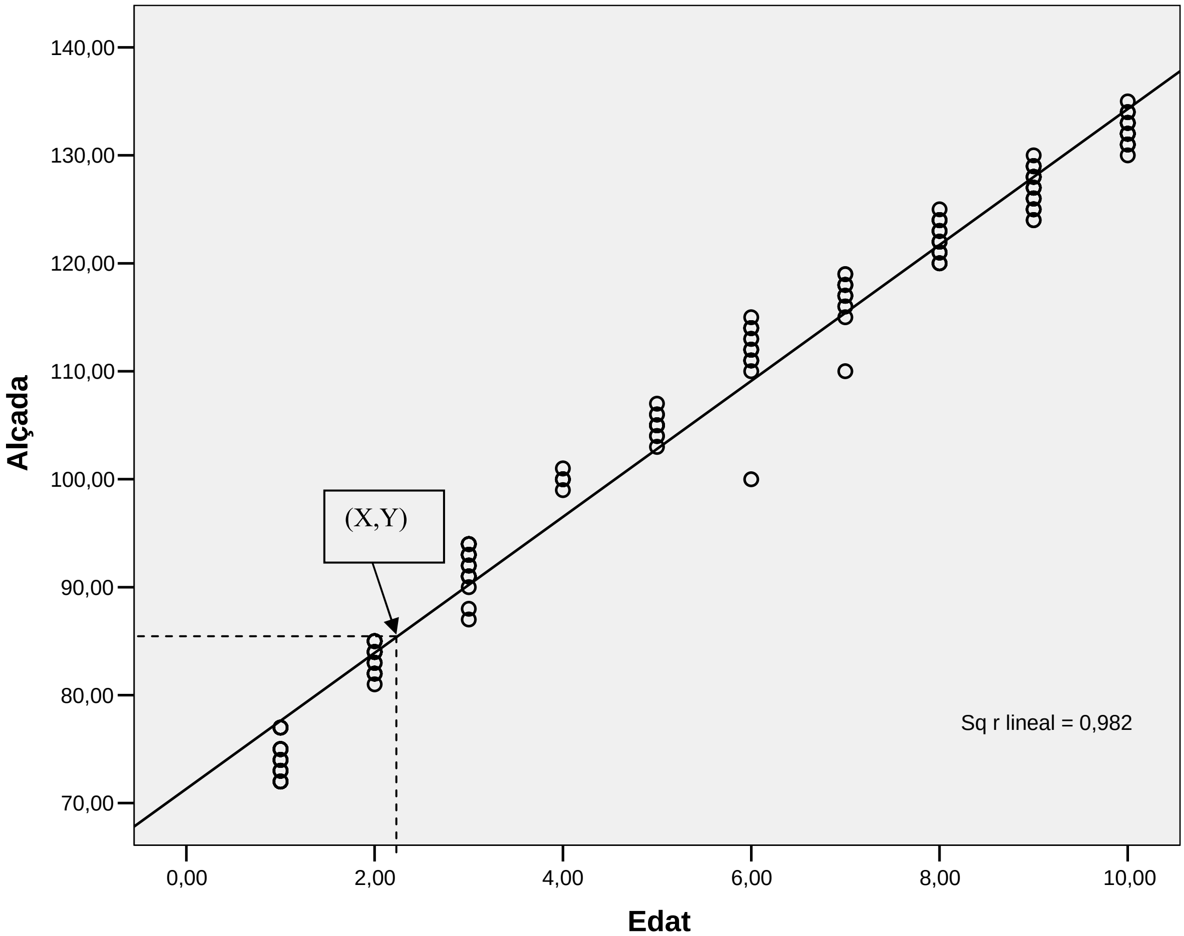
Gràfic 3.2: Diagrama de dispersió edat-alçada amb la recta de regressió



Quan sabem que dues variables estan correlacionades, trobar aquesta línia recta que resumeix la tendència dels punts ens permetrà predir el valor d'una variable en una persona si coneixem el valor de l'altra.

Així, si coneixem la recta de regressió entre edat i alçada, sabent l'edat d'un nen determinat podrem predir amb una certa aproximació l'alçada que té. Per exemple, podrem predir que l'alçada d'un nen de 2 anys i 4 mesos serà, aproximadament, d'uns 86 cm, tal com queda representat al gràfic.

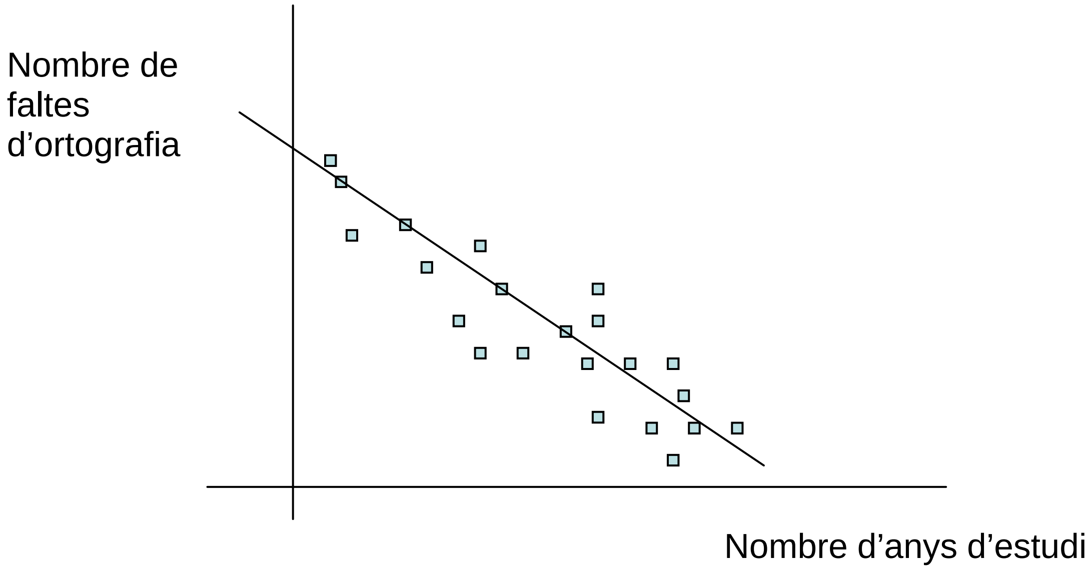
Gràfic 3.3: Representació d'una predicció



Quan dues variables estan correlacionades, podem fer aquesta mena de prediccions. Ara bé, si no ho estan, com seria el cas de l'alçada i el coeficient intel·lectual, no podem fer aquest tipus de prediccions.

No totes les variables tenen el tipus de correlació que hem descrit fins aquí. Posem per cas que hem agafat una mostra de persones d'entre 8 i 20 anys i volem veure si hi ha relació entre el nombre d'anys d'estudis d'aquestes persones i el nombre de faltes d'ortografia que fan quan escriuen un text dictat de 150 paraules. En principi, si hi hagués correlació, probablement seria en el sentit següent: com més anys d'estudis, menys nombre de faltes d'ortografia. En aquest cas, ho podríem representar com es pot observar al gràfic 3.4. La recta de regressió té una inclinació diferent de la que vèiem al gràfic 3.2. En aquest cas, es diu que hi ha correlació, però és negativa, cosa que significa que la tendència de la distribució conjunta entre les dues variables és que, a mesura que augmenten els valors d'una variable, disminueixen els valors de l'altra, mentre que en la correlació positiva la tendència és que, a mesura que augmenten els valors d'una variable, augmenten també els valors de l'altra variable.

Gràfic 3.4: Representació d'una correlació negativa

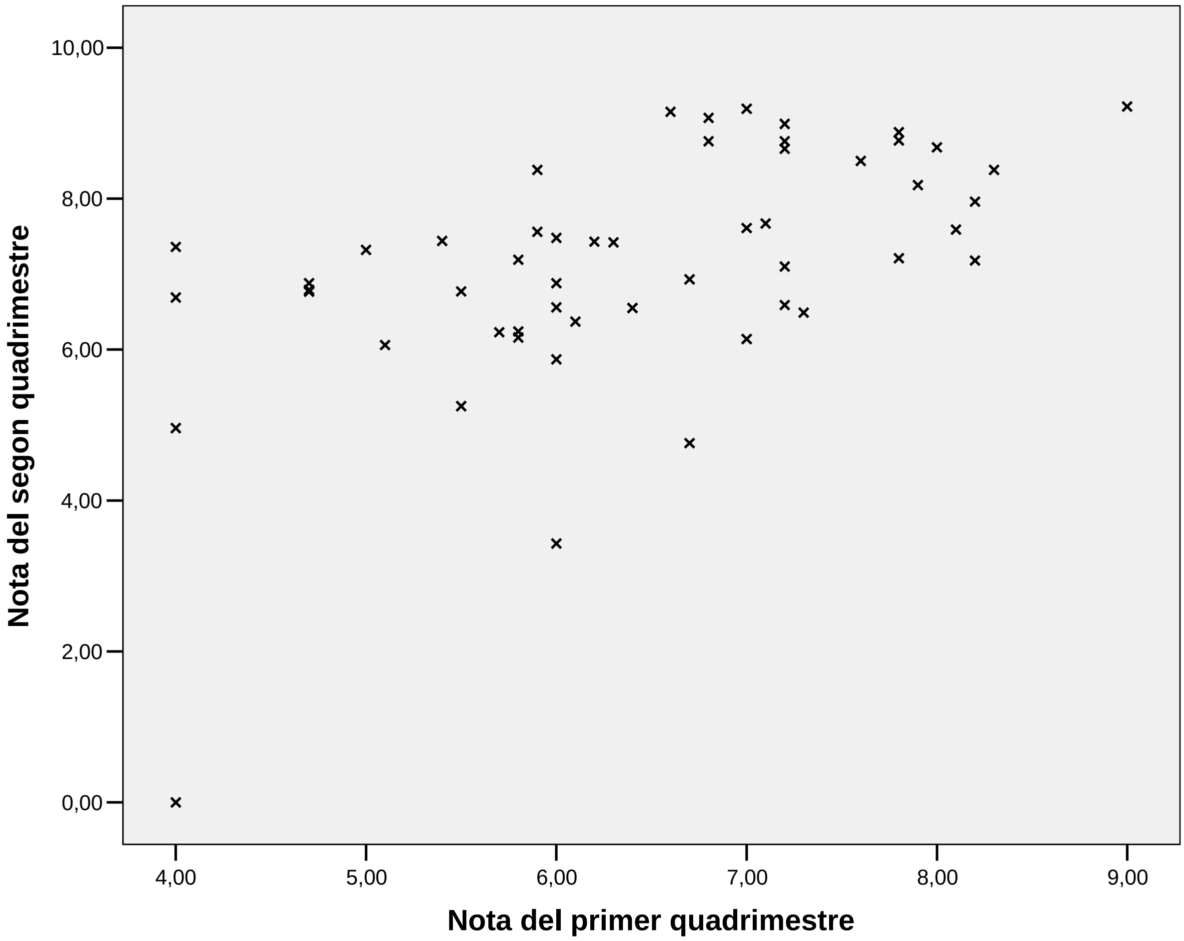


Un altre exemple

Hem recollit les notes de l'assignatura de Bases metodològiques de la investigació educativa del curs 2004-2005 i volem veure si hi ha relació entre les notes que els estudiants treuen en finalitzar el primer quadrimestre i les que treuen en finalitzar el segon. El diagrama de dispersió d'aquestes dades el trobem al gràfic 5. En aquest cas, s'observa una certa tendència dels punts, però ens és molt més difícil que en el cas de les variables edat-alçada determinar, a simple vista, si és probable que hi hagi una relació entre aquestes dues variables.

En general, per calcular el grau de correlació entre dues variables, ens caldrà buscar un coeficient que ens permeti saber si hi ha relació i si aquesta és més o menys intensa. Parlarem d'un tipus de correlació que es coneix amb el nom de *correlació lineal*.

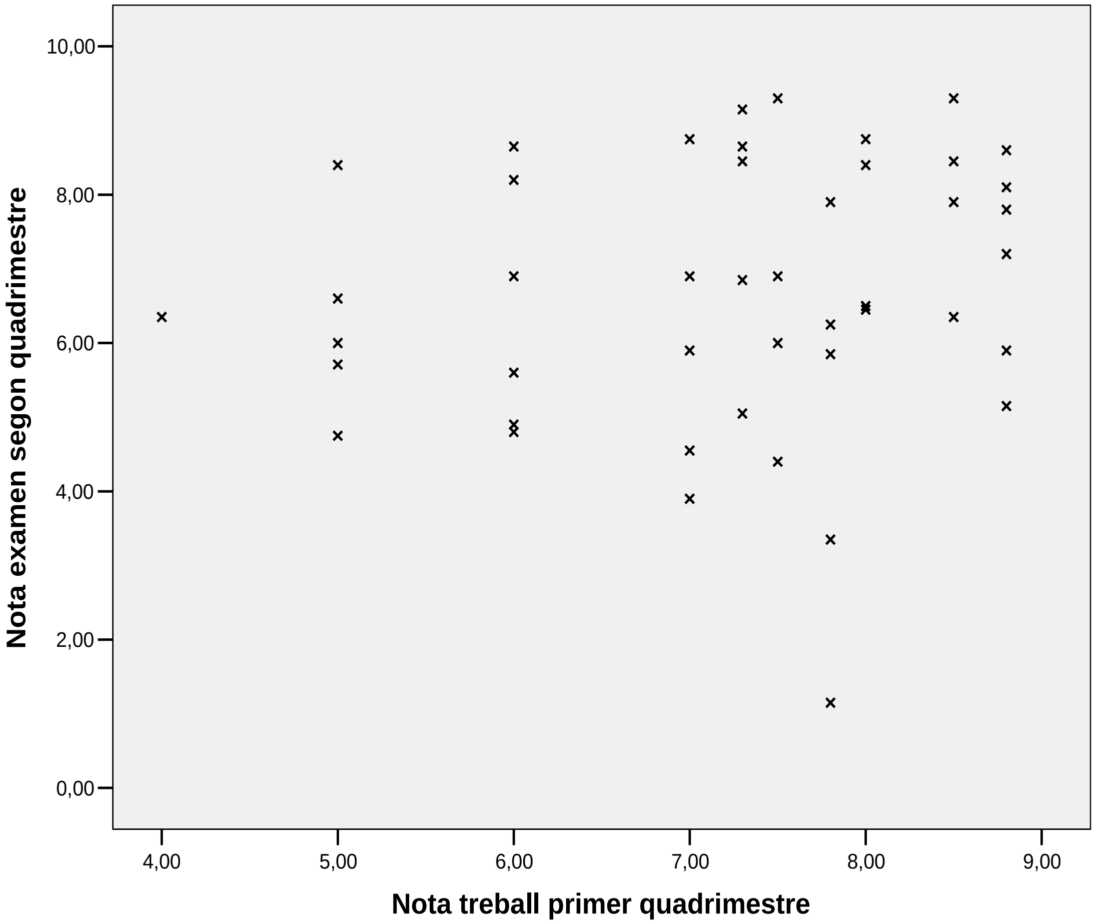
Gràfic 4.5: Diagrama de dispersió per a les qualificacions del primer i del segon quadrimestres de l'assignatura de Bases metodològiques de la investigació educativa. Curs 2004-2005



3.2. Correlació lineal

Tal com diuen Welkowitz, Ewen i Cohen (1981: 204), la relació lineal implica que, si dibuixem un diagrama de dispersió amb els valors de les dues variables, la tendència del núvol de punts obtingut s'ajusta bé a una línia recta. Fixem-nos que els punts que es presenten en els gràfics 3.3 i 3.4 tendeixen a situar-se damunt d'una línia recta. Això és més difícil d'observar en el gràfic 3.5 (nota primer quadrimestre - nota segon quadrimestre) i en el gràfic 3.6 (nota treball primer quadrimestre - nota examen segon quadrimestre).

Gràfic 3.6: Diagrama de dispersió de la relació entre la nota del treball del primer quadrimestre i la nota de l'examen del segon quadrimestre



Que hi hagi correlació lineal entre dues variables no vol dir que tots els punts hagin d'estar situats damunt d'una recta. En l'exemple que estem comentant, l'existència de correlació entre les qualificacions del primer quadrimestre i les qualificacions del segon quadrimestre indicaria que, en general, un estudiant que té una puntuació elevada el primer quadrimestre tendeix a treure una puntuació elevada el segon quadrimestre i que si un estudiant té una puntuació baixa en finalitzar el primer quadrimestre també tendeix a tenir una puntuació baixa en finalitzar el segon quadrimestre. Però això no significaria que tots els estudiants segueixin aquesta tendència. Podria ser que un estudiant hagués tret una puntuació alta en finalitzar el primer semestre i, en canvi, una puntuació baixa en finalitzar el segon, i això no faria variar la tendència de la relació en cas que n'hi hagués. És a dir, el coeficient de correlació lineal ens indicarà quina és *la tendència general del conjunt de valors de les dues variables que s'analitzen*, és a dir, ens indicarà si les dues variables varien conjuntament o no.

El coeficient de correlació lineal oscil·larà entre -1 i +1, segons que la correlació sigui negativa o positiva. Un valor del coeficient molt pròxim a 0 o igual a 0 indicarà que no hi ha correlació.

Així, per exemple, l'índex de correlació corresponent a la relació entre les notes del treball del primer quadrimestre i les notes de l'examen del segon quadrimestre és de 0,163. Aquest índex s'acosta molt a 0, i si observem el gràfic 3.6 veiem que el núvol de punts està molt dispers en el pla configurat pels eixos de coordenades; per tant, en aquest cas no hi ha correlació.

En canvi, tornant a l'exemple de l'edat i l'alçada dels nens d'1 a 10 anys, l'índex de correlació és 0,991. Això indica que és una correlació positiva i que és una correlació elevada. En el diagrama de dispersió corresponent (gràfic 3) veiem que tots els punts tendeixen a situar-se al voltant d'una recta.

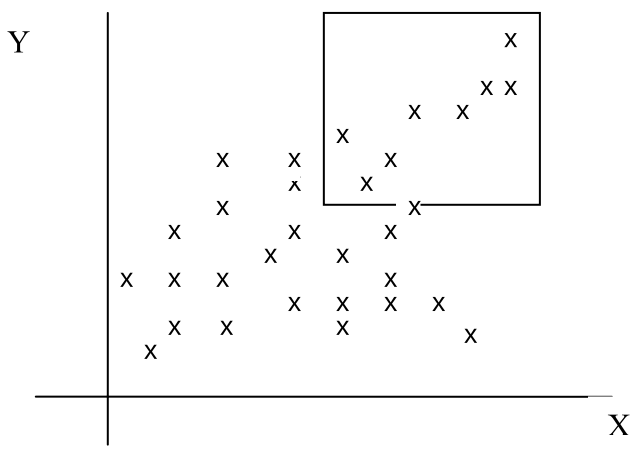
En resum, el coeficient de correlació té les característiques següents:

1. El valor 0 indica que no hi ha relació lineal entre les variables.
2. El valor numèric del coeficient indica la força o intensitat de la relació (els valors absoluts grans indiquen una correlació forta entre les dues variables i els valors absoluts petits indiquen una correlació feble).
3. El signe del coeficient indica la direcció de la correlació.
4. El valor positiu més elevat és +1 i el valor negatiu més elevat és -1.

També cal tenir present que com més gran sigui la mostra més probabilitats hi haurà que el resultat de l'anàlisi correlacional s'acosti al que passa realment a les variables en la població. Si la mostra és petita, pot passar que:

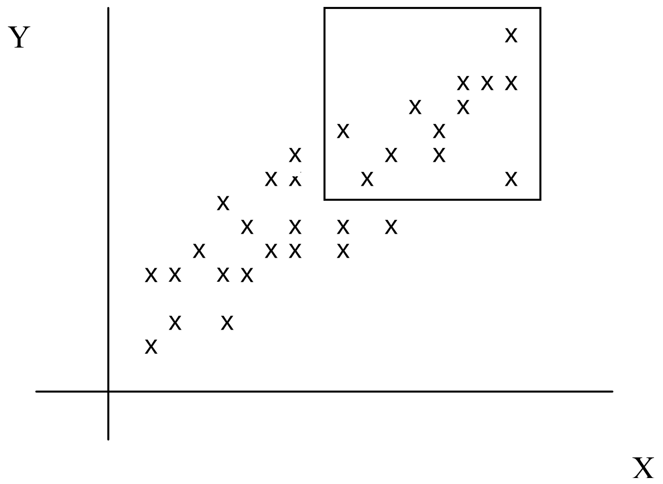
— *Ens surti que hi ha correlació, però en realitat aquesta és deguda a l'atzar*, de manera que, si agaféssim una mostra més gran, veuríem que realment no n'hi ha, de correlació. Al gràfic 7, si ens fixem en els punts que hi ha dins del requadre, veurem que sembla que hi ha correlació. Si féssim l'anàlisi només amb aquests valors, és probable que ens sortís un índex elevat. No obstant això, si agafem una mostra molt més gran, veiem que, en realitat, és molt poc probable que hi hagi correlació entre les dues variables.

Gràfic 3.7



— *Ens surti que no hi ha correlació, però, en realitat, sí que n'hi ha.* Si agafem una mostra més gran, podem veure que el núvol de punts que semblava molt dispers no ho és tant. El gràfic 8 representa aquesta situació.

Gràfic 3.8.



3.2.1. El coeficient de correlació de Pearson

El coeficient de correlació de Pearson ens indica si hi ha correlació lineal entre dues variables. Es representa amb una *r*.
Per poder aplicar el coeficient de Pearson cal que, com a mínim, es compleixin les condicions següents:

1. Les variables han d'estar mesurades en una escala d'interval.
2. Hem de tenir un mínim de 30 casos per analitzar, per tant, 30 parells de valors.
3. Es parteix de la base que la distribució de cada variable segueix el model de la corba normal. Teòricament caldria analitzar si la distribució de valors segueix aquesta distribució abans de decidir aplicar el coeficient de Pearson. Per raons pràctiques, en els exercicis donarem per fet que es compleix aquesta condició.

També cal tenir en compte que el coeficient de Pearson es pot buscar siguin quines siguin les unitats del sistema de mesura que s'hagi utilitzat (sempre que es tracti d'una escala d'interval). És a dir, podem correlacionar una variable mesurada amb metres amb una de mesurada en grams. Podem buscar si hi ha correlació entre els resultats de dues proves encara que una hagi estat puntuada de l'1 al 10 i l'altra de l'1 al 100.

La fórmula és la següent:

$$r_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2]\,[N\Sigma Y^2 - (\Sigma Y)^2]}} \qquad (8)$$

N = nombre de parelles de valors

Interpretació

Ja hem dit que el coeficient de correlació oscil·la entre els valors -1 i +1, sent menor com més s'acosta a 0 i major com més s'acosta a -1 o a +1. Amb tot, per interpretar si la correlació és significativa, hem de recórrer a les taules de la *r* de Pearson, que trobareu a l'annex 1 d'aquest document.

Per interpretar la *r* de Pearson cal buscar el valor crític de *r*. Això es fa tenint en compte el marge d'error (normalment utilitzarem un marge d'error del 0,05 o del 0,01, mirant la significació bilateral) i els graus de llibertat, que són N - 2.

A continuació comparem la *r* que ens ha sortit a nosaltres utilitzant la fórmula amb la *r* que ens surt a les taules. En aquest cas sempre es comparen els valors absoluts, és a dir, no es té en compte el signe que surt a l'índex de correlació. Cal tenir present que el valor de les taules ens indica quin ha de ser el valor mínim de *r* perquè puguem considerar que la correlació és significativa. Per tant:

- Si la *r observada* és més gran que la *r de la taula*, direm que hi ha correlació significativa.

- Si la *r observada* és més petita que la *r de la taula*, direm que no hi ha correlació significativa.

Així, per exemple, imaginem-nos que busquem el coeficient de correlació de Pearson entre l'edat i l'alçada d'una mostra de 37 nens d'entre 1 i 10 anys. Ens surt que $r = 0{,}825$. Podem pensar que la correlació serà alta, però, per estar-ne segurs i veure si la correlació és realment significativa, hem de recórrer a les taules que tenim a l'annex 1. Farem la interpretació per a un marge d'error de 0,05.

Anem a les taules i mirem el valor crític de *r* per a un marge d'error de 0,05 i $N - 2$ graus de llibertat, que en aquest cas són 35. El valor que trobem a les taules és 0,325. Aquest valor ens indica el valor mínim que ha de tenir *r* perquè puguem dir que la correlació és significativa. Com que la *r* observada és més gran que la *r* crítica, direm que hi ha correlació entre l'edat i l'alçada.

$r = 0{,}825$
$r_c = 0{,}325$ $\quad r > r_c$ Per tant, hi ha correlació entre les dues variables.

Un altre exemple
Imaginem-nos que hem recollit informació sobre el nombre d'anys d'estudi d'una mostra de 100 persones de 20 anys i informació sobre el nombre de faltes d'ortografia que han fet en un dictat de 200 paraules. Busquem el coeficient de correlació de Pearson i ens surt que és $r = -0{,}245$. Volem saber si hi ha correlació entre aquestes dues variables per a un nivell de significació del 95,5 % (això és el mateix que dir que es vol saber per a un marge d'error de 0,05).

En aquest cas, veiem que es tracta d'un índex força baix, però, tal com hem fet abans, per saber si la correlació és significativa ens cal mirar les taules.

r_c (α = 0,05 i 98 g. ll.) = 0,205 (No existeix un valor exacte per a 98 graus de llibertat. En aquest cas, podríem decidir agafar el corresponent a 100 o el corresponent a 90. Per norma general, agafem el valor crític que sigui més gran.)

$r > r_c$, és a dir, $|0,245| > |0,205|$. Per tant, podrem concloure que hi ha correlació significativa a un nivell de significació del 95,5 %. La correlació és negativa. Això vol dir que, segons aquest resultat, com més anys d'estudis, menys faltes d'ortografia han comès.

(Recordeu que comparem els valors absoluts de la *r* observada i de la *r* crítica.)

Tal com expliquen Welkowitz, Ewen i Cohen (1991), si una *r* de Pearson és estadísticament significativa, aquesta significació denota un cert grau de relació lineal entre les dues variables de la població. No obstant això, no indica una relació significativament elevada o significativament forta; només indica la improbabilitat d'una relació nul·la entre ambdues variables de la població. Cal observar que, com més gran és la mida de la mostra, més petit és el valor del coeficient de correlació necessari perquè hi hagi significació estadística. És a dir, cal tenir no només una significació estadística, sinó també un valor absolut de *r* prou alt abans de decidir que probablement la correlació a la població és prou gran per indicar una relació prou forta.

És important recordar que la *r* de Pearson ens indica <u>només si hi ha correlació lineal</u> entre dues variables. Amb aquest coeficient, doncs, no es pot veure si la relació és curvilínia o d'un altre tipus. Per això calen altres tipus de coeficients, dels quals no parlarem en aquest curs d'introducció a l'estadística.

3.2.2. Correlació i causalitat

Tal com assenyalen Etxeberria i Tejedor (2005), a vegades es confon l'existència de relació entre variables amb la possible existència d'una relació de causa-efecte entre elles, però això són conceptes molt diferents. Entendre correctament el concepte de correlació implica entendre que el coeficient no permet determinar la causa de la relació entre dues variables. Pot haver-hi una correlació alta entre dues variables per tres possibles raons:

a) *Les dues variables varien de manera simultània.* Per exemple, entre el pes i l'alçada de les persones hi ha una relació. Les persones més altes solen pesar més, però no és que el pes sigui la causa de l'alçada o l'alçada del pes; simplement, varien de manera conjunta.

b) Una és la causa i l'altra, l'efecte. Però perquè sigui així cal que:
    a. Hi hagi relació entre la variable causa i la variable efecte.

b. La *variable causa* s'ha de produir abans que la *variable conseqüència*.
c. No ha d'existir una variable que mediatitzi aquesta relació.

Per exemple, podríem trobar que l'autoestima i el rendiment acadèmic estan relacionats, però difícilment podríem dir que una baixa autoestima pot ser una causa del baix rendiment acadèmic, ja que és molt difícil determinar que l'autoestima s'hagi produït abans del rendiment acadèmic. Senzillament, es tracta de dues característiques que estan associades, però no podrem dir que una sigui conseqüència de l'altra.

c) Tant la variable *X* com la variable *Y* són causades per una tercera variable, que és la causa de totes dues i de la relació entre elles. Per exemple, hi ha correlació entre l'alçada dels nens i els resultats en lectoescriptura, és a dir, els nens més alts tenen millors resultats. Això és així perquè hi ha una tercera variable, l'edat, que és la causa que els nens siguin més alts i, per tant, més grans, i llegeixin i escriguin millor que els més baixets, que són més petits.

El problema de l'atribució de la causalitat és un problema lògic o científic, no estadístic. Cal ser molt conscients que l'existència de relació estadística entre variables no és suficient per parlar de relació causa-efecte entre elles. Tal com diuen Etxeberria i Tejedor (2005), l'existència de correlació és una condició necessària però no suficient per establir la causalitat entre dues variables.

3.2.3. Consideracions finals

La utilització del coeficient de correlació de Pearson implica el supòsit que hi ha una relació lineal entre dues variables. Si la relació que existeix entre dues variables no és lineal sinó curvilínia, la *r* de Pearson no la detectarà.

Quan es comprova la significació estadística d'un coeficient de correlació, se suposa que la distribució implicada és normal bivariada, és a dir, les puntuacions *Y* es distribueixen normalment per cada valor de *X* i, de manera recíproca, els valors de *X* es distribueixen normalment per cada valor de *Y*. No obstant això, quan els graus de llibertat són més de 25 o 30, l'incompliment d'aquest supòsit té poca influència sobre la validesa de la prova (Welkowitz, Ewen i Cohen, 1981: 199-200).

Els coeficients de correlació no es poden interpretar com a percentatges. Per exemple, no es pot dir que una correlació de 0,80 és el 80 % d'una relació perfecta o que equival al doble d'una correlació de 0,40. En canvi, el quadrat del coeficient de correlació sí que permet una interpretació en termes de percentatge de la intensitat de relació entre dues variables.

Tal com diu Calvo (1987), els coeficients de correlació no formen una escala quantitativa d'unitat constant; és a dir, entre una $r = 0,80$ i una $r = 0,70$ no existeix la mateixa diferència que entre una $r = 0,40$ i una $r = 0,50$. Tampoc no és veritat que $r = 0,50$ sigui igual al doble de $r = 0,25$.

Com més heterogènia és la població, més força té el coeficient de correlació de Pearson. Si la població és molt homogènia, *r* serà més petita. Per exemple, la correlació entre l'alçada i el salt de longitud és molt alta, però si només agafem les persones que fan més d'1,80 d'alçada, veurem que en aquest grup, pel fet de ser més homogeni, la correlació de l'alçada amb el salt de longitud serà menor (Calvo, 1987: 110).

Cal conèixer la naturalesa de les variables per deduir la importància de la correlació. Així, per exemple, una $r = 0{,}30$ entre el pes i la capacitat intel·lectual és alta (perquè en principi és difícil que pugui haver-hi una correlació entre aquestes dues variables). En canvi, una $r = 0{,}80$ entre la pressió arterial a les 10 del matí i a les 11 del matí és baixa. També cal analitzar si la correlació diu alguna cosa que és real o no.

Per ajudar-nos a interpretar la correlació podem recórrer als valors de Guilford (citat per Calvo, 1987: 110):

	Correlació	**Relació**
$0 < r < 0{,}20$	Petita	Molt poc intensa
$0{,}20 < r < 0{,}40$	Baixa	Petita però apreciable
$0{,}40 < r < 0{,}60$	Regular	Considerable
$0{,}60 < r < 0{,}80$	Alta	Intensa
$0{,}80 < r < 1{,}00$	Molt alta	Molt intensa

Coeficient de determinació r^2

r^2 indica el tant per u de la variació entre les variables. És a dir, indica quina proporció representa la correlació trobada de la correlació perfecta. O, dit d'una altra manera, quina proporció de la variabilitat total de la variable *Y* queda explicada per la variable *X* (Calvo, 1987: 110).

Exercici de correlació

Ens preguntem si hi ha correlació significativa entre les notes obtingudes al final del primer quadrimestre i al final del segon quadrimestre pels estudiants de l'assignatura Bases metodològiques de la investigació educativa. Farem la interpretació amb un nivell de significació del 95,5 %.

Per fer aquest exercici d'exemple, treballarem amb una mostra de 30 estudiants, escollits a l'atzar, del curs 2004-2005. A continuació, tenim la taula amb els valors de les dues variables.

Haurem d'aplicar la fórmula de la *r* de Pearson, que és:

$$r_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2\,]\,[N\Sigma Y^2 - (\Sigma Y)^2]}}$$

Per tant, els càlculs que ens caldrà fer són els que apareixen a la taula següent, al costat de les dades.

	Nota del primer quadrimestre (X)	Nota del segon quadrimestre (Y)	XY	X^2	Y^2
1	5	7,32	36,60	25	53,58
2	6	6,56	39,36	36	43,03
3	4,7	6,79	31,91	22,09	46,10
4	7	7,61	53,27	49	57,91
5	6,7	6,93	46,43	44,89	48,02
6	8,3	8,38	69,55	68,89	70,22
7	6,4	6,55	41,92	40,96	42,90
8	7,8	7,21	56,24	60,84	51,98
9	7,8	8,77	68,41	60,84	76,91
10	7,3	6,49	47,38	53,29	42,12
11	6,7	4,76	31,89	44,89	22,66
12	7,2	7,1	51,12	51,84	50,41
13	9	9,22	82,98	81	85,01
14	7,1	7,67	54,46	50,41	58,83
15	4	7,36	29,44	16	54,17
26	6,2	7,43	46,07	38,44	55,20
17	5,8	6,16	35,73	33,64	37,95
18	4	6,69	26,76	16	44,76
19	6	7,48	44,88	36	55,95
20	5,5	6,77	37,24	30,25	45,83
21	6	5,87	35,22	36	34,46
22	6,8	8,76	59,57	46,24	76,74
23	5,4	7,44	40,18	29,16	55,35
24	4,7	6,77	31,82	22,09	45,83
25	5,9	8,38	49,44	34,81	70,22
26	5,9	7,56	44,60	34,81	57,15
27	5,5	5,25	28,88	30,25	27,56
28	8,2	7,96	65,27	67,24	63,36
29	8,2	7,18	58,88	67,24	51,55
30	7,8	8,88	69,26	60,84	78,85
	ΣX=192,9	**ΣY=217,3**	**ΣXY=1414,74**	**ΣX^2=1288,95**	**ΣY^2=1604,66**

Apliquem la fórmula (8):

$$r_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2]\,[N\Sigma Y^2 - (\Sigma Y)^2]}} =$$

$$= \frac{30 \,.\, 1414{,}74 - (192{,}9 \,.\, 217{,}3)}{\sqrt{[30 \,.\, 1288{,}95 - (192{,}9)^2]\,[30 \,.\, 1604{,}66 - (217{,}3)^2]}} =$$

$$= \frac{42.442{,}2 - (41.917{,}17)}{\sqrt{[38.668{,}5 - (37.210{,}41)]\,[48.139{,}8 - (47.219{,}29)]}} =$$

$$= \frac{525{,}03}{\sqrt{[1458{,}09]\,[920{,}51]}} = \frac{525{,}03}{1.158{,}53} = 0{,}453$$

Per tant, $r_{xy} = 0{,}453$.

Ara, per fer la interpretació ens cal anar a les taules i buscar la *r* crítica. Per *N*-2 graus de llibertat i $\alpha = 0{,}05$ (el marge d'error corresponent al 95,5 % de nivell de significació).

r_c (g. ll. = 28; $\alpha = 0{,}05$) = 0,361

0,361 < 0,453

Per tant, podem dir que, amb un marge d'error del 0,05, hi ha correlació significativa entre les notes del primer quadrimestre i les notes del segon quadrimestre dels estudiants que el curs 2005-2006 van fer l'assignatura Bases metodològiques de la investigació educativa.

Mirant les taules de Guilford veiem que la correlació és regular, cosa que implica una relació considerable.

Si busquem el coeficient de determinació r^2 veiem que és de 0,205. Això vol dir que un 20,5 % de la variabilitat de les notes del segon quadrimestre s'explica per les notes que s'obtenen el primer quadrimestre.

Un altre exemple

Volem saber si hi ha correlació entre les despeses per alumne i curs a educació primària a diferents països de la UE i les despeses en educació (expressades en percentatge del PIB). Les dades apareixen a la taula següent. Fes la interpretació per a un marge d'error tant del 0,05 com del 0,01.

	Despeses per alumne i curs escolar a ed. primària	Despeses d'educació en percentatge del PIB als països de l'euro
Espanya	3635	4,9
Alemanya	3818	5,3
França	4139	6,1
Portugal	3478	5,7
Irlanda	3018	4,6
Itàlia	5354	4,9
Bèlgica	3952	5,5
Àustria	6568	5,7
Finlàndia	4138	5,6
Països Baixos	4162	4,7
Grècia	2176	4

A continuació tens els càlculs fets, només cal que apliquis la fórmula (8):

	Despeses per alumne i curs escolar a ed. primària	Despeses d'educació en percentatge del PIB als països de l'euro	XY	X^2	Y^2
Espanya	3635	4,9	17811,5	13213225	24,01
Alemanya	3818	5,3	20235,4	14577124	28,09
França	4139	6,1	25247,9	17131321	37,21
Portugal	3478	5,7	19824,6	12096484	32,49
Irlanda	3018	4,6	13882,8	9108324	21,16
Itàlia	5354	4,9	26234,6	28665316	24,01
Bèlgica	3952	5,5	21736	15618304	30,25
Àustria	6568	5,7	37437,6	43138624	32,49
Finlàndia	4138	5,6	23172,8	17123044	31,36
Països Baixos	4162	4,7	19561,4	17322244	22,09
Grècia	2176	4	8704	4734976	16
	$\Sigma X = 44438$	$\Sigma Y = 57$	$\Sigma XY = 233848{,}6$	$\Sigma X^2 = 192728986$	$\Sigma Y^2 = 299{,}16$

Apliquem la fórmula de la *r* de Pearson.

$$r_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2]\,[N\Sigma Y^2 - (\Sigma Y)^2]}} =$$

$$= \frac{(11 \cdot 233848{,}6) - (44438 \,.\, 57)}{\sqrt{[(11 \cdot 192728986) - (44438)^2]\,[(11 \cdot 299{,}16) - (57)^2]}} = 0{,}505$$

Per fer la interpretació busquem la r crítica a les taules:

r_c (g. ll. = 9; α = 0,05) = 0,602

r_c (g. ll. = 9; α = 0,01) = 0,735

Interpretació:

Per als dos marges d'error veiem que la *r* que hem obtingut aplicant la fórmula és menor que la *r* que apareix a les taules. Per tant, podem dir que tant amb un marge d'error de 0,05 com amb un marge d'error de 0,01, no hi ha correlació significativa entre les despeses per alumne i curs escolar i les despeses d'educació en percentatge del PIB. És a dir, no podem afirmar que les despeses per alumne i curs escolar a educació primària tinguin res a veure amb el percentatge del PIB dedicat a educació.

Tot i això, cal fer un advertiment, i és que aquesta vegada hem aplicat el coeficient de Pearson per fer un exercici amb pocs valors. En aquest cas, en què només tenim dades d'onze països, la mostra és excessivament reduïda perquè la informació que extraiem mitjançant la *r* de Pearson sigui realment fiable.

3.3. Predicció i regressió lineal

Tal com dèiem al començament d'aquest capítol, si dues variables estan correlacionades, és a dir, si dues variables varien de forma conjunta, podrem explicar els canvis que es produeixen en una d'aquestes (variable criteri) a partir de les dades de les variables que s'anomenen *variables predictores*.

Per exemple, si la variable criteri que volem analitzar és l'alçada dels nens d'1 a 10 anys, podem tenir diferents variables predictores que poden explicar les variacions d'alçada dels nens d'aquesta edat, entre les quals podem esmentar l'edat, el sexe, el número que calcen, etc. Si treballem només amb una variable predictora, estarem parlant de regressió simple, i si en tenim diverses, de regressió múltiple (Etxeberria i Tejedor, 2005: 211). Nosaltres només parlarem de la *regressió simple.*

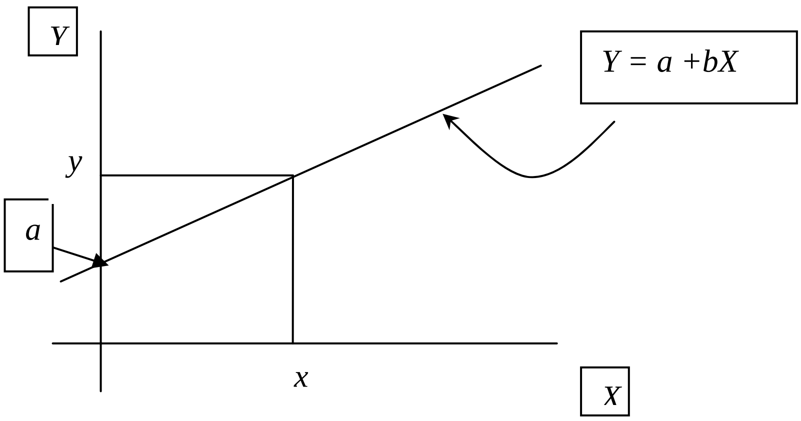
3.3.1. Regressió simple

Per poder fer prediccions d'una variable criteri (*Y*) a partir d'una variable predictora (*X*), el primer que cal és que, una vegada establert que efectivament existeix una correlació lineal significativa entre les dues variables, busquem la *recta de regressió.*

En general, se sap que l'equació d'una recta té l'estructura següent:

$Y = a + bX$

Y = el valor predit de la variable criteri
b = pendent de la recta
X = el valor de la variable predictora
a = valor de Y quan X = 0



Recordem que la recta o línia de regressió és la recta que representa millor la tendència dels punts en un diagrama de dispersió. Per trobar aquesta recta, cal buscar el coeficient *b* de l'equació i després buscar la constant *a*.

La fórmula per trobar el coeficient *b* (que representarem com a b_{yx} per expressar que es tracta de la recta de regressió de *Y* sobre *X*) és la següent:

$$b_{yx} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N\Sigma X^2 - (\Sigma X)^2} \quad (9)$$

Una vegada es té el valor de b_{yx}, es busca el valor de a_{yx} a partir dels valors de la mitjana de la variable X i de la mitjana de la variable Y:

$$a_{yx} = \overline{Y} - b_{yx}\,\overline{X} \qquad (10)$$

Per exemple, abans hem trobat correlació entre les notes del primer quadrimestre i les del segon quadrimestre de l'assignatura Bases metodològiques de la investigació educativa. Per tant, ara buscarem la recta de regressió.

Primer, cal buscar el coeficient b, amb la fórmula (9):

$$b_{yx} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N\Sigma X^2 - (\Sigma X)^2} = \frac{(30 \cdot 1414{,}74) - (192{,}9 \cdot 217{,}3)}{(30 \cdot 1288{,}95) - 192{,}9^2} =$$

$$= \frac{42442{,}2 - 41917{,}17}{38668{,}5 - 37210{,}41} = \frac{525{,}03}{1458{,}09} = \mathbf{0{,}36}$$

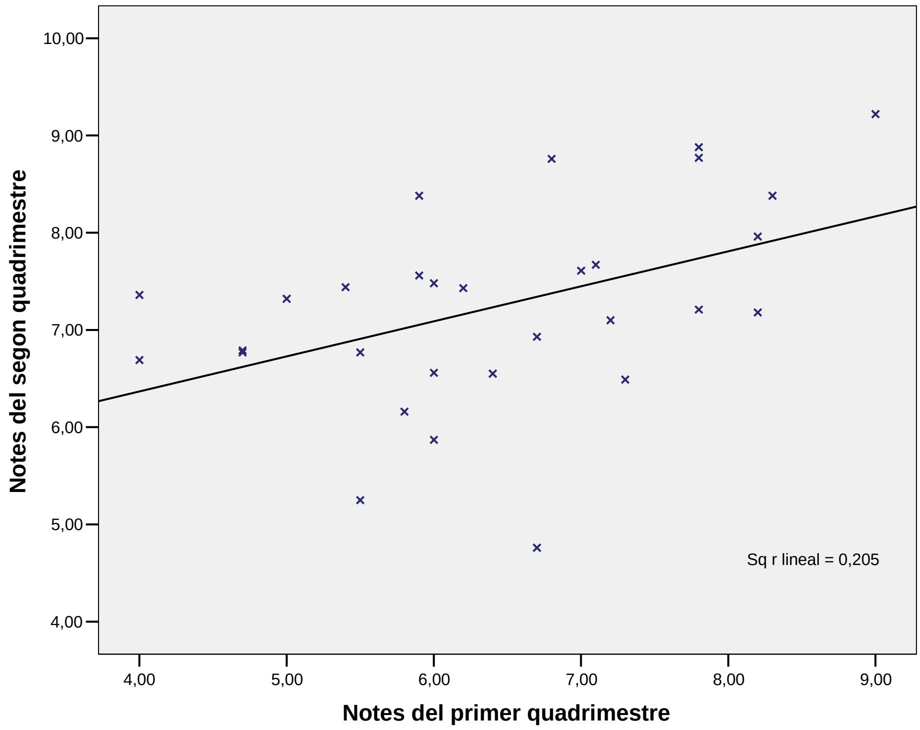
Ara ens cal buscar la a, amb la fórmula (10). Per a això, hem de trobar la mitjana de la variable X i la mitjana de la variable Y.

$\overline{X} = 6{,}43 \qquad \overline{Y} = 7{,}24$

$a_{yx} = \overline{Y} - b_{yx}\,\overline{X} = 7{,}24 - (0{,}36 \cdot 6{,}43) = \mathbf{4{,}93}$

L'equació de la recta és, per tant, $\boldsymbol{Y = 4{,}93 + 0{,}36X}$.

Gràfic 3.9: Diagrama de dispersió i recta de regressió de les notes de les variables "notes del primer quadrimestre" i "notes del segon quadrimestre" de la mostra de 30 casos



Sabent quina és la recta de regressió per a aquestes dues variables, se suposa que si coneixem el valor d'una de les variables podrem predir el valor de l'altra. Així, ens podem preguntar quina serà la puntuació del segon quadrimestre per estudiant que obtingui una puntuació de 5 el primer quadrimestre. L'únic que hem de fer és substituir la X pel valor de 5 i sortirà Y:

$$Y = 4{,}93 + 0{,}36 \cdot 5 = \mathbf{6{,}73}$$

3.3.2. Error típic de predicció

La puntuació de 6,73 seria la puntuació que es troba exactament a sobre de la recta de regressió per al valor de $X = 5$. Ara bé, cal tenir en compte que quan es fa un pronòstic o predicció, sempre es comet el que es coneix com a *error de predicció*. És a dir, si miréssim realment quina puntuació treu del segon quadrimestre un estudiant que hagi tret una puntuació de 5 al final del primer quadrimestre, difícilment coincidiria exactament amb el valor predit de 6,73. Per a cada valor de Y que predim es comet un error de predicció. Els matemàtics ens indiquen com es pot calcular el que es coneix com a *error típic de predicció*, que és aproximadament una mitjana de tots els errors de predicció que es cometen per a cada valor de X observat. La fórmula per calcular l'error típic de predicció és la següent:

$\sigma_{y'} = \sigma_y \sqrt{1-r^2}$ σ_y = desviació típica de la variable Y
r^2 = quadrat del coeficient de correlació

Aquest error típic de l'estimador de *Y* pot ser interpretat com la dispersió de *Y* al voltant de la línia de regressió.

Per al cas que ens ocupa, l'error típic de predicció serà:

$\sigma_{y'} = \sigma_y \sqrt{1-r^2} \quad = \quad 1{,}03\sqrt{1 - 0{,}453^2} = 0{,}92$

Una predicció serà bona com més petit sigui l'error típic de predicció. En termes generals, si l'error típic de predicció és menor que la desviació típica de la variable *Y*, podem dir que la predicció feta serà bastant bona. Si, en canvi, l'error típic de predicció és més gran que la dispersió de la variable *Y*, llavors la predicció no serà gaire fiable. Tal com diuen Welkowitz, Ewen i Cohen (1981), en aquest segon cas el coneixement d'un valor concret de *X* no permetrà fer una bona predicció sobre del valor que tindrà la *Y* que se li associa. La predicció serà bona en la mesura que el coneixement de *X* redueixi la dispersió en la predicció de *Y*.

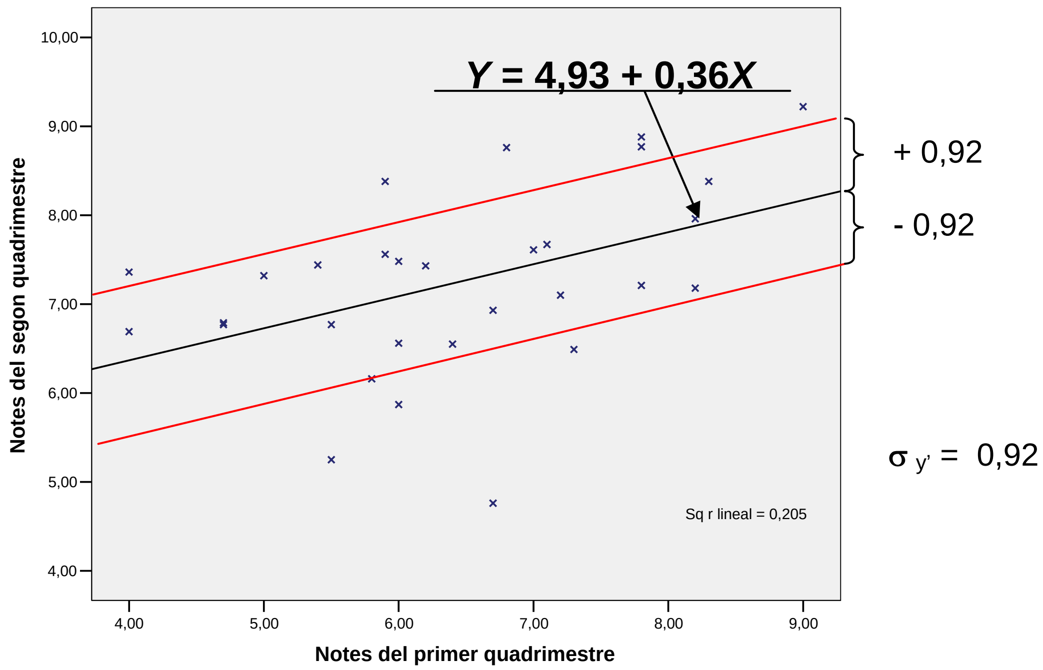
En l'exemple, 0,92 és menor que la desviació típica de la variable *Y*, que és d'1,03, encara que la diferència no és gaire gran.

Si sumem i restem el valor de l'error típic de predicció a la predicció feta per a una *X* = 5, aconseguirem l'interval de valors de *Y* on serà més probable trobar el valor predit. Així, sabent que el valor predit amb la recta de regressió és de 6,73 (aquest seria el valor exacte damunt de la recta), podem sumar i restar l'error de predicció:

6,73 + 0,92 = 7,65
6,73 – 0,92 = 5,81

Per tant, un estudiant que obtingui una puntuació de 5 el primer quadrimestre probablement traurà una puntuació d'entre 5,81 i 7,65 el segon quadrimestre.

Gràfic 3.10: Recta de regressió i error de predicció de les notes de les variables "notes del primer quadrimestre" i "notes del segon quadrimestre" de la mostra de 30 casos

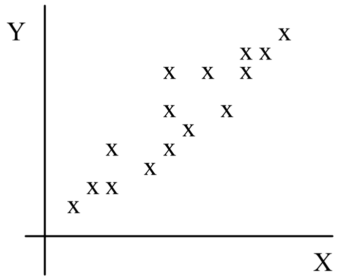


3.4. Exercicis de correlació lineal

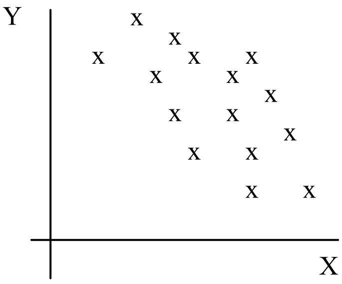
1. Entre els coeficients de correlació següents, escull els més adients als diagrames de dispersió que es presenten:

$r = 0{,}44$ $r = -0{,}12$ $r = -0{,}54$ $r = -0{,}89$ $r = 0{,}92$

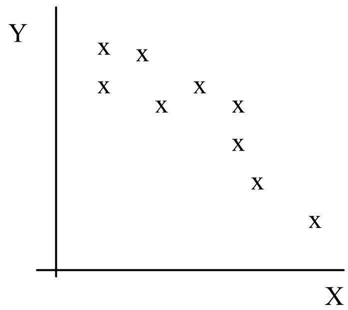
1.1



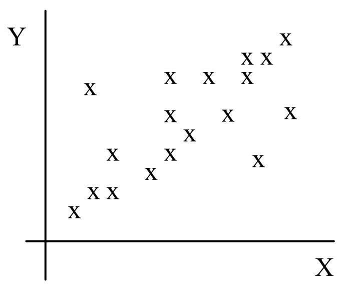
1.2



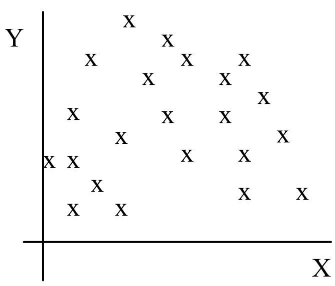
1.3



1.4



1.5



2. A una mostra escollida a l'atzar de 10 alumnes d'una escola se'ls ha passat un test estandarditzat sobre relacions lògiques (test 1) i un altre de càlcul mental (test 2). Els resultats obtinguts són els següents:

Cas	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Test 1	22	20	18	17	16	14	14	12	10	7
Test 2	14	18	12	16	14	12	11	10	9	4

Volem saber si existeix algun tipus de relació entre ambdues proves. Suposem, malgrat el nombre d'elements de la mostra, que es donen les condicions per aplicar proves paramètriques.

2.1. Quina prova aplicaràs? Per què?

2.2. Aquest coeficient és estadísticament significatiu per a un nivell de confiança del 95 %?

2.3. Interpreta'l.

3. Volem saber si hi ha relació entre les habilitats socials i l'acceptació per part del grup d'iguals en nois i noies que realitzen el segon curs d'ESO. Per això es passa un test d'habilitats socials a una mostra de 30 persones escollides a través d'un procediment de mostreig aleatori simple. Per cada alumne també es mesura el

grau d'acceptació del grup d'iguals a partir d'una escala que es passa al grup classe. Contesta les preguntes següents:

3.1. Per saber si hi ha relació entre habilitats socials i acceptació per part del grup d'iguals, quina prova has aplicat?

3.2. És estadísticament significativa? Per què? (Analitza-ho tant per a un marge d'error de $\alpha = 0{,}05$ com per a un marge d'error de $\alpha = 0{,}01$.)

3.3. Interpreta el resultat.

3.4. Es pot buscar la recta de regressió? Si es pot, busca-la i també calcula l'error típic de predicció. (Busca la recta considerant les habilitats socials com a variable predictora (X) i el grau d'acceptació com a variable criteri (Y).)

3.5. Quina predicció podem fer per al cas d'un estudiant que tingui una puntuació de 6,5 en la prova d'habilitats socials?

Cas	**Habilitats socials**	**Grau d'acceptació**
1	8	10
2	7	12
3	6	9
4	5	15
5	9	14
6	3	6
7	4	7
8	5	11
9	7	12
10	8	13
11	8	10
12	4	6
13	9	12
14	7	7
15	8	12
16	6	9
17	5	10
18	5	14
19	3	5
20	2	4
21	4	8
22	5	9
23	8	10
24	6	9
25	8	11
26	9	12
27	10	10
28	9	9
29	8	11
30	3	7

4. A la matriu de la pàgina següent tens les dades de 100 casos. Són 100 persones adultes amb discapacitat psíquica profunda a les quals s'ha passat dues proves: una avalua les destreses socials i comunicatives i l'altra, les destreses de vida personal. Una puntuació alta en aquestes proves significa un nivell més alt de destreses. Es vol saber si hi ha alguna relació entre les puntuacions de destreses socials i comunicatives i les puntuacions de destreses de vida personal. Suposem que la distribució de les dues variables segueix el model de la corba normal. S'aplica una determinada prova mitjançant el programa SPSSx i s'obtenen els resultats que apareixen a continuació.

		Puntuació de destreses socials i comunicatives	Puntuació de destreses de vida personal
Puntuació de destreses socials i comunicatives	Correlació de Pearson	1	,819
Puntuació de destreses de vida personal	Correlació de Pearson	,819	1

4.1. Quina prova s'ha aplicat?

4.2. Quin índex s'ha obtingut? Interpreta'l segons un nivell de confiança del 95 % (signe, intensitat).

4.3. En cas que hi hagi correlació, busca el coeficient de determinació i explica què significa.

4.4. Si hi ha correlació, busca la recta de regressió. (Per fer això, utilitza els càlculs de la taula de la pàgina següent.)

4.5. Una persona que obtingui una puntuació de 20 en la prova de destreses socials i comunicatives, quina puntuació tindrà en la prova de destreses de vida personal?

Destreses socials i comunicatives	Destreses de vida personal	***XY***	X^2	Y^2
18	16	288	324	256
3	22	66	9	484
16	17	272	256	289
21	30	630	441	900
9	10	90	81	100
15	16	240	225	256
17	19	323	289	361
26	39	1014	676	1521
21	33	693	441	1089
12	25	300	144	625
1	13	13	1	169
17	20	340	289	400

6	11	66	36	121
12	29	348	144	841
32	41	1312	1024	1681
24	30	720	576	900
1	2	2	1	4
14	18	252	196	324
14	21	294	196	441
20	28	560	400	784
24	38	912	576	1444
0	13	0	0	169
32	37	1184	1024	1369
31	38	1178	961	1444
41	43	1763	1681	1849
32	35	1120	1024	1225
28	33	924	784	1089
29	30	870	841	900
30	29	870	900	841
35	34	1190	1225	1156
20	29	580	400	841
28	38	1064	784	1444
13	28	364	169	784
18	28	504	324	784
8	16	128	64	256
27	41	1107	729	1681
15	23	345	225	529
8	14	112	64	196
21	31	651	441	961
22	13	286	484	169
11	8	88	121	64
15	24	360	225	576
33	29	957	1089	841
26	35	910	676	1225
24	28	672	576	784
13	34	442	169	1156
42	49	2058	1764	2401
13	16	208	169	256
3	7	21	9	49
3	5	15	9	25
17	26	442	289	676
9	4	36	81	16
41	38	1558	1681	1444
40	40	1600	1600	1600
25	35	875	625	1225
12	21	252	144	441
19	16	304	361	256
19	24	456	361	576
15	17	255	225	289
17	17	289	289	289
12	17	204	144	289
16	24	384	256	576

15	22	330	225	484
31	27	837	961	729
25	37	925	625	1369
32	35	1120	1024	1225
18	21	378	324	441
9	13	117	81	169
30	27	810	900	729
34	33	1122	1156	1089
28	36	1008	784	1296
32	38	1216	1024	1444
21	38	798	441	1444
32	40	1280	1024	1600
22	29	638	484	841
22	18	396	484	324
22	32	704	484	1024
22	22	484	484	484
22	27	594	484	729
8	8	64	64	64
29	24	696	841	576
46	49	2254	2116	2401
8	10	80	64	100
9	10	90	81	100
2	6	12	4	36
1	16	16	1	256
12	30	360	144	900
4	25	100	16	625
13	19	247	169	361
1	7	7	1	49
20	26	520	400	676
32	41	1312	1024	1681
24	37	888	576	1369
4	33	132	16	1089
9	9	81	81	81
37	41	1517	1369	1681
8	26	208	64	676
3	1	3	9	1
13	19	247	169	361
18	29	522	324	841
ΣX = 1904	ΣY = 2506	ΣYX = 57474	ΣX^2 = 47834	ΣY^2 = 75076

5. S'ha passat una prova sobre conducta adaptativa a una mostra de 239 persones adultes amb discapacitat psíquica. Aquesta prova té un bloc compost per les puntuacions següents: destreses motores, destreses socials i comunicatives, destreses de vida personal i destreses de vida en comunitat. Per cada persona s'ha obtingut una puntuació de cada una d'aquestes variables. A continuació hi ha una taula amb l'explicació de cada un d'aquests grups de destreses.

Escala	Aspectes avaluats
Destreses motores	• Destreses de motricitat fina i bàsica. • Destreses relatives a la mobilitat. • Forma física. • Coordinació motora general. • Coordinació visuomotora. • Precisió de moviments.
Destreses socials i comunicatives	• Destreses implicades en la interacció social de diferents entorns. • Comprensió i expressió del llenguatge transmès a través de signes, de forma escrita o de forma oral.
Destreses de vida personal	• Capacitat del subjecte per satisfer les seves necessitats d'autonomia personal, principalment a la llar, però també en altres entorns socials: destreses relacionades amb el menjar i la seva preparació; destreses relacionades amb l'ús del servei; vestit; cura de si mateix; habilitats domèstiques.
Destreses de vida en comunitat	• Habilitats necessàries per a un ús adequat dels recursos i serveis de la societat. • Capacitat per respondre adequadament als requeriments econòmics i socials del món laboral i altres situacions socials: ús del rellotge, capacitat per ser puntual, diners i valor de les coses, destreses relacionades amb l'àmbit laboral, sentit de l'orientació a la llar i a la comunitat.

S'ha aplicat el coeficient de correlació de Pearson i s'ha obtingut la matriu de correlacions següent:

		Puntuació escala destreses motores	Puntuació escala destreses socials	Puntuació escala destreses de vida personal	Puntuació escala destreses de vida en comunitat
Puntuació escala destreses motores	Correlació de Pearson	1	,709	,856	,753
Puntuació escala destreses socials	Correlació de Pearson	**,709**	1	,725	,793
Puntuació escala destreses de vida personal	Correlació de Pearson	**,856**	**,725**	1	,753
Puntuació escala destreses de vida en comunitat	Correlació de Pearson	**,753**	**,793**	**,753**	1

Digues quines correlacions resulten significatives per a un marge d'error del = 0,05.

6. Seguint amb el mateix estudi, es vol analitzar si hi ha relació entre el nivell de conductes adaptatives i els problemes de conducta. Per a això, s'aplica la prova de correlació de Pearson entre les diferents variables de conducta adaptativa (destreses motores, destreses socials, destreses de vida personal, destreses de vida en comunitat) i els índexs de problemes de conducta (índex intern, índex extern, índex asocial i índex general). L'*índex intern* inclou els problemes de conducta autolesiva, hàbits atípics i falta d'atenció; *l'índex extern* inclou la conducta heteroagressiva, la destrucció d'objectes i la conducta disruptiva, i

l'índex asocial inclou la conducta social ofensiva i la conducta no col·laboradora.

També cal tenir molt present, perquè afecta la interpretació de la correlació, que en el cas de les puntuacions de les variables de conducta adaptativa, un valor alt indica un nivell alt de destreses. *En el cas dels índexs de problemes de conducta, els valors més alts indiquen menys problemes de conducta i, doncs, els valors més negatius indiquen més problemes de conducta*. Una vegada puntualitzat això, podem observar en la taula següent la matriu de correlacions:

		Índex intern de problemes de conducta	Índex extern de problemes de conducta	Índex asocial de problemes de conducta
Puntuació escala destreses motores	Correlació de Pearson	,148	-,234	-,169
Puntuació escala destreses socials	Correlació de Pearson	,313	-,163	-,078
Puntuació escala destreses de vida personal	Correlació de Pearson	,112	-,201	-,134
Puntuació escala destreses de vida en comunitat	Correlació de Pearson	,269	-,123	-,061

Interpreta quines resulten significatives a un nivell de confiança del 95 % i digues quin tipus de relació és, així com el seu significat.

3.5. Respostes als exercicis de correlació i regressió

Exercici 1

1.1. 0,92

1.2. -0,54

1.3. -0,89

1.4. 0,44

1.5. -0,12

Exercici 2

2.1. Apliquem la prova de la *r* de Pearson tot i que, de fet, aquesta no seria la prova adequada, ja que tenim massa pocs valors. Aquí es tracta de fer un exercici per comprovar si sabem aplicar la fórmula de la *r* de Pearson, i per això posem pocs valors.

Cas	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	
Test 1	22	20	18	17	16	14	14	12	10	7	$\Sigma X = 150$
Test 2	14	18	12	16	14	12	11	10	9	4	$\Sigma Y = 120$
XY	308	360	216	272	224	168	154	120	90	28	$\Sigma YX = 1940$
X^2	484	400	324	289	256	196	196	144	100	49	$\Sigma X^2 = 2438$
Y^2	196	324	144	256	196	144	121	100	81	16	$\Sigma Y^2 = 1578$

$r = 0{,}869$

2.2. Per saber si és significatiu, ens cal mirar la *r* crítica a les taules de l'annex 1, per a *N* - 2 graus de llibertat i per a un marge d'error del 0,05 (corresponent al 95,5 % de nivell de significació). Mirant les taules veiem que

$$r_c \text{ (g. ll. 8; } \alpha = 0{,}05) = 0{,}632$$

$$0{,}632 < 0{,}869$$

Per tant, podem dir que hi ha correlació significativa entre les dues variables, per a un nivell de confiança del 95,5 %.

2.3. Suposant, doncs, que hàgim aplicat la prova correcta, podem dir que, amb un nivell de confiança del 95,5 %, existeix correlació significativa entre les puntuacions de la prova de relacions lògiques i les puntuacions de la prova de càlcul mental. Com que la correlació és positiva, podem dir que els alumnes que treuen una puntuació elevada en la prova de relacions lògiques tendeixen a treure també una puntuació elevada en la prova de càlcul mental, i que els alumnes que treuen puntuacions baixes en la prova sobre relacions lògiques, treuen normalment puntuacions baixes també en la prova de càlcul mental.

Exercici 3

3.1. Apliquem la prova del coeficient de correlació de Pearson.

3.2.

Cas	Habilitats socials	Grau d'acceptació	*XY*	X^2	Y^2
1	8	10	80	64	100
2	7	12	84	49	144
3	6	9	54	36	81
4	5	15	75	25	225
5	9	14	126	81	196
6	3	6	18	9	36
7	4	7	28	16	49
8	5	11	55	25	121
9	7	12	84	49	144
10	8	13	104	64	169
11	8	10	80	64	100
12	4	6	24	16	36
13	9	12	108	81	144
14	7	7	49	49	49
15	8	12	96	64	144
16	6	9	54	36	81
17	5	10	50	25	100
18	5	14	70	25	196
19	3	5	15	9	25
20	2	4	8	4	16
21	4	8	32	16	64
22	5	9	45	25	81
23	8	10	80	64	100
24	6	9	54	36	81
25	8	11	88	64	121
26	9	12	108	81	144
27	10	10	100	100	100
28	9	9	81	81	81
29	8	11	88	64	121
30	3	7	21	9	49
	$\Sigma X = 189$	$\Sigma Y = 294$	$\Sigma YX = 1959$	$\Sigma X^2 = 1331$	$\Sigma Y^2 = 3098$

$r = 0{,}612$

r_c (g. ll. 28; $\alpha = 0{,}05$) = 0,361

r_c (g. ll. 28; $\alpha = 0{,}01$) = 0,463

En ambdós casos la *r* que hem observat és més gran que la *r* de les taules. Per tant, podem dir que tant per a un marge d'error de 0,05 com per a un marge d'error de 0,01, hi ha correlació significativa entre aquestes dues variables.

3.3. Per tant, podem dir que, tant amb un nivell de significació del 95 % com amb un nivell de significació del 99 %, hi ha correlació entre les proves d'habilitats socials i d'acceptació per part del grup d'iguals en nois i noies que fan segon d'ESO. El fet que l'índex surti positiu significa que una puntuació alta en habilitats socials implica una puntuació alta en el grau d'acceptació. Generalitzant una mica més, podríem dir que, com més nivell d'habilitats socials, més grau d'acceptació per part del grup d'iguals. I a l'inrevés, com menys nivell d'habilitats socials, menys nivell d'acceptació per part del grup d'iguals.

3.4. Es pot buscar la recta de regressió perquè ja hem vist que hi ha correlació significativa.

$$b_{yx} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N\Sigma X^2 - (\Sigma X)^2} = \frac{30 \cdot 1959 - (189 \cdot 294)}{(30 \cdot 1331) - (189)^2} =$$

$$= \frac{58770 - 55566}{39930 - 35721} = \frac{3204}{4209} = 0{,}761$$

$a_{yx} = \overline{Y} - b_{yx}\overline{X} = 9{,}8 - 0{,}761(6{,}3) = 5{,}01$

La recta de regressió serà ***Y* = 5,01 + 0,761*X***

Per calcular l'error de predicció necessitem la desviació típica de *Y*, que en aquest cas és 2,73.

$\sigma_{y'} = \sigma_y \sqrt{1-r^2} = 2{,}73 \sqrt{1-0{,}612^2} = 2{,}16$

2,16 és menor que la desviació típica de la variable *Y*, per la qual cosa la predicció que podem fer és bastant ajustada.

3.5. Per fer la predicció cal substituir el valor de *X* a la recta de regressió. En aquest cas la recta és:

Y* = 5,01 + 0,761*X $\quad Y = 5{,}01 + 0{,}761 \cdot 6{,}5 = 9{,}96$

El valor *Y* = 9,96 seria la predicció exacta, és a dir, és el valor que es troba damunt la recta quan *X* = 6,5. Però hem de tenir en compte l'error típic de predicció que podem cometre en aquest cas.

9,96 + 2,16 =12,12
9,96 – 2,16 = 7,8

Per tant, en aquest cas, un alumne que hagi tret una puntuació de 6,5 en la prova d'habilitats socials és probable que obtingui una puntuació d'entre 7,8 i 12,12 en la prova de grau d'acceptació per part del grup d'iguals.

Exercici 4

4.1. S'ha aplicat el coeficient de correlació de Pearson.

4.2. S'ha obtingut un índex de $r = 0{,}819$. Per interpretar-lo hem de buscar el valor de r crítica.

r_c (g. ll. 98; a = 0,05) = 0,205
$0{,}205 < 0{,}819$

Per tant, podem concloure que la correlació entre les dues variables és significativa. Es tracta d'una correlació positiva. Si mirem els valors de Guilford, observem que l'índex és molt alt; per tant, la relació és molt intensa entre les dues variables. Dit d'una altra manera, amb un nivell de confiança del 95 % podem afirmar que hi ha una correlació positiva entre les puntuacions de destreses socials i comunicatives i les de destreses de vida personal en les persones adultes amb discapacitat psíquica. Com que la correlació és positiva, una puntuació alta en la prova de destreses socials i comunicatives es correspon amb una puntuació alta en la prova de destreses de vida personal, i a l'inrevés, una puntuació baixa en la prova de destreses socials i comunicatives s'associa a puntuacions baixes de la prova de destreses de vida personal.

4.3. $r^2 = 0{,}671$

Significa que el 67,1 % de la variació de les puntuacions de la prova de destreses de vida personal s'explica per la variable "destreses socials i comunicatives"; per tant, hi ha una elevada dependència entre elles.

4.4. L'equació de la recta de regressió és $\boldsymbol{Y = 9{,}015 + 0{,}843X}$.

4.5. Cal substituir la X de l'equació de la recta pel valor 20.

$Y = 9{,}015 + 0{,}843 \cdot 20 = 25{,}86$

Fixem-nos, però, que el valor 25,86 seria el valor de l'ordenada que es troba sobre la recta de regressió. Cal recordar que per fer prediccions hem de buscar l'error típic de predicció. Per això necessitem la desviació típica de la variable Y (destreses de vida personal), que és 11,14:

$$\sigma_{y'} = \sigma_y\sqrt{1-r^2} = 11{,}14\sqrt{1-0{,}819^2} = 6{,}39$$

$25{,}86 + 6{,}39 = 32{,}25$
$25{,}86 - 6{,}39 = 19{,}47$

Per un valor de 25,86 de destreses socials i comunicatives podem trobar valors entre 19,47 i 32,25 en destreses de vida personal.

Exercici 5

Per a un marge d'error del 0,05 i 237 graus de llibertat, el valor de la *r* crítica és 0,195. Mirant la taula observem que totes les correlacions són significatives i de caràcter positiu. Això vol dir que una puntuació alta de destreses socials i comunicatives es relaciona amb puntuacions altes de destreses motores, de destreses de vida personal i de destreses de vida en la comunitat. I així amb totes les relacions. La correlació més elevada és la correlació entre la puntuació de destreses de vida personal i la puntuació de destreses motores ($r = 0{,}856$). La resta són correlacions intenses. Això ens indica que hi ha una elevada dependència entre aquests tipus de destreses en les persones adultes amb discapacitat psíquica.

Exercici 6

Mirant les taules, per a = 0,05 i 237 graus de llibertat la *r* crítica és 0,195. Ens cal mirar quins valors de *r* són més grans que aquest 0,195. A la taula següent posem els índexs que indiquen una correlació significativa per a un nivell de confiança del 95 %.

Correlacions significatives entre índexs de problemes de conducta i puntuacions de les escales de conducta adaptativa

	Índex intern	**Índex extern**	**Índex asocial**
Destreses motores		$r = -0{,}234$	
Destreses socials i comunicatives	$r = 0{,}313$		
Destreses de vida personal		$r = -0{,}201$	
Destreses de vida en comunitat	$r = 0{,}269$		

Observem que les correlacions entre les variables referides a conducta adaptativa i l'índex extern de problemes de conducta resulten negatives. A la taula següent apareix la interpretació que podem fer per cada una d'aquestes correlacions.

Hem marcat en negreta les correlacions que resulten més significatives segons els resultats de les proves. Així, pel que fa a la correlació entre les puntuacions en destreses motores i els diferents índexs de problemes de conducta, observem que la relació més significativa es dóna entre les destreses motores i l'índex extern de problemes de conducta, de manera que un nivell alt en destreses motores es correspon amb un nivell alt en els problemes de conducta relacionats amb l'índex extern, que —recordem-ho— engloba l'heteroagressivitat, la destrucció d'objectes i la conducta disruptiva.

Correlació entre índexs de problemes de conducta i puntuacions de les escales de conducta adaptativa. Interpretació.

	Índex intern (c. autolesiva, hàbits atípics i falta d'atenció)	**Índex extern (heteroagressivitat, destrucció d'objectes i c. disruptiva)**	**Índex asocial (c. social ofensiva i conducta no col·laboradora)**
Destreses motores		La relació que s'obté és negativa, és a dir, les puntuacions altes en la variable "destreses motores" es relacionen amb puntuacions baixes en la variable "índex intern". Per tant, un **nivell alt de destreses motores** es correspon amb un **nivell alt també dels problemes de conducta** inclosos en aquest índex**, i un nivell baix de destreses motores** tendeix a correspondre's **amb nivells baixos dels problemes de conducta** inclosos en aquest índex**.**	
Destreses socials i comunicatives	Un **nivell alt de destreses socials i comunicatives** es relaciona amb un **nivell baix de problemes de conducta** d'aquest grup.		
Destreses de vida personal		La *r* és negativa. Per tant, una puntuació alta en destreses de vida personal es correlaciona amb puntuacions baixes de l'índex extern, la qual cosa significa que **nivells alts de destreses de vida personal** es corresponen amb **nivells alts de problemes de conducta** d'aquest grup (ja que una puntuació baixa indica l'existència d'altres problemes de conducta).	
Destreses de vida en comunitat	**Com més destreses de vida en comunitat menys problemes de conducta** d'aquest grup.		

En el cas de la relació entre les puntuacions de destreses socials i comunicatives i els diferents índexs de problemes de conducta, trobem que únicament hi ha relació entre aquests tipus de destreses i l'índex intern. Per tant, un nivell alt en destreses socials i comunicatives es relaciona amb un nivell baix de problemes de conducta vinculats amb l'índex intern (conducta autolesiva, hàbits atípics, retraïment i falta d'atenció).

Quant a les destreses de vida personal, es correlacionen amb l'índex extern de problemes de conducta, la qual cosa significa que nivells alts de destreses de vida personal es corresponen amb nivells alts en els problemes de conducta inclosos en l'índex extern (heteroagressivitat, destrucció d'objectes i conducta disruptiva).

Pel que fa a les destreses de vida en comunitat, l'única correlació significativa la trobem entre les puntuacions d'aquest tipus de destreses i l'índex intern de problemes de conducta. Per tant, un nivell alt de destreses de vida en comunitat es correspon amb una puntuació alta en l'índex intern, i recordem que puntuacions altes en aquest índex signifiquen menys problemes de conducta autolesiva, conductes atípiques i falta d'atenció.

En conjunt observem que els nivells més alts de destreses dels diferents tipus es corresponen amb nivells baixos dels problemes de conducta relacionats amb l'índex intern, és a dir, sembla que tenir un nivell més alt de conducta adaptativa implica menor nivell de problemes de conducta relacionats amb la conducta autolesiva, els hàbits atípics i la falta d'atenció.

4. Introducció a l'estadística inferencial

Recordem que quan es fa una investigació, les poblacions estadístiques resulten sovint massa àmplies per poder-les estudiar. Una opció és recollir dades només d'una part —una mostra— de la població, però és possible que els resultats obtinguts de l'estudi d'aquesta part no siguin aplicables al conjunt de la població. Aquest problema de no poder abastar tota la població que es vol estudiar és comú en totes les ciències socials i humanes. La solució és utilitzar un conjunt de tècniques que permeten fer inferències sobre una població completa basant-se en l'estudi de dades d'una mostra d'aquesta població. Aquest conjunt de tècniques és el que es coneix com a *estadística inferencial.*

Seguint Welkowitz, Ewen i Cohen (1986), els *objectius de l'estadística inferencial* són:

— Estimar el valor de la mitjana i la variància de la població a partir de les dades d'una mostra de la qual es coneix la mitjana i la variància. Qualsevol estadístic calculat sobre una sola mostra, com la mitjana o la desviació típica ($\bar{X}$ o S), per produir una estimació del paràmetre poblacional corresponent (μ o σ), s'anomena *estimador puntual*.

— Buscar un *interval de confiança*, que és un conjunt de valors amb una probabilitat coneguda de contenir el veritable valor del paràmetre poblacional, ja que a vegades l'estimador puntual és insuficient perquè no ens indica fins a quin punt s'acosta al paràmetre poblacional real.

— Avaluar la probabilitat d'obtenir certs tipus de resultats mostrals quan es compleixen determinades condicions poblacionals. És a dir, conegudes les característiques d'una població (μ o σ), estimar la probabilitat que una mostra extreta d'aquesta població tingui unes determinades característiques ($\bar{X}$ o S).

4.1. Estratègia general de l'estadística inferencial

Per explicar quina és l'estratègia que segueix l'estadística inferencial, posarem un exemple que podeu trobar a Welkowitz, Ewen i Cohen (1986:108-112).

Imaginem-nos que dos amics —en Joan i la Fina—juguen a fer apostes amb una moneda. En Joan aposta 6 euros al resultat de llançar una moneda i la Fina demana l'opció "cara". Llancen la moneda a l'aire i, efectivament, surt cara, de manera que en Joan perd els 6 euros que havia apostat. Decideix que vol recuperar els euros i proposa a la Fina tornar a jugar. La Fina segueix demanant cara i segueix sortint cara en el llançament següent. Ho proven sis vegades seguides i cada vegada surt cara, que és l'opció que demana la Fina.

Després de sis llançaments, en Joan comença a sospitar. La Fina sempre demana cara i sempre li surt el mateix. Si la moneda fos correcta, podria continuar jugant i, a la llarga, aniria recuperant els diners perduts (que ja són 36 euros, encara que, per al que volem explicar, això és el menys important). Però en Joan pensa que si la Fina està fent trampes i la moneda està trucada, hauria de plantar-se i dir a la Fina que és una tramposa.

Per tant, arribats a aquest punt del joc, tenim dues hipòtesis possibles:

La hipòtesi 0: els resultats són deguts a la sort, a l'atzar. La moneda és correcta i l'únic que passa és que la Fina té una bona ratxa.

La hipòtesi 1: la moneda està trucada i té tendència que surti cara. En aquest cas, doncs, els resultats no serien deguts a l'atzar sinó a un factor (la moneda té truc) que fa que els resultats difereixin dels que serien esperables.

A la taula següent podem veure un esquema de la situació en què es troba en Joan, pel que fa a les dues possibles decisions que pot prendre.

		Situació real o estat de la població	
		La hipòtesi 0 és certa: els resultats són deguts a la sort	**La hipòtesi 1 és certa: la moneda està trucada**
Decisió d'en Joan	No rebutja la hipòtesi 0 (segueix jugant)	Decisió correcta	Error: continua donant diners a una amiga que fa trampes
	Rebutja la hipòtesi 0 (abandona el joc i acusa la Fina de fer trampes)	Error: calúmnia injusta a una amiga honesta	Decisió correcta

La situació real o *estat de la població* no es pot conèixer només amb els sis llançaments. Una possibilitat seria observar la població completa. El problema és que la població, en aquest cas, és el conjunt de tots els llançaments de moneda que es poden fer, és a dir, observar infinits llançaments. Si la moneda és correcta, trobaríem que en la meitat dels llançaments surt cara i en l'altra meitat surt creu. Però resulta impossible, en aquest cas, incloure tota la població.

Una altra possibilitat seria observar una mostra de molts llançaments, 100 o 500. Això permetria tenir una mostra prou gran per observar el comportament de la moneda. No obstant això, imaginem-nos que en Joan ha de prendre una decisió a partir d'aquests sis llançaments de la moneda. Per això diem que hi ha dues possibles hipòtesis sobre quin pot ser l'estat de la població. En Joan ha de decidir quina de les dues hipòtesis accepta, és a dir, ha de prendre una decisió sense saber l'estat real de la població.

Per resoldre aquest problema s'han d'explicar —o recordar— algunes definicions i càlculs bàsics sobre la probabilitat d'un succés.

La probabilitat

La probabilitat d'un succés donat es defineix amb la fracció següent:

$$P(\text{succés}) = \frac{\text{Nombre de casos en què el succés pot ocórrer (o nombre de casos favorables)}}{\text{Nombre total de successos (o casos) possibles.}}$$

Per exemple, la probabilitat que surti cara en llançar una moneda a l'aire és la següent:

$$P\ (\text{cara}) = \frac{\text{Nombre de casos en què pot sortir cara en un llançament}}{\text{Nombre de successos possibles en un llançament}} = \frac{1\ (\text{cara})}{2\ (\text{cara o creu})} = 0{,}5$$

Per tant, la probabilitat que surti cara és 0,5, és a dir, del 50 %. Per tant, si la moneda és correcta, sortirà cara la meitat de les vegades i creu l'altra meitat.

Més exemples:

Suposem que tenim un joc de 48 cartes, de les quals 12 són d'or, 12 de bastons, 12 de copes i 12 d'espases.

- Quina és la probabilitat que agafant una carta a l'atzar surti una copa?

$$P\ (\text{copes}) = \frac{12\ \text{copes}}{48\ \text{cartes}} = \frac{1}{4} = 0{,}25$$

Passat a percentatge, podem dir que tenim el 25 % de probabilitat que surti copa.

- Quina és la probabilitat que surti l'as d'or?

$$P\ (\text{as d'or}) = \frac{1}{48} = 0{,}021$$

Tenim un 2,1 % de probabilitat que, agafant una carta del joc a l'atzar, la que ens surti sigui l'as d'oros.

- Quina és la probabilitat que agafant una carta a l'atzar surti un rei?

$$P\ (\text{rei}) = \frac{4}{48} = \frac{1}{12} = 0{,}083$$

Tenim, doncs, una probabilitat del 8,3 % que en extreure una carta del joc a l'atzar ens surti un rei.

La probabilitat no pot ser més petita que 0 ni més gran que 1. Si un succés té probabilitat 0 vol dir que no és possible que ocorri; si té probabilitat 1 vol dir que aquest succés segur que passarà, i no en pot passar cap altre.

Fins ara hem parlat de la probabilitat d'un sol succés cada vegada. Ara parlarem de dos successos que puguin passar alhora.

La probabilitat de A o B

La probabilitat que s'esdevingui un dels dos successos es calcula amb la fórmula següent:

$P(A \text{ o } B) = P(A) + P(B) - P\ (A \text{ i } B)$

Per exemple, si volem saber quina és la probabilitat que en extreure a l'atzar una carta del joc surti un rei o una espasa, la probabilitat serà:

P (rei o espasa) = P(rei) + P(espasa) - P (rei d'espases) = 4/48 + 12/48 – 1/48 = 0,31

Vegem per què és així. Quants casos favorables tenim?

- Per una banda, els 4 reis: rei d'or, rei de copes, rei d'espases i rei de bastons.
- Per l'altra, tenim 12 espases: l'as, el dos, el tres, el quatre, el cinc, el sis, el set, el vuit, el nou, el deu, el cavall i el rei.
- Sembla, doncs, que hi ha 16 casos favorables: 4 reis i 12 espases. No obstant això, fixem-nos que el rei d'espases l'hem comptat dues vegades, una com a rei i l'altra com a espasa. Per tant, per calcular la probabilitat real, a la suma dels dos successos hem de restar-hi la probabilitat dels successos que s'han comptat dues vegades.

Si els successos A i B no poden ocórrer de manera simultània (és a dir, són mútuament excloents), llavors la $P\ (A \text{ i } B)$ és igual a 0. Per exemple, si volem calcular la probabilitat que en treure una carta a l'atzar surti un cavall o un rei, tenim que :

P (cavall o rei) = P (cavall) + P (rei) – P (cavall i rei) = 4/48 +4/48 – 0 = 0,17

La probabilitat que surti alhora un cavall i un rei és 0. Les cartes, o són cavalls o són reis, però no les dues figures al mateix temps.

La probabilitat de A i després B

Suposem que extraiem una carta del joc, la mirem i la tornem al joc. Després barregem les cartes i n'extraiem una altra. Volem saber la probabilitat que en la primera extracció surti un rei i que en la segona extracció surti l'as d'or. Les dues extraccions són independents, és a dir, que tornem la primera carta al joc, de manera que el que hagi passat en la primera extracció no afecta el que pugui passar en la segona. La solució es representa amb la fórmula següent:

P (A i després B) = P (A) · P (B)

En l'exemple, la resposta és:

P (rei i després as d'or) = 4/48 · 1/48 = 4/2304 = 0,0017

Si els dos successos no fossin independents, és a dir, si després de la primera extracció la primera carta no fos retornada al joc, la probabilitat seria:

P (rei i després as d'or) = 4/48 · 1/47 = 4/2256 =0,0018

En aquest cas hem de restar una carta dels casos possibles relatius a l'extracció d'una carta que sigui l'as d'or perquè la primera carta no s'ha retornat, i per tant, al joc només hi queden 47 cartes.

Quina és la probabilitat que en extreure dues cartes a l'atzar la primera sigui un or i la segona un cavall? (Tornem la carta al joc després de la primera extracció.)

P (or i després cavall) = P (or) · P (cavall) = 12/48 · 4/48 = 48/2304 =0,021

Ara tornem als dos amics que estaven jugant a tirar la moneda a l'aire. Havíem dit que en Joan havia de decidir si seguia jugant o no, després de sis llançaments en els quals havia sortit sempre cara. El que pot fer en Joan és calcular la probabilitat que en tirar la moneda surti sis vegades seguides cara.

Es tracta de buscar la probabilitat que ocorrin sis successos independents, és a dir, que surti cara al primer llançament, cara al segon, cara al tercer... i així fins a sis llançaments:

P (cara i després cara i després cara i després cara i després cara i després cara) =

= ½ · ½ · ½ · ½ · ½ · ½ = 1/64 = 0,016

Ara ja sabem que, si la moneda no està trucada, la probabilitat que surti cara sis vegades seguides és de 0,016. Per tant, o bé la moneda és correcta i s'ha esdevingut un succés amb probabilitat 0,016 (hipòtesi 0) o bé la moneda està trucada i la probabilitat que surti cara no és de ½ i, per tant, la probabilitat que surtin 6 cares consecutives no serà del 0,016 (hipòtesi 1). Ara el problema és saber si una probabilitat de 0,016 és prou elevada per acceptar que el fet que hagin sortit 6 cares seguides és fruit de l'atzar, o bé és massa petita i llavors podem sospitar que la moneda està trucada.

Tal com diuen Welkowitz, Ewen i Cohen (1986: 117-118), "el científic segueix l'arbitrària però raonable pràctica de considerar els successos amb probabilitats menors de 0,05 com a 'suficientment improbables per justificar el rebuig de la hipòtesi 0' i els successos amb probabilitats més grans de 0,05 com a 'no suficientment improbables' per justificar aquest rebuig".

Seguint aquesta norma, doncs, haurem de concloure que una probabilitat de 0,016 és massa petita perquè sigui fruit de l'atzar el fet que surtin 6 cares en 6 llançaments consecutius. Per tant, aquest succés és molt improbable i haurem de rebutjar la hipòtesi 0. La millor decisió que pot prendre en Joan davant d'aquest resultat és abandonar el joc i no arriscar-se a seguir perdent. No obstant això, cal fer notar que el que fa en Joan és prendre una decisió a partir de les dades que té. L'estadística li fa veure que el succés és poc probable, però realment ell no sap si la moneda està trucada. En Joan pot estar cometent un error. L'estadística indica el camí cap a la millor decisió sobre la població, a partir dels resultats obtinguts en la mostra.

Fins aquí hem vist l'estratègia que segueix l'estadística inferencial per arribar a decisions sobre la població a partir de les dades d'una mostra. Ho hem vist amb un exemple senzill, que és el llançament d'una moneda. En situacions més complexes, l'estratègia és la mateixa, si bé el càlcul de la probabilitat que ocorri un succés no sol ser tan simple. Ens interessarà buscar, per exemple:

- La probabilitat que sigui degut a l'atzar el fet que la mitjana de les notes d'estadística d'un curs sigui més elevada que la del curs anterior.

- La probabilitat que, aplicant un programa educatiu per millorar les habilitats socials en adolescents, la mitjana d'habilitats d'aquests joves augmenti significativament.

4.2. El model de la corba normal

En el cas de la moneda que hem vist fins ara, és relativament fàcil calcular les probabilitats perquè sabem que a cada llançament la probabilitat que surti cara o creu és de ½. Però què passa quan volem saber quina és la probabilitat que, en extreure una mostra a l'atzar d'una població, ens surti una mitjana determinada?

Una manera de fer-ho —encara que gens pràctica— seria extreure un nombre gran de mostres, totes de la mateixa mida, de la població definida, calcular la mitjana per a cada una i comptar les freqüències amb què surten els valors de cada mitjana. Les mostres han de ser aleatòries i representatives de la població. És a dir, cal que en seleccionar la mostra cada element de la població tingui la mateixa probabilitat de ser escollit i que totes les mostres possibles de la mateixa mida tinguin la mateixa probabilitat de ser escollides. Si calculem la mitjana de cada mostra podríem dibuixar la distribució de les freqüències. Això ens ajudaria a fer-nos una idea de la freqüència amb què podem esperar una mitjana donada.

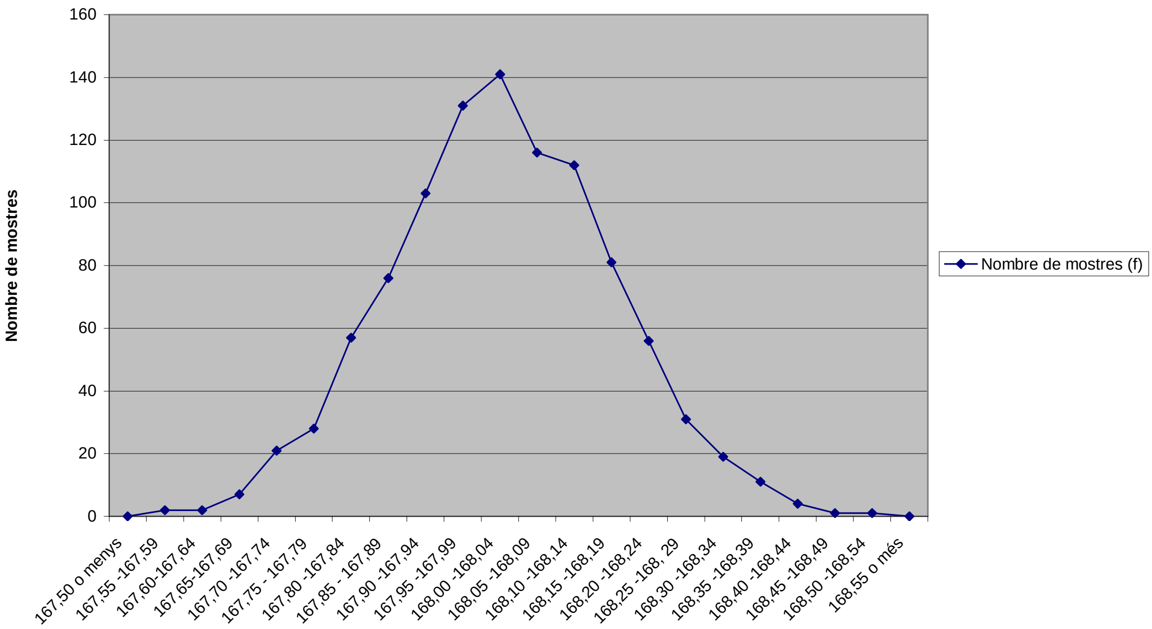
Imaginem-nos que hem buscat 1.000 mostres de la població, cada una de N = 100, i hem buscat la mitjana d'alçada de cadascuna. A la taula 4.1 podríem tenir la distribució de freqüències de les alçades mitjanes.

Taula 4.1:Distribució hipotètica de freqüències de les alçades mitjanes en 1.000 mostres de 100 casos cada una[6]

Alçada mitjana (cm)	Nombre de mostres (f)	Proporció de mostres f/N
167,50 o menys	0	0
167,55 -167,59	2	0,002
167,60-167,64	2	0,002
167,65-167,69	7	0,007
167,70 -167,74	21	0,021
167,75 - 167,79	28	0,028
167,80 -167,84	57	0,057
167,85 - 167,89	76	0,076
167,90 -167,94	103	0,103
167,95 -167,99	131	0,131
168,00 -168,04	141	0,141
168,05 -168,09	116	0,116
168,10 -168,14	112	0,112
168,15 -168,19	81	0,081
168,20 -168,24	56	0,056
168,25 -168, 29	31	0,031
168,30 -168,34	19	0,019
168,35 -168,39	11	0,011
168,40 -168,44	4	0,004
168,45 -168,49	1	0,001
168,50 -168,54	1	0,001
168,55 o més	0	0
Total	1000	1

[6] Aquest exemple s'ha extret de Welkowitz, Ewen i Cohen (1986), pàgina 124.

Gràfic 4.1: Distribució hipotètica de le alçades mitjanes de 1000 mostres de N = 100

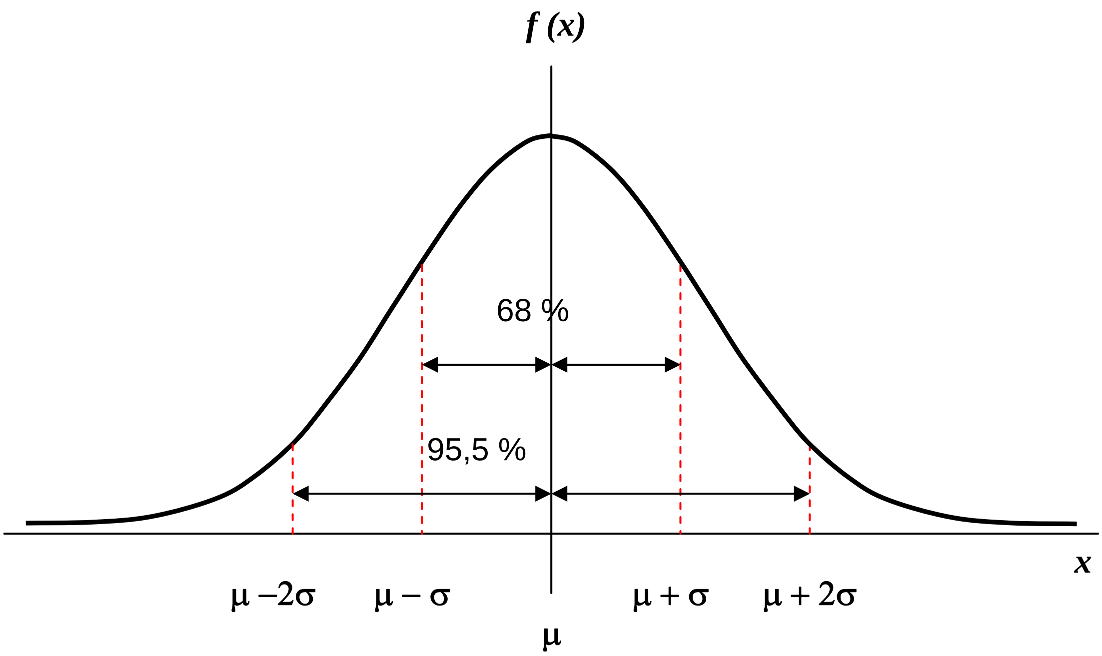


La distribució empírica de mostres serveix com a model estadístic per assignar probabilitats a qualsevol observació d'alçada mitjana. La columna "Proporció de mostres" de la taula 4.1 ens indica la probabilitat que tenim d'obtenir una mostra d'una alçada mitjana determinada. És a dir que, si hem extret de la població 76 mostres amb mitjana d'alçada de 167,85 i 167,89 cm (casos favorables), del total de 1.000 mostres (casos possibles), la probabilitat que qualsevol mostra de mida *N* = 100 tingui una alçada d'entre 167,85 i 167,89 cm és de 76/1.000 = 0,076. Dit d'una altra manera, si d'aquesta població haguéssim extret moltes mostres de mida *N* = 100, podríem esperar que aproximadament el 7,6 % d'aquestes mostres tinguessin mitjanes d'entre 167,85 i 167,89 cm. Basant-nos en aquesta distribució, també podríem saber que la probabilitat d'obtenir una mostra amb una mitjana de 168,00 cm o menys és de 0,141 + 0,131 + 0,103 + 0,076 + 0,057 + 0,028 + 0,021+ 0,007 + 0,002 + 0,002 = 0,568, és a dir, del 56,8 %.

De totes maneres, si haguéssim de seguir aquest procediment per cada variable que es vol estudiar seria molt costós. A la pràctica s'utilitzen distribucions mostrals teòriques i no empíriques. És a dir, els matemàtics han pogut mostrar que moltes distribucions mostrals tendeixen a seguir *el model de la corba normal*. És molt freqüent, doncs, que la distribució mostral de les mitjanes segueixi aquest model. Això permet utilitzar el model de la corba normal per calcular probabilitats en molts estudis estadístics. És a dir, no caldrà calcular cap model empíric per calcular probabilitats, sinó que podrem basar-nos en el model de la corba normal que ja ha estat descrit i estudiat pels matemàtics i estadístics. Això ens permetrà comparar els resultats d'una sola mostra amb els del model teòric.

Aquest model de la corba normal ja l'havíem començat a descriure. Com en el cas de qualsevol distribució de freqüències, l'eix horitzontal de la distribució normal representa els valors observats i a l'eix vertical es representen les freqüències observades.

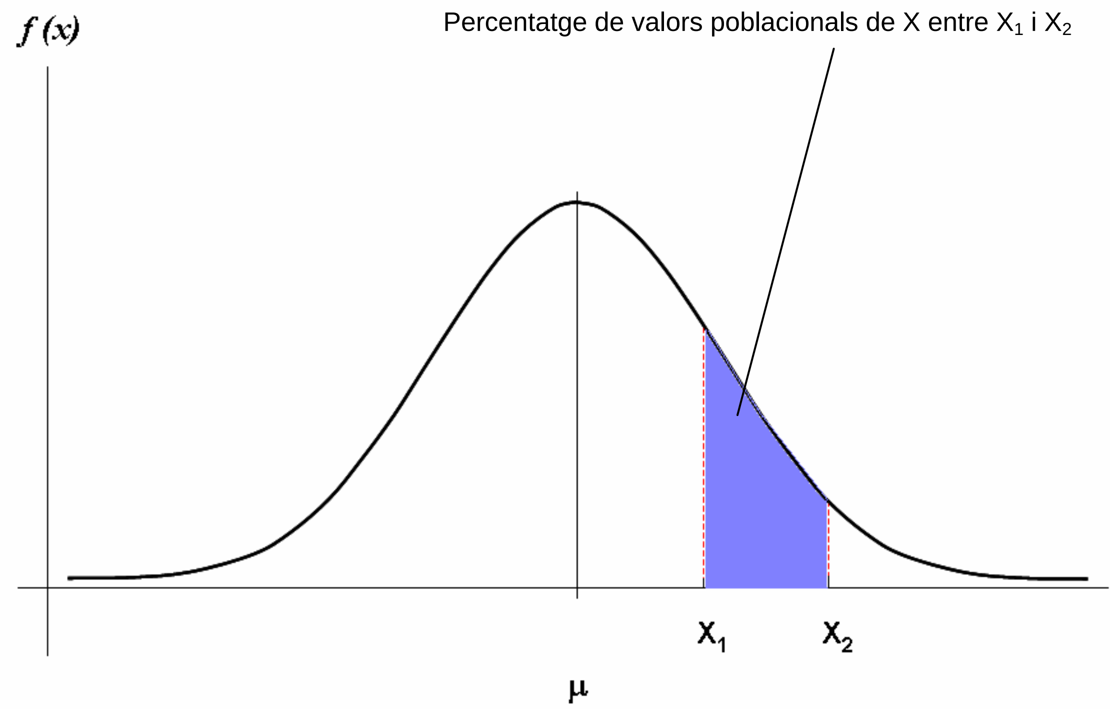
Gràfic 4.2: El model de la corba normal



L'àrea total sota de la corba representa la població total. Aquesta àrea es defineix com el 100 % dels casos de la població.

Suposem que ara ens demanen que trobem la proporció de la població amb puntuacions entre els punts X_1 i X_2 (gràfic 3). La representació gràfica d'aquesta proporció és l'àrea que queda sota la corba i entre aquests dos punts, que equival al percentatge de freqüències d'aquestes puntuacions.

Gràfic 4.3: Àrees sota la corba normal



Els matemàtics calculen les àrees entre dos valors recorrent al càlcul integral, és a dir, fent una integral de l'equació de la corba normal entre aquests dos valors. Per sort, les integrals necessàries ja estan calculades i recollides en una taula, la taula d'àrees sota la corba normal, que trobareu a l'annex 2 d'aquest document. El que cal fer és convertir els valors que volem analitzar en puntuacions típiques *Z*.

A continuació en tenim un fragment que ens servirà per explicar com s'interpreten els valors que surten en aquestes taules. La columna vertical de l'esquerra representa les puntuacions *Z*, expressades amb una xifra decimal. Els valors de dins de la taula representen el percentatge de l'àrea entre la mitjana i el valor *Z*, expressat amb dues xifres decimals.

Per exemple, en la taula 4.2, per trobar el percentatge de casos entre la mitjana i una $Z = 0{,}94$ hem de seleccionar el valor 0,9 de la primera columna i el valor 0,04 de la primera fila. En el punt on es creuin aquests dos valors trobarem el valor de l'àrea, que serà del 32,64 %.

Taula 4.2: Fragment de la taula d'àrees (percentatges) de la corba normal, entre les puntuacions típiques i la mitjana

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	00,00	00,40	00,80	01,20	01,60	01,99	02,39	02,79	03,19	03,59
0,1	03,98	04,38	04,78	05,17	05,57	05,96	06,36	06,75	07,14	07,53
0,2	07,93	08,32	08,71	09,10	09,48	09,87	10,26	10,64	11,03	11,41
0,3	11,79	12,17	12,55	12,93	13,31	13,68	14,06	14,43	14,80	15,17
0,4	15,54	15,91	16,28	16,64	17,00	17,36	17,72	18,08	18,44	18,79
0,5	19,15	19,50	19,85	20,19	20,54	20,88	21,23	21,57	21,90	22,24
0,6	22,57	22,91	23,24	23,57	23,89	24,22	24,54	24,86	25,17	25,49
0,7	25,80	26,11	26,42	26,73	27,04	27,34	27,64	27,94	28,23	28,52
0,8	28,81	29,10	29,39	29,67	29,95	30,23	30,51	30,78	31,06	31,33
0,9	31,59	31,86	32,12	32,38	32,64	32,89	33,15	33,40	33,65	33,89
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.

La taula només proporciona els percentatges a un costat de la mitjana, per la qual cosa el percentatge més gran que surt a la taula és el 50 %, és a dir, la meitat de l'àrea sota la corba normal. Però com que sabem que la corba normal és simètrica, és a dir, que l'alçada de la corba per un valor positiu de *Z* és exactament la mateixa pel mateix valor però negatiu, coneixent els valors de la meitat de la corba podem saber totes les àrees que ens interessin.

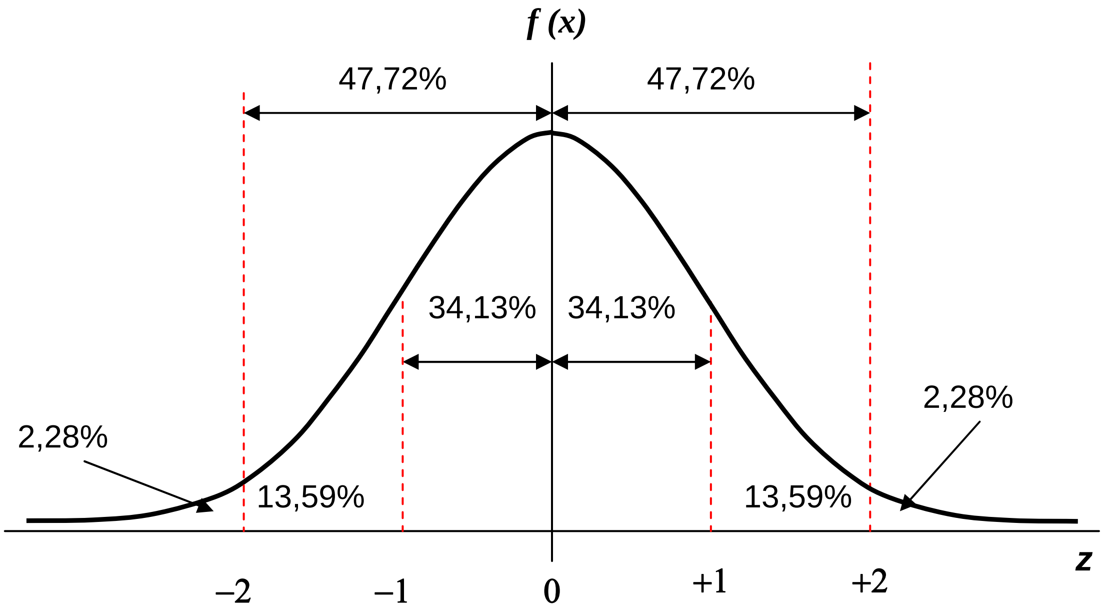
Cal observar que l'àrea entre *Z* i la mitjana per a un valor de $Z = 1$ és 34,13 %, i que per a $Z = 2$ és 47,72 %. Si recordeu que les *Z* representen desviacions típiques, el percentatge de casos que es troben entre la mitjana i +/- una desviació típica és del 68,26 % i el percentatge de casos que es poden trobar entre la mitjana i +/- dues desviacions típiques és del 95,5 %. Si en lloc de $Z = 2$ agafem $Z = 1{,}96$ veurem que el percentatge de casos entre la mitjana i +/- 1,96 és del 95 %. Podeu veure aquests valors representats al gràfic 4.

Per tant, segons el model de la corba normal, es pot esperar que aproximadament el 95 % de tots els successos possibles se situïn entre la mitjana i +/- 2 desviacions típiques. Per exemple, si suposem que l'alçada de la mitjana dels adults és de 169 cm i la desviació típica és de 6, es pot esperar que el 95 % dels adults tinguin alçades entre

$$169 \pm 2 \cdot 6 = 169 \pm 12$$

és a dir, cal esperar que tinguin alçades entre 157 cm i 181 cm. Només un 2,28 % tindrien alçades superiors a 181 cm i el mateix percentatge —un 2,28 %— tindrien alçades inferiors a 157cm.

Gràfic 4.4: Característiques de la corba normal

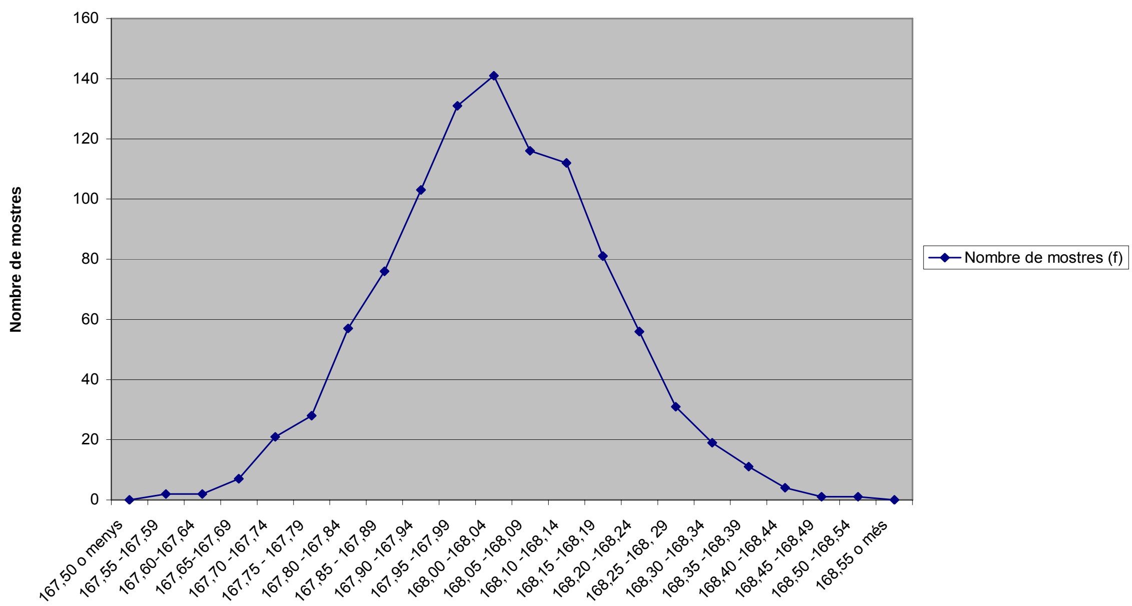


5. Inferències sobre la mitjana d'una població

Tot el que hem explicat al capítol d'introducció a l'estadística inferencial ens serveix per entendre com podem fer inferències a la població a partir de l'estudi d'una mostra. Recordem que al capítol 4 ja hem vist la distribució mostral de molts estadístics, entre els quals ens fixarem en les mitjanes. Així, hem observat que, si escollim un nombre elevat de mostres d'una població, totes de la mateixa mida, busquem la mitjana de cadascuna i representem gràficament la distribució de freqüències d'aquestes mitjanes, la gràfica resultant s'aproxima molt a la corba normal. El valor que té més freqüències d'aquesta distribució el podem considerar la mitjana de la població.

Per exemple, hem vist que representant gràficament la distribució mostral de les mitjanes d'alçada de 1.000 mostres de 100 persones cada una, ens sortia el gràfic següent:

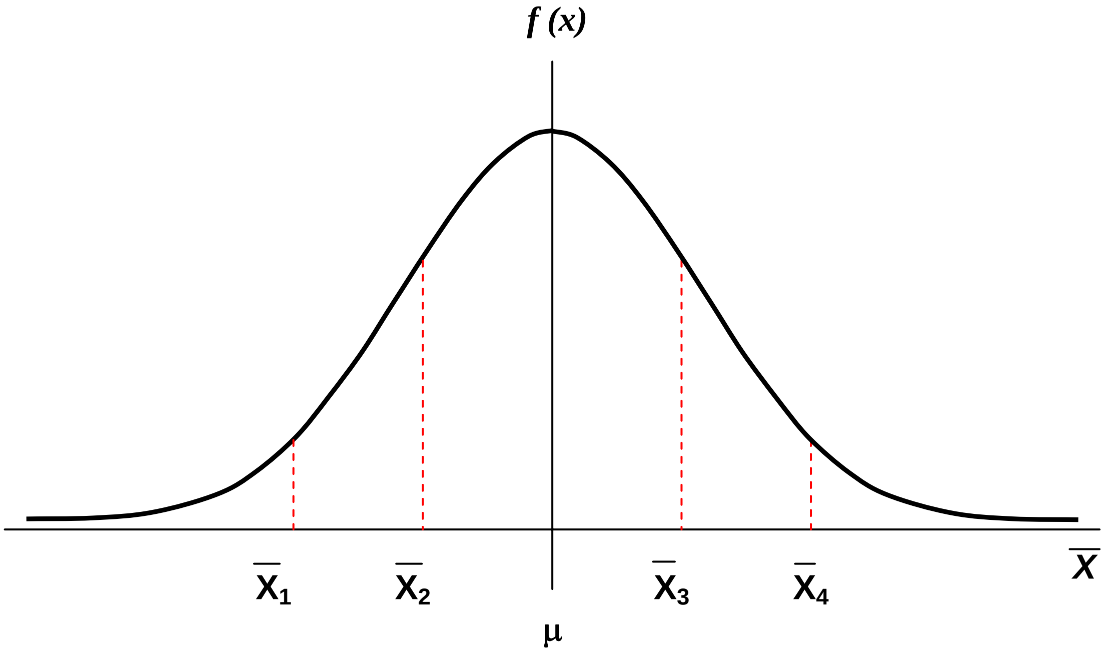
Gràfic 5.1: Distribució hipotètica de le alçades mitjanes de 1000 mostres de N = 100



Si en lloc d'extreure mil mostres n'extraiéssim moltes més —fins a un nombre infinit—, la forma de la distribució mostral de mitjanes seguiria el model de la corba normal. El nombre més elevat de freqüències correspondria a la mitjana de les mitjanes, que podríem considerar la mitjana de la població.

Però abans d'efectuar inferències sobre la mitjana de la població, cal tenir una mesura de dispersió d'aquesta distribució. Si la variable que estem analitzant és l'alçada de les persones adultes, i sabem que la mitjana és de 168 cm, podem entendre que és molt poc probable que una persona seleccionada a l'atzar tingui una alçada de 157,00 cm. Però encara és molt menys probable que una mostra de N = 100 seleccionada tingui una mitjana d'alçada de 157 cm. És a dir, *la distribució mostral de mitjanes és molt menys dispersa que la distribució de freqüències de les puntuacions directes d'una variable*. Dit d'una altra manera, un valor extrem d'una mitjana d'una mostra és menys probable que el mateix valor d'una puntuació directa. Com més gran sigui la *N* de les mostres amb què treballem per fer la distribució mostral, menys dispersió hi haurà en aquesta distribució.

Gràfic 5 2. Distribució mostral de mitjanes



Per calcular la dispersió de la distribució mostral de mitjanes podríem utilitzar la fórmula que ja coneixem de la desviació típica, tenint en compte que els valors serien les diferents mitjanes. No obstant això, normalment els investigadors només disposen d'una mostra, i és aquesta mostra la que ha de servir per fer inferències. Les teories estadístiques mostren que la dispersió d'una distribució de mitjanes mostrals a partir d'una sola mostra escollida aleatòriament, de mida *N*, es pot calcular de la manera següent:

$$S_x = \frac{S}{\sqrt{N}} \qquad (11)$$

S = desviació típica de la mostra
N = nombre de casos de la mostra

Aquesta fórmula s'anomena *error típic de la mitjana*.

L'error típic de la mitjana (S_x) és menor que la desviació típica de les puntuacions directes (*S*). Com més gran sigui la mostra més petit serà l'error típic, és a dir, menor serà la dispersió de la distribució mostral de mitjanes. Es fa servir la paraula *error* perquè expressa la idea que la diferència entre la mitjana poblacional i la mitjana d'una mostra de la població que s'escull aleatòriament és un error produït pel mostreig, és a dir, per l'atzar.

Seguint amb l'exemple de les alçades:

Imaginem-nos que tenim una mostra de N = 100 homes dels països escandinaus i que s'ha calculat que la seva alçada mitjana és de 169 cm i que la desviació típica d'aquesta mostra és de 7 cm. Llavors, l'error típic de la mitjana el podem calcular de la manera següent:

$$S_x = \frac{S}{\sqrt{N}} = \frac{7}{\sqrt{100}} = 0,7$$

Se sap també que la mitjana de l'alçada de la població d'homes d'Europa és de 168 cm.

Ara ens pot interessar esbrinar si l'alçada mitjana dels homes dels països escandinaus és diferent de la població completa d'homes adults d'Europa. Sense mesurar l'alçada de tots els homes dels països escandinaus no podem respondre amb certesa a aquesta pregunta, però podem utilitzar la mitjana mostral i l'error típic de la mitjana per fer inferències sobre la població. Hi ha dues possibilitats:

a) La mostra de mitjana 169 cm prové d'una població amb mitjana 168 cm, i per tant la diferència d'1 cm és deguda a l'error de mostreig.

b) La mostra no prové d'una població amb mitjana 168 cm sinó d'una amb qualsevol altra mitjana.

Per decidir entre aquestes dues alternatives ens caldria saber fins a quin punt és estrany obtenir una mostra de mitjana $\overline{X} = 169$ d'una població amb una $\mu = 168$. Quina de les dues alternatives —*a* o *b*— és la més encertada?

5.1. Comprovació d'hipòtesis

En l'exemple de la moneda vàrem veure que calia calcular la probabilitat que per atzar pogués sortir 6 vegades cara en 6 llançaments consecutius. Vàrem deduir que això era bastant improbable prenent com a regla de decisió el 0,05, és a dir, vàrem rebutjar la hipòtesi 0 (la que establia que la moneda era correcta) i vàrem acceptar la hipòtesi 1 (la que establia que la moneda estava trucada). El procediment és molt semblant en el cas de les alçades: cal determinar la probabilitat d'escollir una mostra amb una mitjana de 169 cm, sent la mitjana poblacional de 168, i comparar aquesta probabilitat amb una regla de decisió.

5.1.1. Les hipòtesis nul·la i alternativa

El primer pas consisteix a establir la hipòtesi nul·la, que simbolitzarem amb H_0, que especifica la suposició sobre el paràmetre poblacional. En l'exemple que ens ocupa la hipòtesi nul·la implica que la mostra, amb mitjana = 169 cm, és una mostra aleatòria d'una població amb μ = 168 cm.

A continuació es formula la hipòtesi alternativa, que simbolitzarem amb una H_1. Aquesta hipòtesi estableix que la mitjana poblacional de la qual s'ha extret la mostra amb mitjana 169 cm, no és 168 sinó qualsevol altre valor. És a dir, la diferència d'1 punt entre la mitjana de la mostra i la de la població no pot ser deguda a l'atzar.

Cal tenir en compte que les proves estadístiques no determinen mai si la hipòtesi nul·la és veritable o és falsa. Una prova estadística indica si la H_0 és prou probable o no, segons la regla de decisió que s'estableixi. Sense mesurar la població completa és impossible provar res, i la decisió final que es pren pot ser correcta o incorrecta. De fet, existeixen quatre possibilitats:

1. La mitjana poblacional és realment 168 (és a dir, la mostra procedeix d'una població amb μ = 168). El fet de rebutjar una H_0 veritable s'anomena *error de tipus I,* i la probabilitat (o risc) de cometre aquest error se simbolitza amb la lletra grega alfa (α).

2. La mitjana poblacional no és realment 168 i incorrectament acceptem H_0. L'error comès en acceptar una hipòtesi nul·la que en realitat és falsa s'anomena *error de tipus II*, i la probabilitat (risc) de cometre'l se simbolitza amb la lletra grega beta (β).

3. La mitjana poblacional no és realment 168 i es rebutja correctament H_0. La probabilitat d'assolir aquesta decisió correcta s'anomena *potència* de la prova estadística i és igual a 1 - β.

4. La mitjana poblacional és realment 168 i s'accepta correctament H_0. La probabilitat de prendre aquesta decisió correcta és igual a 1 - α .

Aquestes quatre possibilitats estan resumides a la taula 5.1.

Taula 5.1: Model de comprovació d'hipòtesis estadístiques

		Situació real o estat de la població	
		H_0 és realment certa	**H_0 és realment falsa**
El resultat de l'experiment aconsella:	Acceptar H_0	*Decisió correcta* La probabilitat d'acceptar la H_0 certa és 1 - α.	*Error de tipus II* La probabilitat (risc) d'acceptar una H_0 falsa és β.
	Rebutjar H_0	*Error de tipus I* La probabilitat (risc) de rebutjar una H_0 certa és α.	*Decisió correcta* La probabilitat de rebutjar una H_0 falsa (potència) és 1 - β.

Aquests dos tipus d'error no es poden eliminar ja que són complementaris, és a dir, encara que reduíssim al màxim la probabilitat de cometre un error de tipus I, el que ens passaria és que augmentaria la probabilitat de fer un error de tipus II, i a l'inrevés. Si disminuïm el risc de rebutjar una hipòtesi nul·la que és correcta, augmenta el risc d'acceptar una hipòtesi nul·la que és falsa.

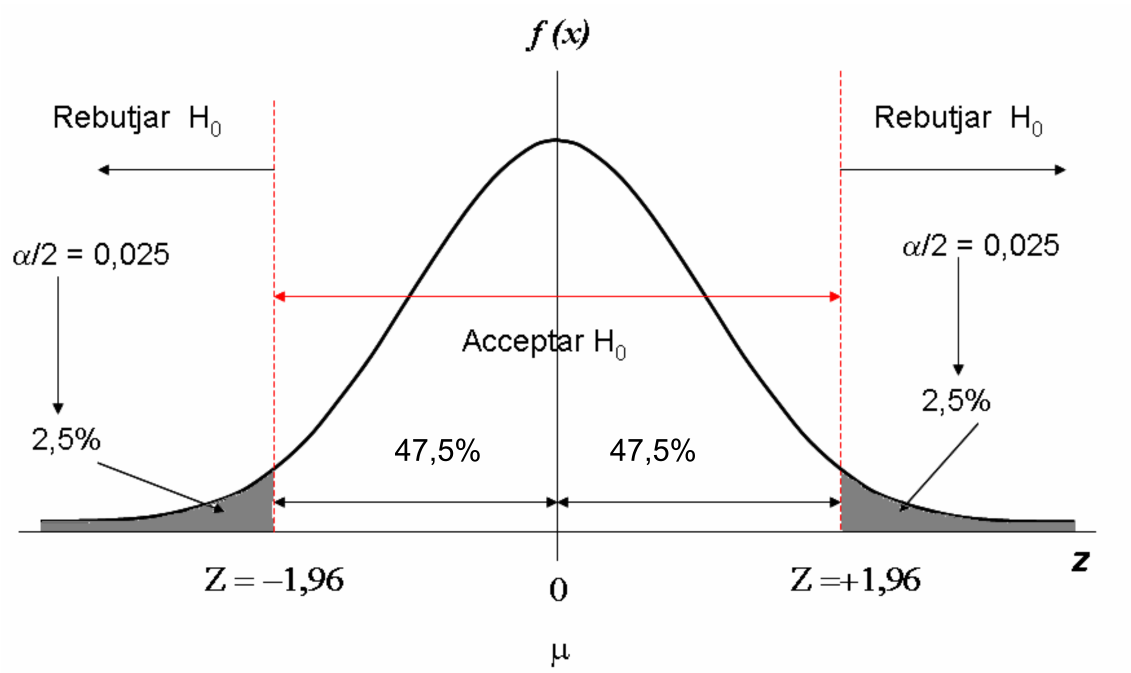
5.1.2. Nivell de significació

Una vegada establertes les hipòtesis nul·la i alternativa, es necessita una regla que indiqui com és d'improbable que una determinada mitjana mostral es produeixi, per rebutjar la H_0 i acceptar la H_1. Ja hem vist, en parlar de l'estratègia general de l'estadística inferencial, que no hi ha regles absolutes per establir valors a partir dels quals les mitjanes mostrals s'hagin de considerar prou diferents per rebutjar la H_0.

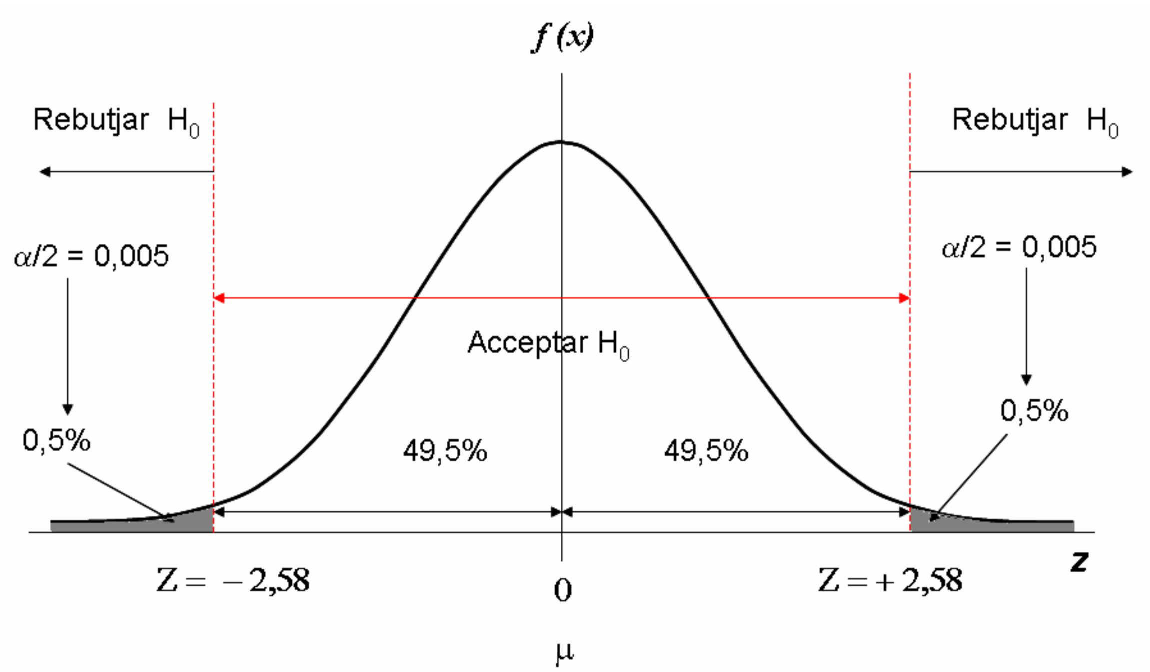
El valor numèric que s'especifica com a regla de decisió rep el nom de *nivell de significació*. Així, quan s'escull la "regla de decisió del 0,05", s'està emprant un nivell de significació del 5 %, és a dir, s'està dient que, si s'obté una probabilitat igual o menor al 0,05, es considerarà prou improbable i, per tant, es rebutjarà la H_0. El nivell de significació s'expressa amb $\alpha = 0{,}05$. Tal com hem dit, aquest valor α indica la probabilitat de cometre un error de tipus I, és a dir, indica el risc que tenim de rebutjar una H_0 que sigui veritable.

Als gràfics 3 i 4 trobem representats els errors de tipus I per $\alpha = 0{,}05$ i $\alpha = 0{,}01$. Cal tenir en compte que l'error α es reparteix a ambdós costats de la mitjana, a cada una de les cues de la corba normal.

Gràfic 5.3: Nivell de significació per α = 0,05



Gràfic 5.4: Nivell de significació per α = 0,01



El nivell de significació de l'l,1 % és més restrictiu en el rebuig de H_0 que el del 5 %, ja que només les mitjanes mostrals molt extremes es consideraran prou improbables. A més, té l'avantatge de reduir l'error de tipus I. No obstant això, com a conseqüència, el nivell de significació de l'l,1 % és menys exigent en l'acceptació de H_0, per la qual cosa augmenta la probabilitat de l'error de tipus II (és a dir, augmenta el risc d'acceptar una H_0 que és falsa).

5.2. Prova estadística per a la mitjana d'una població quan σ és coneguda

Tornem ara a l'exemple de les alçades. Suposem que la desviació típica de la població és 6 i que volem utilitzar la regla de decisió del 0,05, és a dir, un nivell de significació de 0,05.

El primer que cal fer és tipificar el valor de la mitjana mostral. Recordem que la mitjana mostral que havíem suposat era de 169 cm, per a una mostra de N = 100. Recordem també que tipificar significa convertir els valors observats als valors Z. Com que no es tracta ara de tipificar una puntuació directa d'alçada sinó de tipificar la mitjana d'una mostra, haurem d'utilitzar l'error típic de la mitjana com a mesura de dispersió. La fórmula que haurem d'emprar és la següent:

$$Z = \frac{\overline{X} - \mu}{\sigma_x} \qquad (12)$$

$\overline{X}$ = valor observat de la mitjana mostral
μ = valor de la mitjana poblacional
σ_x = error típic de la mitjana quan la mitjana de la població és coneguda ($= \sigma / \sqrt{N}$)

Com que $\sigma = 6$

$$\sigma_x = \frac{6}{\sqrt{100}} = 0{,}6$$

$$Z = \frac{169 - 168}{0{,}6} = 1{,}67$$

Veiem que el valor +1,67 cau entre els valors -1,96 i +1,96, que són els valors crítics quan $\alpha = 0{,}05$. És a dir, que una $Z = 1{,}67$ té una probabilitat que cau dins de l'àrea d'acceptació de la hipòtesi nul·la. Per tant, és probable obtenir, per a un nivell de significació del 0,05, una mostra de 100 casos amb una mitjana de 169 d'una població que té una mitjana de 168.

Suposem ara que la desviació típica de la població fos 4 i la mitjana, 168. És probable obtenir una mostra de 100 casos amb una mitjana de 169 d'aquesta població?

Busquem primer l'error típic de la mitjana:

$$\sigma_x = \frac{4}{\sqrt{100}} = 0{,}4$$

Ara tipifiquem el valor de la mitjana:

$$Z = \frac{169 - 168}{0{,}4} = 2{,}5$$

En aquest cas, el valor +2,5 queda fora de l'àrea d'acceptació compresa entre els valors -1,96 i +1,96. Per tant, en aquest cas rebutgem la H_0 i deduïm que és molt poc probable obtenir una mostra de 100 casos amb una mitjana de 169 d'una població que té una $\mu = 168$ i una $\sigma = 4$. Una altra manera de dir-ho és que la diferència entre la mitjana de la mostra i la de la població és estadísticament significativa per a un nivell de significació del 0,05.

5.3. Prova estadística per a la mitjana d'una població quan σ és desconeguda

Cal tenir en compte que quan es treballa amb mostres el més habitual és que no coneguem la desviació típica de la població. La majoria de vegades és necessari calcular una estimació de l'error típic de la mitjana a partir de la desviació típica de la mostra. En aquest cas, cal utilitzar la fórmula (11):

$$S_x = \frac{S}{\sqrt{N}}$$

I per tipificar els valors de les mitjanes mostrals s'utilitza :

$$t = \frac{\overline{X} - \mu}{S_x} \qquad (13)$$

Els valors obtinguts amb aquesta fórmula no estan distribuïts seguint el model de la corba normal, encara que la corba normal constitueix una bona aproximació si la mida de la mostra és més gran de 30 casos. El procediment de la corba normal podria portar a respostes errònies perquè en aquesta fórmula s'està utilitzant un estimador de l'error típic de la mitjana, calculat a partir de la desviació típica de la mostra.

El model que s'utilitza en aquest cas, és a dir, quan no coneixem σ de la població i tenim menys de 30 casos en la mostra, és la *distribució t*. Existeix una distribució *t* per a cada mida de les mostres. Els matemàtics ja han tabulat les diferents probabilitats per les corbes corresponents a cada possible mida de la mostra.

La distribució t

La distribució *t* la trobem en unes taules, que estan expressades en funció dels graus de llibertat (g. ll.) i no segons la mida de la mostra. Els graus de llibertat són, en cada cas, iguals a la mida de la mostra menys 1:

$$\text{g. ll.} = N - 1$$

A la taula 5.2 es mostra un fragment de la taula de valors crítics de *t*. Cada fila horitzontal representa una distribució diferent, que correspon als graus de llibertat. A l'annex 3 trobareu la taula de valors crítics de *t* completa.

Tabla C. *Valores críticos de t (*)*

	Nivel de significación para una prueba unilateral					
gl	0,10	0,05	0,025	0,01	0,005	0,0005
	Nivel de significación para una prueba bilateral					
	0,20	0,10	0,05	0,02	0,01	0,001
1	3,078	6,314	12,706	31,821	63,657	636,619
2	1,886	2,920	4,303	6,965	9,925	31,598
3	1,638	2,353	3,182	4,541	5,841	12,941
4	1,533	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,859
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,449	5,405
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	$\boxed{2,262}$	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587

Taula 5.2: Fragment de la taula de valors crítics de t

Exemple

Imaginem-nos que agafem una mostra de 10 persones adultes i els mesurem l'alçada. Trobem que la mitjana és de 166 cm i la desviació típica, de 7 cm. Volem saber si aquesta mostra pot pertànyer a la població en què la mitjana d'alçada és de 168 cm.

El primer que farem serà una tipificació, però, com que N = 10, no podem convertir-ho a un valor de Z, i per tant utilitzarem els valors t.

- En primer lloc, formularem les hipòtesis nul·la i alternativa:

H_0: la mostra pertany a la població. Podem simbolitzar-ho dient que $\bar{X} = \mu$
H_1: la mostra no pertany a la població. $\bar{X} \neq \mu$

- Tot seguit buscarem l'error típic de la mitjana, utilitzant l'expressió (11):

$$S_x = \frac{S}{\sqrt{N}} = \frac{7}{\sqrt{10}} = 2{,}21$$

- A continuació tipificarem la mitjana mostral, utilitzant la fórmula (13):

$$t = \frac{\overline{X} - \mu}{S_x} = \frac{166 - 168}{2{,}21} = -0{,}90$$

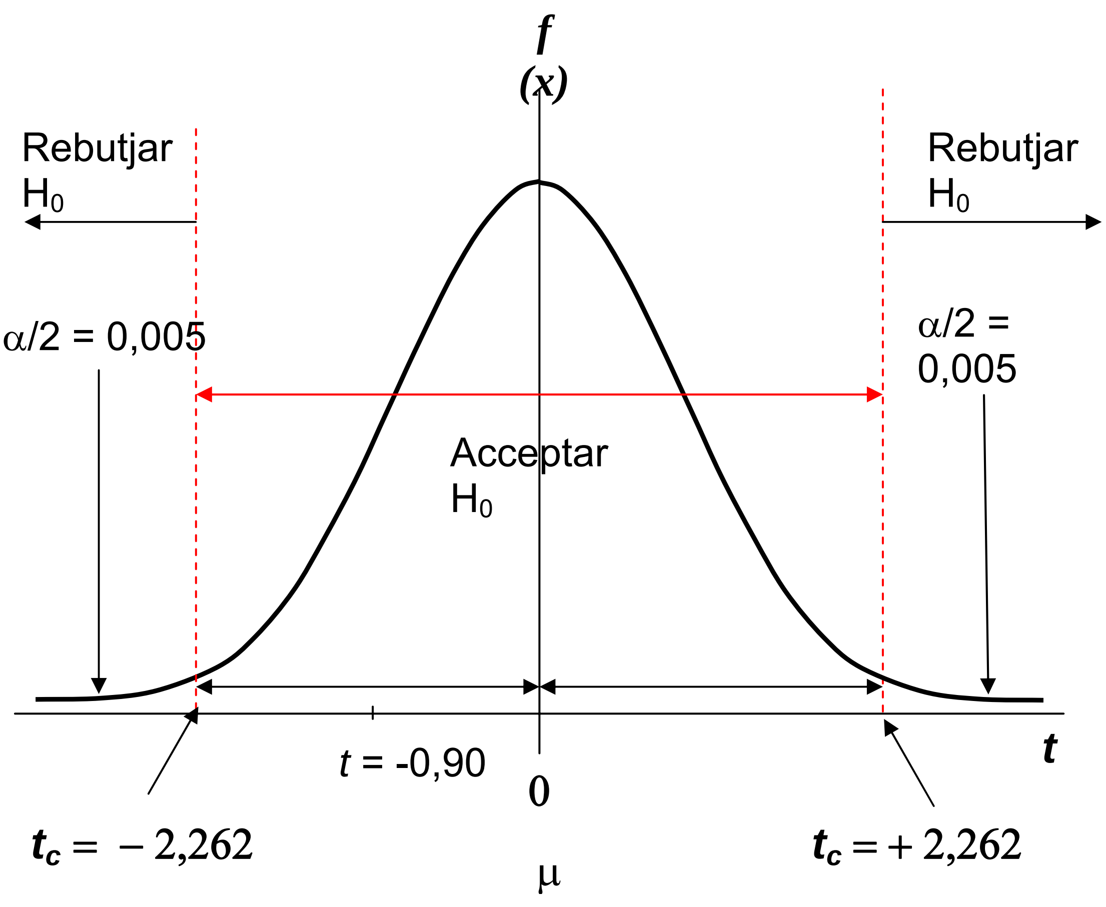
- Ara hem de buscar els valors crítics de *t*, per poder comparar el valor que ens ha sortit. Farem la interpretació per a un nivell de significació de $\alpha = 0{,}05$.

Mirem les taules i busquem el valor crític:

t_c (g. ll. = N - 1; $\alpha = 0{,}05$)
t_c (g. ll. = 9; $\alpha = 0{,}05$) = 2,262

El gràfic 5.5 ens pot ajudar en la interpretació. Com que el valor de la *t* que hem trobat és dins de l'interval definit per $-t_c$ i $+t_c$, podem dir que és dins de la zona d'acceptació de la H_0.

Gràfic 5.5: Situació dels valors crítics de t *i de la* t *observada en l'exemple*



- Per tant, amb un nivell de significació de 0,05, acceptem la hipòtesi nul·la, és a dir, que és possible obtenir una mostra aleatòria de N = 10 amb una mitjana de 166 cm d'una població que té una mitjana de 168 cm. Dit d'una altra manera, tenim un 95 % de probabilitat que, en escollir una mostra a l'atzar de 10 persones adultes d'una població que té de mitjana d'alçada 168 cm, ens surti que la mitjana de la mostra és de 166 cm.

Si observem les taules de valors crítics de t veurem que ,com més petita és la mida de la mostra, més gran ha de ser t per rebutjar H_0, amb el mateix nivell de significació. També es veu que, quan la mida de la mostra és gran, les distribucions Z i t són equivalents. Així, per exemple, per 40 graus de llibertat, la t_c és 2,02 (per α = 0,05), mentre que per a 120 graus de llibertat, t = 1,98. Aquest valor ja s'aproxima molt al de Z = 1,96.

RESUM

La tipificació de valors de les mitjanes mostrals ens permet saber si una mostra pot pertànyer a una població determinada, amb mitjana μ.

Quan μ i σ són conegudes

- Tipificarem el valor de la mitjana de la mostra amb la fórmula:

$$Z = \frac{\overline{X} - \mu}{\sigma_x} \qquad (12)$$

$\overline{X}$ = valor observat de la mitjana mostral
μ = valor de la mitjana poblacional
σ_x = error típic de la mitjana quan la mitjana de la població és coneguda ($= \sigma / \sqrt{N}$)

- El valor de Z que surti el compararem amb els valors corresponents de la corba normal. Per α = 0,05 utilitzarem el valor Z = 1,96; i per α = 0,01 utilitzarem els valors de Z = 2,58. No és necessari fer el dibuix cada vegada.
    - Per a α = 0,05, si el valor de Z que ens surt, en valor absolut, és més petit que 1,96, acceptarem H_0, i si és més gran rebutjarem H_0 i acceptarem H_1.
    - Per a α = 0,01, si el valor de Z que ens surt, en valor absolut, és més petit que 2,58, acceptarem H_0 i si és més gran rebutjarem H_0 i acceptarem H_1.

Quan μ és coneguda i σ és desconeguda

- En general, tendirem a utilitzar els valors de la taula *t* per decidir si acceptem la H_0 o si la rebutgem. Utilitzarem la fórmula:

$$t = \frac{\overline{X} - \mu}{S_x} \qquad (13)$$

$\overline{X}$ = valor observat de la mitjana mostral
μ = valor de la mitjana poblacional
S_x = error típic de la mitjana quan la mitjana de la població és desconeguda

$$S_x = S / \sqrt{N} \qquad (11)$$

- A l'hora de fer la interpretació, no cal que fem el dibuix cada vegada, sinó que compararem el valor de *t* que ens ha sortit amb el valor de la t_c, en valor absolut. Tenint en compte que els valors de les taules ens indiquen el nivell a partir del qual, per marge d'error determinat, la diferència entre la mitjana mostral i la poblacional comença a ser significativa, la interpretació la farem de la manera següent:
  - Si t < o = t_c, acceptarem H_0
  - Si $t > t_c$, rebutjarem H_0

5.4. Interval de probabilitat o interval de confiança

Una altra manera de resoldre el problema d'acceptar o rebutjar la H_0 consisteix a estimar un interval en el qual és probable que trobem la mitjana de la població. Si el valor de μ expressat per la hipòtesi nul·la no cau dins de l'interval, es rebutja H_0. *L'interval de probabilitat o de confiança* és l'interval de valors en què és més probable trobar la mitjana de la població de la qual s'ha extret la mostra. Així, l'exemple que hem posat a l'apartat anterior el podem resoldre també buscant la mitjana de la població de la qual s'ha extret la mostra de 10 casos, amb una mitjana de 166 cm. No podem trobar el valor exacte de la mitjana poblacional, però podem estimar un interval de valors en què és més probable —en què podem tenir més confiança de— trobar la mitjana de la població. Si la mitjana coneguda de la població $\mu = 168$ es troba dins de l'interval, podrem acceptar la H_0, és a dir, podrem concloure que la mostra pertany a una població amb una mitjana de 168 cm.

Vegem el procediment que cal seguir, posant un altre supòsit.

Exemple

Volem saber si és possible obtenir una mostra de 20 casos amb una mitjana de 167 cm d'alçada i una desviació típica S = 5, d'una població que té de mitjana μ = 168 cm. Volem interpretar-ho amb un nivell de significació del 95 %.

Aquesta vegada optarem pel procediment de l'interval de probabilitat. Buscarem la mitjana de la població de la qual s'ha extret la mostra.

1. Com que la mostra és petita, anirem a les taules de valors crítics de t per buscar el valor corresponent a 20-1 graus de llibertat i α = 0,05:

   t_c (19 g. ll.; μ = 0,05) = 2,093

2. Busquem l'error típic de la mitjana:

$$S_x = \frac{S}{\sqrt{N}} = \frac{5}{\sqrt{20}} = 1{,}12$$

3. Substituïm la fórmula que ja coneixem pels valors que sabem. Com que estem buscant el valor de la mitjana de la població de la qual s'ha extret la mostra, la µ constitueix la incògnita, el valor que busquem.

$$t = \frac{\overline{X} - \mu}{S_x}$$

   Cal tenir en compte que, com que estem treballant sempre amb proves bilaterals, el valor de t_c pot adoptar el signe positiu o el negatiu. Per tant,

$$+2{,}093 = \frac{167 - \mu_1}{1{,}12}$$

$$-\ 2{,}093 = \frac{167 - \mu_2}{1{,}12}$$

4. Si aïllem la mitjana de la població en cada una d'aquestes fórmules, tindrem el següent:

   $\mu_1 = 167 – 1{,}12 \cdot 2{,}093 = 164{,}66$
   $\mu_2 = 167 + 1{,}12 \cdot 2{,}093 = 169{,}34$

5. Fem la interpretació:

   164,66 < 168 < 169,34

   Com que la mitjana suposada de la població es troba dins de l'interval, amb un nivell de confiança del 95 % podem acceptar la H_0. És a dir, és probable que una mostra de 20 casos, amb una mitjana de 167 i una desviació típica de 5, l'hàgim tret d'una població amb una mitjana de 168.

En general, la fórmula que hem d'utilitzar per calcular intervals de probabilitat quan σ és desconeguda és:

$$\boxed{\overline{X} - t_c\, S_x < \mu < \overline{X} + t_c\, S_x} \qquad (14)$$

t_c = valor crític de t per N-1 graus de llibertat i α
$\overline{X}$ = valor observat de la mitjana mostral
S_x = error típic de la mitjana ($= S/\sqrt{N}$)

Qualsevol valor per sobre o per sota d'aquest interval es considera improbable. Per això, si el valor que comparem cau dins de l'interval, acceptarem la hipòtesi nul·la, i si cau fora, rebutjarem aquesta hipòtesi.

Si les mostres són grans

$$\boxed{\overline{X} - Z\, S_x < \mu < \overline{X} + Z\, S_x} \qquad (15)$$

Z = valor de Z segons el marge d'error α
$\overline{X}$ = valor observat de la mitjana mostral
S_x = error típic de la mitjana ($= S/\sqrt{N}$)

5.5. Exercicis d'inferències sobre la mitjana d'una població

1. Es vol saber quina és la mitjana d'edat dels alumnes que assisteixen a la Universitat. Per això s'ha agafat una mostra aleatòria de 400 alumnes i ens dóna una mitjana de 24,3 anys i una desviació típica de 12,8 anys. Calcula els resultats en els nivells de confiança del 95 % i del 99 %.

2. Sabem que la nota mitjana del primer quadrimestre dels estudiants de l'assignatura Bases metodològiques de la investigació educativa del curs 2005-2006 és de 6,02 i que la desviació típica és d'1,58. Aquesta mostra està composta per 63 estudiants. Quina serà la mitjana de puntuació de les notes del primer

quadrimestre que podem inferir per als estudiants d'aquesta assignatura dels cursos vinents? Utilitza un nivell de significació del 0,05.

3. Es vol conèixer la mitjana d'euros que es gasten els estudiants de la UdG per esmorzar cada setmana. S'ha seleccionat una mostra de 250 estudiants i se'ls ha demanat aquesta dada. La mitjana d'euros que es gasten a la setmana és de 5 euros, amb una desviació típica d'1,5. Respon a les qüestions següents:

   3.1. Amb una confiança del 95 %, entre quins valors es troba la mitjana d'euros que els estudiants es gasten per esmorzar cada setmana?
   3.2. Quina és la probabilitat d'obtenir una mitjana mostral de 5 euros/setmana o més quan la mitjana de la població és igual a 4,8?
   3.3. Podem admetre a un nivell de confiança del 0,05 que la nostra mostra pertany a la població amb una mitjana de 4,8?
   3.4. Calcula de nou l'apartat 1 a un nivell de confiança del 0,01. Amb aquest nivell de confiança, podem dir que la mostra pot pertànyer a una població amb una mitjana de 4,8?

4. Tenim una mostra de 200 estudiants de la FEP, amb una mitjana d'edat de 24 anys. En el nivell de significació del 0,05, podem admetre que aquesta mostra pertany a la població d'estudiants de la FEP amb una mitjana d'edat de 22 i una desviació típica de 4?

5. S'ha agafat una mostra de 300 persones que viuen en pisos de lloguer. Investigada la quantitat que paguen de lloguer, surt una mitjana de 400 € i una desviació típica de 25 €. A partir d'aquestes dades, podem afirmar amb un 95 % de confiança que la mitjana de lloguer per a la població és de 390 €?

6. Se sap que la mitjana d'ingressos anuals dels titulats universitaris de grau mitjà (diplomats) és de 16.000 €. Hem seleccionat una mostra de 30 educadores socials i els hem preguntat sobre els seus ingressos anuals. La mitjana d'ingressos d'aquesta mostra d'educadores és de 15.000 € anuals, amb una desviació típica de 1.500 €. Podem dir, basant-nos en aquest resultat, que els diplomats en Educació Social estan més mal pagats en comparació del conjunt de diplomats universitaris? Interpreta el resultat per a un marge d'error del 0,05.

5.6. Respostes als exercicis d'inferències sobre la mitjana d'una població

Exercici 1

Per a un nivell de confiança del 95 % la mitjana d'edat estaria entre els 23,05 i els 25,55 anys.
Per a un nivell de confiança del 99 % la mitjana d'edat se situaria entre els 22,64 i els 25,95 anys.

Exercici 2

Aquí se'ns demana que busquem l'interval de probabilitat agafant 63 estudiants del curs 2005-2006 com a mostra representativa de la població. A més, no coneixem la desviació típica de la població. Per tant, buscarem l'interval de probabilitat amb la fórmula:

$\overline{X}$ +/- t. $S/\sqrt{N}$

t (g. ll. N -1 = 62; α = 0,05) = 2 (Utilitzem g. ll. = 60, ja que el següent ja és 120).

$6{,}02 + 2.\ 1{,}58/\sqrt{63} = 6{,}42$

$6{,}02 - 2.\ 1{,}58/\sqrt{63} = 5{,}62$

Per tant, amb un marge d'error del 0,05, es pot esperar per als anys vinents que la mitjana del primer quadrimestre estarà entre una puntuació de 5,62 i 6,42.

Exercici 3

3.1. $\overline{X} = 5$ $S = 1{,}5$ $N = 250$
Per α = 0,05 la Z = 1,96 (fem servir Z perquè N és molt gran.)

$\overline{X}$ +/- $Z \cdot$ S / $\sqrt{N}$

$\overline{X} + Z. \cdot S/\sqrt{N} = 5 + 1{,}96 \cdot (1{,}5/\sqrt{250}) = 5{,}19$

$\overline{X} - Z \cdot S/\sqrt{N} = 5 - 1{,}96 \cdot (1{,}5/\sqrt{250}) = 4{,}81$

Resposta: a un nivell de confiança del 95 %, els estudiants de la UdG es gasten entre 4,81 i 5,19 euros de mitjana cada setmana per esmorzar.

3.2. Per saber la probabilitat exacta haurem de tipificar el valor de la mitjana mostral. Les dades que tenim són les següents:

$\overline{X} = 5$ $\mu = 4{,}8$ $S = 1{,}5$ $N = 250$

Desconeixem la desviació típica de la població, per la qual cosa haurem d'utilitzar la desviació típica de la mostra.

$$Z = \frac{\overline{X} - \mu}{S / \sqrt{N}} = \frac{5 - 4{,}8}{1{,}5 / \sqrt{250}} = \frac{0{,}2}{0{,}095} = 2{,}11$$

Ara cal buscar en les taules de la corba normal el valor 2,11. Veiem que per a una Z = 2,17 l'àrea entre aquest valor i la mitjana és del 48,26 %. Com que la pregunta és la probabilitat de tenir una mitjana de 5 euros o més a la setmana, i el que tenim és el percentatge entre la mitjana i Z, hem de restar aquest valor de 50 %:

P (5 o més) = 50 % - 48,26 % = 1,74 %

Per tant, la probabilitat de tenir una mostra de 250 casos amb mitjana de 5 euros o més a la setmana és de l'1,5 %.

3.3. Primer formulem les hipòtesis estadístiques:

H_0: la mostra de 250 casos amb una mitjana de 5 i una desviació típica d'1,5 pertany a una població amb una mitjana de 4,8.
H_1: la mostra no pertany a una població amb una mitjana de 4,8.

En aquest cas, d'entrada veiem que si l'interval de probabilitat es troba entre 4,81 i 5,19, una μ de 4,8 se situa, per poc, fora d'aquest interval. Per tant, rebutgem la hipòtesi nul·la i diem que no és probable que la mostra pertanyi a una població amb una mitjana de 4,8.

3.4.

$\overline{X} = 5 \qquad S = 1{,}5 \qquad N = 250$
Per $\alpha = 0{,}01$ la $Z = 2{,}58$ (fem servir Z perquè N és molt gran.)

$$\overline{X} +/- Z \cdot S / \sqrt{N}$$

$$\overline{X} + Z. \cdot S / \sqrt{N} = 5 + 2{,}58 \cdot (1{,}5/\sqrt{250}) = 5{,}24$$

$$\overline{X} - Z \cdot S / \sqrt{N} = 5 - 2{,}58 \cdot (1{,}5/\sqrt{250}) = 4{,}76$$

Amb aquest nivell de confiança sí que podem admetre que la mostra de 250 casos amb una mitjana de 5 i S = 1,5 pot ser extreta d'una població amb μ = 4,8 ja que aquest valor queda dins de l'interval.

Exercici 4

En aquest cas tenim les dades següents:

$\overline{X} = 24$ $\mu = 22$ $\sigma = 4$ $N = 200$ Per $\alpha = 0{,}05$

Formulem les hipòtesis:

H_0: la mostra de 200 estudiants amb una mitjana de 24 pertany a la població amb una mitjana de 22.
H_1: la mostra de 200 estudiants amb una mitjana de 24 no pertany a la població amb una mitjana de 22.

Tipifiquem el valor de la mitjana mostral, tenint en compte que σ és coneguda.

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{N}} = \frac{24 - 22}{4/\sqrt{200}} = \frac{2}{0{,}28} = 7{,}14$$

Fixem-nos que el valor de Z que surt està molt per sobre d'1,96, que és el valor de *Z* per al nivell de confiança del 0,05. Per tant, rebutgem la H_0 i acceptem la hipòtesi alternativa. Això vol dir que a un nivell de confiança del 0,05 és molt poc probable que es pugui extreure una mostra de 200 estudiants de la FEP que tinguin una mitjana d'edat de 24 anys.

I seria probable obtenir una mostra de 20 estudiants amb una mitjana d'edat de 24 anys?

$$t = \frac{\overline{X} - \mu}{\sigma / \sqrt{N}} = \frac{24 - 22}{4 \sqrt{20}} = \frac{2}{0{,}89} = 2{,}25$$

En aquest cas, com que la mostra és petita, hem de fer una comparació amb els valors de *t*. Busquem la *t* crítica en les taules:

t_c (a 0,05; g. ll. 19) = 2,093

Com que 2,093 < 2,25, podem concloure que també és poc probable obtenir una mostra a l'atzar de 20 casos i que surti una mitjana de 24 anys.

Exercici 5

$\overline{X} = 400$ € $S = 25$ $N = 300$ Per $\alpha = 0{,}05$ la $Z = 1{,}96$ (fem servir *Z* perquè *N* és molt gran.)

Formulem les hipòtesis:

H_0: és molt probable que la mitjana de lloguer de la població sigui de 390 euros. És a dir, la mitjana de la població de la qual s'ha extret la mostra de 300 persones pot ser de 390 euros.
H_1: la mostra de 300 persones amb una mitjana de 400 euros no pertany a una població amb una mitjana de 390 euros.

El que fem és buscar l'interval de probabilitat de la mitjana de la població de la qual s'ha extret la mostra de 300 persones amb una mitjana de 400 € i una desviació típica de 25 €.

$\overline{X} \pm Z \cdot S / \sqrt{N}$

$\overline{X} + Z. \cdot S/\sqrt{N} = 400 + 1{,}96 \cdot (25 / \sqrt{300}) = 400 + 1{,}96 \cdot 1{,}44 = 402{,}82$

$\overline{X} - Z \cdot S / \sqrt{N} = 400 - 1{,}96 \cdot (25/\sqrt{300}) = 400 - 1{,}96 \cdot 1{,}44 = 397{,}18$

Per tant, a un nivell de confiança del 95 %, rebutgem la H_0. Per tant, és molt poc probable que la mitjana de la població sigui de 390 euros, ja que aquest valor es troba per sota del límit inferior de l'interval de probabilitat.

També es pot resoldre aquest exercici tipificant el valor de la mitjana de la mostra, agafant la mitjana de 390 com a mitjana de la població.

$$Z = \frac{\overline{X} - \mu}{S / \sqrt{N}} = \frac{400 - 390}{25 / \sqrt{300}} = \frac{10}{1{,}44} = 6{,}94$$

Tenint en compte que ens demanen que utilitzem un nivell de confiança del 95 %, el valor de *Z* corresponent a aquest nivell de confiança és d'1,96. El valor de *Z* que ens ha sortit és molt més gran que 1,96. Per tant, haurem de concloure que cal rebutjar la hipòtesi nul·la i que, per tant, acceptem que no és probable, tenint en compte les dades obtingudes amb l'estudi de la mostra, que la mitjana de la població pugui ser de 390 €.

Exercici 6

H_0: La mostra pertany a la població. És a dir, la diferència entre la mitjana de la mostra i la de la població és deguda a l'atzar. Això significa que aquesta diferència no és estadísticament significativa.

H_1: La mostra no pertany a la població. La diferència entre les dues mitjanes no és deguda a l'atzar. És a dir, la diferència entre la mitjana de la mostra i la de la població és estadísticament significativa.

$$t = \frac{\overline{X} - \mu}{\sigma / \sqrt{N}} = \frac{15.000 - 16.000}{1.500 \sqrt{30}} = \frac{-1.000}{273,8} = -3,65$$

t_c (g. ll. 29; a 0,05) = 2,045

-3,65<-2,045

Rebutgem la H_0. La mostra no pertany a la població.

Per tant, segons aquests resultats, podem dir amb un 95 % de confiança que la mitjana d'ingressos anuals de les educadores socials és significativament menor que la mitjana d'ingressos del conjunt de diplomats universitaris.

6. Exercicis de recopilació i respostes

Un grup d'educadores ha dissenyat un programa d'intervenció educativa per millorar les habilitats de comunicació de persones adultes amb discapacitat psíquica, usuàries d'un servei de teràpia ocupacional. Per tal d'esbrinar si el programa contribueix a millorar aquestes habilitats, les educadores plantegen realitzar un estudi pilot. El que fan és el que s'explica a continuació.

Seleccionen aleatòriament 40 persones que són distribuïdes en dos grups. Amb un dels grups (el grup 1) treballen el programa elaborat, i amb l'altre (grup 2) segueixen realitzant les activitats habituals. Les educadores recullen dades referents a les habilitats de comunicació mitjançant una prova específica que avalua aquest tipus d'habilitats, la qual puntua de l'1 al 20. La prova es passa als dos grups abans de l'aplicació del programa (Comunic1) i una altra vegada, després de 6 mesos d'aplicació del programa (Comunic2). També recullen informació sobre les habilitats socials que tenen aquestes persones amb una prova específica que puntua de l'1 al 10 (hsocials).

La matriu de dades obtingudes per als 40 casos s'adjunta a aquest document.

A partir d'aquesta situació respon a les qüestions que es plantegen tot seguit.

1. A continuació tens una taula de freqüències de la variable "Puntuació d'habilitats socials" (hsocials) del grup de dones.

HSOCIALS Puntuació d'habilitats socials

Puntuació d'habilitats socials

Puntuació	f
3,20	2
4,50	1
5,00	2
5,30	1
5,40	2
5,50	1
5,60	1
6,20	1
6,30	1
7,00	1
7,20	1
7,90	1
8,00	3
8,30	1
9,00	1
Total	20

1.1. Busca les freqüències acumulades, les freqüències relatives i els percentatges per a cada valor.
1.2. Calcula la mitjana, la mediana i la desviació típica i explica què signifiquen.

2. Resumeix la variable “sexe” en una taula de freqüències. Quin tipus de gràfica resulta més adequada per a aquesta variable? Justifica la resposta.

3. Una altra cosa que volen saber és si hi ha dependència entre les proves sobre habilitats de comunicació que passen abans i després de l’aplicació del programa. Hem de realitzar una prova per comprovar si hi ha relació entre les dues variables. A continuació tens les dades que necessites per poder fer aquest càlcul:

	Comunic1 (X)	**Comunic2 (Y)**
ΣX	332,6	358,9
ΣX^2	3.187,08	3.656,95
ΣXY	3.367,6	

3.1. Quina prova has d’aplicar?
3.2. Aplica-la.
3.3. Interpreta el resultat amb un marge d’error del 0,05.
3.4. Què significa?
3.5. Es pot buscar la recta de regressió? Si pots, busca-la.
3.6. Quin valor podem predir per a una puntuació en habilitats de comunicació (en la segona prova) per a una persona que hagi obtingut una puntuació de 9 en la primera prova?

4. Les educadores es pregunten si el fet que obtinguin una puntuació elevada en habilitats de comunicació (Comunic1) implica que també tinguin una puntuació elevada en habilitats socials (Hsocials). Per això apliquen una prova entre les puntuacions de la prova de comunicació que es passa a l’inici de la recerca i les puntuacions de la prova d’habilitats socials. Obtenen un coeficient de 0,3447.
   4.1. Quina prova creus que han aplicat? Per què?
   4.2. Interpreta el resultat tant per a un nivell de confiança del 95 % com per a un del 99 %.

5. Suposant que la mostra de 40 casos és normal i és representativa de la població de la qual s’ha extret, quina mitjana podem estimar en la població per a la variable “habilitats de comunicació”? Busca-la amb la variable Comunic1 i utilitza un marge d’error de $\alpha = 0{,}05$.

6. Una persona ha tret una puntuació de 10 punts tant en la prova d’habilitats de comunicació (Comunic1) com en la prova d’habilitats socials. Suposant que les distribucions en la població d’aquestes dues variables segueixin el model de la corba normal, en quina de les dues proves aquesta persona ha tret, comparativament, una millor puntuació?

Els estudis que s’han realitzat aplicant aquesta mateixa prova d’habilitats socials a persones amb discapacitat moderada donen una mitjana en habilitats socials de 6,7. Podem dir que la mostra de 40 persones del servei de teràpia ocupacional tenen una mitjana significativament inferior? Interpreta el resultat amb un marge d’error tant del 0,05 com del 0,01.

Cas	Sexe	Grup	Comunic1	Comunic2	Hsocials
1	dona	1	6	8,3	4,5
2	dona	1	6,5	6,7	6,2
3	dona	1	7,2	9,4	5,4
4	dona	1	8,5	11	7
5	dona	1	5,3	5,6	5
6	dona	1	10	10	8
7	dona	1	11,5	11,4	5,3
8	dona	1	9,3	14	7,2
9	dona	1	6	13	8,3
10	dona	1	13	12,8	5,5
11	dona	2	15	15,1	3,2
12	dona	2	12	11,5	9
13	dona	2	6	7	5,4
14	dona	2	8,7	8,5	6,3
15	dona	2	9,9	8	5,6
16	dona	2	11,2	11,8	8
17	dona	2	14,3	16	7,9
18	dona	2	5,6	5,4	8
19	dona	2	7,6	7,6	3,2
20	dona	2	8,8	8,8	5
21	home	1	4,3	5,2	3
22	home	1	5,8	6	7,1
23	home	1	7,9	7,8	4,2
24	home	1	4,5	5,3	6,5
25	home	1	6,8	8,7	7,3
26	home	1	18,3	18,9	9
27	home	1	10,4	11	7
28	home	1	4,9	6	4
29	home	1	6,5	7,2	3,6
30	home	1	7,2	7,2	7,6
31	home	2	4,3	4,4	6,7
32	home	2	6,8	6,7	8,1
33	home	2	6	6,2	5,1
34	home	2	5,9	5,7	4
35	home	2	6,3	6,1	3,2
36	home	2	7,4	7,4	6,7
37	home	2	11,9	11	8
38	home	2	13,2	12	7
39	home	2	6,1	7	6
40	home	2	5,7	7,2	5,1
		Σx	332,6	358,9	243,2
		Σx^2	3187,08	3656,95	1593,98
		Mitjana	8,32	8,97	6,08
		S	3,25	3,30	1,70

Respostes

Exercici 1

Puntuació d'habilitats socials

		Freqüència	Freqüències acumulades	Percentatge	Percentatge acumulat
Vàlids	3,20	2	2	10,0	10,0
	4,50	1	3	5,0	15,0
	5,00	2	5	10,0	25,0
	5,30	1	6	5,0	30,0
	5,40	2	8	10,0	40,0
	5,50	1	9	5,0	45,0
	5,60	1	10	5,0	50,0
	6,20	1	11	5,0	55,0
	6,30	1	12	5,0	60,0
	7,00	1	13	5,0	65,0
	7,20	1	14	5,0	70,0
	7,90	1	15	5,0	75,0
	8,00	3	18	15,0	90,0
	8,30	1	19	5,0	95,0
	9,00	1	20	5,0	100,0
	Total	20		100,0	

Estadístics descriptius

	N	Mínim	Màxim	Suma	Mitjana	Mediana	Desv. típica
Puntuació d'habilitats socials	20	3,20	9,00	124,00	6,2000	52,5	1,67363
N vàlid (segons llista)	20						

Exercici 2

sexe

		Freqüència	Percentatge	Percentatge vàlid	Percentatge acumulat
Vàlids	dona	20	50,0	50,0	50,0
	home	20	50,0	50,0	100,0
	Total	40	100,0	100,0	

Exercici 3

3.2.

Surt r = 0,89348163

Comunic1	**Comunic2**	X^2	Y^2	***XY***
6	8,3	36	68,89	49,8
6,5	6,7	42,25	44,89	43,55
7,2	9,4	51,84	88,36	67,68
8,5	11	72,25	121	93,5
5,3	5,6	28,09	31,36	29,68
10	10	100	100	100
11,5	11,4	132,25	129,96	131,1
9,3	14	86,49	196	130,2
6	13	36	169	78
13	12,8	169	163,84	166,4
15	15,1	225	228,01	226,5
12	11,5	144	132,25	138
6	7	36	49	42
8,7	8,5	75,69	72,25	73,95
9,9	8	98,01	64	79,2
11,2	11,8	125,44	139,24	132,16
14,3	16	204,49	256	228,8
5,6	5,4	31,36	29,16	30,24
7,6	7,6	57,76	57,76	57,76
8,8	8,8	77,44	77,44	77,44
4,3	5,2	18,49	27,04	22,36
5,8	6	33,64	36	34,8
7,9	7,8	62,41	60,84	61,62
4,5	5,3	20,25	28,09	23,85
6,8	8,7	46,24	75,69	59,16
18,3	18,9	334,89	357,21	345,87
10,4	11	108,16	121	114,4
4,9	6	24,01	36	29,4
6,5	7,2	42,25	51,84	46,8
7,2	7,2	51,84	51,84	51,84
4,3	4,4	18,49	19,36	18,92
6,8	6,7	46,24	44,89	45,56
6	6,2	36	38,44	37,2
5,9	5,7	34,81	32,49	33,63
6,3	6,1	39,69	37,21	38,43
7,4	7,4	54,76	54,76	54,76
11,9	11	141,61	121	130,9
13,2	12	174,24	144	158,4
6,1	7	37,21	49	42,7
5,7	7,2	32,49	51,84	41,04
332,6	**358,9**	**3187,08**	**3656,95**	**3367,6**

3.3. r_c (38 g. ll. i $\alpha = 0{,}05$) = 0,325
Com que la r observada és més gran que 0,325, direm que amb un marge d'error del 0,05 hi ha correlació significativa entre les dues variables. La correlació és positiva. A més, segons els valors de Guilford, és molt elevada.

3.4. Això vol dir que la persona que treu una puntuació alta en la prova d'habilitats de comunicació (comunic1) tendeix a treure una puntuació alta en la prova d'habilitats de comunicació feta després de l'aplicació del programa, i a l'inrevés, és a dir, una persona que obtingui una puntuació baixa en la variable Comunic1 tendeix a treure també una puntuació baixa en la variable Comunic2.

3.5. Recta de regressió

$Y = a + bX$

$Y = 1{,}41 + 0{,}909X$

3.6.

Per a $X = 9$ $Y = 0{,}909 \cdot 9 + 1{,}41 = 9{,}6$

Tenint en compte que l'error típic de predicció és 1,52.

Podem predir que el més probable és que, si una persona treu una puntuació de 9 en la primera prova d'habilitats de comunicació, en la segona prova la puntuació estarà entre 8,08 i 11,12. Aquests dos valors es troben sumant i restant l'error típic de predicció (1,52) a la puntuació que hem predit per a la variable *Y* (9,6).

Exercici 4

En aquest cas, com que r_c (g. ll. 38; i α 0,05) = 0,325 i r = 0,3447, podem dir que hi ha correlació significativa entre les dues variables.
Però per a r_c (g. ll. 38; α 0,01) = 0,418. No hi ha correlació significativa.

Exercici 5

Comunic1

Mitjana = 8,32 t_c (g. ll. 39; α = 0,05) = 2,021
S = 3,25
N = 40

$\overline{X} \pm t \cdot S / \sqrt{N} = 8{,}32 \pm 2{,}021 \cdot 3{,}25 / \sqrt{40} = 8{,}32 \pm 1{,}04$

Els dos valors que surten són 7,28 i 9,36.

Això vol dir que la mitjana que podem estimar per a la població amb un 95 % de probabilitat està entre 7,28 i 9,36.

Exercici 6

$$Z_1 = \frac{X - \overline{X_1}}{S_1} = \frac{10 - 8{,}32}{3{,}25} = 0{,}52$$

$$Z_2 = \frac{X - \overline{X_2}}{S_2} = \frac{10 - 6{,}08}{1{,}7} = 2{,}31$$

És millor una puntuació de 10 en la prova d'habilitats socials que una puntuació de 10 en la prova d'habilitats de comunicació (Comunic1).

Exercici 7

H_0: la mostra de 40 casos amb una mitjana de 6,08 en la prova d'habilitats socials pertany a una població amb una mitjana de 6,7. És a dir, no hi ha diferències significatives entre la mitjana de la mostra i la mitjana de la població. O dit d'una altra manera encara: la mostra de 40 casos, amb una mitjana de 6,08 i una desviació típica d'1,7 és representativa d'una població amb una mitjana de 6,7.

H_1: la mostra de 40 casos amb una mitjana en habilitats socials de 6,08 no pertany a una població amb una mitjana de 6,7. És a dir, hi ha diferències significatives entre la mitjana de la mostra i la mitjana de la població. La mostra de 40 casos, amb una mitjana de 6,08 i una desviació típica d'1,7 no és representativa d'una població amb una mitjana de 6,7.

(N'hi ha prou de formular-les d'una d'aquestes maneres.)

$$t = \frac{\overline{X} - \mu}{S_x} = \frac{6{,}08 - 6{,}7}{1{,}7/\sqrt{40}} = \frac{-0{,}62}{0{,}27} = -2{,}29$$

t_c (g. ll. 39; α = 0,05) = 2,021

Per tant, amb un marge d'error del 0,05 (o amb un nivell de confiança del 95 %), podem rebutjar la hipòtesi nul·la. És a dir, la mostra no pertany a la població amb una mitjana de 6,7. També podem dir que la diferència entre la mitjana de la mostra i la de la població és significativa (vol dir que no és deguda a l'atzar). Segons aquest resultat, doncs, la mitjana en habilitats socials de la mostra d'aquest servei de teràpia ocupacional és significativament inferior a la mitjana en habilitats socials de les persones amb discapacitats moderades.

t_c (g. ll. 39 α = 0,01) = 2,704

Si agafem un marge d'error del 0,01, com que la *t* que hem obtingut ens queda dins de l'interval format pels valors -2,704 i +2,704, acceptem la hipòtesi nul·la. Això vol dir que, amb aquest marge d'error, les diferències entre les dues mitjanes són degudes a l'atzar. Dit d'una altra manera, és probable que la mostra de 40 casos amb una mitjana de 6,08 i una desviació típica d'1,7 pertanyi a una població amb una mitjana de 6,7. Amb un nivell de confiança del 99 %, doncs, la mitjana d'habilitats socials de la mostra del servei de teràpia ocupacional no és significativament inferior a la mitjana de la població.

Per tant, en funció del marge d'error, tindrem una interpretació o una altra.

7. Referències bibliogràfiques

CALVO, Félix. *Estadística aplicada*. Bilbao: Ediciones Deusto, 1987.

ETXEBERRIA, Juan; TEJEDOR, Javier. *Análisis descriptivo de datos en educación*. Madrid: La Muralla, 2005.

WELKOWITZ, Joan; EWEN, Robert B.; COHEN, Jacob. *Estadística aplicada a las ciencias de la educación*. Madrid: Santillana, 1981.

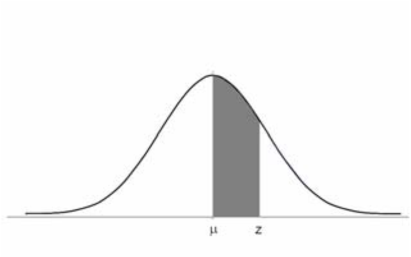
8. Annexos

Annex 1. Taula de valors crítics de la r *de Pearson*

gl (= *N* − 2; *N* = número de pares)	Nivel de significación para una prueba unilateral			
	0,05	0,025	0,01	0,005
	Nivel de significación para una prueba bilateral			
	0,10	0,05	0,02	0,01
1	0,988	0,997	0,9995	0,9999
2	0,900	0,950	0,980	0,990
3	0,805	0,878	0,934	0,959
4	0,729	0,811	0,882	0,917
5	0,669	0,754	0,833	0,874
6	0,622	0,707	0,789	0,834
7	0,582	0,666	0,750	0,798
8	0,549	0,632	0,716	0,765
9	0,521	0,602	0,685	0,735
10	0,497	0,576	0,658	0,708
11	0,476	0,553	0,634	0,684
12	0,458	0,532	0,612	0,661
13	0,441	0,514	0,592	0,641
14	0,426	0,497	0,574	0,623
15	0,412	0,482	0,558	0,606
16	0,400	0,468	0,542	0,590
17	0,389	0,456	0,528	0,575
18	0,378	0,444	0,516	0,561
19	0,369	0,433	0,503	0,549
20	0,360	0,423	0,492	0,537
21	0,352	0,413	0,482	0,526
22	0,344	0,404	0,472	0,515
23	0,337	0,396	0,462	0,505
24	0,330	0,388	0,453	0,496
25	0,323	0,381	0,445	0,487
26	0,317	0,374	0,437	0,479
27	0,311	0,367	0,430	0,471
28	0,306	0,361	0,423	0,463
29	0,301	0,355	0,416	0,456
30	0,296	0,349	0,409	0,449
35	0,275	0,325	0,381	0,418
40	0,257	0,304	0,358	0,393
45	0,243	0,288	0,338	0,372
50	0,231	0,273	0,322	0,354
60	0,211	0,250	0,295	0,325
70	0,195	0,232	0,274	0,302
80	0,183	0,217	0,256	0,283
90	0,173	0,205	0,242	0,267
100	0,164	0,195	0,230	0,254

Extret de Welkowitz, J.; Ewen, R.B.; Cohen, J. (1986): Estadística aplicada a las Ciencias de la Educación. Madrid: Santillana. p. 377

Annex 2. Taula d'àrees sota la corba normal



Percentatge de l'àrea a la corba normal entre la mitjana i z

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	00,00	00,40	00,80	01,20	01,60	01,99	02,39	02,79	03,19	03,59
0,1	03,98	04,38	04,78	05,17	05,57	05,96	06,36	06,75	07,14	07,53
0,2	07,93	08,32	08,71	09,10	09,48	09,87	10,26	10,64	11,03	11,41
0,3	11,79	12,17	12,55	12,93	13,31	13,68	14,06	14,43	14,80	15,17
0,4	15,54	15,91	16,28	16,64	17,00	17,36	17,72	18,08	18,44	18,79
0,5	19,15	19,50	19,85	20,19	20,54	20,88	21,23	21,57	21,90	22,24
0,6	22,57	22,91	23,24	23,57	23,89	24,22	24,54	24,86	25,17	25,49
0,7	25,80	26,11	26,42	26,73	27,04	27,34	27,64	27,94	28,23	28,52
0,8	28,81	29,10	29,39	29,67	29,95	30,23	30,51	30,78	31,06	31,33
0,9	31,59	31,86	32,12	32,38	32,64	32,89	33,15	33,40	33,65	33,89
1,0	34,13	34,38	34,61	34,85	35,08	35,31	35,54	35,77	35,99	36,21
1,1	36,43	36,65	36,86	37,08	37,29	37,49	37,70	37,90	38,10	38,30
1,2	38,49	38,69	38,88	39,07	39,25	39,44	39,62	39,80	39,97	40,15
1,3	40,32	40,49	40,66	40,82	40,99	41,15	41,31	41,47	41,62	41,77
1,4	41,92	42,07	42,22	42,36	42,51	42,65	42,79	42,92	43,06	43,19
1,5	43,32	43,45	43,57	43,70	43,82	43,94	44,06	44,18	44,29	44,41
1,6	44,52	44,63	44,74	44,84	44,95	45,05	45,15	45,25	45,35	45,45
1,7	45,54	45,64	45,73	45,82	45,91	45,99	46,08	46,16	46,25	46,33
1,8	46,41	46,49	46,56	46,64	46,71	46,78	46,86	46,93	46,99	47,06
1,9	47,13	47,19	47,26	47,32	47,38	47,44	47,50	47,56	47,61	47,67
2,0	47,72	47,78	47,83	47,88	47,93	47,98	48,03	48,08	48,12	48,17
2,1	48,21	48,26	48,30	48,34	48,38	48,42	48,46	48,50	48,54	48,57
2,2	48,61	48,64	48,68	48,71	48,75	48,78	48,81	48,84	48,87	48,90
2,3	48,93	48,96	48,98	49,01	49,04	49,06	49,09	49,11	49,13	49,16
2,4	49,18	49,20	49,22	49,25	49,27	49,29	49,31	49,32	49,34	49,36
2,5	49,38	49,40	49,41	49,43	49,45	49,46	49,48	49,49	49,51	49,52
2,6	49,53	49,55	49,56	49,57	49,59	49,60	49,61	49,62	49,63	49,64
2,7	49,65	49,66	49,67	49,68	49,69	49,70	49,71	49,72	49,73	49,74
2,8	49,74	49,75	49,76	49,77	49,77	49,78	49,79	49,79	49,80	49,81
2,9	49,81	49,82	49,82	49,83	49,84	49,84	49,85	49,85	49,86	49,86
3,0	49,87									
3,5	49,98									
4,0	49,997									
5,0	49,99997									

Extret de Welkowitz, J.; Ewen, R.B.; Cohen, J. (1986): *Estadística aplicada a las Ciencias de la Educación*. Madrid: Santillana. p. 37

An*nex 3. Taula dels valors crítics de la distribució t*

gl	Nivel de significación para una prueba unilateral					
	0,10	0,05	0,025	0,01	0,005	0,0005
	Nivel de significación para una prueba bilateral					
	0,20	0,10	0,05	0,02	0,01	0,001
1	3,078	6,314	12,706	31,821	63,657	636,619
2	1,886	2,920	4,303	6,965	9,925	31,598
3	1,638	2,353	3,182	4,541	5,841	12,941
4	1,533	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,859
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,449	5,405
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
11	1,363	1,796	2,201	2,718	3,106	4,437
12	1,356	1,782	2,179	2,681	3,055	4,318
13	1,350	1,771	2,160	2,650	3,012	4,221
14	1,345	1,761	2,145	2,624	2,977	4,140
15	1,341	1,753	2,131	2,602	2,947	4,073
16	1,337	1,746	2,120	2,583	2,921	4,015
17	1,333	1,740	2,110	2,567	2,898	3,965
18	1,330	1,734	2,101	2,552	2,878	3,922
19	1,328	1,729	2,093	2,539	2,861	3,883
20	1,325	1,725	2,086	2,528	2,845	3,850
21	1,323	1,721	2,080	2,518	2,831	3,819
22	1,321	1,717	2,074	2,508	2,819	3,792
23	1,319	1,714	2,069	2,500	2,807	3,767
24	1,318	1,711	2,064	2,492	2,797	3,745
25	1,316	1,708	2,060	2,485	2,787	3,725
26	1,315	1,706	2,056	2,479	2,779	3,707
27	1,314	1,703	2,052	2,473	2,771	3,690
28	1,313	1,701	2,048	2,467	2,763	3,674
29	1,311	1,699	2,045	2,462	2,756	3,659
30	1,310	1,697	2,042	2,457	2,750	3,646
40	1,303	1,684	2,021	2,423	2,704	3,551
60	1,296	1,671	2,000	2,390	2,660	3,460
120	1,289	1,658	1,980	2,358	2,617	3,373
∞	1,282	1,645	1,960	2,326	2,576	3,291

Extret de Welkowitz, J.; Ewen, R.B.; Cohen, J. (1986): *Estadística aplicada a las Ciencias de la Educación*. Madrid: Santillana. P. 376