

Exemple Les taules de freqüències de les variables LLEN i SOC són:

LLEN	freqüència absoluta	freqüència relativa
1	149	0,4585
2	105	0,3231
3	64	0,1969
4	7	0,0215
Totals	325	1

SOC	freq. abs.	freq. rel.
0	5	0,0154
1	1	0,0031
2	2	0,0062
3	5	0,0154
4	20	0,0615
5	59	0,1815
6	33	0,1015
7	77	0,2369
8	86	0,2646
9	18	0,0554
10	19	0,0585

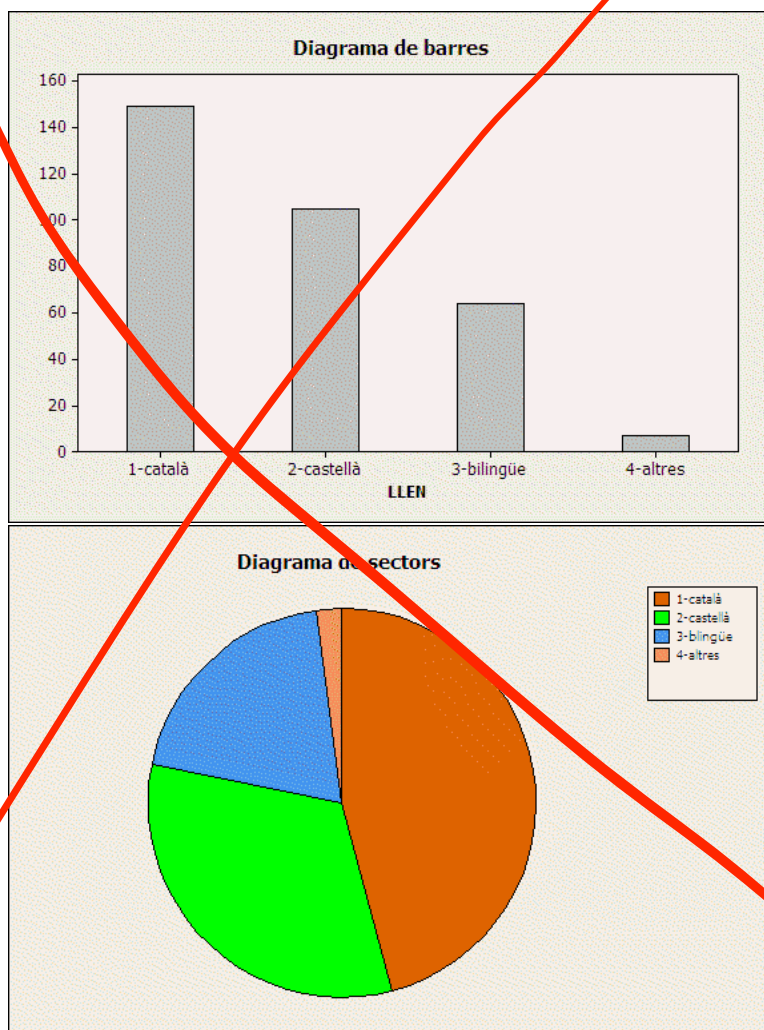
Exemple Per la variable LLEN no tindria sentit fer una taula de freqüències acumulades. En canvi, per la variable SOC sí que la podem fer:

SOC	freq. abs.	freq. abs. acum.	freq. rel.	freq. rel. acum.
0	5	5	0,0154	0,0154
1	1	6	0,0031	0,0185
2	2	8	0,0062	0,0246
3	5	13	0,0154	0,04
4	20	33	0,0615	0,1015
5	59	92	0,1815	0,2831
6	33	125	0,1015	0,3846
7	77	202	0,2369	0,6215
8	86	288	0,2646	0,8862
9	18	306	0,0554	0,9415
10	19	325	0,0585	1

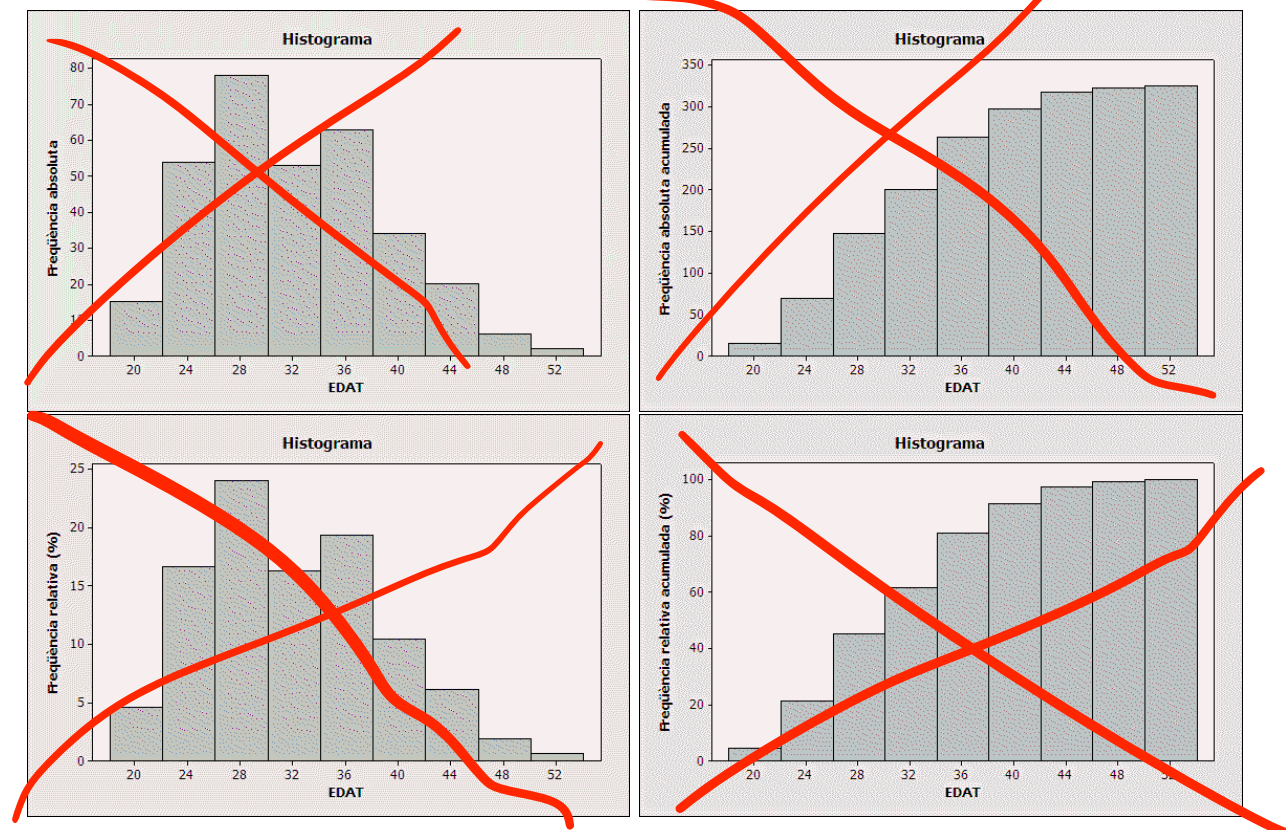
Exemple La variable EDAT pot prendre masses valors diferents. Per fer la taula hem d'agrupar els valors en intervals. Els agafem tots de la mateixa longitud. El punt mig de cada interval serà la marca de classe. Per exemple, 24 serà la marca de classe de l'interval (22, 26].

SOC	freq. abs.	freq. abs. acum.	freq. rel.	freq. rel. acum.
(18, 22]	15	15	0,0462	0,0462
(22, 26]	54	69	0,1661	0,2123
(26, 30]	78	147	0,2400	0,4523
(30, 34]	53	200	0,1631	0,6154
(34, 38]	63	263	0,1938	0,8092
(38, 42]	34	297	0,1046	0,9138
(42, 46]	20	317	0,0616	0,9754
(46, 50]	6	323	0,0184	0,9938
(50, 54]	2	325	0,0062	1

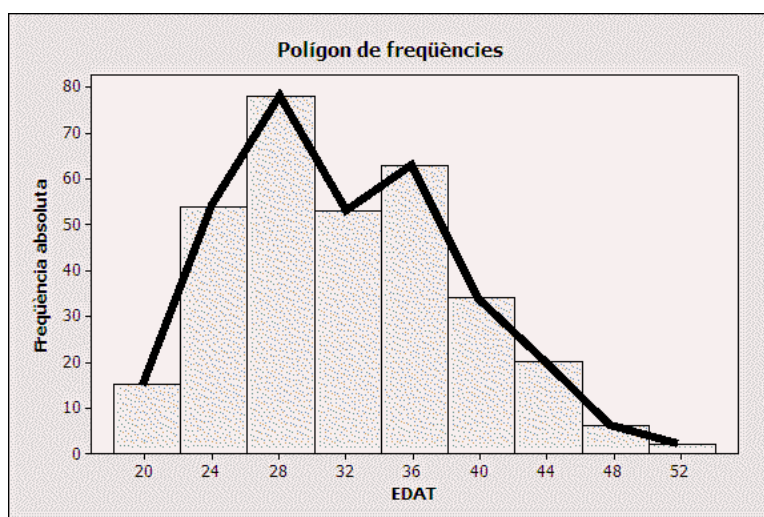
Exemple Per la variable LLEN dibuixem els dos gràfics comentats:



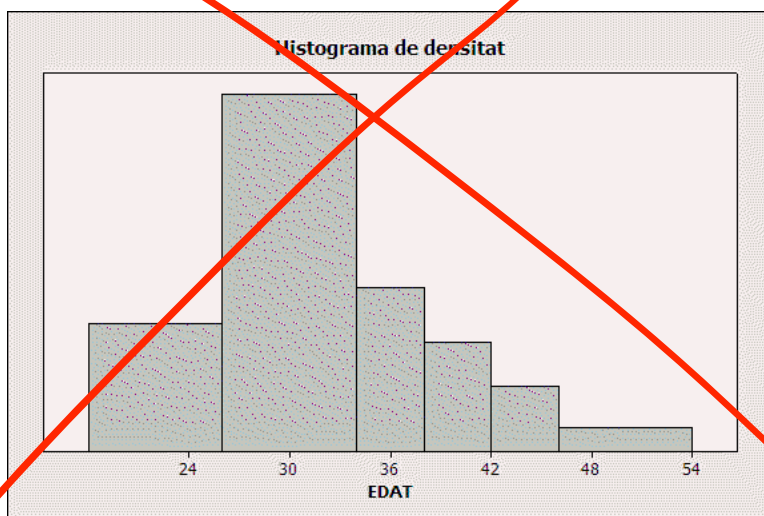
Exemple. Recuperem la variable EDAT per fer els histogrames de freqüències a partir de la taula que havíem vist a l'apartat anterior.



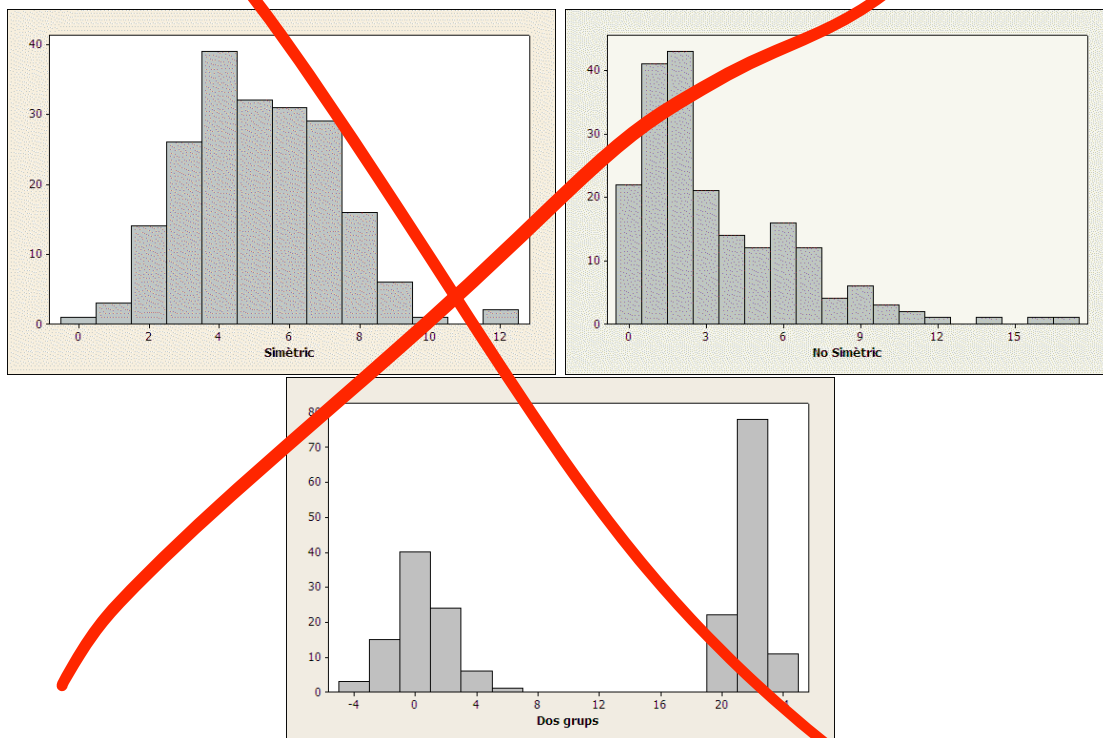
Si unim els punts superiors dels rectangles amb una línia obtenim l'anomenat **polígon de freqüències**.

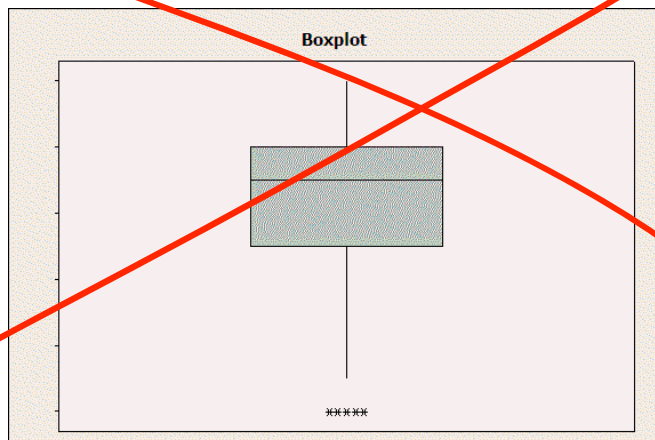


Exemple. També amb la variable EDAT podem agafar intervals de distintes longituds:

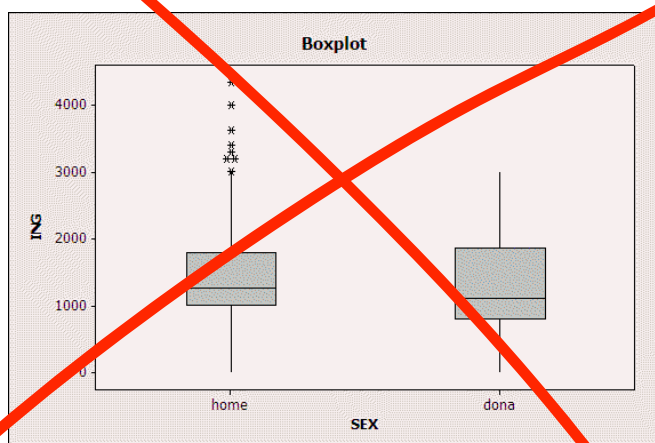


La interpretació dels histogrames ens ha de permetre, entre d'altres coses, detectar: simetries, l'existència de classes buides que separin la població en grups..





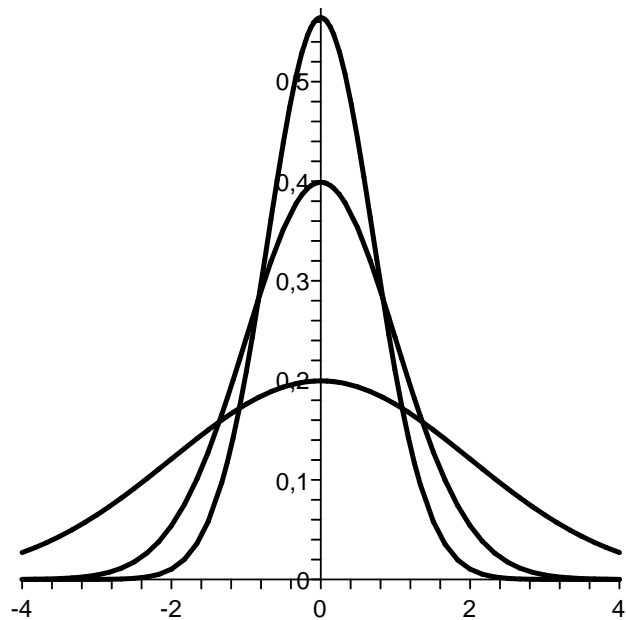
Example: Fem els diagrames de caixa de la variable ING -ingressos mensuals en euros- en dos grups segons la variable SEX (home-dona).



Observem que en els homes hi ha molt outliers, individus amb ingressos molt alts.

Exemple. Fem el gràfic de tija i fulles de la variable EDAT.

3	1	999
7	2	1111
22	2	2222222233333333
49	2	4444444444445555555555555555
87	2	66666666666666666666667777777777777777
127	2	88888888888888888888889999999999999999
(39)	3	00000000000000000000001111111111111111111
159	3	22222222222233333333333
139	3	4444444444444444555555555555555555555555
101	3	66666666666666666666667777777777777777
76	3	88888888888888888888889999999999
54	4	00000000000000000000001111111111
34	4	222222223333
24	4	44444444444455555555
9	4	6
8	4	8899
4	5	001
1	5	3

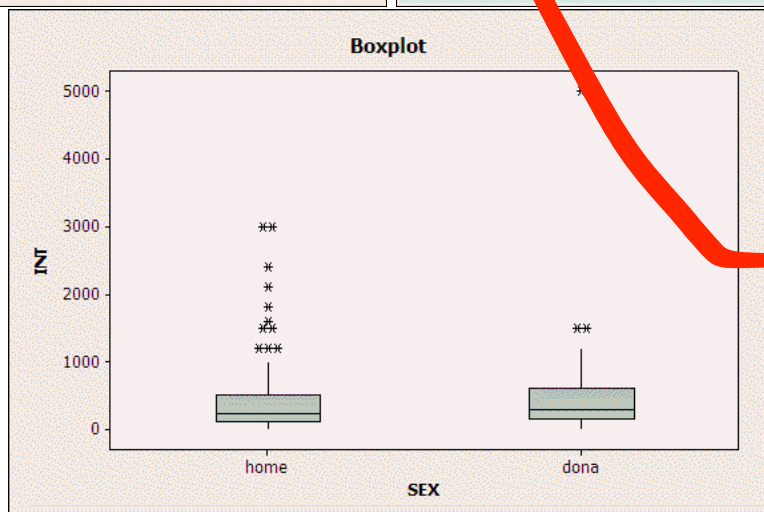
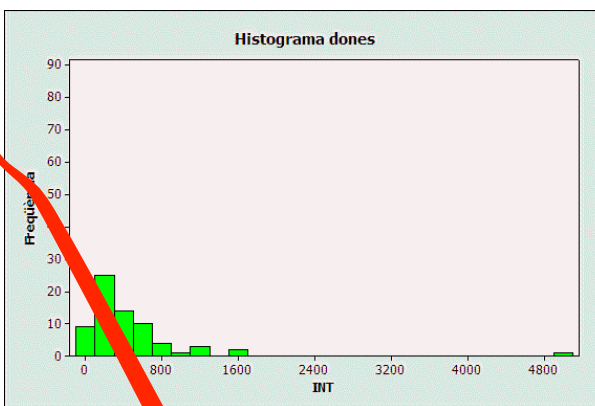
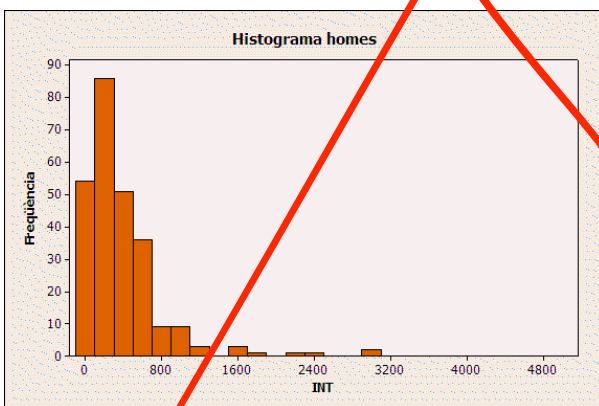


S'aprecien: la corba normal al mig, la corba més plana (curtosi < 3) i la corba més apuntada (curtosi > 3).

Exemple. Fem un estudi descriptiu de la variable INT -temps de connexió a internet- distingint segons la variable SEX -home i dona. Recollim els estadístics numèrics.

	Tots	Homes	Dones
mitjana	381,4	363,5	447,8
mediana	240	240	309
moda	600	600	300
desviació	480,6	419,5	659,4
variància	230947,4	175983,9	434773,7
coeficient d variació	126	115,41	147,24
mínim	0	0	0
Q1	120	120	150
Q3	520	500	600
màxim	5000	3000	5000
rang	5000	3000	5000
rang interquar.	400	380	450
skewness	4,50	3,23	5,13
curtosis	31,89	14,67	33,70

I els gràfics:



Exemple. Fem la taula de contingència de les variables SEX i LLEN.

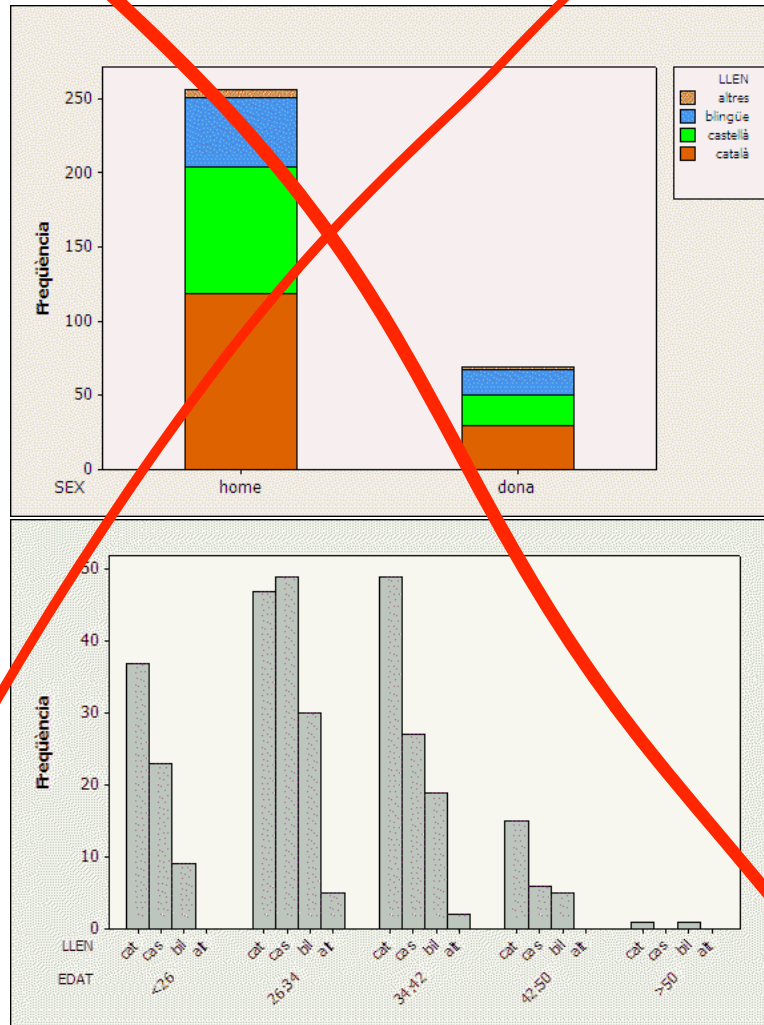
	català	castellà	bilingüe	altres	Totals
home	119	85	47	5	256
dòna	30	20	17	2	69
Totals	149	105	64	7	325

Interpretem alguns valors: la freqüència absoluta de ("home", "català") és 119, la freqüència absoluta de "home" és 256, la freqüència relativa de "home" és $\frac{256}{325} = 0,7877$, la freqüència relativa de (home,català) és $\frac{119}{325} = 0,4648$, la freqüència relativa de "català" dins dels homes és de $\frac{119}{256} = 0,4648$, la freqüència relativa de "home" dins dels que parlen català és de $\frac{119}{149} = 0,7987$.

Exemple. Fem la taula de contingència de les variables EDAT i LLEN. Com que la variable EDAT pren molts valors diferents, l'agruparem en classes.

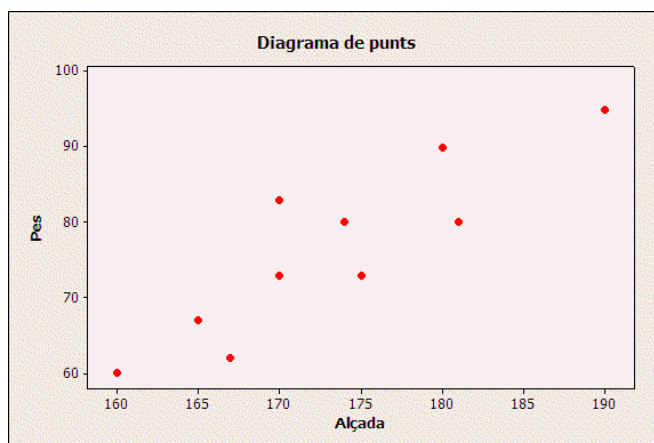
	català	castellà	bilingüe	altres	Totals
menys de 26	37	23	9	0	69
entre 26 i 34	47	49	30	5	131
entre 34 i 42	49	27	19	2	97
entre 42 i 50	15	6	5	0	26
entre 50 i 58	1	0	1	0	2
Totals	149	105	64	7	325

Hi ha alguns gràfics -tipus diagrames de barres- que volen representar aquestes taules de contingència. Vegem alguns exemples:

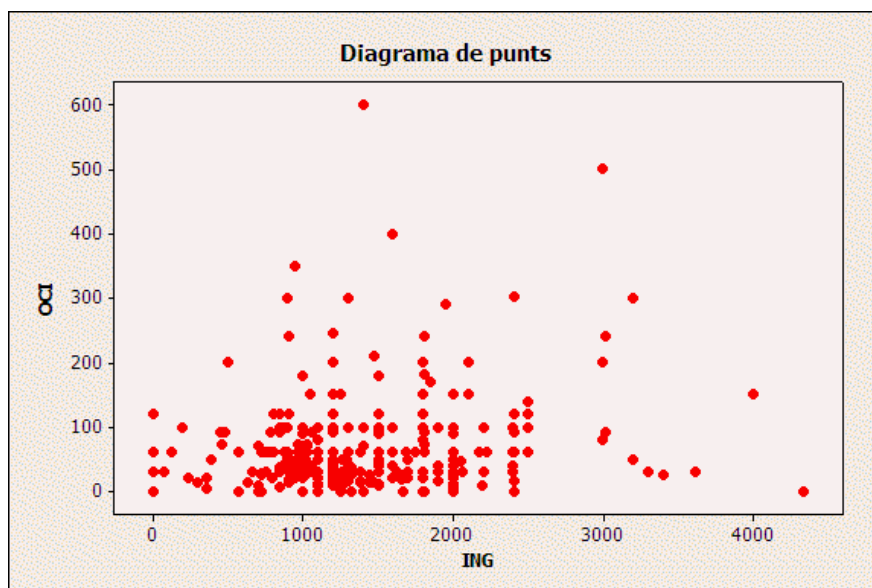


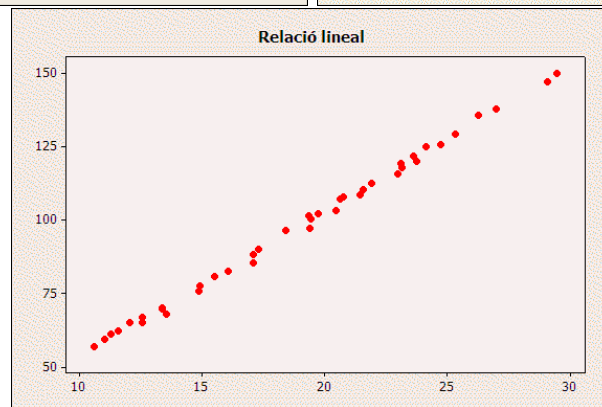
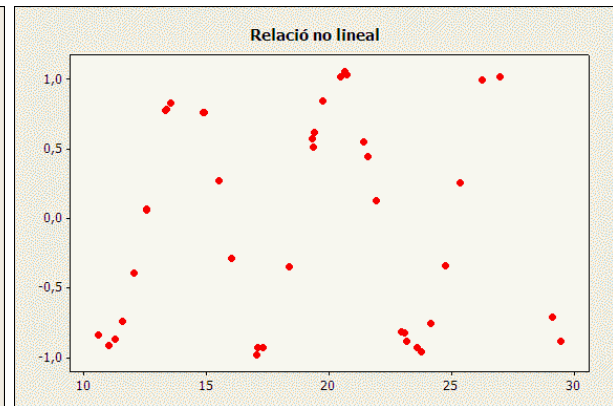
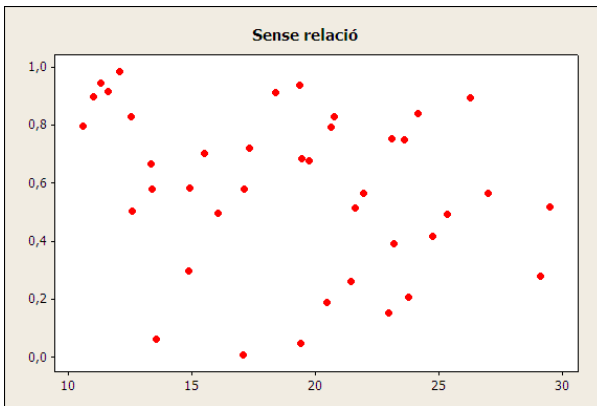
Exemple. Tenim un mostra de 10 homes dels quals volem estudiar el seu pes i la seva alçada. Tenim les dades següents:

alçada: 180 170 165 190 170 181 175 174 167 160
pes: 90 73 67 95 83 80 73 80 62 60. El núvol de punts és el següent:



Exemple. Fem el núvol de punts de les variables ING -ingressos mensuals en euros- i OCI -despeses mensuals en oci en euros.

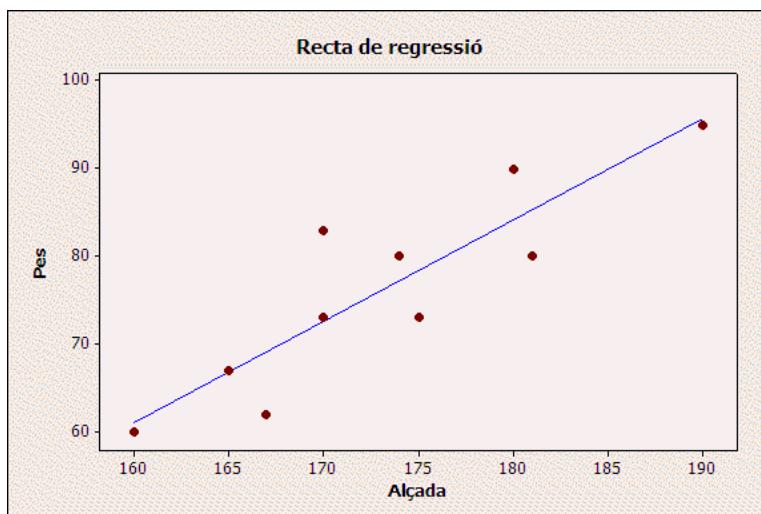




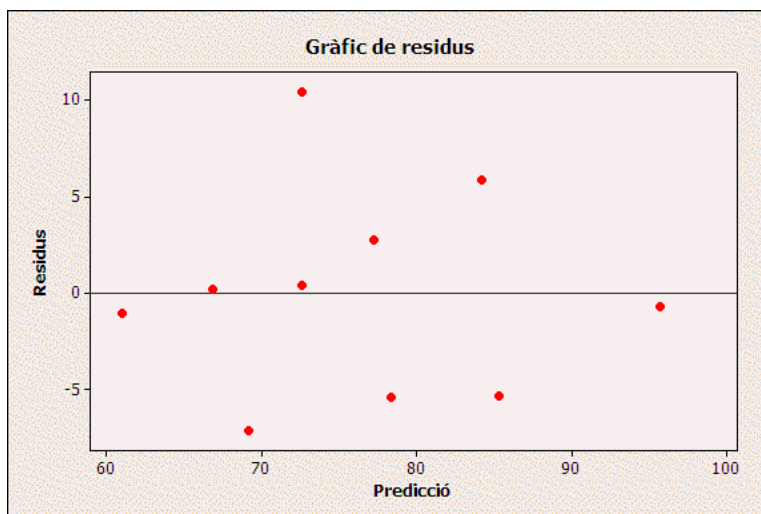
Tornem a l'exemple dels 10 homes. Tenim la recta

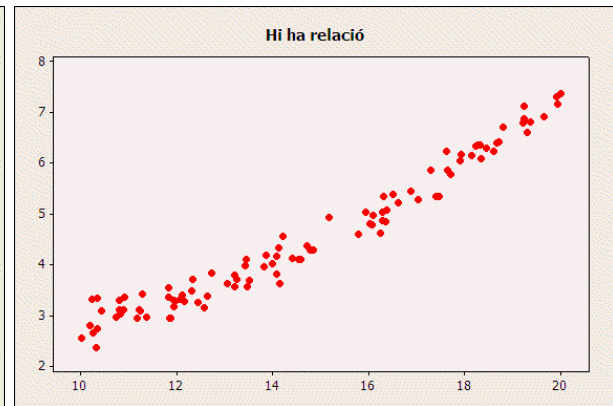
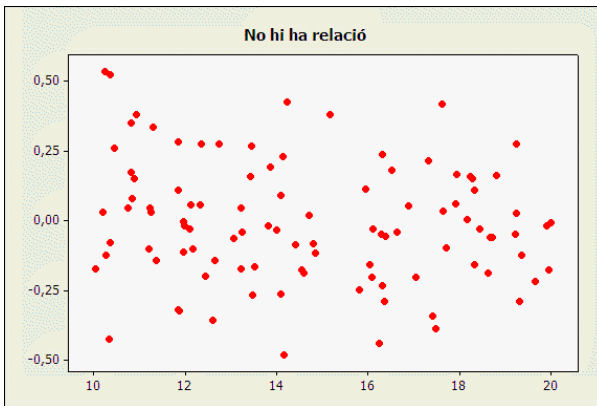
$$y = -124,16 + 1,15686x.$$

Si la dibuixem tenim



L'anàlisi dels residus :



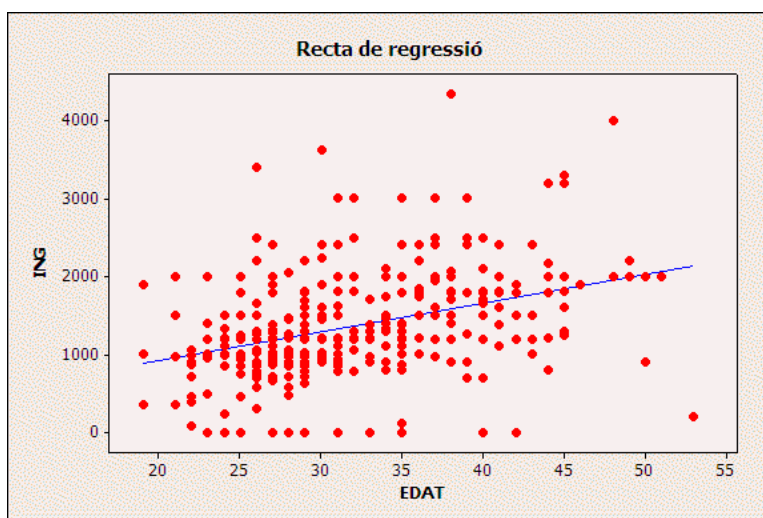


Exemple. Del nostre fitxer de dades volem explicar la variable ING -ingressos mensuals en euros- en termes de la variable EDAT. ING serà la variable resposta y i EDAT serà la variable predictiva x .

Podem calcular la recta de regressió i obtenim que

$$y = 174 + 37,1x$$

amb el gràfic



Dibuixem el gràfic dels residus

