

Estadística descriptiva¹

*Lluís F. Sanjuan i Nebot
Departament d'Infermeria
Universitat de València*

¹ *Aquests materials han rebut un incentiu del Servei de Política Lingüística*

ESTADÍSTICA DESCRIPTIVA

1. Conceptes generals

Estadística descriptiva

Descriu les regularitats o característiques existents en un conjunt de dades, organitzant-les en taules i representacions gràfiques i analitzant-les mitjançant l'obtenció de valors numèrics representatius.

Població

Conjunt d'elements que presenten uns trets comuns i en els quals volem estudiar una o diverses característiques.

Ingressos en la Unitat d'Obstetrícia d'un hospital durant l'any 2012
 Alumnes matriculats a la Facultat d'Infermeria i Podologia el curs 2014-2015
 Històries d'infermeria obertes en un centre de salut al llarg d'un any
 Persones empadronades a la Comunitat Valenciana a data 1 de gener de 2015-09-07
 Xeringues fabricades en una indústria

Individu/unitat

Cada un dels elements que pertanyen a la població.

Grandària de la població (N)

Nombre d'individus que hi ha a la població. Segons la seua dimensió les poblacions es classifiquen en *finites* (formada per un nombre limitat d'unitats) o *infinites* (si el nombre d'elements és il·limitat o no pot establir-se).

Mostra

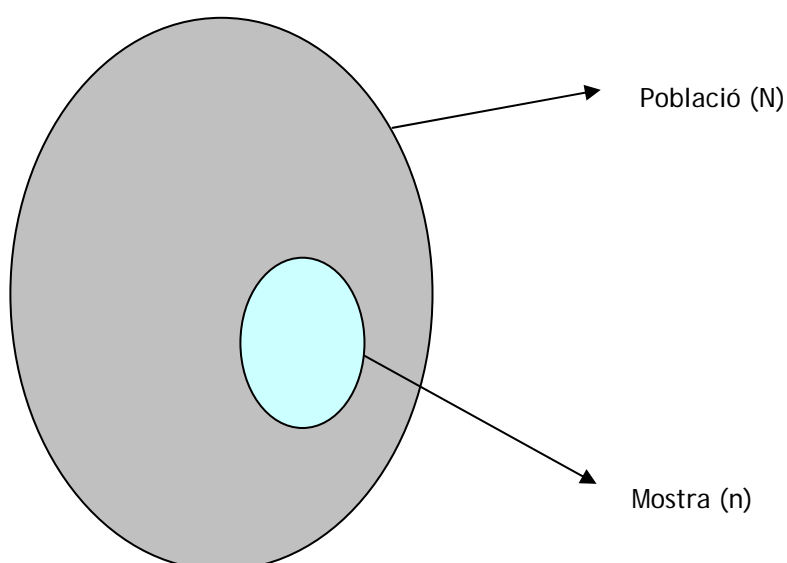
Subconjunt representatiu de la població que s'estudia quan no és possible accedir a tots els elements de la població.

Grandària de la mostra (n)

Nombre d'individus que hi ha a la mostra (generalment ve determinat pel tipus d'estudi que realitzem, les característiques de la població estudiada, el grau d'error que estiguem disposats a assumir, etc.).

Mostratge

És el procés seguit per obtenir una mostra.



2. Variable estadística

Variable

Qualsevol aspecte de la població que ens interessa estudiar. Cada una de les possibles manifestacions de la variable es denomina *modalitat* i aquestes han de constituir una llista *exhaustiva* (tot element de la població presenta una de les modalitats de la llista) i *exclusiva* (cada individu presenta una única modalitat).

Variable qualitativa

Les modalitats són qualitats, atributs o categories, no poden mesurar-se.

Motiu d'ingrés; Sala on ingressa; Sexe del nounat; Grup sanguini

Variable quantitativa

A cada modalitat li podem associar un número. El conjunt de valors numèrics de la variable es denomina *recorregut o rang*. Es poden classificar en discretes i contínues.

Variable quantitativa discreta

Si el recorregut està constituït per un nombre finit o infinit numerable d'elements. Generalment són variables que només admeten valors enters $\langle 1;2;3;4;\dots \rangle$.

Nombre de parts anteriors $\langle 0;1;2;3;4;\dots \rangle$

Assignatures matriculades en el curs actual $\langle 1;2;3;4;\dots \rangle$

Variable quantitativa contínua

Si el rang és un conjunt format per infinits valors o bé si la variable pot prendre *a priori* qualsevol valor dins d'un interval. Les modalitats en aquest tipus de variables són intervals o classes de possibles valors.

Pes en nàixer (grams); Edat en el moment d'ingressar (anys); Potassi en sang (mEq/l)

Si la variable és contínua caldrà establir un mètode que permeti la construcció d'intervals o classes per agrupar les dades observades.

Construcció d'intervals o classes

Per poder construir els intervals o classes cal definir els següents conceptes:

Sensibilitat (s)

És la unitat més petita que l'instrument que mesura la variable és capaç d'apreciar.

Variable	Sensibilitat (s)
Pes en nàixer	1 gram
Potassi en sang	0,1 mEq/l

Valor observat (X_o)

És el resultat obtingut amb el *procés de mesura* de la variable.

Variable	X_o
Pes en nàixer	2935 grams
Potassi en sang	4,6 mEq/l

Valor exacte (X_e)

És el valor *real* de la variable. Aquest valor es troba dins l'interval $X_o \pm \frac{s}{2}$. Els límits d'aquest interval s'anomenen límits exactes del valor observat.

Variable	X_o	s	X_e
----------	-------	---	-------

Pes en naixer	2935 grams	1	2934,5-2935,5
Potassi en sang	4,6 mEq/l	0,1	4,55-4,65

Amplitud de la distribució (A)

És la diferència entre el límit exacte superior del valor observat màxim i el límit exacte inferior del valor observat mínim.

$$A = (X_{\max} + \frac{s}{2}) - (X_{\min} - \frac{s}{2}) = X_{\max} - X_{\min} + s$$

Variable	s	X _{min}	X _{max}	A
Pes en naixer	1	1425	4540	4540-1425+1=3116
Potassi en sang	0,1	3,5	5,5	5,5-3,5+0,1=2,1

Classes o intervals

Són les possibles agrupacions dels valors de la variable.

Nombre d'intervals (k)

El nombre d'intervals depèn de la grandària de la població o de la mostra que estem estudiant. No hi ha un criteri fix per determinar-lo. Alguns dels criteris que podem trobar en la literatura són:

$$k = 1 + 3,322 * \log(N) \quad (\text{proposada per Herbert Sturges en 1926})$$

$$k = \sqrt{N}$$

$$\text{Si } N = 792 \rightarrow k = 1 + 3,322 * \log(792) = 10,63 \approx 11$$

$$\text{Si } N = 235 \rightarrow k = \sqrt{235} = 15,33 \approx 15$$

Longitud de l'interval (l)

S'obté dividint l'amplitud de la distribució (A) pel nombre d'intervals (k). Després s'aproxima el resultat, per excés, a un *múltiple de la sensibilitat* (l'aproximació tindrà les mateixes xifres decimals que la sensibilitat).

$$l \cong \frac{A}{k}$$

Variable	A	$l = \frac{A}{11}$	$l = \frac{A}{10}$
Pes en naixer	3116	283,27 \cong 284	311,6 \cong 312
Potassi en sang	2,1	0,191 \cong 0,2	0,21 \cong 0,3

Construcció de les classes o intervals

Per construir els intervals es fixa el límit exacte inferior del primer interval ($X_{\min} - \frac{s}{2}$) i a partir d'ell s'obtenen les diferents classes sumant la longitud (l).

Interval	l_i	L_i		l_i	L_i
1	$X_{\min} - \frac{s}{2} = 1424,5$	1424,5 + l		$X_{\min} - \frac{s}{2} = 1424,5$	1424,5 + l
2	1708,5	1708,5 + l		1424,5 + l	1424,5 + 2 · l
...
i	$l_i = L_{i-1}$	$L_i = L_{i-1} + l$		$1424,5 + (i - 1) \cdot l$	$1424,5 + i \cdot l$
...
11	4264,5	4264,5 + l		$1424,5 + 10 \cdot l$	$1424,5 + 11 \cdot l$

Límits aparents

Són els que tenen per extrems valors observables.

Marca o punt mitjà d'una classe (M_i)

És el punt mitjà de l'interval (límits exactes o límits aparents).

$$M_i = \frac{\text{Lím Sup} + \text{Lím Inf}}{2}$$

Intervals i marca de classe coneixent la sensibilitat, els valors mínim i màxim i el nombre d'interval.

s	0,1		Límits exactes		Límits aparents		Marca
X mín	121,1		121,05	125,85	121,1	125,8	123,45
X màx	154,4		125,85	130,65	125,9	130,6	128,25
A	33,4		130,65	135,45	130,7	135,4	133,05
k	7		135,45	140,25	135,5	140,2	137,85
l	4,7714	4,8	140,25	145,05	140,3	145,0	142,65
			145,05	149,85	145,1	149,8	147,45
			149,85	154,65	149,9	154,6	152,25

$$A = (154,4 + 0,05) - (121,1 - 0,05) = 154,4 - 121,1 + 0,1 = 33,4$$

$$\text{Si } k = 7 \rightarrow l = \frac{33,4}{7} = 4,7714 \cong 4,8$$

IMPORTANT

Categorització d'una variable quantitativa

Podem transformar una variable quantitativa en una variable qualitativa codificant els possibles valors de la variable i generant unes noves modalitats categòriques. Aquesta transformació implicarà sempre una pèrdua d'informació.

Pes del nadó [grams]

Baix pes (pes inferior a 2500 gr.)
 Pes normal (pes igual o superior a 2500 gr. i inferior a 4000 gr.)
 Pes elevat (pes igual o superior a 4000 gr.)

Nombre de parts anteriors a l'ingrés

0;1;2;3;més de tres

Edat

Població infantil, adulta i anciana

Sensibilitat

Si tenim les dades d'una variable contínua i no en coneixem la sensibilitat podem determinar-la amb la fórmula:

$s = 10^{-d}$, on d = núm. xifres decimals que s'observen a les dades

10,25;11,36;15,57;... $\rightarrow s = 10^{-2} = 0,01$

Variables relacionades amb el temps

Tindran un *tractament especial*. Les classes han de construir-se sense perdre la continuïtat.

Edat: entre 0 i 4 anys; entre 5 i 9 anys; 74 anys

Classes: [0;5); [5;9); [74;75)

Relació entre límits exactes (LE) i límits aparents (LA)

Si tenim límits aparents la sensibilitat (s) és la distància entre el límit superior d'un interval i el límit inferior de l'interval següent.

$$s = L_{i+1} - l_i$$

Si tenim límits exactes i no en coneixem la sensibilitat podem determinar-la amb la fórmula:

$s = 10^{-(d-1)}$, on d = núm. xifres decimals dels límits

Límits exactes		Límits aparents		Marca
4,5	9,5	5	9	7
9,5	14,5	10	14	12
14,5	19,5	15	19	17
19,5	24,5	20	24	22
24,5	29,5	25	29	27

$$\text{Sensibilitat} = s = 10^{-(d-1)} = 10^0 = 10 - 9 = 1$$

$$\text{Longitud} = l = 9,5 - 4,5 = 9 - 5 + s = 5$$

$$\text{Marca} = M_j = 7 = \frac{9,5 + 4,5}{2} = \frac{9 + 5}{2}$$

La longitud de l'interval amb límits exactes és igual a la longitud de l'interval amb límits aparents més la sensibilitat.

La relació entre límits exactes i aparents és la següent:

Límits exactes		Límits aparents	
4,5	9,5	5	9

$$LE_{\text{inf}} = LA_{\text{inf}} - \frac{s}{2} \rightarrow 4,5 = 5 - 0,5$$

$$LE_{\text{sup}} = LA_{\text{sup}} + \frac{s}{2} \rightarrow 9,5 = 9 + 0,5$$

$$LA_{\text{inf}} = LE_{\text{inf}} + \frac{s}{2} \rightarrow 5 = 4,5 + 0,5$$

$$LA_{\text{sup}} = LE_{\text{sup}} - \frac{s}{2} \rightarrow 9 = 9,5 - 0,5$$

3. Distribucions de freqüència

Les distribucions de freqüència permeten una millor presentació i interpretació de les dades observades.

Freqüència/freqüència absoluta ordinària/nombre de casos ($f_i; n_i$)

La freqüència d'una modalitat M_i és el nombre d'individus que presenten la modalitat. La suma de les freqüències es sempre igual al nombre total d'observacions (N).

$$\sum f_i = N$$

Proporció/freqüència relativa ordinària ($h_i; p_i$)

La freqüència relativa d'una modalitat M_i és la proporció d'individus que presenten la modalitat. La suma de les proporcions es sempre igual a u.

$$h_i = \frac{f_i}{N} \quad \sum h_i = 1$$

Percentatge ($\%_i$)

El percentatge d'una modalitat M_i és el percentatge d'individus que presenten la modalitat. La suma dels percentatges es sempre igual a cent.

$$\%_i = 100 * h_i = \frac{100 * f_i}{N} \quad \sum \%_i = 100$$

Variable [N=110]

Modalitat	f_i	h_i	$\%_i$
M_1	63	0,5727	57,27
M_2	2	0,0181	1,81
M_3	45	0,4090	40,90
Total	110	1	100

Si la variable és quantitativa podem definir els conceptes de freqüència, proporció i percentatge acumulat.

Suposem que la variable presenta k possibles modalitats $\langle M_1; M_2; \dots; M_k \rangle$ on M_i serà un valor si la variable és discreta o un interval si la variable és contínua.

Freqüència acumulada/freqüència absoluta acumulada/casos acumulats ($F_i; n_i$)

La freqüència acumulada d'una modalitat M_i és el nombre d'individus en els quals observem un valor igual o inferior a M_i si la variable és discreta. Definim la variable contínua com el nombre d'individus en els quals observem un valor igual o inferior al límit superior de l'interval.

$$F_i = \sum_{j \leq i} f_j \quad F_i = F_{i-1} + f_i \quad F_k = N$$

Proporció acumulada/freqüència relativa acumulada ($H_i; P_i$)

La proporció acumulada d'una modalitat M_i és la proporció d'individus en els quals observem un valor igual o inferior a M_i si la variable és discreta. Definim la variable contínua com la proporció d'individus en els quals observem un valor igual o inferior al límit superior de l'interval.

$$H_i = \sum_{j \leq i} h_j \quad H_i = H_{i-1} + h_i \quad H_k = 1 \quad H_i = \frac{F_i}{N}$$

Percentatge acumulat ($\%_i^a$)

El percentatge acumulat d'una modalitat M_i és el percentatge d'individus en els quals observem un valor igual o inferior a M_i si la variable és discreta. Definim la variable contínua com el percentatge d'individus en els quals observem un valor igual o inferior al límit superior de l'interval.

$$\%_i^a = \sum_{j \leq i} \%_j \qquad \%_i^a = \%_{i-1}^a + \%_i \qquad \%_k^a = 100 \qquad \%_i^a = 100 * H_i = 100 * \frac{F_i}{N}$$

Determinació de la freqüència acumulada [coneguda la freqüència]

Modalitat	f_i	$F_i = \sum_{j \leq i} f_j$	$F_i = F_{i-1} + f_i$
M_1	655	655	655
M_2	123	655+123=778	655+123=778
M_3	16	655+123+16=797	778+16=797
M_4	1	655+123+16+1=798	797+1=798

Determinació de la freqüència [coneguda la freqüència acumulada]

Modalitat	F_i	$f_i = F_i - F_{i-1}$
M_1	655	655
M_2	778	778-655=123
M_3	797	797-778=16
M_4	798	798-797=1

Interpretació de la freqüència (variable discreta)

Modalitat	f_i	Interpretació	F_i	Interpretació
3	9	9 casos amb valor observat igual a 3	9	9 casos \leq 3
4	18	18 casos amb valor observat igual a 4	27	27 casos \leq 4

Interpretació de la freqüència (variable contínua)

Modalitat	f_i	Interpretació	F_i	Interpretació
10,25-20,25	10	10 casos en l'interval 10,25 i 20,25	10	10 casos \leq que 20,25
20,25-30,25	23	23 casos en l'interval 20,25 i 30,25	33	33 casos \leq que 30,25

Taula de distribució de freqüències (TDF)

Si el nombre de dades observades és molt petit es presentaran en ordre de magnitud. En cas contrari les dades es resumiran mitjançant una taula de distribució de freqüències. La taula mostrarà les diferents modalitats i les seues freqüències.

En tota taula de distribució de freqüències hem de considerar:

1. El títol, que serà clar, concís i informatiu.
2. La taula: cel·les disposades en files i columnes. Generalment la primera fila indicarà a què es refereixen les dades i la primera columna mostrarà la llista de modalitats.
3. Les notes explicatives, que ens indiquen d'on s'han obtingut les dades.

També podem elaborar taules de distribució de freqüències on la primera fila mostra la llista de modalitats i la primera columna indica a què es refereixen les dades.

Títol

M	f	h	$\%$	F	H	$\%^a$
M_1	f_1	h_1	$\%_1$	F_1	H_1	$\%_1^a$
.....
M_k	f_k	h_k	$\%_k$	N	1	100
TOTAL	N	1	100			

Notes explicatives

Títol

M	M_1	M_k	TOTAL
f	f_1		f_k	N
h	h_1		h_k	1
$\%$	$\%_1$		$\%_k$	100
F	F_1		N	
H	H_1		1	
$\%^a$	$\%_1^a$		100	

Notes explicatives

IMPORTANT

- Si la variable és qualitativa les freqüències, les proporcions i els percentatges acumulats no es poden calcular.
- A les taules de distribució de freqüències de variables quantitatives les modalitats es mostraran en ordre creixent.
- Quan la variable estiga agrupada en intervals es treballarà amb límits exactes.
- La taula de distribució de freqüències resumirà tota la informació proporcionada per les dades observades si s'hi mostra alguna de les següents combinacions:
 - Modalitats i freqüències
 - Modalitats i freqüències acumulades
 - Modalitats, proporcions i nombre de dades observades
 - Modalitats, proporcions acumulades i nombre de dades observades
 - Modalitats, percentatges i nombre de dades observades
 - Modalitats, percentatges acumulats i nombre de dades observades
- Les proporcions s'expressaran amb quatre decimals, els percentatges amb dos.
- La referència al TOTAL pot ometre's.

Trasplantaments d'òrgans en hospitals de la Comunitat Valenciana. Any 2009

Tipus	Casos (f_i)	% casos ($\%_i$)
Pàncrees-Renyó	5	1,29
Lòbul hepàtic	6	1,55
Pulmó	24	6,20
Cor	36	9,30
Fetge	112	28,94
Renyó	204	52,71
Total	387	100

Font: Generalitat Valenciana. Conselleria de Sanitat

**Casos declarats de sida. Comunitat Valenciana.
Edat entre 20 i 49 anys. Situació a 30/09/2010**

Edat (anys)	Casos (f_i)	% acum. ($\%_i^a$)
20-24	406	7,17
25-29	1.355	31,11
30-34	1.697	61,10
35-39	1.217	82,60

40-44		644	93,98
45-49		341	100,00

Font: Registre de casos de sida a la Comunitat Valenciana

Pes dels nadons. Hosp. Dr. Peset. Any 1996

Pes (gr)	Casos (f_i)	Percentatge ($\%_i$)
1.424,5-1.708,5	2	0,25
1.708,5-1.999,5	2	0,25
1.992,5-2.276,5	18	2,27
2.276,5-2.560,5	38	4,80
2.560,5-2.844,5	110	13,89
2.844,5-3.128,5	188	23,74
3.128,5-3.412,5	189	23,86
3.412,5-3.696,5	135	17,05
3.696,5-3.980,5	86	10,86
3.980,5-4.264,5	19	2,40
4.264,5-4.548,5	5	0,63
Total	792	100

Avortaments anteriors. Ingressos a la Unitat d'Obstetrícia de l'Hospital Dr. Peset. Any 1996.

Avortaments	0	1	2	3
Casos	655	126	16	1
Casos acumulats	655	781	797	798

Sèries temporals o sèries cronològiques

Són aquelles en què la variable s'estudia en diferents moments (casos observats en un mes, any...) i es pretén mostrar com varia en relació amb el 'temps'. En aquests casos la presentació de les dades és semblant a una taula de distribució de freqüències, però no té sentit (generalment) el càlcul de proporcions i percentatges. A més si la característica estudiada es refereix a allò que ha ocorregut en una població en la qual el nombre d'individus varia al llarg del temps, és convenient incloure en la taula una columna que mostre la relació entre les dades i el nombre d'individus (generalment aquesta relació s'expressa com a *nombre de vegades que va ocórrer el fenomen per cada 10^k individus*).

Podòlegs col·legiats. Espanya 2003-2009

Any	Podòlegs	Població	Índex ¹
2003	3.362	42.717.064	7,87
2004	3.719	43.197.684	8,61
2005	4.158	44.108.530	9,43
2006	4.554	44.708.964	10,19
2007	4.909	45.200.737	10,86
2008	5.114	46.157.822	11,08
2009	5.263	46.745.807	11,26

Font: Podòlegs col·legiats a data 31/12; Població a data 01/01 (INE)

¹Podòlegs per 100.000 habitants

**Població de la província de València
1996-2006**

Data	Població	Índex ¹
01/05/1996	2.172.840	100,00
01/01/1999	2.187.633	100,68
01/01/2000	2.201.200	101,31
01/01/2001	2.227.170	102,50
01/01/2002	2.267.503	104,36
01/01/2003	2.320.297	106,79
01/01/2004	2.358.919	108,56
01/01/2005	2.416.628	111,22
01/01/2006	2.463.592	113,38

Font: Padró Municipal d'Habitants (INE)

Taules de contingència

Si estudiem dues variables [X amb m modalitats i Y amb k modalitats] la variable conjunta tindrà m · k modalitats i la presentació de les dades la farem mitjançant una taula de doble entrada [taula de contingència] que contindrà les modalitats i el nombre (proporció o percentatge) de casos que observem de cada una.

Taula de contingència

X / Y	M ₁	M ₂	M ₃	M ₄	Total
M ₁	8	25	80	17	130
M ₂	4	20	81	15	f _{2•} = 120
M ₃	3	f ₃₂ = 19	65	14	101
M ₄	3	54	40	10	107
M ₅	1	10	53	9	73
M ₆	1	26	49	9	85
M ₇	0	17	46	8	71
Total	20	171	f _{•3} = 414	82	687

En aquest exemple la variable X té 7 modalitats i la variable Y, 4. Per tant la variable conjunta tindrà 7 · 4 = 28 modalitats. Així la presentació de les dades la farem mitjançant una taula de doble entrada [taula de contingència] amb 28 cel·les. Cada cel·la conté el nombre (proporció o percentatge) de casos que presenten a la vegada una modalitat de X i una de Y. Així el valor 19 (f₃₂) indica que en 19 casos el valor de la variable X és M₃ el valor de la variable Y és M₂. Aquets valors s'anomenen freqüències (proporcions o percentatges) conjuntes.

Freqüències (proporcions o percentatges) marginals

Si sumem les files obtenim el nombre (proporció o percentatge) de casos de les modalitats de la variable X.

$$f_{i•} = \sum_j f_{ij} \qquad h_{i•} = \sum_j h_{ij} \qquad \%_{i•} = \sum_j \%_{ij}$$

Al nostre exemple:

$$f_{2•} = \sum_j f_{2j} = 4 + 20 + 81 + 15 = 120$$

Si sumem les columnes obtenim el nombre (proporció o percentatge) de casos de les modalitats de la variable Y.

$$f_{•j} = \sum_i f_{ij} \qquad h_{•j} = \sum_i h_{ij} \qquad \%_{•j} = \sum_i \%_{ij}$$

Al nostre exemple:

$$f_{•3} = \sum_i f_{i3} = 80 + 81 + 65 + 40 + 53 + 49 + 46 = 414$$

A aquests valors se'ls denomina freqüències (proporcions o percentatges) marginals i ens proporcionen informació sobre cadascuna de les dues variables estudiades.

Algunes propietats

Si sumem totes les freqüències (proporcions o percentatges) obtindrem el nombre total de casos (1 o 100) i el mateix ocorre si sumem les freqüències marginals de qualsevol de les dues variables estudiades.

$$\sum_{ij} f_{ij} = \sum_j f_{•j} = \sum_i f_{i•} = N$$

$$\sum_{ij} h_{ij} = \sum_j h_{\bullet j} = \sum_i h_{i\bullet} = 1$$

$$\sum_{ij} \%_{ij} = \sum_j \%_{\bullet j} = \sum_i \%_{i\bullet} = 100$$

Altres taules de contingència

En ocasions ens interessa presentar les dades observades en estudiar dues variables condicionant-ne una a l'altra. La interpretació de les dades en aquests casos varia significativament en funció del tipus de taula.

València. PMH 2012 (casos)			
	Gènere		
Edat	Homes	Dones	
0014	57.349	54.291	111.640
1564	267.043	272.180	539.223
6500	58.853	89.472	148.325
	383.245	415.943	799.188

$f_{0014,H} = 57.349$; hi ha 57.349 homes amb edat inferior als 15 anys.

València. PMH 2012 (%)			
	Gènere		
Edat	Homes	Dones	
0014	7,18	6,79	13,97
1564	33,41	34,06	67,47
6500	7,36	11,20	18,56
	47,95	52,05	100

$$\%_{ij} = \frac{f_{ij}}{N} \rightarrow \%_{0014,H} = \frac{57.349}{799.188} = 7,18 ; \text{ el } 7,18\% \text{ de}$$

la població són homes amb edat inferior als 15 anys.

València. PMH 2012 (Gènere Edat)			
	Gènere		
Edat	Homes	Dones	
0014	51,37	48,63	100
1564	49,52	50,48	100
6500	39,68	60,32	100
	47,95	52,05	100

$$\%_{j|i} = \frac{f_{ij}}{f_i} \rightarrow \%_{H|0014} = \frac{57.349}{111.640} = 51,37 ; \text{ al grup}$$

d'edat entre 0 i 14 anys, el 57,37% són homes.

València. PMH 2012 (Edat Gènere)			
	Gènere		
Edat	Homes	Dones	
0014	14,96	13,05	13,97
1564	69,68	65,44	67,47
6500	15,36	21,51	18,56
	100	100	100

$$\%_{j|ij} = \frac{f_{ij}}{f_j} \rightarrow \%_{0014|H} = \frac{57.349}{383.245} = 14,96 ; \text{ al grup}$$

d'homes el 14,96% tenen menys de 15 anys.

4. Gràfics

Encara que la taula de distribució de freqüències conté tota la informació disponible, de vegades resulta necessari presentar-la mitjançant un gràfic per poder tenir una millor visió de conjunt. En general els gràfics es basen en el principi de proporcionalitat entre àrees i freqüències (proporcions o percentatges).

Gràfics per a variables qualitatives

Diagrama de sectors

Associa a cada modalitat un sector circular amb angle central proporcional a la seua freqüència (proporció o percentatge) [en l'exemple es representen les freqüències de les modalitats de la variable *Trasplantaments en la CV. Any 2009*].

A cada modalitat M_i se li assigna un angle α_i proporcional a la freqüència de la modalitat

$$\alpha_i \propto f_i$$

$$\sum \alpha_i = 360^\circ = \sum cte \cdot f_i = cte \cdot N \rightarrow cte = \frac{360^\circ}{N}$$

$$\alpha_i = \frac{360^\circ}{N} \cdot f_i = 360^\circ \cdot h_i = 3,6^\circ \cdot \%_i$$

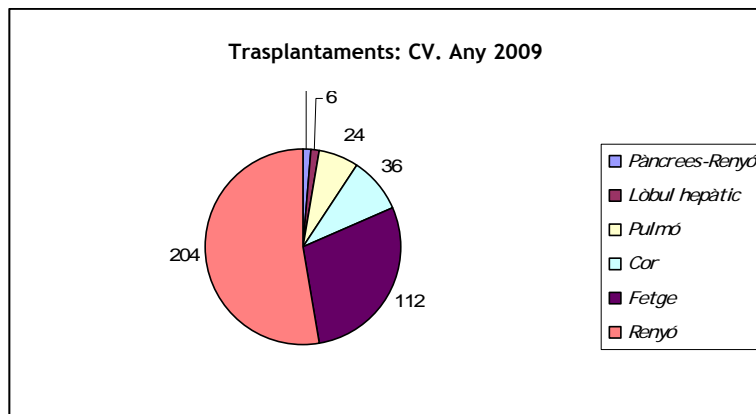


Diagrama de barres

Sobre uns eixos de coordenades marquem en l'eix d'abscisses les possibles modalitats i sobre l'eix d'ordenades la freqüència (proporció o percentatge). Sobre cada modalitat dibuixem rectangles de base constant i altura igual a la freqüència (proporció o percentatge) corresponent [en l'exemple es representen les proporcions de les modalitats de la variable *Lloc d'ingrés*].

Si a cada modalitat M_i li associem un rectangle de base igual a b i altura f_i

$$\text{Àrea del rectangle} = b \cdot f_i \rightarrow \text{Àrea del rectangle} \propto f_i$$

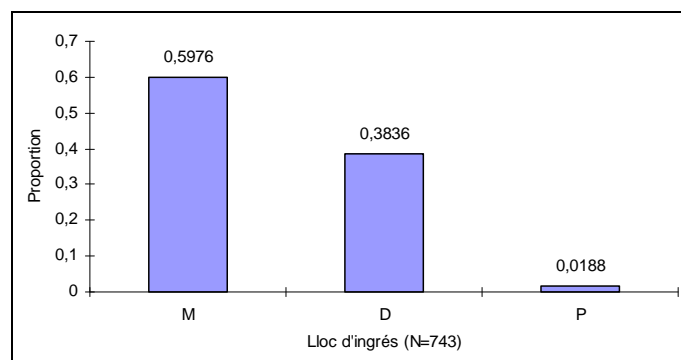


Diagrama de barres dobles

Semblant al diagrama de barres, s'utilitza per a representar conjuntament dues o més variables qualitatives.

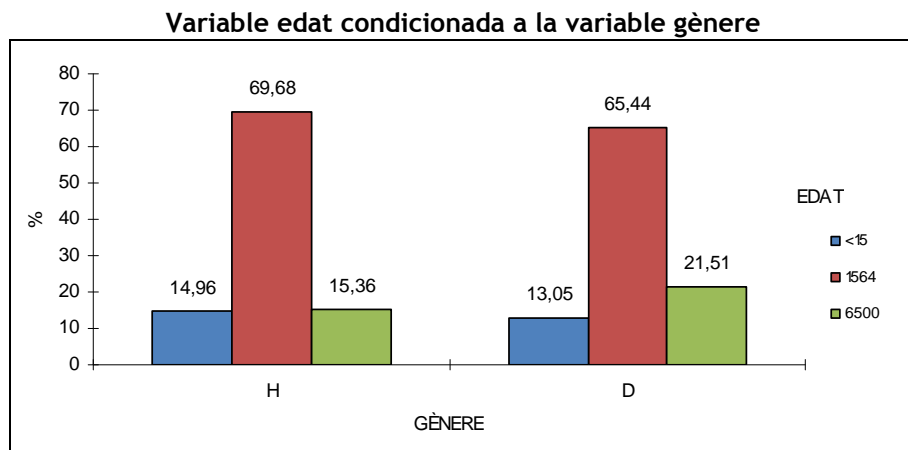
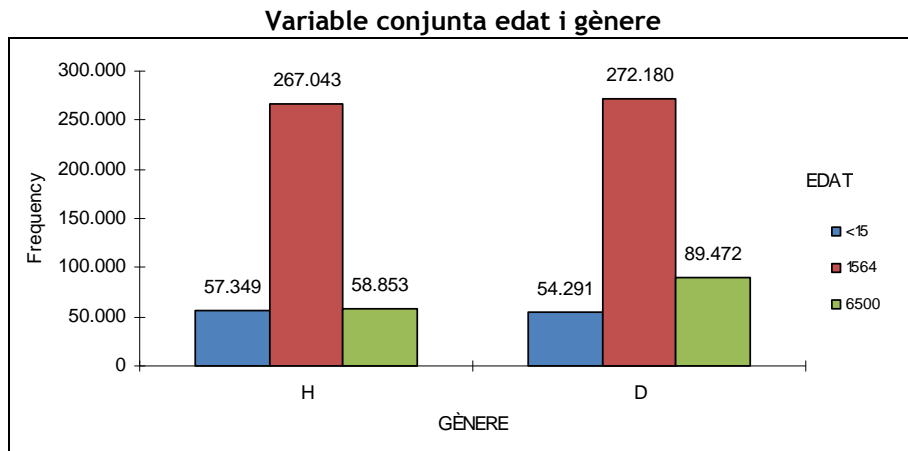
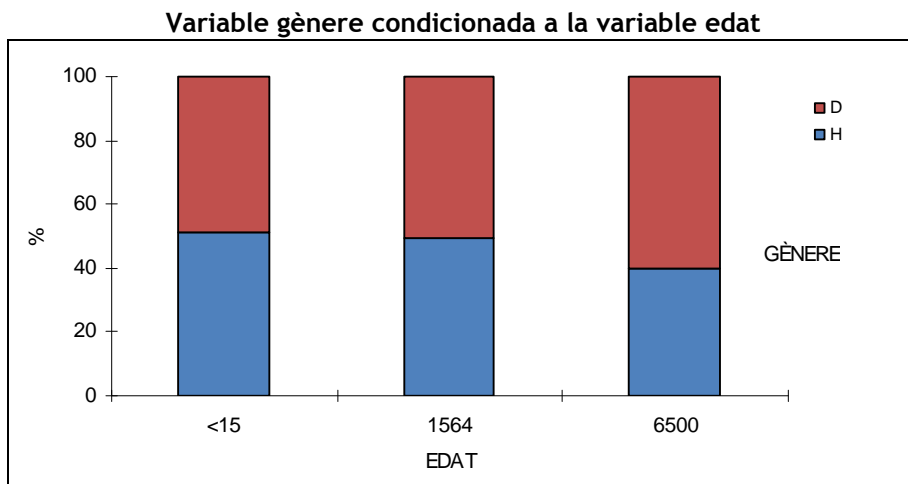


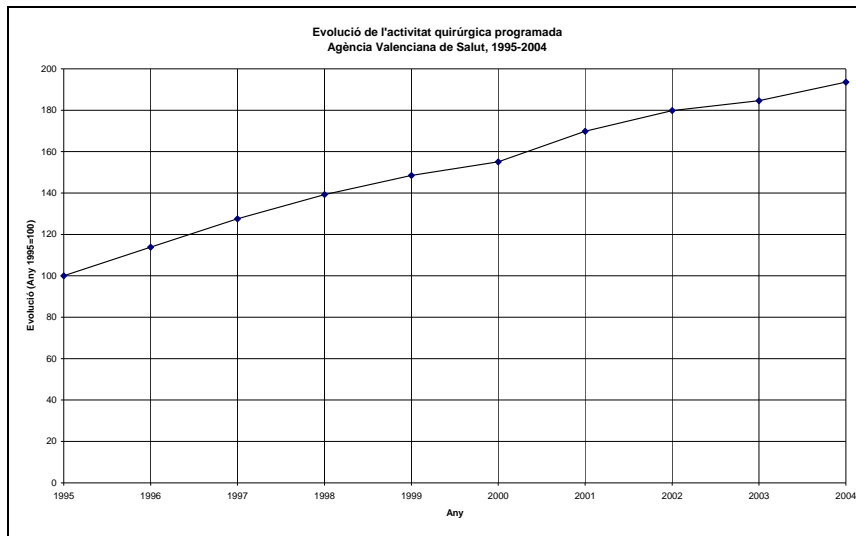
Diagrama de barres estratificat

Presenta les modalitats d'una variable condicionades a una segona variable. Normalment s'expressa en percentatges.



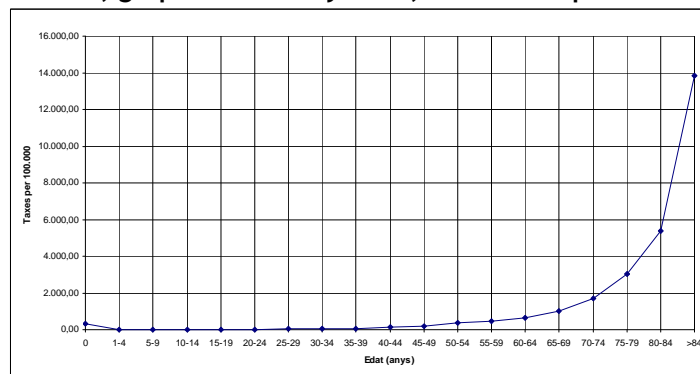
***Sèrie temporal o perfil ortogonal**

Es representen en absccisses les possibles modalitats i en ordenades les corresponents freqüències (de vegades utilitzarem les xifres relatives calculades: índexs, taxes...). Unint els punts obtenim el perfil ortogonal o sèrie temporal [en l'exemple l'activitat quirúrgica programada en el període 1995-2004].

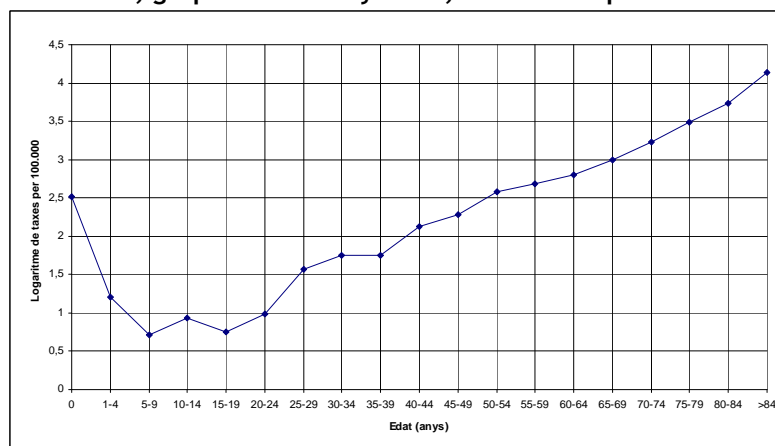


En alguns casos cal recórrer a una escala semilogarítmica perquè els perfils es puguem apreciar [en l'exemple les taxes de mortalitat per totes les causes (València, any 2112, defuncions per 100.000)].

Taxes de mortalitat per totes les causes
València, grups d'edat. Any 2012, defuncions per 100.000



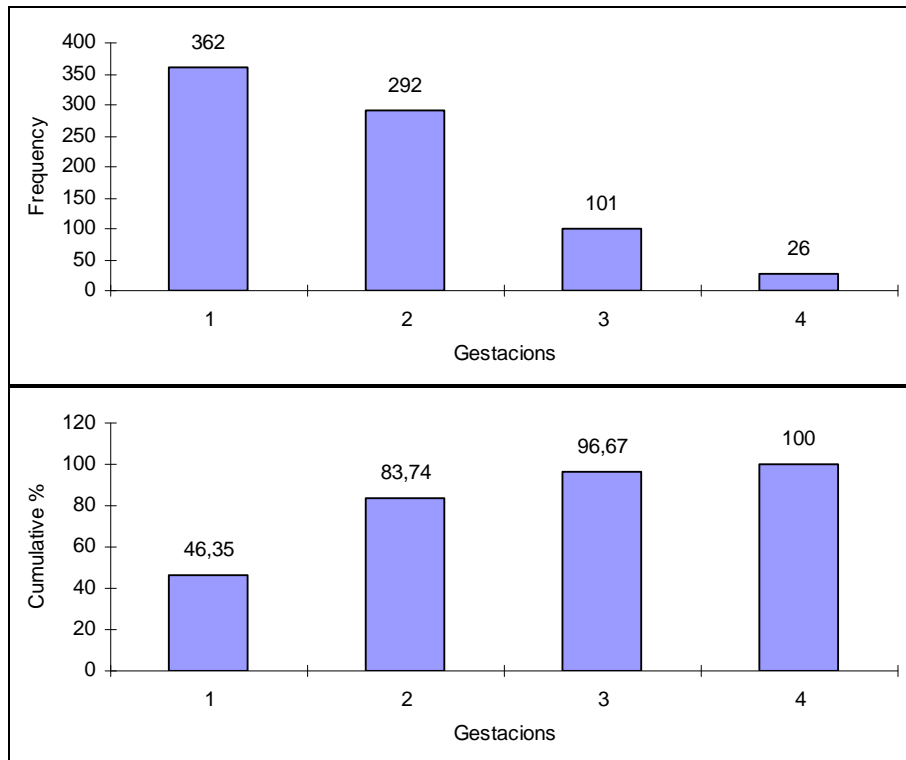
Logaritme de les taxes de mortalitat per totes les causes
València, grups d'edat. Any 2012, defuncions per 100.000



Gràfics per a variables discretes

Diagrama de barres

Sobre uns eixos de coordenades marquem sobre l'eix d'abscisses les possibles modalitats i sobre l'eix d'ordenades la freqüència (proporció/percentatge). Sobre cada modalitat dibuixem rectangles de base constant i altura igual a la freqüència (proporció/percentatge) corresponent [en l'exemple es representen el 'nombre de gestacions' i el percentatge acumulat de gestacions]



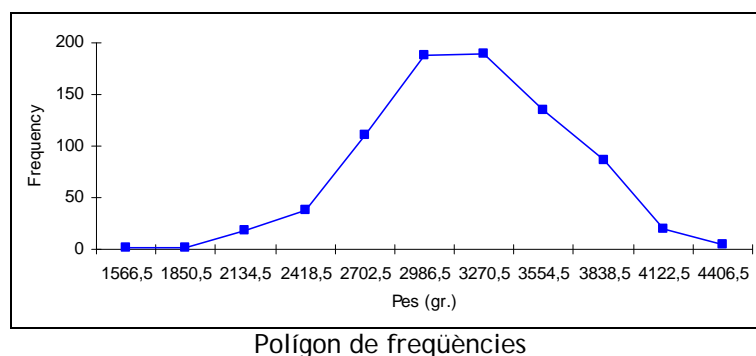
Gràfics per a variables contínues

Histograma

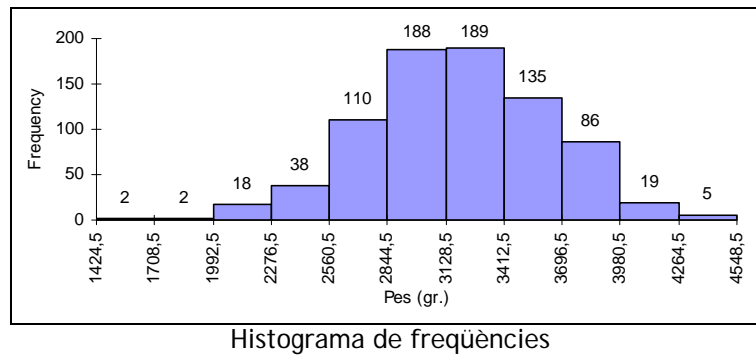
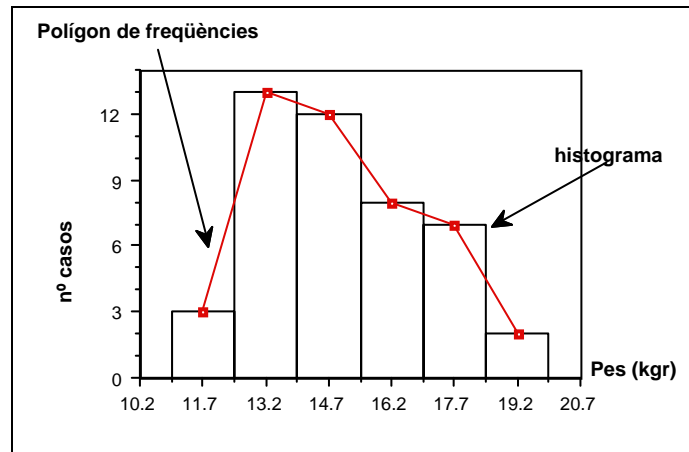
Associem a cada classe un rectangle amb una base igual a la longitud de la classe (treballarem SEMPRE amb límits exactes) i una altura igual a la freqüència (proporció/percentatge) de la classe. Aquesta definició és vàlida només si totes les classes tenen la mateixa longitud.

*Polígon de freqüències

Si considerem les parelles formades per la marca de classe i la seua freqüència corresponent (proporció/percentatge) i representem en uns eixos de coordenades aquests punts unint dos consecutius per mitjà d'una línia recta obtenim el polígon de freqüències.



Polígon de freqüències

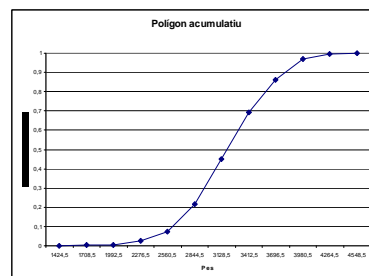
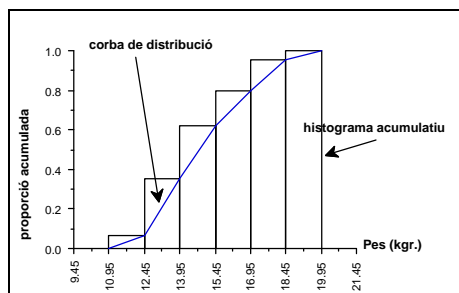


Polígon acumulatiu o corba de distribució

Si considerem la proporció (percentatge) d'una classe uniformement repartida, podem definir la proporció acumulada (percentatge acumulat) en un punt de l'interval com:

$$H(x) = \begin{cases} 0 & x < l_1 \\ H_{i-1} + \frac{x - l_i}{L_i - l_i} \cdot h_i & l_i \leq x \leq L_i \\ 1 & x > L_k \end{cases}$$

Aquesta funció és monòtona creixent i la seua representació gràfica s'anomena polígon acumulatiu o corba de distribució.



5. Paràmetres i estadístics

Són valors representatius d'un conjunt de dades. Si estudiem tota la població, anomenarem aquests valors paràmetres, i si estudiem una mostra aleatòria, estadístics. Si estudiem una mostra d'una població, els estadístics s'usaran per realitzar inferències sobre els paràmetres de la població.

Propietats desitjables per a una mesura

- Ha d'estar definida de forma objectiva.
- Ha d'usar totes les observacions.
- Ha de tenir un significat concret.
- Ha de ser senzilla de calcular.
- Ha de ser poc sensible a les fluctuacions mostrals.

Mesures de tendència central i de posició

Són valors al voltant dels quals s'agrupen les dades observades. Els principals són la moda, la mitjana i els percentils.

Moda (Mo)

La moda d'un conjunt d'observacions és la modalitat que presenta la freqüència (proporció o percentatge) més gran. Si la variable és discreta la moda serà una modalitat. Si és contínua, serà una classe o interval (interval modal) i establirem que la moda és el punt mitjà (marca de classe) de l'interval.

Propietats

- Existeix sempre i pot no ser única.
- Té la mateixa unitat de mesura que la variable estudiada.

Mitjana aritmètica (\bar{x})

La mitjana d'un conjunt d'observacions $\langle x_1, x_2, \dots, x_N \rangle$ és igual a la suma de totes elles, dividida pel seu nombre.

$$\bar{x} = \frac{\sum x_i}{N}$$

Si observen els valors 85;96;110;97;106;99;74;103;109;85; 75;82 la mitjana serà igual a 93,417

$$\bar{x} = \frac{85 + 96 + 110 + 97 + 106 + 99 + 74 + 103 + 109 + 85 + 75 + 82}{12} = 93,417$$

Si tenim les dades agrupades $\langle x_1, x_2, \dots, x_k \rangle$ amb freqüències $\langle f_1, f_2, \dots, f_k \rangle$ on x_i és la modalitat (variable discreta) o la marca de classe (variable contínua) i f_i la seua corresponent freqüència, la mitjana serà igual a:

$$\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \sum x_i \cdot h_i = \frac{\sum x_i \cdot \%_i}{100}$$

Límits exactes		Marca	Casos
10,5	15,5	13	10
15,5	20,5	18	30
20,5	25,5	23	15

$$\bar{x} = \frac{13 \cdot 10 + 18 \cdot 30 + 23 \cdot 15}{10 + 30 + 15} = 18,455$$

Propietats

La unitat de mesura és la de la variable que estem estudiant.

Els valors extrems influeixen en la mitjana i en alguns casos poden alterar-la molt. En aquests casos no resulta convenient com a mesura de tendència central.

Si la variable és discreta, la mitjana no sempre coincideix amb algun dels possibles valors de la variable.

Si la variable és contínua, la mitjana calculada amb les dades sense agrupar i la calculada amb les dades agrupades no sempre són iguals.

$$\frac{\sum x_i}{N} \cong \frac{\sum x_i \cdot f_i}{\sum f_i}$$

El seu valor està entre el menor i el major dels valors observats.

$$x_{\min} \leq \bar{x} \leq x_{\max}$$

Si observem els valors 85;96;110;97;106;99;74;103;109;85; 75;82 la mitjana és igual a 93,417 i els valors mínim i màxim són 74 i 110 respectivament

$$74 \leq \bar{x} \leq 110$$

La suma de les desviacions de cada observació x_i respecte de la mitjana és igual a zero. Aquesta propietat ens permet interpretar la mitjana com una mena de 'centre de gravetat' dels valors observats.

$$\sum (x_i - \bar{x}) = 0$$

La funció $g(k) = \sum (x_i - k)^2$ és mínima per a $k = \bar{x}$

$$\sum (x_i - \bar{x})^2 \leq \sum (x_i - k)^2$$

Si totes les observacions x_i les transformem mitjançant una relació de tipus lineal $y_i = A \cdot x_i + B$, la mitjana de les observacions transformades serà $\bar{y} = A \cdot \bar{x} + B$

Si coneixem la mitjana d'una variable per a diversos grups la mitjana global serà la mitjana ponderada de les mitjanes, sent el pes de cada grup igual a la proporció que representen les observacions del grup en el total.

si $\langle \bar{x}_1, \bar{x}_2, \dots, \bar{x}_g \rangle$ són les mitjanes i $\langle N_1, N_2, \dots, N_g \rangle$ les grandàries dels grups

$$\bar{x} = \frac{\sum \bar{x}_i \cdot N_i}{\sum N_i} = \sum \bar{x}_i \cdot \frac{N_i}{\sum N_i}$$

El quadre adjunt presenta l'estada mitjana d'un grup de malalts ingressats al servei d'hematooncologia d'un hospital. L'estada mitjana del total de malalts ingressats serà de 12,73 dies.

Grup d'edat	00-14	15-44	45-64	≥65
Estada mitjana (dies)	13,94	11,37	11,36	14,45
núm. de casos	16	211	408	481

$$\bar{x} = \frac{13,94 \cdot 16 + 11,37 \cdot 211 + 11,36 \cdot 408 + 14,45 \cdot 481}{16 + 211 + 408 + 481} = 12,73$$

Percentils (P_r)

El seu nom deriva de "per cent". El percentil d'ordre r (P_r) indica que almenys el $r\%$ de les observacions presenten valors menors o iguals a ell i el $(100 - r)\%$ són majors o iguals a ell.

$$\begin{aligned} \%^a(P_r) \geq r\% & \quad \%(X \geq P_r) \geq (100 - r)\% & r=1;2;3;\dots;99 \\ H(P_r) \geq \frac{r}{100} & \quad H(X \geq P_r) \geq \frac{(100 - r)}{100} & r=1;2;3;\dots;99 \end{aligned}$$

Pel seu interès destaquem els següents percentils:

Mediana (Md) o percentil 50 (P_{50})

La mediana és el valor que divideix el conjunt en dues parts iguals de manera que la proporció de valors iguals o majors que la mitjana i la proporció de valors menors o iguals a ella siga en ambdós casos major o igual a 0,50.

$$H(Md) \geq 0,50 \quad H(X \geq Md) \geq 0,50$$

Quartils (Q_i)

Els quartils són tres i corresponen als percentils d'ordre 25, 50 i 75.

$$Q_1 = P_{25} \quad Q_2 = P_{50} \quad Q_3 = P_{75}$$

Càlcul dels percentils amb dades sense agrupar (*Hi ha diferents criteris per a calcular percentils. De fet, diferents programes de computació estadística presenten resultats diferents pel fet que usen criteris semblants, però no iguals. No ha de causar sorpresa, llavors, trobar aquestes diferències originades per la falta d'un procediment universalment acceptat.*)

Ordenarem les dades de menor a major valor

$$x_1 \leq x_2 \leq \dots \leq x_{N-1} \leq x_N$$

Determinarem el valor de **k** i **f**

$$k = Ent\left(\frac{r \cdot N}{100}\right) \quad Ent = \text{part entera}$$

$$f = Frac\left(\frac{r \cdot N}{100}\right) \quad Frac = \text{part fraccionària}$$

Localitzarem les dades que ocupen els llocs k i $k+1 \rightarrow x_k$ i x_{k+1} i el valor del percentil serà

$$\begin{aligned} P_r &= \frac{x_k + x_{k+1}}{2} & \text{si } f = 0 \\ P_r &= x_{k+1} & \text{si } f \neq 0 \end{aligned}$$

Amb les dades de la taula adjunta determinarem els percentils d'ordre 80 i la mediana i el primer quartil

Obs nº	X	Obs nº	X	Obs nº	X	Obs nº	X	Obs nº	X	Obs nº	X	Obs nº	x
1	28	10	34	19	38	28	47	37	53	46	57	55	61
2	29	11	34	20	40	29	47	38	53	47	58	56	62
3	30	12	35	21	41	30	47	39	54	48	59	57	64
4	32	13	35	22	42	31	47	40	54	49	59	58	64
5	32	14	35	23	43	32	48	41	54	50	59	59	66
6	32	15	36	24	44	33	51	42	54	51	60	60	66
7	33	16	37	25	45	34	53	43	55	52	60	61	66
8	33	17	37	26	46	35	53	44	56	53	61	62	69
9	33	18	37	27	46	36	53	45	57	54	61		

$$P_{80} \rightarrow r = 80$$

$$\frac{r \cdot N}{100} = \frac{80 \cdot 62}{100} = 49,6$$

$$f \neq 0 \rightarrow P_{80} = x_{49+1} = 59$$

$$\text{El } P_{80} \text{ compleix que } H(59) = \frac{50}{62} = 0,806 \geq 0,8 \text{ i } H(X \geq 59) = \frac{15}{62} = 0,242 \geq (1 - 0,8)$$

Mediana

$$P_{50} \rightarrow r = 50$$

$$\frac{r \cdot N}{100} = \frac{50 \cdot 62}{100} = 31$$

$$f = 0 \rightarrow P_{50} = \frac{x_{31} + x_{31+1}}{2} = \frac{47 + 48}{2} = 47,5$$

$$\text{El } P_{50} \text{ compleix que } H(47,5) = \frac{31}{62} = 0,50 \geq 0,50 \text{ i } H(X \geq 47,5) = \frac{31}{62} = 0,50 \geq (1 - 0,50)$$

Primer quartil

$$P_{25} \rightarrow r = 25$$

$$\frac{r \cdot N}{100} = \frac{25 \cdot 62}{100} = 15,5$$

$$f \neq 0 \rightarrow P_{25} = x_{15+1} = 37$$

$$\text{El } P_{25} \text{ compleix que } H(37) = \frac{18}{62} = 0,29 \geq 0,25 \text{ i } H(X \geq 37) = \frac{47}{62} = 0,758 \geq (1 - 0,25)$$

Càlcul dels percentils amb dades agrupades (variable discreta)

Si tenim una variable discreta i la corresponent taula de distribucions de freqüències el percentil d'ordre r (P_r) serà igual a

$$P_r = \frac{x_i + x_{i+1}}{2}$$

si existeix un valor una proporció acumulada $H_i = \frac{r}{100}$

Si no existeix un valor una proporció acumulada $H_i = \frac{r}{100}$ llavors $P_r = x_i$ on x_i és el valor més petit amb

proporció acumulada major o igual a $\frac{r}{100}$

núm. visites	0	1	2	3	4
prop. acum.	0,15	0,35	0,65	0,90	1

$$Md = 2$$

$$P_{35} = \frac{1+2}{2} = 1,5$$

$$P_{90} = \frac{3+4}{2} = 3,5$$

Càlcul dels percentils amb dades agrupades (variable contínua)

El percentil d'ordre r (P_r) quan tenim les dades d'una variable contínua agrupades en intervals és igual al valor de la variable que té una proporció acumulada igual a $r/100$. En la pràctica, per calcular el percentil aplicarem la fórmula:

$$P_r = l_i + \frac{\frac{r}{100} - H_{i-1}}{h_i} * (L_i - l_i)$$

L'interval $l_i - L_i$ s'anomena interval crític i és el primer interval amb proporció acumulada major o igual a $\frac{r}{100}$.

Les fórmules per calcular els quartils són:

Quartil 1

$$Q_1 = P_{25} = l_i + \frac{0,25 - H_{i-1}}{h_i} * (L_i - l_i)$$

$l_i - L_i$ és el primer interval amb proporció acumulada major o igual a 0,25.

Quartil 2 o mediana

$$Md = Q_2 = P_{50} = l_i + \frac{0,50 - H_{i-1}}{h_i} * (L_i - l_i)$$

$l_i - L_i$ és el primer interval amb proporció acumulada major o igual a 0,50.

Quartil 3

$$Q_3 = P_{75} = l_i + \frac{0,75 - H_{i-1}}{h_i} * (L_i - l_i)$$

$l_i - L_i$ és el primer interval amb proporció acumulada major o igual a 0,75.

Hem agrupat les dades de l'exemple anterior en cinc intervals de mida nou. La taula de distribució de freqüències resultants és:

Límits exactes		h_i	H_i
27,5	36,5	0,2419	0,2419
36,5	45,5	0,1613	0,4032
45,5	54,5	0,2742	0,6774
54,5	63,5	0,2258	0,9032
63,5	72,5	0,0968	1

Amb les dades de la taula el percentil 80, la mediana i el primer quartil seran:

P_{80}

El primer interval amb proporció acumulada major o igual a 0,80 és l'interval 54,5;63,5

$$P_{80} = 54,5 + \frac{0,80 - 0,6774}{0,2258} * (63,5 - 54,5) = 59,387$$

Md

El primer interval amb proporció acumulada major o igual a 0,50 és l'interval 45,5;54,5

$$Md = 45,5 + \frac{0,50 - 0,4032}{0,2742} * (54,5 - 45,5) = 48,677$$

Q_1

El primer interval amb proporció acumulada major o igual a 0,25 és l'interval 36,5;45,5

$$Q_1 = 36,5 + \frac{0,25 - 0,2419}{0,1613} * (45,5 - 36,5) = 36,952$$

Propietats

- La unitat de mesura del percentil és la de la variable.
- Si la variable és continua el percentil d'ordre r divideix l'histograma en dues parts de manera que la superfície d'una d'elles és el r% del total i l'altra el (100-r)%.

Mesures de dispersió

Són aquelles que quantifiquen la dispersió de les dades observades.

Rang o recorregut (R)

És la diferència entre el menor i el major dels valors observats

$$R = X_{\max} - X_{\min}$$

Propietats

- És fàcil de calcular i interpretar i només utilitza dues de les observacions.

Interval interquartílic (IIQ)

Quan tenim valors extrems el recorregut és molt gran. En aquestes ocasions calcularem l'interval interquartílic, que és igual a la diferència entre el tercer i el primer quartil.

$$IIQ = Q_3 - Q_1 = P_{75} - P_{25}$$

Propietats

- En l'interval interquartílic es troben el 50% de les observacions.
- És una mesura de la concentració de les observacions. Quan menor siga la seua longitud major serà la concentració de les observacions al voltant de la mediana.

Variància (S^2) i desviació típica (S)

Són mesures de dispersió que indiquen que disperses o agrupades que estan les observacions respecte de la mitjana. El seu valor és igual a la mitjana de les desviacions al quadrat de cada una de les dades respecte de la mitjana aritmètica del conjunt d'observacions.

$$\langle x_1, x_2, \dots, x_N \rangle \quad \bar{x} = \frac{\sum x_i}{N}$$

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{N} = \frac{\sum x_i^2}{N} - \bar{x}^2$$

Si observem els valors 85;96;110;97;106;99;74;103;109;85; 75;82 la mitjana és 93,417 i la variància

$$S^2 = \frac{85^2 + 96^2 + 110^2 + 97^2 + 106^2 + 99^2 + 74^2 + 103^2 + 109^2 + 85^2 + 75^2 + 82}{12} - 93,417^2 = 150,576$$

Si tenim les dades agrupades en k modalitats $\langle x_1, x_2, \dots, x_k \rangle$ amb freqüències $\langle f_1, f_2, \dots, f_k \rangle$ on x_i és la modalitat (variable discreta) o la marca de classe (variable contínua) i f_i la seua corresponent freqüència, la variància és igual a

$$S^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i} = \frac{\sum x_i^2 \cdot f_i}{\sum f_i} - \bar{x}^2$$

Amb les dades de la taula adjunta la mitjana és igual a 18,455 i la variància

Límits exactes		Marca	Casos
10,5	15,5	13	10
15,5	20,5	18	30
20,5	25,5	23	15

$$S^2 = \frac{13^2 \cdot 10 + 18^2 \cdot 30 + 23^2 \cdot 15}{10 + 30 + 15} - 18,455^2 = 11,157 \rightarrow S = \sqrt{S^2} = 3,34$$

Propietats

- La variància i la desviació típica són sempre números no negatius.
- La variància i la desviació típica són iguals a zero quan tots els valors observats són iguals.
- La unitat de mesura de la desviació és la de la variable, en el cas de la variància en el quadrat de la dita unitat.
- La variància i la desviació típica poden no coincidir si es calculen amb les dades agrupades i sense agrupar.
- Si totes les observacions x_i les transformem mitjançant una relació de tipus lineal $y_i = A \cdot x_i + B$, la variància de les observacions transformades serà $S_y^2 = A^2 \cdot S_x^2$
- Si coneixem la mitjana i les variàncies d'una variable per a diversos grups la variància global pot ser calculada mitjançant la fórmula

si $\langle \bar{x}_1, \bar{x}_2, \dots, \bar{x}_g \rangle$ són les mitjanes, $\langle S_1^2, S_2^2, \dots, S_g^2 \rangle$ les variàncies i $\langle N_1, N_2, \dots, N_g \rangle$ les mides dels grups

$$S^2 = \frac{\sum (S_i^2 + \bar{x}_i^2) \cdot N_i}{\sum N_i} - \bar{x}^2$$

Amb les dades del quadre adjunt obtenim una mitjana igual a 41,636 i una variància igual a

	G1	G2
N	254	300
Mitjana	40,38	42,7
Desviació	21,827	22,912

$$S^2 = \frac{(21,827^2 + 40,48^2) * 254 + (22,912^2 + 42,7^2) * 300}{254 + 300} - 41,636^2 = 504,034$$

Coefficient de variació (CV)

Si mesurem dues variables en un grup o una mateixa variable en diferents grups i calculem la mitjana, mediana, desviació, etc, sabem que aquests valors estan afectats per la unitat de mesura de la variable. Definirem el **coeficient de variació (CV)** d'una variable com el quocient entre la seua desviació típica i la seua mitjana:

$$CV = \frac{S}{\bar{x}}$$

A causa del seu caràcter adimensional, el coeficient de variació ens permet comparar variables mesurades en unitats distintes o diferents mostres d'una mateixa variable. En general direm que una variable X és més homogènia, menys dispersa o menys variable que una variable Y si el coeficient de variació de X és menor que el coeficient de variació de Y

Variable	Mitjana	Desviació	Coef Var
Núm. parts ant.	0,57	0,86	1,51
Núm. avortaments ant.	0,20	0,46	2,30

S'observa que la desviació és més gran a la variable nombre d'avortaments anteriors però en calcular el coeficient de variació concloem que la variable més homogènia és el nombre de parts anterior.

ESTADÍSTICA DESCRIPTIVA AMB FULL DE CÀLCUL

Matrius

Una **matriu** és un conjunt de caselles indexades. La indexació mostrarà la primera i la darrera de les caselles que integren la matriu. Tota matriu té x files i y columnes i conté $x \cdot y$ dades.

	A	B	C	D	E	F	G	H	I
1									
2									
3		2850	2850	2850	2850				
4		2850	2850	2850	2850				
5		2855	2855	2855	2855				
6		2855	2855	2855	2855				
7		2860	2860	2860	2860				
8		2860	2860	2860	2860				
9		2860	2860	2860	2860	B3:E9			
10									
11									
12					A	B	A	C	E12:H12
13		7							
14		9							
15		7							
16		8							
17		6							
18		4							
19		3							
20		2	B13:B20						
21									
22									

A la figura adjunta observem els tres tipus bàsics de matrius i la seua notació:

B3:E9

És una matriu d'ordre 7 per 4 (7 files i 4 columnes) i hi podem allotjar un màxim de $7 \cdot 4 = 28$ dades.

E12:H12

És una matriu d'ordre 1 per 4 (1 fila i 4 columnes) i hi podem allotjar un màxim de $1 \cdot 4 = 4$ dades. Aquesta matriu s'anomena matriu fila.

B13:B20

És una matriu d'ordre 8 per 1 (8 files i 1 columna) i hi podem allotjar un màxim de $8 \cdot 1 = 8$ dades. Aquesta matriu s'anomena matriu columna.

Càlcul d'estadístics bàsics amb un full de càlcul

Concepte	Fórmula
Nombre de dades (N)	=CONTARA(MATRIU)
Valor mínim	=MIN(MATRIU)
Valor màxim	=MAX(MATRIU)
Mitjana	=PROMEDIO(MATRIU)
Desviació (N)	=DESVESTP(MATRIU)
Desviació (N-1)	=DESVEST(MATRIU)
Percentil d'ordre r (P_r)	=PERCENTIL(MATRIU;r/100)

XLStatistics - Excel Workbooks for Statistical Analysis (XLS5/XLS6)

XLStatistics - Excel Workbooks for Statistical Analysis (XLS6)

XLStatistics és un conjunt de llibres de treball (fitxers) de Microsoft Excel que poden ser utilitzats per a la majoria de les anàlisis estadístiques clàssiques. Aquest material ha estat elaborat pel professor Rodney Carr de la Deakin University de Austràlia ([XLStatistics](#))

1Cat.xls (Data & Description)

Analitza les dades d'una variable qualitativa

Dades

- A2 Nom de la variable quantitativa
- A3:A* Dades
- E5:*5 Modalitats. Llista de possibles modalitats (*cal escriure-les*)
- C4:*8 Taula de distribució de freqüències

Data		Descriptives		
Answer		Category Labels, Counts (Frequencies) and other numerical summaries		
		Answer		
		All	Yes	No
Count	13	8	5	
Proportion	1	0,6154	0,3846	
Percentage	100	61,538	38,462	

Diagrama de barres

Podem dibuixar el diagrama de barres de freqüències (*Counts*), proporcions (*Proportion*) o percentatges (*Percentage*). En tots tres podem mostrar els corresponents valors (*Show values*) amb els decimals que ens interessin (*Dec places*). Cal recordar que s'aconsellen quatre decimals per a les proporcions i dos per als percentatges.

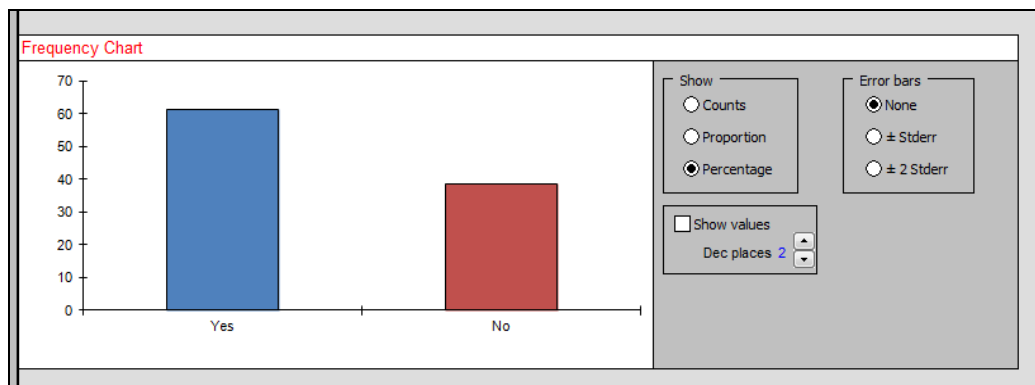


Diagrama de barres (percentatges)

Diagrama de sectors

El diagrama de sectors es presenta amb percentatges sense decimals. Podem canviar el format de les dades fent clic amb el botó dret sobre les xifres del gràfic. Al menú que s'obrirà farem clic sobre *Formato de etiqueta de datos* i se'ns mostrarà la finestra que ens permetrà el canvi.

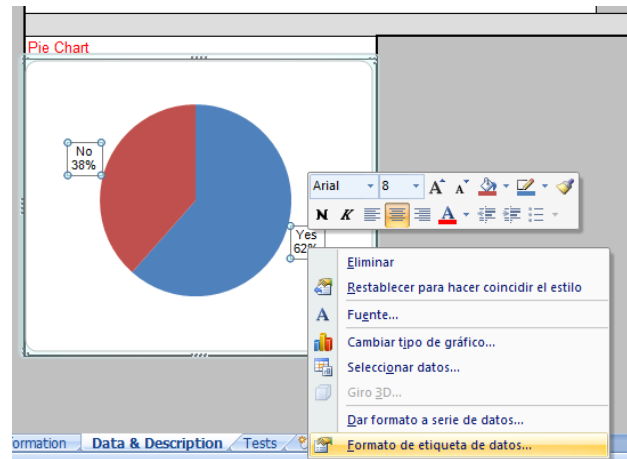
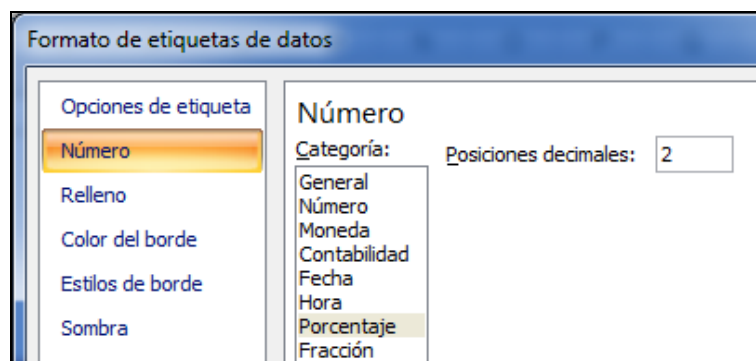
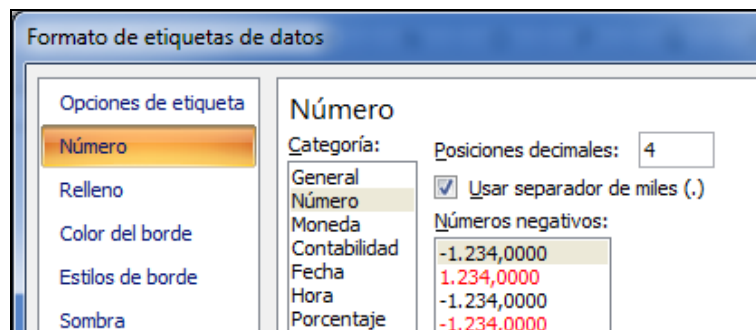


Diagrama de sectores (percentatges)



Percentatge amb dues xifres decimals



Proporció amb quatre xifres decimals

1Num.xls (Data & Description)

Analitza les dades d'una variable quantitativa.

Dades

A2 Nom de la variable quantitativa
 A3:A* Dades

H6		fx =D6/RAIZ(D4)									
	A	B	C	D	E	F	G	H	I	J	K
1	Data		Descriptives								
2		X									
3		88	Numerical Summaries for X								
4		86	Number	92	Kurtosis	-0,46481	Min	75			
5		90	Mean	96,91304	10 % Tr mean	96,49412	Q ₁	88			
6		120	St Dev	10,96613	StdErr Mean	1,143298	Median	95			
7		112	Coeff of Var	0,113154			Q ₃	104			
8		92	Skew	0,451314			Max	125			
9		88									

Valors calculats

		Valor mínim	K4
Nombre de dades	D4	Primer quartil	K5
Mitjana	D5	Mediana	K6
Desviació (n-1)	D6	Tercer quartil	K7
Coefficient de variació	D7	Valor màxim	K8

Taula de distribució de freqüències

La construirem amb les dades que ens proporcionen els gràfics quan hi mostrem els valors.

Diagrama de barres

Per dibuixar el diagrama de barres caldrà introduir els següents valors i marcar *Bar*, *Classes* i *Midpoints*.

classes Nombre de modalitats (O12)= Valor màxim -Valor mínim+1
 From *Valor d'ajuda* (O13)= Valor mínim-0,5
 Up to *Valor d'ajuda* (O14)= Valor màxim +0,5
 Dec places Nombre de xifres decimals (O15)=0

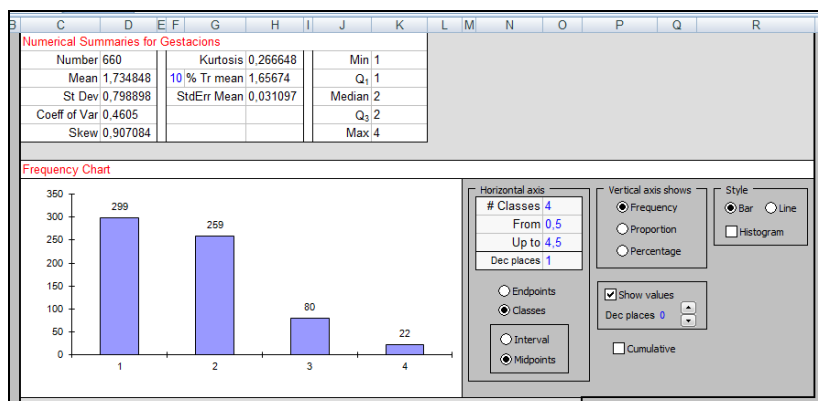


Diagrama de barres (freqüències)

Podem dibuixar el diagrama de barres de freqüències (*Frequency*), proporcions (*Proportion*) o percentatges (*Percentage*). En tots tres podem mostrar els corresponents valors (*Show values*) amb els

decimals que ens interessin (*Dec places*). També podem dibuixar diagrames de barres acumulats (*Cumulative*).

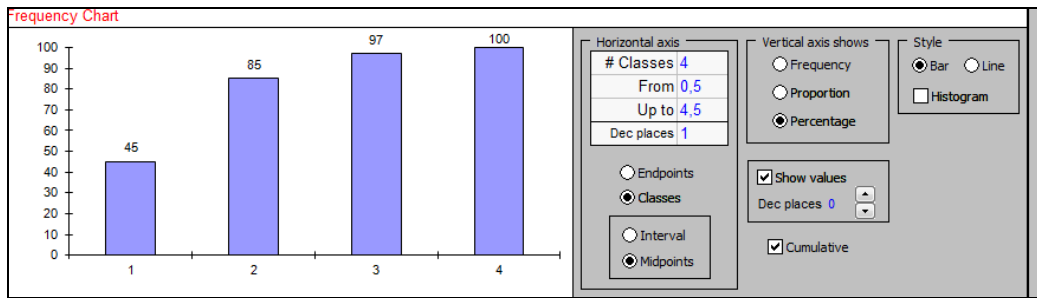
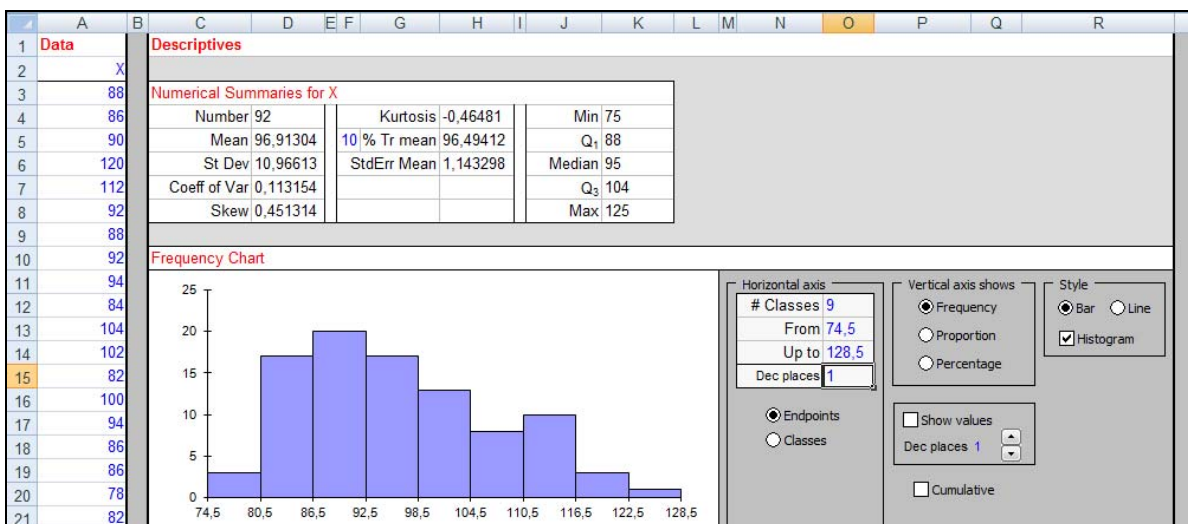


Diagrama de barres acumulatiu (percentatges)

Histograma

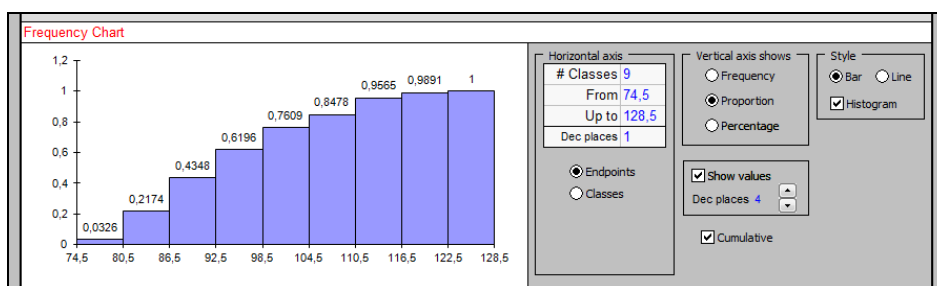
Per dibuixar l'histograma caldrà introduir els següents valors i marcar *Endpoints*, *Bar*, i *Histogram*.

- # classes Nombre d'interval·s (O12)
- From Límit exacte inferior del primer interval (O13)
- Up to Límit exacte superior de l'últim primer interval (O14)
- Dec places Xifres decimals dels límits exactes (O15)



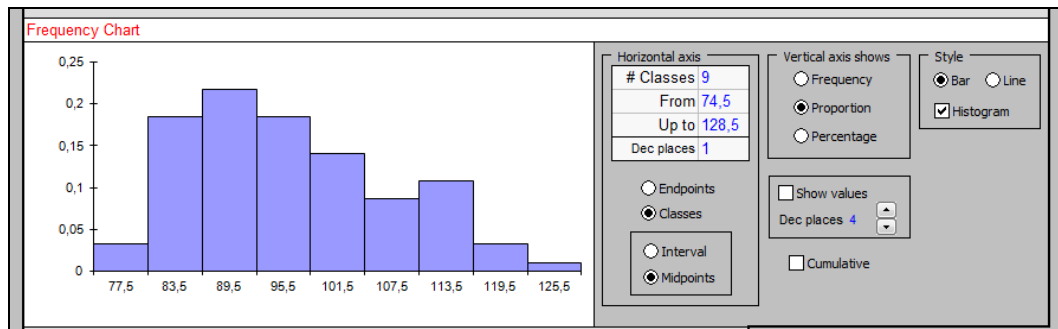
Histograma (freqüència i límits exactes)

Podem dibuixar l'histograma de freqüències (*Frequency*), proporcions (*Proportion*) o percentatges (*Percentage*). En tots tres podem mostrar els corresponents valors (*Show values*) amb els decimals que ens interessin (*dec places*). També podem dibuixar histogrames acumulats (*Cumulative*).



Histograma acumulatiu (proporcions i límits exactes)

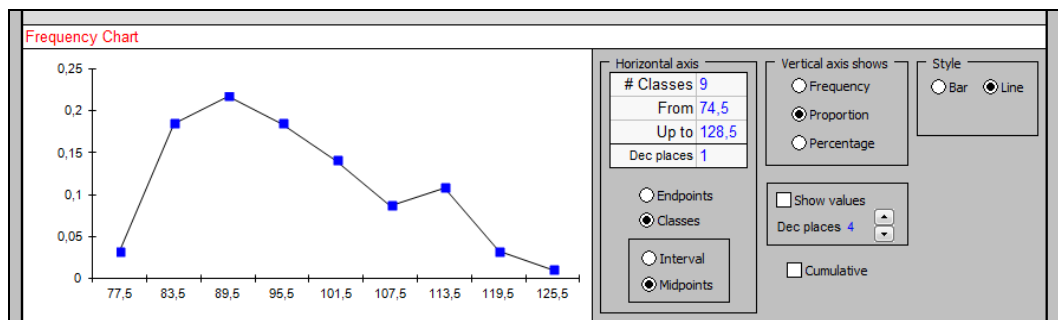
Si en un histograma volem mostrar a l'eix d'abscisses les marques de classe caldrà marcar *Bar*, *Classes* i *Midpoints*.



Histograma (proporcions i marques de classe)

Polígon de freqüències

Per dibuixar el polígon de freqüències cal marcar *Line*, *Classes* i *Midpoints*. Podem dibuixar el polígon de freqüències (*Frequency*), proporcions (*Proportion*) o percentatges (*Percentage*). En tots tres podem mostrar els corresponents valors (*Show values*) amb els decimals que ens interessen (*dec places*).



Polígon de freqüències (proporcions)

Diagrama de caixes (Boxplot)

El diagrama de caixes ens presenta un rectangle on es representen els valors dels quartils. L'amplària de la caixa és igual a l'interval interquartílic [$IIQ=(Q_3-Q_1)$]. El valor inferior (A) i el valor superior (B) que observem al diagrama (bigot inferior i superior) corresponen als valors mínim i màxim a l'interval [$Q_1-1,5*IIQ; Q_3+1,5*IIQ$]. Si hi ha valors fora d'aquest rang es marcaran amb una creu.

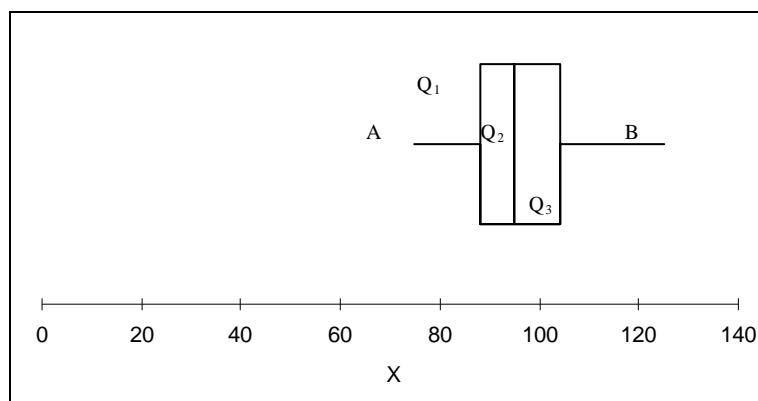


Diagrama de caixes

1Num.xls (Extra Tools / Grouped Data / Data & Description)

Analitza les dades d'una variable quantitativa contínua a partir de la taula de distribució de freqüències absolutes. Els intervals han de tenir tots la mateixa mida i els seus extrems han de presentar una continuïtat.

Dades

- C1 Nom de la variable quantitativa
- A3:A* Límits inferiors dels intervals
- B3:B* Límits superiors dels intervals
- C3:C* Freqüència

	A	B	C	D	E	F	G	H	I
1	Data	Variable X			Description				
2	Lower	Upper	Frequency						
3	12,75	15,75	10	Numerical Summaries					
4	15,75	18,75	23	Number	112		Min	12,75	
5	18,75	21,75	42	Mean	20,33036		Q ₁	18,09783	
6	21,75	24,75	28	St Dev	3,207925		Median	20,39286	
7	24,75	27,75	9	Coeff of Var	0,15779		Q ₃	22,71429	
8							Max	27,75	

Variable contínua. Dades agrupades

Valors calculats

Nombre de dades	F4	Valor mínim	I4
Mitjana	F5	Primer quartil	I5
Desviació (n-1)	F6	Mediana	I6
Coefficient de variació	F7	Tercer quartil	I7
		Valor màxim	I8

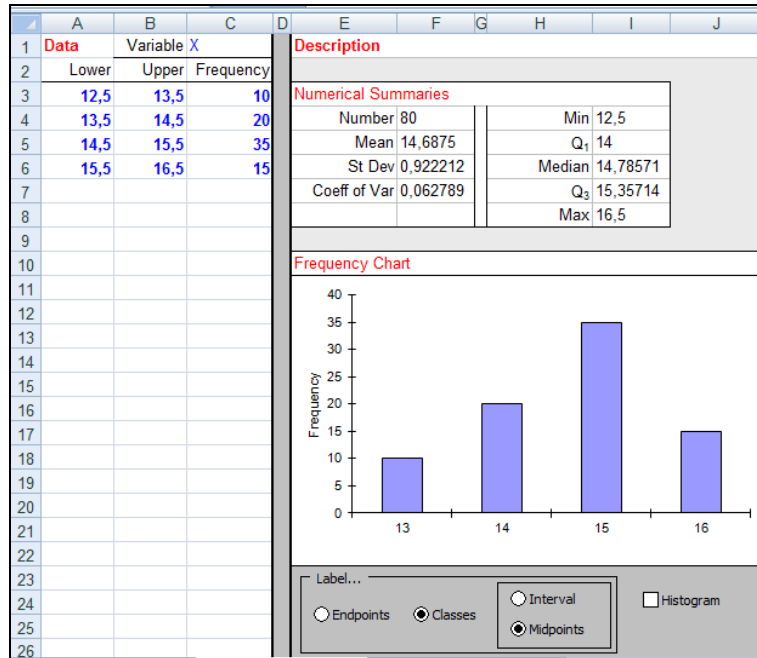
Gràfics

Seguint el mateixos criteris que en el cas d'una variable quantitativa, amb les dades sense agrupar podem dibuixar els diferents gràfics. En aquest cas alguns dels gràfics es mostren al llibre *Summaries*.

NOTA

Si volem descriure les dades d'una variable discreta amb dades agrupades podem "transformar" cada modalitat M en un interval d'extremes $M-0,5; M+0,5$ i treballar com si fóra una variable contínua agrupada. D'aquesta manera "enganyarem" l'aplicació i podrem dibuixar els diagrames de barres que ens interessin [cal marcar *Classes*, *Midpoints* i deixar sense marcar *Histogram*].

Dels valors calculats tan sols són correctes el total d'observacions (F4), la mitjana (F5), la desviació (F6) i el coeficient de variació (F7).



Variable discreta. Dades agrupades

PROBLEMES

- 1) Classifiqueu les següents variables en qualitatives, quantitatives discretes o quantitatives contínues:
 - a. Grup sanguini
 - b. Pressió arterial diastòlica
 - c. Sexe
 - d. Nombre de fills d'una família
 - e. Diagnòstic d'un malalt ingressat a la sala de cardiologia de l'Hospital Clínic
 - f. Accidents mortals diaris en la ciutat de València
 - g. Longitud del fèmur esquerre de les dones de 22 anys
 - h. Temperatura axil·lar d'un grup de pacients

- 2) Es vol estudiar la variable *temperatura* en un grup de lactants. Establiu una llista de modalitats perquè:
 - a. La variable siga qualitativa
 - b. La variable siga quantitativa

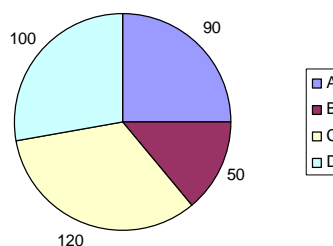
- 3) Es vol estudiar la variable *edat* de les dones ingressades al Servei de Cirurgia de l'Hospital Clínic.
 - a. És únic el llistat de possibles modalitats que pot presentar la variable *edat*?
 - b. Elaboreu una llista de possibles modalitats

- 4) S'ha estudiat la variable *estat civil* en una mostra de 50 persones adultes amb els següents resultats (*C:Casat/da, S:Fadrí/na, V:Vidu/a, D:Divorciat/da*):

C	C	V	C	C	S	C	C	C	S	C	C	C	S	S	C	C
C	C	S	C	S	V	C	S	C	V	D	V	S	C	D	C	S
C	C	C	S	C	D	S	C	S	C	S	S	C	S	S	C	

- a. Construïu la taula de distribució de freqüències
 - b. Visualitzeu la taula de l'apartat a) mitjançant un gràfic
- 5) Realitzat un estudi sobre un grup de 50 xiquets que van requerir intubació i/o traqueotomia en una unitat de cures intensives pediàtriques s'observa que el motiu d'ingrés en la dita unitat fou degut en un 30% dels casos a una insuficiència aguda respiratòria o crònica aguditzada (IAR), en un 40% a un quadre neurològic agut (QNA) i la resta a causes de diverses entitats nosològiques (DCN).
 - a. Construïu la taula de distribució de freqüències
 - b. Dibuixeu el diagrama de barres

 - 6) Un caràcter qualitatiu presenta quatre possibles modalitats: A, B, C i D. Estudiada una mostra de grandària 72 s'ha dibuixat el diagrama de sectors (en cada sector circular es mostra l'angle central corresponent en graus).



- a) Construïu la taula de distribució de freqüències.
- 7) El nombre de brots declarats d'hepatitis en el període 1982-1986 en tot l'Estat es reflecteix en el quadre adjunt. (*Dades del Butlletí Epidemiològic Setmanal, núm. 1791. Min. de Sanitat i Consum*).

Any	1982	1983	1984	1985	1986
Núm. brots	10	5	17	23	27
Núm. afectats	232	217	433	512	850

 - a) Calculeu una mesura per a valorar la intensitat dels diferents brots declarats
 - b) Dibuixeu el perfil ortogonal

8) El nombre de fills de 20 famílies seleccionades a l'atzar és:

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

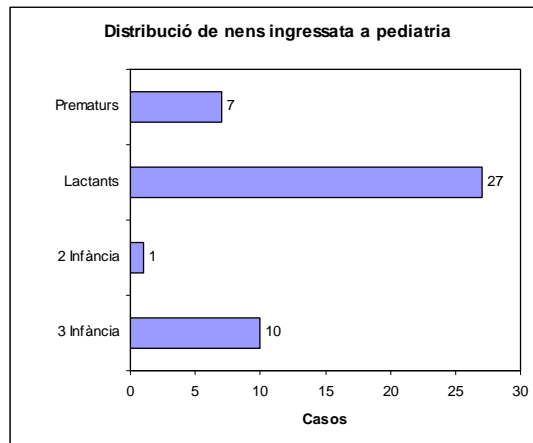
- a) Construiu la taula de distribució de freqüències
- b) Calculeu el percentil 40

9) En un grup de 50 embarassades d'alt risc s'ha estudiat la variable *nombre d'embarassos anteriors*, amb els resultats següents:

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

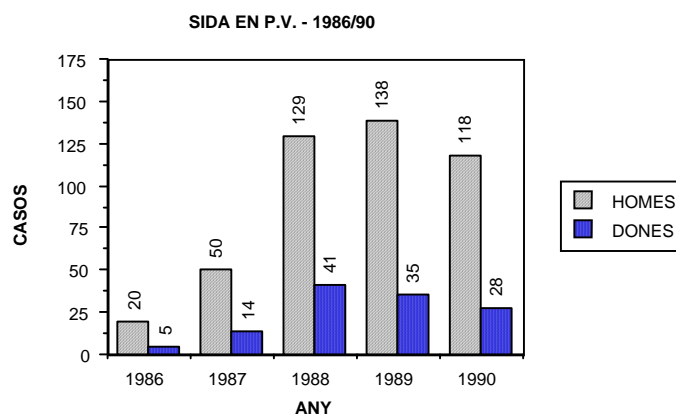
- a) Construiu la taula de distribució de freqüències
- b) Dibuixeu el diagrama de barres
- c) Calculeu la mitjana i la mitjana

10) Hem classificat els xiquets ingressats en el Servei de Pediatria d'un hospital en prematurs, lactants, segona infància i tercera infància i hem dibuixat el corresponent diagrama de barres.



- a) Quants xiquets hi havia ingressats en el Servei de Pediatria?
- b) Construiu la taula de distribució dels percentatges

11) El gràfic adjunt ens mostra la distribució per sexes dels casos de SIDA declarada al País Valencià en el període que comprèn des de l'any 1986 a l'any 1990. Determineu:



- a) El nombre total d'afectades de SIDA durant aquest període
- b) La proporció que representen els homes sobre el total d'afectats
- c) El percentatge (sobre el total) de casos que es van detectar l'any 1988

12) Les freqüències absolutes acumulades obtingudes en estudiar una variable estadística discreta són:

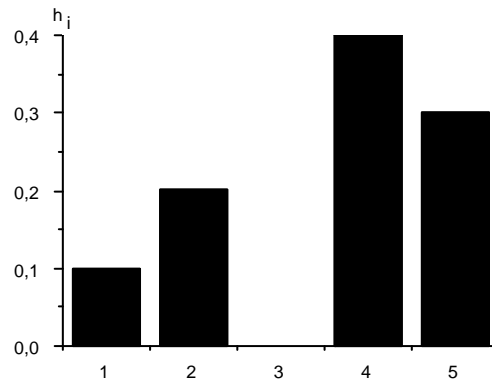
X	0	1	2	3	4	5	6	7
F	1	3	6	10	14	18	19	20

- Determineu les freqüències
- Calculeu la mitjana i la desviació típica
- Calculeu els percentils 25, 50 i 75

13) El nombre de germans de 20 alumnes de primer curs d'infermeria, seleccionats aleatòriament, és el següent: 3;1;2;2;1;5;2;2;0;6;3;2;4;3;4;2;3;1;5;6.

- Calculeu el percentatge acumulat al valor "3 germans"

14) La gràfica adjunta ens mostra el diagrama de barres d'una variable discreta.



- Calculeu la mitjana
- Si el nombre d'observacions és igual a 40, construïu la taula de distribució de freqüències

15) S'ha estudiat en una mostra de 60 històries clíniques de pacients atesos en el Servei d'Endocrinologia de l'Hospital Clínic Universitari de València la variable 'nombre de visites' amb els següents resultats:

1	2	2	4	4	6	8	6	7	3
5	5	6	1	4	2	4	6	8	5
2	5	3	5	4	3	6	7	3	6
5	1	4	5	6	7	6	3	5	7
5	8	7	3	7	2	3	5	2	3
6	5	2	6	2	5	2	6	8	7

- Construïu la taula de distribució de freqüències
- Dibuixeu el diagrama de barres
- Calculeu el nombre mitjà de visites

16) Els pesos (kg) de 45 xiquets d'entre 2 i 3 anys són:

13,7	14,0	12,8	14,6	14,8	11,0	14,5	18,2	14,9
13,7	15,3	13,8	17,0	12,9	15,9	13,5	12,6	14,3
14,3	15,8	11,8	18,0	18,5	15,5	13,8	17,8	17,5
15,8	12,8	11,5	14,7	15,9	19,7	16,7	15,7	15,0
13,3	12,6	14,3	15,5	13,7	17,5	13,9	14,2	17,2

- Agrupeu les dades en 7 classes de longitud 1,3 prenent com a límit exacte inferior de la primera classe 10,85
- Construïu la taula de distribució de freqüències
- Dibuixeu l'histograma de percentatges

17) S'han efectuat 20 determinacions de glucosa (mg/100 ml).

85	96	110	97	106	99	74	103	109	85
75	82	83	92	77	79	89	98	96	88

- Agrupeu les dades en classes de longitud 10, sent el límit exacte inferior de la primera classe 72,5
- Construïu la taula de distribució de freqüències

18) Determineu els límits exactes dels intervals en què s'ha agrupat una variable contínua sabent que els límits aparents són:

12,1-12,8 12,9-13,6 13,7-14,4 14,5-15,2

19) Completeu la següent taula de distribució de freqüències:

Classes	3,5;9,5	9,5;___	___;21,5	___;27,5
f	3	___	8	___
H	___	0,25	___	___
F	___	___	16	20

20) Les dades de la taula corresponen a les quantitats de clor (mEq/l) en mostres d'orina de 39 persones (21 dones i 18 homes) normals.

18,0	22,0	31,0	35,5	57,0	58,0	66,0	67,0
68,0	70,0	74,0	76,4	80,0	81,0	85,1	85,5
90,0	90,0	93,0	95,0	96,0	102,0	112,3	115,0
115,0	122,0	125,0	135,0	139,0	153,0	158,1	166,0
168,0	172,0	173,0	176,0	178,2	179,3	180,0	

a) Determineu sis (6) intervals (límits exactes) per a poder agrupar els valors observats.

21) Després de determinar la longitud del fèmur d'un grup de 200 homes amb edats entre els 12 i els 18 anys es va observar que el valor mínim era de 42,5 cm. i el valor màxim de 51,1cm.

a) Establiu 6 classes (límits exactes) de la mateixa grandària per a agrupar les dades observades

22) Determineu el límit exacte superior d'un interval sabent que la marca de classe és 45 i el límit exacte inferior 40,5.

23) La taula adjunta mostra la distribució de freqüències corresponent a la quantitat de Na (mEq/l) en mostres d'orina d'un grup de persones normals.

Mod.	23,0-56,0	56,1-89,1	89,2-122,2	122,3-155,3	155,4-188,4
%	10	8	38	26	18

a) Calculeu el percentil 40

24) Moltes persones experimenten reaccions al·lèrgiques sistèmiques a les picadures d'insectes. Aquestes reaccions difereixen de pacient a pacient no sols per la gravetat sinó també en el temps d'aparició de la reacció. S'han registrat els temps d'aparició (en minuts) de la reacció en un grup de pacients que van experimentar una reacció al·lèrgica sistèmica a la picadura d'una abella, i s'ha obtingut la següent taula de distribució de freqüències:

Temps	3,75-5,95	5,95-8,15	8,15-10,35	10,35-12,55	12,55-14,75	14,75-16,95
H	0,05	0,10	0,25	0,40	0,15	0,05

a) Calculeu el percentil d'ordre 80 i el percentatge acumulat al valor 13,65

25) S'ha realitzar un estudi de la malaltia de Hodgkin restringit a pacients amb edats inferiors als quaranta anys. Un propòsit de l'estudi va ser comparar la distribució de casos per edat en homes i dones. Les dades agrupades es mostren en el quadre següent

Edat (anys)	Homes	Dones
04-14	5	2
15-24	30	17
24-34	23	8
35-39	10	2

a) Calculeu l'edat mitjana del grup de persones estudiat

b) Calculeu la proporció d'homes amb una edat igual o inferior a 24,5 anys

26) S'han determinat els valors en grams, d'albumina total circulant, en un grup d'homes entre 20 i 30 anys.

Classe	Hi
99,5;109,5	0,16
109,5;119,5	0,20
119,5;129,5	0,30
129,5;139,5	0,24
139,5;149,5	0,10

- a) Calculeu la mitjana i la mediana i la desviació típica
- b) Calculeu el tercer quartil

27) Després d'estudiar les estatures (cm.) d'un grup d'estudiants universitaris s'ha obtingut la següent taula de distribució de freqüències (límits aparents)

Classe	%
150;159	10
160;169	25
170;179	35
180;189	15
190;199	15

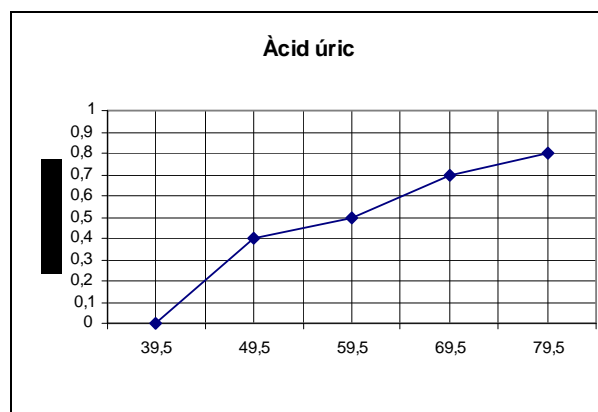
- a) Calculeu la mitjana i el percentil 80.

28) La utilització de serveis hospitalaris de la població que per domicili els correspon acudir a l'Hospital La Fe i a l'Hospital Clínic Universitari va generar les estades hospitalàries que s'arreglen al quadre adjunt:

Grup	Casos	Mitjana	Desviació
Homes	17.278	11,38	11,78
Dones	22.122	9,27	10,03

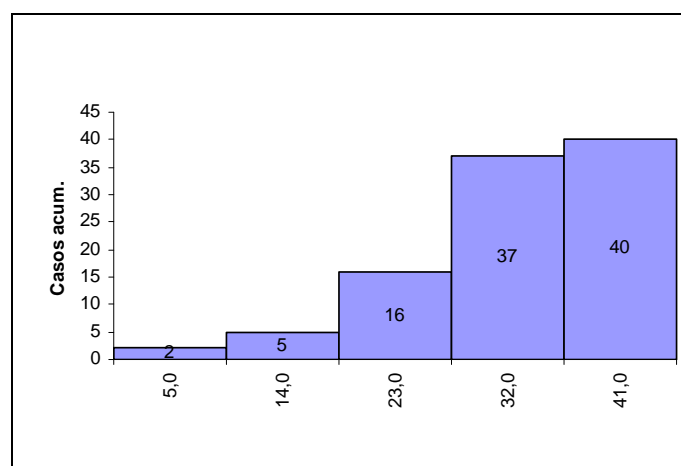
- a) Calculeu l'estada mitjana hospitalària que va generar tota la població que per domicili els corresponia acudir a l'Hospital La Fe i a l'Hospital Clínic Universitari
- b) Determineu el sexe on la variable 'estada hospitalària' està més dispersa

29) S'ha determinat la quantitat d'àcid úric (mg/l) en un grup de 100 persones normals, obtenint el polígon de proporcions acumulades adjunt.



- a) Calculeu la mitjana, la mediana i la proporció acumulada al valor 64,5

30) Amb l'histograma adjunt, calculeu la mitjana i el percentatge acumulat al valor 32



31) S'ha observat una variable X en dos grups d'individus.

Grup	Casos	Mitjana	Desviació
Grup A	50	9,70	6,2
Grup B	60	10,09	7,1

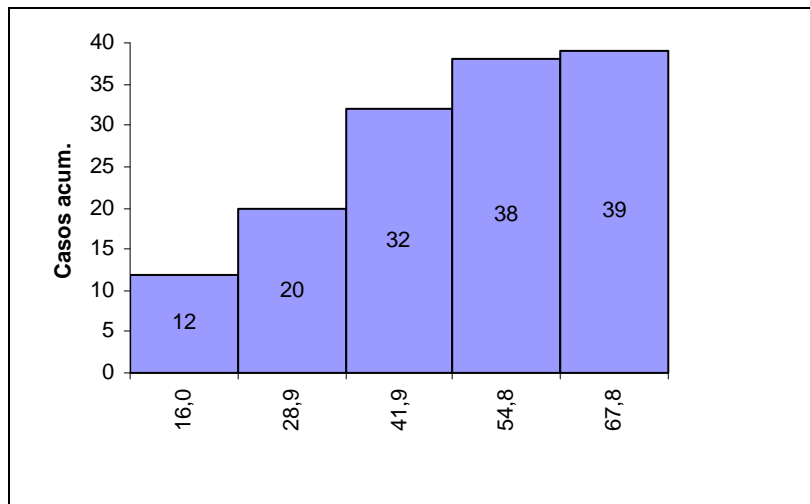
a) Calculeu la mitjana i la desviació típica del grup format per les persones dels grups A i B

32) El consum mitjà nacional de ferro en la dieta diària és de 13,14 mil·ligrams; atès que les necessitats fixades internacionalment per a aquest nutrient són d'11,50 mg/dia, es considera que la població espanyola està ben alimentada segons el factor citat. A fi de comprovar aquest fet i establir comparacions s'ha realitzat un estudi en dues ciutats A i B, amb els resultats següents:

	Ciutat A	Ciutat B
Casos	157	168
Mitjana	13,16	13,10
Desviació	1,76	1,98

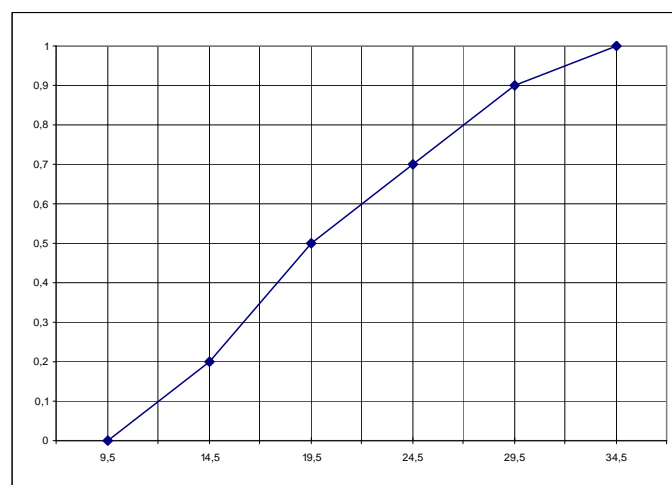
a) Calculeu la mitjana i la desviació que obtindriem si estudiàrem les ciutats A i B conjuntament

33) L'histograma de freqüències acumulades mostra la distribució de les quantitats de potassi (mEq/l) en mostres d'orina de 39 persones normals.



- a) Calculeu la mitjana i la desviació típica
 b) Calculeu la proporció acumulada al valor 64,35 mEq/l

34) Amb la informació que ens proporciona el polígon acumulatiu adjunt



a) Calculeu la mitjana, la mediana, el primer i el tercer quartil

35) Un grup d'investigadors del càncer de mama va reunir els següents resultats sobre la grandària de dos tipus de tumors (cm)

Tipus de tumor	Casos	Mitjana	Desviació
A	16	3,85	1,95
B	21	2,80	1,70

a) Calculeu la grandària mitjana dels 37 tumors estudiats

36) En estudiar una variable contínua obtenim les dades següents:

casos=175 valor mínim observat=15 valor màxim observat=42

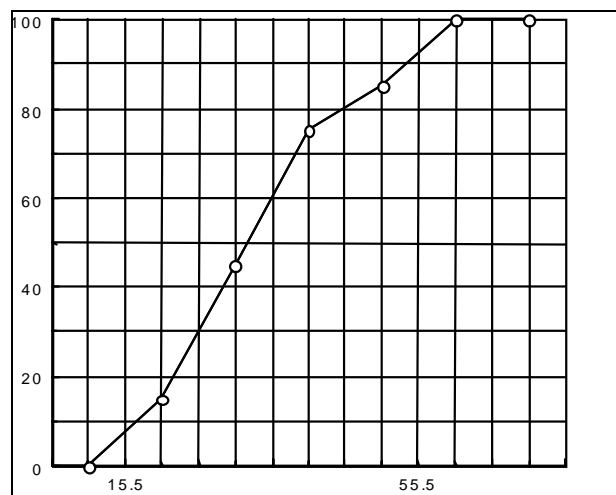
a) Construiu 8 intervals de la mateixa longitud (límits exactes) per a agrupar els valors observats

37) Els 52 alumnes que, després de realitzar dos exàmens parcials, van superar el curs van obtenir una nota mitjana de 6,1 punts en el primer parcial i 7,7 punts en el segon. Si la seua qualificació final es va obtenir mitjançant la fórmula:

$$\text{Nota final} = (2 \cdot \text{nota primer parcial} + \text{nota segon parcial}) / 3$$

a) Calculeu la nota mitjana final dels 52 alumnes. Raoneu la resposta.

38) El polígon acumulatiu adjunt mostra els percentatges acumulats (N=100)



- a) Construiu una taula de distribució de freqüències
- b) Calculeu la mitjana
- c) Calculeu l'interval interquartílic

39) Amb les dades de la taula adjunta

18,0	22,0	31,0	35,5	57,0	58,0	66,0	67,0
68,0	70,0	74,0	76,4	80,0	81,0	85,1	85,5
90,0	90,0	93,0	95,0	96,0	102,0	112,3	115,0
122,0	125,0	135,0	139,0	153,0	158,1	166,0	168,0
172,0	173,0	176,0	178,2	179,3	180,0	181,2	24,9

Clor (mEq/l) en orina de persones normals

- a) Determineu 7 intervals per agrupar les dades
- b) Construiu la taula de distribució de freqüències
- c) Calculeu, amb les dades sense agrupar, la mitjana, la desviació, el P_{70} el Q_1 i la Md
- d) Calculeu, amb les dades agrupades, la mitjana, la desviació, el P_{70} el Q_1 i la Md
- e) Dibuixeu l'histograma i el polígon acumulatiu

40) Coneguda la distribució de les defuncions de les dones menors de 45 anys residents al País Valencià durant el període 1991-1995 causades per accidents de vehicles de motor [codis E810-E825 de la CIM 9-Revisió].

Grup d'edat	0000	0104	0509	1014	1519	2024	2529	3034	3539	4044
Defuncions	2	7	10	20	105	76	55	41	15	24

- Calculeu l'edat mitjana de les defuncions, la mediana i el primer i tercer quartil
- Determineu el nombre total de defuncions d'aquest grup d'edat
- Construiu la taula de freqüències acumulada
- Dibuixeu l'histograma de freqüències

41) Amb la taula de distribució de freqüències adjunta determineu la sensibilitat i els límits aparents

Classes
12,5;15,5
15,5;18,5
18,5;21,5
21,5;24,5

42) Amb la taula de distribució de freqüències (N=50) adjunta determineu

Classe	H _i
15;19	0,1
20;24	0,3
25;29	0,6
30;34	0,9
35;39	1,0

- Els límits exactes
- Les freqüències de cadascuna de les classes
- La proporció acumulada al valor 27

43) Completeu la taula de distribució de freqüències adjunta

	3,5;9,5	9,5;A	B;21,5	C;27,5
f _i	3		8	
h _i		0,25		0,20
F _i			16	20

44) Amb la informació proporcionada en la taula de distribució de freqüències adjunta, determineu:

Clase	H _i
15;19	0,1
20;24	0,3
25;29	0,6
30;34	0,9
35;39	1,0

- La mitjana
- La mediana
- Els quartils

45) Els pesos (kg) de 90 xiquets/es amb edats entre els 2 i els 3 anys són:

13,7	10,9	12,8	14,6	14,8	14,0	14,5	18,2	14,9
13,7	15,9	13,8	17,0	12,9	15,3	13,5	12,6	14,3
14,3	15,5	11,8	18,0	18,5	15,8	13,8	17,8	17,5
15,8	19,6	11,5	14,7	15,9	12,8	16,7	15,7	15,0
13,3	17,5	14,3	15,5	13,7	12,6	13,9	14,2	17,2
13,7	14,0	12,8	14,6	14,8	10,9	14,5	18,2	14,9
13,7	15,3	13,8	17,0	12,9	15,9	13,5	12,6	14,3
14,3	15,8	11,8	18,0	18,5	15,5	13,8	17,8	17,5
15,8	12,8	11,5	14,7	15,9	19,6	16,7	15,7	15,0
13,3	12,6	14,3	15,5	13,7	17,5	13,9	14,2	17,2

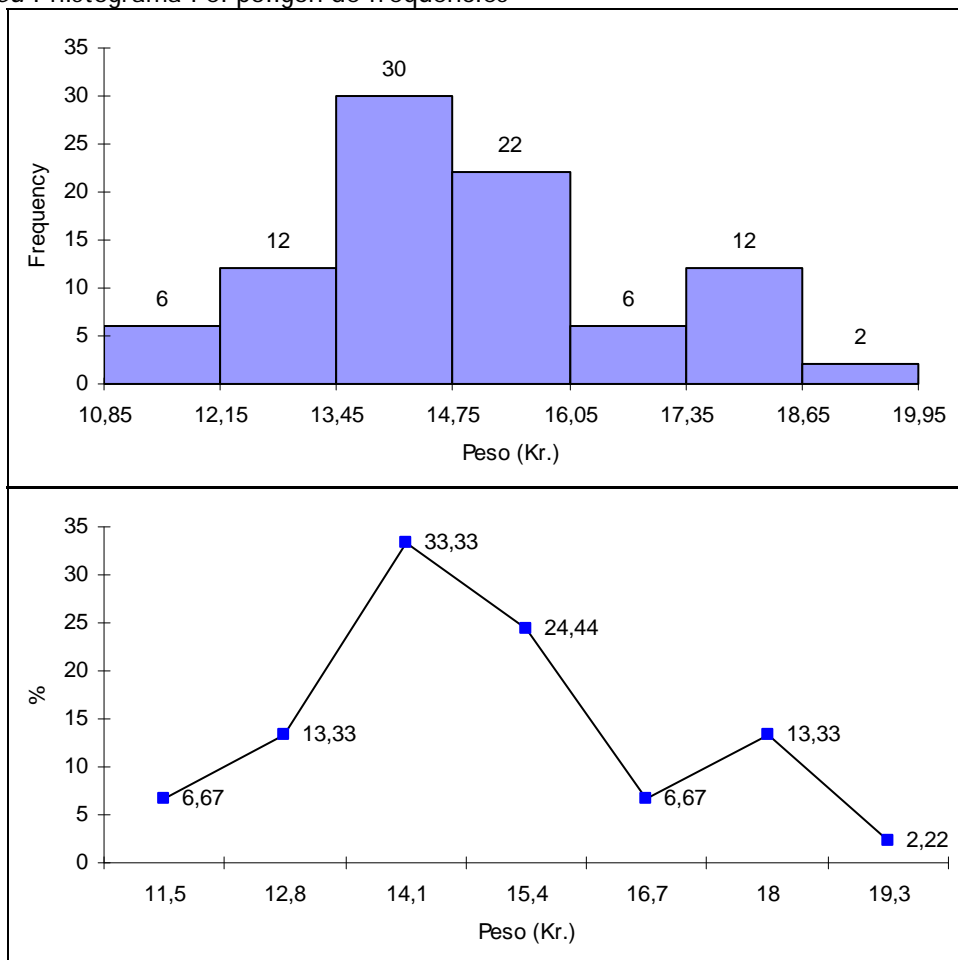
- Determineu els intervals per agrupar les dades en 7 classes de la mateixa mida

s	0,1	1,00	Límits exactes	
X mín	10,9		10,85	12,15
X màx	19,6		12,15	13,45
A	8,8		13,45	14,75
k	7		14,75	16,05
l	1,257143	1,3	16,05	17,35
			17,35	18,65
			18,65	19,95

b) Completeu la taula de distribució de freqüències

Intervals	Marca de classe	Freqüències	Proporcions	Proporcions acumulades
10,85 - 12,15	11,5	6	0,0667	0,0667
12,15 - 13,45	12,8	12	0,1333	0,2000
13,45 - 14,75	14,1	30	0,3333	0,5333
14,75 - 16,05	15,4	22	0,2444	0,7778
16,05 - 17,35	16,7	6	0,0667	0,8444
17,35 - 18,65	18,0	12	0,1333	0,9778
18,65 - 19,95	19,3	2	0,0222	0,0667

c) Dibuixeu l'histograma i el polígon de freqüències



d) Calculeu la mitjana i la desviació típica (realitzeu el càlcul amb les dades agrupades)

Numerical Summaries for Pes (Kr.)	
Casos	90

Mitjana	14,90
Desviació	1,96

e) Calculeu i interpreteu el percentil 70 i la proporció acumulada al valor $x=16,4$ kg. (realitzeu el càlcul amb les dades sense agrupar)

$P_{70} = \text{PERCENTIL}(\text{matriu de dades}; 0,7) = 15,8$

El 70% dels pesos observats són menors o iguals a 15,8 Kg.

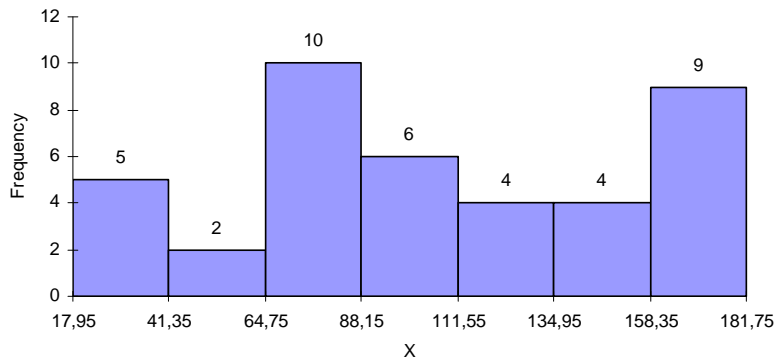
Si ordenem les dades observem que hi ha 70 valors menors o iguals que 16,4 i per tant

$H(16,4) = 70/90 = 0,7778$

46) Amb les dades de la taula adjunta [Clor (mEq/l) en orina de persones normals] *agrupeu les dades en 7 intervals i calculeu, amb les dades sense agrupar, la mitjana, el P_{70} i els tres quartils*

18,0	22,0	31,0	35,5	57,0	58,0	66,0	67,0
68,0	70,0	74,0	76,4	80,0	81,0	85,1	85,5
90,0	90,0	93,0	95,0	96,0	102,0	112,3	115,0
122,0	125,0	135,0	139,0	153,0	158,1	166,0	168,0
172,0	173,0	176,0	178,2	179,3	180,0	181,2	24,9

Casos	40	Mín	18
Mitjana	104,9625	Q_1	69,5
Desviació (n-1)	50,11998	Mediana	94
Coef. Var.	0,477504	Q_3	154,275
		Màx	181,2

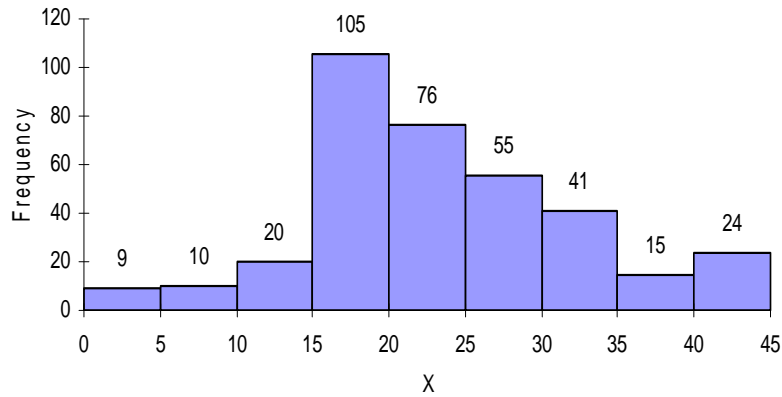


47) Coneguda la distribució de les defuncions de dones menors de 45 anys causades per accidents de vehicles de motor [codis E810-E825 de la CIM 9-Revisió] al País Valencià durant el període 1991-1995. *Calculeu l'edat mitjana de les defuncions, la mediana i el primer i tercer quartil. Determineu el nombre total de defuncions d'aquest grup d'edat, la taula de freqüències acumulada i l'histograma de freqüències*

Grup d'edat	0000	0104	0509	1014	1519	2024	2529	3034	3539	4044
Defuncions	2	7	10	20	105	76	55	41	15	24
Defuncions acum.	2	9	19	39	144	220	275	316	331	355

En el període 1991-1995 moren al PV un total de 355 dones menors de 45 anys per accidents de vehicles de motor

Casos	355	Mín	0
Mitjana	23,44366	Q_1	17,36905
Desviació (n-1)	8,989773	Mediana	22,20395
Coef. Var.	0,383463	Q_3	29,20455
		Màx	45



48) Variable X: Nivell de pols de cadmi i fum d'òxid de cadmi en l'aire (mg de cadmi/m³ d'aire)

n = 35

$x_{\max} = 0,070$ $x_{\min} = 0,020$

Calculeu la sensibilitat, l'amplitud, la longitud de la classe si el nombre d'intervalos és 7, els límits exactes i aparents dels intervals

Sensibilitat (S) = 0,001

Amplitud (A) = $0,070 - 0,020 + 0,001 = 0,051$

Longitud de la classe (interval) si el nombre d'intervalos és 7

$l = 0,051 / 7 = 0,00728571 \rightarrow l = 0,008$ (múltiple de la sensibilitat per excés)

Límits exactes dels 7 intervalos (comencem amb $x_{\min} - S/2$)

Límits exactes		Límits aparents	
0,0195	0,0275	0,020	0,027
0,0275	0,0355	0,028	0,035
0,0355	0,0435	0,036	0,043
0,0435	0,0515	0,044	0,051
0,0515	0,0595	0,052	0,059
0,0595	0,0675	0,060	0,067
0,0675	0,0755	0,068	0,075

49) Amb la següent TDF calculeu la sensibilitat, l'amplitud, la longitud de l'interval i els límits aparents

classes
 12,5 - 15,5
 15,5 - 18,5
 18,5 - 21,5
 21,5 - 24,5
 24,5 - 27,5

Sensibilitat (S)=1

Amplitud (A)= $27,5 - 12,5 = 15$

Longitud de la classe (interval)=3

Límits aparents dels intervalos

13 - 15
 16 - 18
 19 - 21
 22 - 24
 25 - 27

50) Amb la taula de distribució de freqüències (sols proporcions) adjunta determineu els límits exactes, les freqüències de cada classe i la proporció acumulada al valor 27

Classes	H_i	f_i	Límits exactes
15 - 19	0,1	5	14,5 - 19,5
20 - 24	0,3	10	19,5 - 24,5
25 - 29	0,6	15	24,5 - 29,5
30 - 34	0,9	15	29,5 - 34,5
35 - 39	1,0	5	34,5 - 39,5

N=50

La proporció acumulada al valor 27 (punt mitjà de l'interval 24,5-29,5) serà: $(5+10+15/2)/50$