

ESTADÍSTICA DESCRIPTIVA II: ESTADÍSTICA DESCRIPTIVA BIVARIANT

Curs d'Estadística Bàsica per a la Recerca Biomèdica

UEB - VHIR

Ricardo Gonzalo Sanz ricardo.gonzalo@vhir.org

Miriam Mota Foix miriam.mota@vhir.org













23/01/2019



TABLE OF CONTENTS

1. From univariate to bivariate analysis

2. Bivariate analysis

- 1. Qualitative vs Qualitative
- 2. Qualitative vs Quantitative
- 3. Quantitative vs Quantitative

3. Correlation

- 1. Definition
- 2. Types of correlation (Pearson, Spearman)



TABLE OF CONTENTS

1. From univariate to bivariate analysis

2. Bivariate analysis

- 1. Qualitative vs Qualitative
- 2. Qualitative vs Quantitative
- 3. Quantitative vs Quantitative

3. Correlation

- 1. Definition
- 2. Types of correlation (Pearson, Spearman)

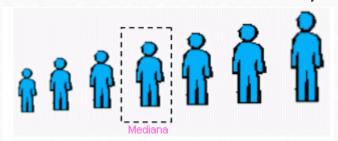


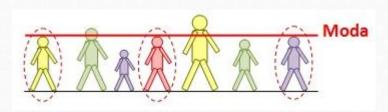
On Monday we learned...

• We can analyse and describe each variable one by one:

1. With some measures:

Measures of central tendency



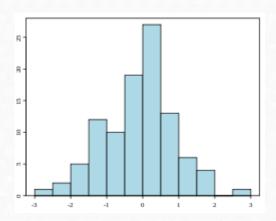


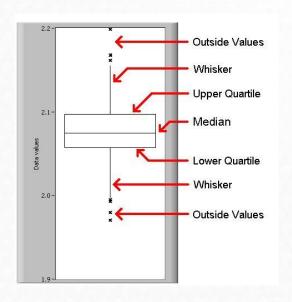
Measures of dispersion

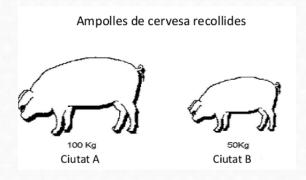




2. Some graphics









• In univariate analysis only one variable is analyzed each time



the purpose of the analysis is descriptive

- If there are more than one variable in the dataset it could be interesting to guess if:
 - Does exist a relation between the two variables?
 - How important is this relation?
 - Which is the direction of the relation?



₩ os	teo											×
	registro	area	f_nac	edad	grupedad	peso	talla	bua	imc	clasific	me	
1	3	10	11659420800	57	55 - 59	70.0	168.0	69	24.80159	OSTEOPENIA		۸
2	4	10	11671689600	46	45 - 49	53.0	152.0	73	22.93975	OSTEOPENIA		
3	10	10	11721024000	45	45 - 49	64.0	158.0	81	25.63692	NORMAL		
4	11	10	11464416000	53	50 - 54	78.0	161.0	58	30.09143	OSTEOPENIA		
5	12	10	11690784000	46	45 - 49	56.0	157.0	89	22.71897	NORMAL		
6	15	10	11716012800	45	45 - 49	63.5	170.0	76	21.97232	NORMAL		
7	16	10	11623737600	48	45 - 49	86.0	161.0	87	33.17773	NORMAL		
8	17	10	11562307200	50	50 - 54	61.5	164.0	74	22.86585	NORMAL		
9	18	10	11538028800	51	50 - 54	60.5	158.0	58	24.23490	OSTEOPENIA		
10	20	10	11332483200	57	55 - 59	64.0	149.0	61	28.82753	OSTEOPENIA		
11	21	10	11631945600	48	45 - 49	70.3	160.0	67	27.46094	OSTEOPENIA		
12	22	10	11425536000	55	55 - 59	74.4	160.0	68	29.06250	OSTEOPENIA		
13	23	10	11553235200	50	50 - 54	55.5	154.5	73	23.25070	OSTEOPENIA		
14	24	10	11367302400	56	55 - 59	89.0	166.0	61	32.29787	OSTEOPENIA		
15	25	10	11585635200	49	45 - 49	50.6	157.0	68	20.52822	OSTEOPENIA		
16	26	10	11572156800	50	50 - 54	71.4	152.0	74	30.90374	NORMAL		
17	27	10	11590992000	49	45 - 49	78.0	157.0	62	31.64429	OSTEOPENIA		
18	28	10	11293516800	58	55 - 59	72.0	162.0	65	27.43484	OSTEOPENIA		
19	29	10	11215238400	61	60 - 64	68.0	155.5	65	28.12212	OSTEOPENIA		
20	30	10	11405664000	55	55 - 59	75.0	161.0	92	28.93407	NORMAL		
21	31	10	11633155200	48	45 - 49	66.5	153.0	11	28.40788	OSTEOPOROSIS		
22	32	10	11287728000	59		101.0	156.0	82	41.50230	NORMAL		
23	34	10	10992758400	68	65 – 69	66.5	145.0	57	31.62901	OSTEOPENIA		
24	35	10	10909382400	69	65 – 69	70.0	168.0	48	24.80159	OSTEOPOROSIS		
25	36	10	11643868800	48	45 - 49	60.1	153.0	86	25.67389	NORMAL		
26	37	10	11551420800	50	50 - 54	67.0	159.0	105	26.50212	NORMAL		
27	38	10	11043907200	66	65 – 69	67.0	144.0	79	32.31096	NORMAL		
28	39	10	10948089600	69	65 – 69	70.5	148.5	40	31.96953	OSTEOPOROSIS		
29	40	10	11051251200	66	65 – 69	66.5	147.0	48		OSTEOPOROSIS		
30	41	10	11333692800	57	55 - 59	58.5	142.0	80	29.01210	NORMAL		٧
	<	Ш									>	



🖙 os	teo									ال	×
	registro	area	f_nac	edad	grupedad	peso tall:	a bua	imc	clasific	me	
1	3	10	11659420800	57	55 - 59	70.0 168.	69	24.80159	OSTEOPENIA		^
2	4	10	11671689600	46	45 - 49	53.0 152.	73	22.93975	OSTEOPENIA		
2 3	10	10	11721024000	45	45 - 49	64.0 158.	81	25.63692	NORMAL		
4	11	10	11464416000	53	50 - 54	78.0 161.	58	30.09143	OSTEOPENIA		
5	12	10	11690784000	46	45 - 49	56.0 157.	89	22.71897	NORMAL		
6	15	10	11716012800	45	45 - 49	63.5 170.	76	21.97232	NORMAL		
7	16	10	11623737600	48	45 - 49	86.0 161.	87	33.17773	NORMAL		
8	17	10	11562307200	50	50 - 54	61.5 164.	74	22.86585	NORMAL		
9	18	10	11538028800	51	50 - 54	60.5 158.	58	24.23490	OSTEOPENIA		
10	20	10	11332483200	57	55 - 59	64.0 149.	0 61	28.82753	OSTEOPENIA		
11	21	10	11631945600	48	45 - 49	70.3 160.	5 67	27.46094	OSTEOPENIA		
12	22	10	11425536000	55	55 - 59	74.4 160.	5 68	29.06250	OSTEOPENIA		
13	23	10	11553235200	50	50 - 54	55.5 154.	5 73	23.25070	OSTEOPENIA		
14	24	10	11367302400	56	55 - 59	89.0 166.	0 61	32.29787	OSTEOPENIA		
15	25	10	11585635200	49	45 - 49	50.6 157.	5 68	20.52822	OSTEOPENIA		
16	26	10	11572156800	50	50 - 54	71.4 152.	74	30.90374	NORMAL		
17	27	10	11590992000	49	45 - 49	78.0 157.	5 62	31.64429	OSTEOPENIA		
18	28	10	11293516800	58	55 - 59	72.0 162.	5 65	27.43484	OSTEOPENIA		
19	29	10	11215238400	61	60 – 64	68.0 155.	5 65	28.12212	OSTEOPENIA		
20	30	10	11405664000	55	55 - 59	75.0 161.	92	28.93407	NORMAL		
21	31	10	11633155200	48	45 - 49	66.5 153.	11	28.40788	OSTEOPOROSIS		
22	32	10	11287728000	59	55 - 59	101.0 156.	82	41.50230	NORMAL		
23	34	10	10992758400	68	65 – 69	66.5 145.	57	31.62901	OSTEOPENIA		
24	35	10	10909382400	69	65 – 69	70.0 168.	48	24.80159	OSTEOPOROSIS		
25	36	10	11643868800	48	45 - 49	60.1 153.	86	25.67389	NORMAL		
26	37	10	11551420800	50	50 - 54	67.0 159.	105	26.50212	NORMAL		
27	38	10	11043907200	66	65 – 69	67.0 144.	79	32.31096	NORMAL		
28	39	10	10948089600	69	65 – 69	70.5 148.	5 40	31.96953	OSTEOPOROSIS		
29	40	10	11051251200	66	65 – 69	66.5 147.	3 48	30.77421	OSTEOPOROSIS		
30	41	10	11333692800	57	55 - 59	58.5 142.	80	29.01210	NORMAL		V
	<									>	



🕠 os	teo											×
	registro	area	f_nac	edad	grupedad	peso	talla	bua	imc	clasific	me	
1	3	10	11659420800	57	55 - 59	70.0	168.0	69	24.80159	OSTEOPENIA		^
2	4	10	11671689600	46	45 - 49	53.0	152.0	73	22.93975	OSTEOPENIA		Ī
}	10	10	11721024000	45	45 - 49	64.0	158.0	81	25.63692	NORMAL		
Ł	11	10	11464416000	53	50 - 54	78.0	161.0	58	30.09143	OSTEOPENIA		
5	12	10	11690784000	46	45 - 49	56.0	157.0	89	22.71897	NORMAL		
5	15	10	11716012800	45	45 - 49	63.5	170.0	76	21.97232	NORMAL		
7	16	10	11623737600	48	45 - 49	86.0	161.0	87	33.17773	NORMAL		
	17	10	11562307200	50	50 - 54	61.5	164.0	74	22.86585	NORMAL		
)	18	10	11538028800	51	50 - 54	60.5	158.0	58	24.23490	OSTEOPENIA		
LO	20	10	11332483200	57	55 - 59	64.0	149.0	61	28.82753	OSTEOPENIA		
11	21	. 10	11631945600	48	45 - 49	70.3	160.0	67	27.46094	OSTEOPENIA		
L2	22	10	11425536000	55	55 - 59	74.4	160.0	68	29.06250	OSTEOPENIA		
L3	23	10	11553235200	50	50 - 54	55.5	154.5	73	23.25070	OSTEOPENIA		
.4	24	10	11367302400	56	55 - 59	89.0	166.0	61	32.29787	OSTEOPENIA		
.5	25	10	11585635200	49	45 - 49	50.6	157.0	68	20.52822	OSTEOPENIA		
.6	26	10	11572156800	50	50 - 54	71.4	152.0	74	30.90374	NORMAL		
.7	27	10	11590992000	49	45 - 49	78.0	157.0	62	31.64429	OSTEOPENIA		
.8	28	10	11293516800	58	55 - 59	72.0	162.0	65	27.43484	OSTEOPENIA		
.9	29	10	11215238400	61	60 - 64	68.0	155.5	65	28.12212	OSTEOPENIA		
0	30	10	11405664000	55	55 - 59	75.0	161.0	92	28.93407	NORMAL		
1	31	. 10	11633155200	48	45 - 49	66.5	153.0	11	28.40788	OSTEOPOROSIS		
22	32	10	11287728000	59	55 - 59	101.0	156.0	82	41.50230	NORMAL		
3	34	10	10992758400	68	65 – 69	66.5	145.0	57	31.62901	OSTEOPENIA		
24	35	10	10909382400	69	65 – 69	70.0	168.0	48	24.80159	OSTEOPOROSIS		
25	36	10	11643868800	48	45 - 49	60.1	153.0	86	25.67389	NORMAL		
26	37	10	11551420800	50	50 - 54	67.0	159.0	105	26.50212	NORMAL		
27	38	10	11043907200	66	65 – 69	67.0	144.0	79	32.31096	NORMAL		
28	39	10	10948089600	69	65 – 69	70.5	148.5	40	31.96953	OSTEOPOROSIS		
29	40	10	11051251200	66	65 – 69	66.5	147.0	48	30.77421	OSTEOPOROSIS		
30	41	. 10	11333692800	57	55 - 59	58.5	142.0	80	29.01210	NORMAL		-
	<										>	I



TABLE OF CONTENTS

1. From univariate to bivariate analysis

2. Bivariate analysis

- 1. Qualitative vs Qualitative
- 2. Qualitative vs Quantitative
- 3. Quantitative vs Quantitative

3. Correlation

- 1. Definition
- 2. Types of correlation (Pearson, Spearman)



Bivariate analysis

• Involves the analysis of **two** variables for the purpose of determining the empirical relationship between them.



easiest way is to measure how those two variables simultaneously change together



Bivariate analysis

• Involves the analysis of **two** variables for the purpose of determining the empirical relationship between them.



easiest way is to measure how those two variables simultaneously change together

• Major differentiating point between univariate and bivariate analysis (a part from the number of variables implicated) is that bivariate analysis goes beyond simply descriptive, since it study the relationship between the two variables.



Why bivariate analysis?

Let's begin by asking if:

People tend to marry other people of about the same age?

Our experience tells us "yes", but how good is the correspondence?

Husband	36	72	37	36	51	50	47	50	37	41
Wife	35	67	33	35	50	46	47	42	36	41

Sample of spousal ages of 10 White American Couples



Why bivariate analysis?

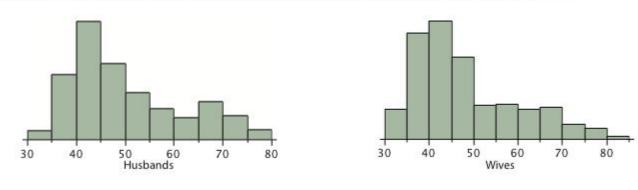


Figure 1. Histograms of spousal ages.

	Mean	Standard Deviation
Husbands	49	11
Wives	47	11



Why bivariate analysis?

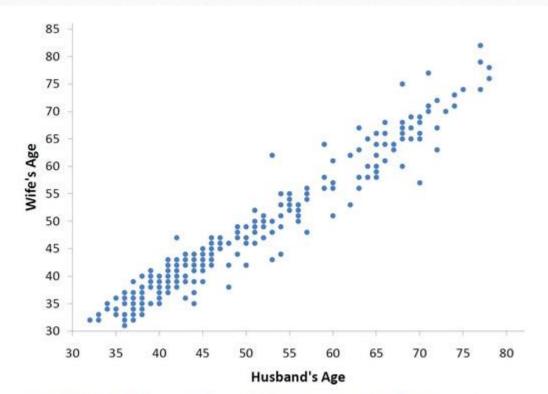
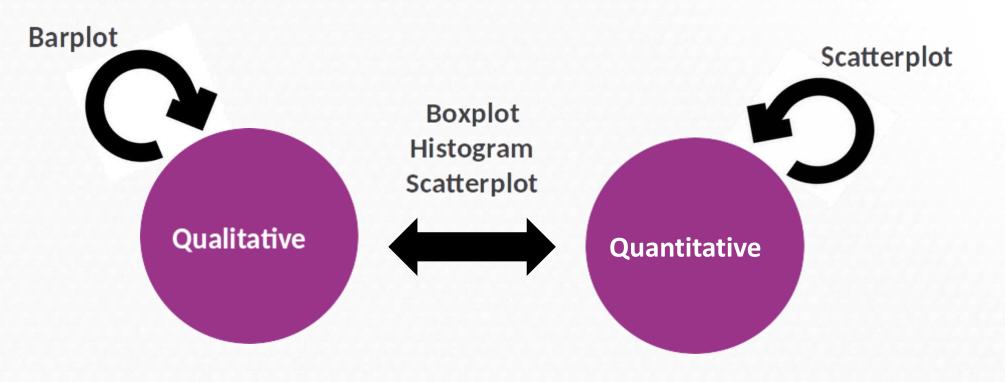


Figure 2. Scatter plot showing wife's age as a function of husband's age.

- The older the husband the older the wife.
- It is possible to know age of wives for an husband age.



Some plots to study the relationship between two variables...





os 🐺	teo									 	<
	registro	area	f_nac	edad	grupedad	peso talla	bua	imc	clasific	me	
1	3	10	11659420800	57	55 - 59	70.0 168.0	69	24.80159	OSTEOPENIA		^
2	4	10	11671689600	46	45 - 49	53.0 152.0	73	22.93975	OSTEOPENIA	l į	
3	10	10	11721024000	45	45 - 49	64.0 158.0	81	25.63692	NORMAL		
4	11	10	11464416000	53	50 - 54	78.0 161.0	58	30.09143	OSTEOPENIA		
5	12	10	11690784000	46	45 - 49	56.0 157.0	89	22.71897	NORMAL		
6	15	10	11716012800	45	45 - 49	63.5 170.0	76	21.97232	NORMAL		
7	16	10	11623737600	48	45 - 49	86.0 161.0	87	33.17773	NORMAL		
8	17	10	11562307200	50	50 - 54	61.5 164.0	74	22.86585	NORMAL		
9	18	10	11538028800	51	50 - 54	60.5 158.0	58	24.23490	OSTEOPENIA		
10	20	10	11332483200	57	55 - 59	64.0 149.0	61	28.82753	OSTEOPENIA		
11	21	10	11631945600	48	45 - 49	70.3 160.0	67	27.46094	OSTEOPENIA		
12	22	10	11425536000	55	55 - 59	74.4 160.0	68	29.06250	OSTEOPENIA		
13	23	10	11553235200	50	50 - 54	55.5 154.5	73	23.25070	OSTEOPENIA		
14	24	10	11367302400	56	55 - 59	89.0 166.0	61	32.29787	OSTEOPENIA		
15	25	10	11585635200	49	45 - 49	50.6 157.0	68	20.52822	OSTEOPENIA		
16	26	10	11572156800	50	50 - 54	71.4 152.0	74	30.90374	NORMAL		
17	27	10	11590992000	49	45 - 49	78.0 157.0	62	31.64429	OSTEOPENIA		
18	28	10	11293516800	58	55 - 59	72.0 162.0	65	27.43484	OSTEOPENIA		
19	29	10	11215238400	61	60 – 64	68.0 155.5	65	28.12212	OSTEOPENIA		
20	30	10	11405664000	55	55 - 59	75.0 161.0	92	28.93407	NORMAL		
21	31	10	11633155200	48	45 - 49	66.5 153.0	11	28.40788	OSTEOPOROSIS		
22	32	10	11287728000	59	55 - 59	101.0 156.0	82	41.50230	NORMAL		
23	34	10	10992758400	68	65 - 69	66.5 145.0	57	31.62901	OSTEOPENIA		
24	35	10	10909382400	69	65 – 69	70.0 168.0	48	24.80159	OSTEOPOROSIS		
25	36	10	11643868800	48	45 - 49	60.1 153.0	86	25.67389	NORMAL		
26	37	10	11551420800	50	50 - 54	67.0 159.0		26.50212	NORMAL		
27	38	10	11043907200	66	65 – 69	67.0 144.0		32.31096	NORMAL		
28	39	10	10948089600	69	65 – 69	70.5 148.5	40	31.96953	OSTEOPOROSIS		
29	40	10	11051251200	66	65 – 69	66.5 147.0	48	30.77421	OSTEOPOROSIS		
30	41	10	11333692800	57	55 - 59	58.5 142.0	80	29.01210	NORMAL		Y
	<	IIII								>	



The way to study the relation will depend on the variable types:

• Two qualitative variables: contingency table



Used for organizing categorical variables and testing hypothesis with the chisquared test for independence



The way to study the relation will depend on the variable types:

• Two qualitative variables: contingency table



Used for organizing categorical variables and testing hypothesis with the chisquared test for independence

 Count of individuals that simultaneously presents variable 1 (x) and variable 2 (y)

	<i>y</i> ₁	<i>y</i> ₁		y_p	n _{i.}
<i>X</i> ₁	n ₁₁	n ₁₂		n_{1p}	<i>n</i> _{1.}
<i>X</i> ₂	n ₁₁ n ₂₁	n_{22}	• • •	n_{2p}	$n_{2.}$
÷	:	÷	÷	: : : :	÷
x_k	n_{k1}	n_{k2}		n_{kp}	$n_{k.}$
$n_{.j}$	n. ₁	n. ₂		n. _p	Ν

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

	<i>y</i> ₁	<i>y</i> ₁		Уp	$f_{i.}$
<i>X</i> ₁	f ₁₁	f_{12}		f_{1p}	$f_{1.}$
<i>X</i> ₂	f ₂₁	f_{22}		f_{2p}	$f_{2.}$
:	:	:	:	:	:
X_k	f_{k1}	f_{k2}	· · · · · · · · · · · · · · · · · · ·	f_{kp}	f_k
$f_{i,j}$	f. ₁	f. ₂		f. _D	1
,				,-	

Absolute

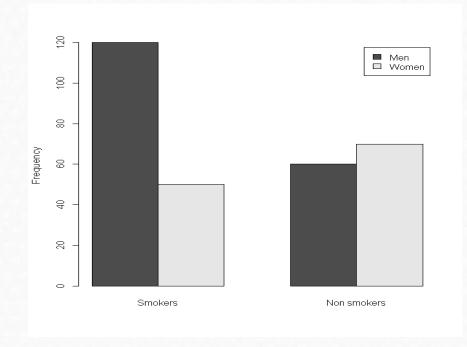
relative



An study wants to know if there are differences about smoking habits in men and women.

Gender	Smoking habits	
1	1	
2	1	
1	0	
1	0	
1	0	
1	1	
2	1	
•••		

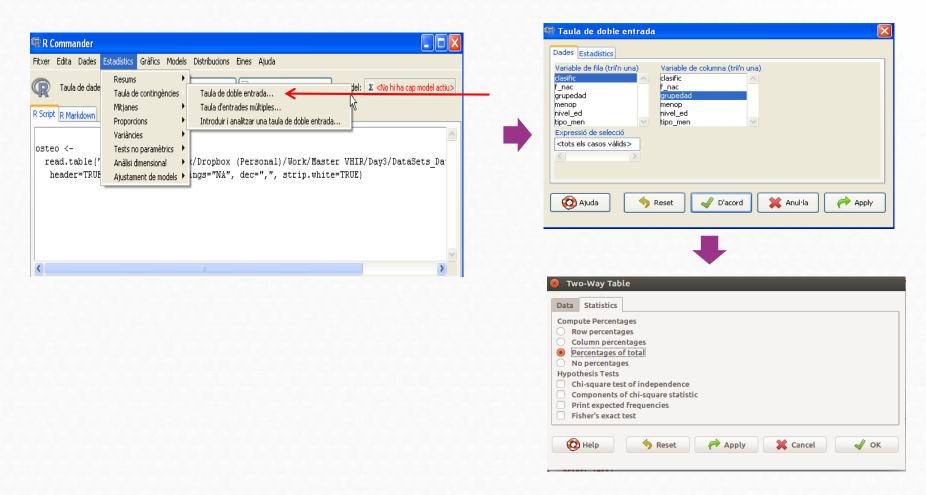
	Smokers	Non Smoking	Total
Men	120	60	180
Women	50	70	120
Total	170	130	300





Let's do in R Commander: Osteoporosis dataset

Study if the group age (grupedad) of patients, influence in the illness type (classific):





Frequency table:

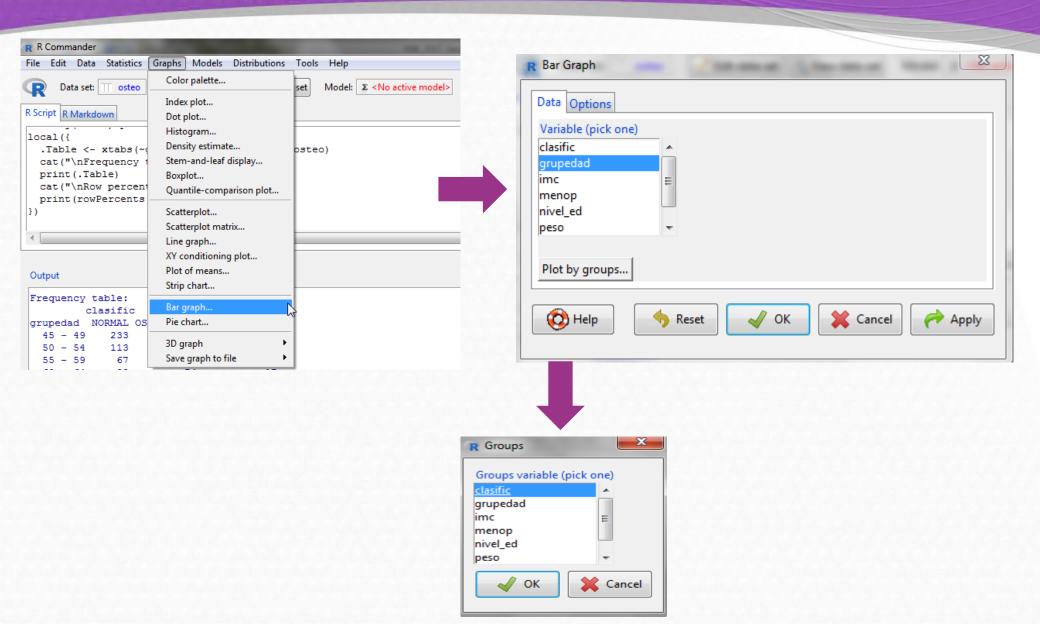
clasific

TEOPOROSIS	OSTEOPENIA	NORMAL	edad	grupe
7	138	233	- 49	45
7	113	113	- 54	50
9	100	67	- 59	55
17	74	38	- 64	60
24	42	18	- 69	65

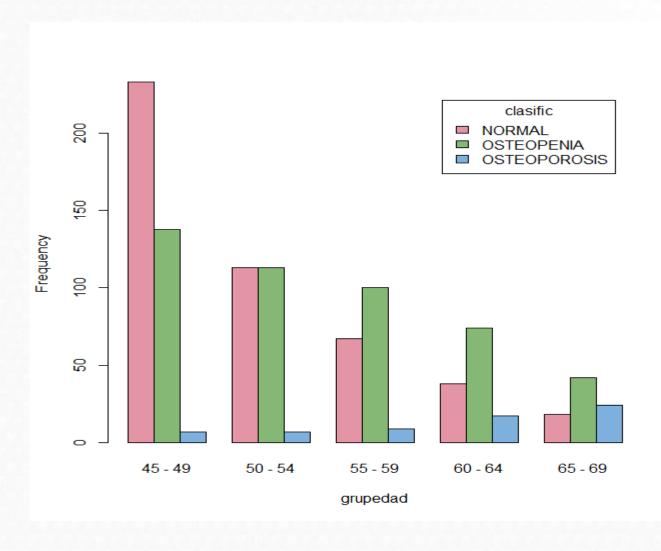
Total percentages:

	NORMAL	OSTEOPENIA	OSTEOPOROSIS	Total
45 - 49	23.3	13.8	0.7	37.8
50 - 54	11.3	11.3	0.7	23.3
55 - 59	6.7	10.0	0.9	17.6
60 - 64	3.8	7.4	1.7	12.9
65 - 69	1.8	4.2	2.4	8.4
Total	46.9	46.7	6.4	100.0





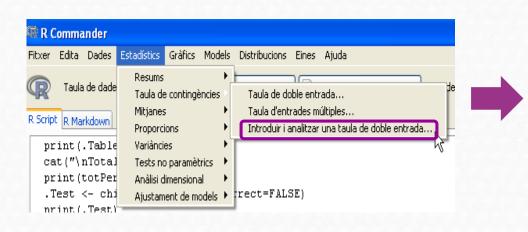


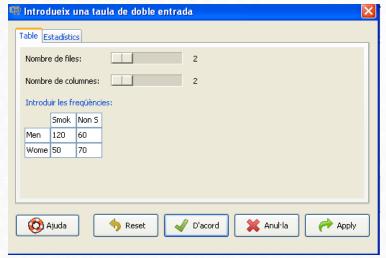




Another way to introduce the data:

	Smokers	Non Smoking	Total
Men	120	60	180
Women	50	70	120
Total	170	130	300







Another way to introduce the data:

	Smokers	Non Smoking	Total
Men	120	60	180
Women	50	0	120
Total	170	130	300

```
> .Table # Counts

1 2

1 120 60

2 50 70

> totPercents(.Table) # Percentage of Total

1 2 Total

1 40.0 20.0 60

2 16.7 23.3 40

Total 56.7 43.3 100
```



egistro area 3 10 4 10 10 10 11 10 12 10 15 10 16 10 17 10 18 10 20 10	11659420800 11671689600 11721024000 11464416000 11690784000 11716012800 11623737600 11562307200	edad 57 46 45 53 46 45 48 50	grupedad 55 - 59 45 - 49 45 - 49 45 - 49 45 - 49	peso talla 70.0 168.0 53.0 152.0 64.0 158.0 78.0 161.0 56.0 157.0 63.5 170.0 86.0 161.0	69 73 81 58 89 76	imc 24.80159 22.93975 25.63692 30.09143 22.71897 21.97232	clasific OSTEOPENIA OSTEOPENIA NORMAL OSTEOPENIA NORMAL NORMAL	me
4 10 10 10 11 10 12 10 15 10 16 10 17 10 18 10	11671689600 11721024000 11464416000 11690784000 11716012800 11623737600 11562307200	46 45 53 46 45 48	45 - 49 45 - 49 50 - 54 45 - 49 45 - 49	53.0 152.0 64.0 158.0 78.0 161.0 56.0 157.0 63.5 170.0	73 81 58 89 76	22.93975 25.63692 30.09143 22.71897 21.97232	OSTEOPENIA NORMAL OSTEOPENIA NORMAL	^
10 10 11 10 12 10 15 10 16 10 17 10 18 10	11721024000 11464416000 11690784000 11716012800 11623737600 11562307200	45 53 46 45 48	45 - 49 50 - 54 45 - 49 45 - 49 45 - 49	64.0 158.0 78.0 161.0 56.0 157.0 63.5 170.0	81 58 89 76	25.63692 30.09143 22.71897 21.97232	NORMAL OSTEOPENIA NORMAL	
11 10 12 10 15 10 16 10 17 10 18 10	11464416000 11690784000 11716012800 11623737600 11562307200	53 46 45 48	50 - 54 45 - 49 45 - 49 45 - 49	78.0 161.0 56.0 157.0 63.5 170.0	58 89 76	30.09143 22.71897 21.97232	OSTEOPENIA NORMAL	
12 10 15 10 16 10 17 10 18 10	11690784000 11716012800 11623737600 11562307200	46 45 48	45 - 49 45 - 49 45 - 49	56.0 157.0 63.5 170.0	89 76	22.71897 21.97232	NORMAL	
15 10 16 10 17 10 18 10	11716012800 11623737600 11562307200	45 48	45 - 49 45 - 49	63.5 170.0	76	21.97232		
16 10 17 10 18 10	11623737600 11562307200	48	45 - 49				NORMAL	
17 10 18 10	11562307200			86.0 161.0	87			
18 10		50	FO F4			33.17773	NORMAL	
	11538028800		50 - 54	61.5 164.0	74	22.86585	NORMAL	
20 10		51	50 - 54	60.5 158.0	58	24.23490	OSTEOPENIA	
	11332483200	57	55 - 59	64.0 149.0	61	28.82753	OSTEOPENIA	
21 10	11631945600	48	45 - 49	70.3 160.0	67	27.46094	OSTEOPENIA	
22 10	11425536000	55	55 - 59	74.4 160.0	68	29.06250	OSTEOPENIA	
23 10	11553235200	50	50 - 54	55.5 154.5	73	23.25070	OSTEOPENIA	
24 10	11367302400	56	55 - 59	89.0 166.0	61	32.29787	OSTEOPENIA	
25 10	11585635200	49	45 - 49	50.6 157.0	68	20.52822	OSTEOPENIA	
26 10	11572156800	50	50 - 54	71.4 152.0	74	30.90374	NORMAL	
27 10	11590992000	49	45 - 49	78.0 157.0	62	31.64429	OSTEOPENIA	
28 10	11293516800	58	55 - 59	72.0 162.0	65	27.43484	OSTEOPENIA	
29 10	11215238400	61	60 - 64	68.0 155.5	65	28.12212	OSTEOPENIA	
30 10	11405664000	55	55 - 59	75.0 161.0	92	28.93407	NORMAL	
31 10	11633155200	48	45 - 49	66.5 153.0	11	28.40788	OSTEOPOROSIS	
32 10	11287728000	59	55 - 59	101.0 156.0	82	41.50230	NORMAL	
34 10	10992758400	68	65 – 69	66.5 145.0	57	31.62901	OSTEOPENIA	
35 10	10909382400	69	65 – 69	70.0 168.0	48	24.80159	OSTEOPOROSIS	
36 10	11643868800	48	45 - 49	60.1 153.0	86	25.67389	NORMAL	
37 10	11551420800	50	50 - 54	67.0 159.0	105	26.50212	NORMAL	
38 10	11043907200	66	65 – 69	67.0 144.0	79	32.31096	NORMAL	
39 10	10948089600	69	65 – 69	70.5 148.5	40	31.96953	OSTEOPOROSIS	
40 10	11051251200	66	65 – 69	66.5 147.0	48	30.77421	OSTEOPOROSIS	
41 10	11333692800	57	55 - 59	58.5 142.0	80	29.01210	NORMAL	~
								>
	21 10 22 10 23 10 24 10 25 10 26 10 27 10 28 10 29 10 30 10 31 10 32 10 34 10 35 10 36 10 37 10 38 10 39 10 40 10 41 10	20 10 11332483200 21 10 11631945600 22 10 11425536000 23 10 11553235200 24 10 11367302400 25 10 11585635200 26 10 11572156800 27 10 11590992000 28 10 11293516800 29 10 11215238400 30 10 11405664000 31 10 11633155200 32 10 11287728000 34 10 10992758400 35 10 10992758400 36 10 11643868800 37 10 11551420800 38 10 11043907200 39 10 10948089600 40 10 11051251200 41 10 11333692800	20 10 11332483200 57 21 10 11631945600 48 22 10 11425536000 55 23 10 11553235200 50 24 10 11367302400 56 25 10 11585635200 49 26 10 11572156800 50 27 10 11590992000 49 28 10 11293516800 58 29 10 11215238400 61 30 10 11405664000 55 31 10 11633155200 48 32 10 11287728000 59 34 10 10992758400 68 35 10 10909382400 69 36 10 11643868800 48 37 10 11551420800 50 38 10 11043907200 66 39 10 10948089600 69 40 10 11051251200 66 41 <	20 10 11332483200 57 55 - 59 21 10 11631945600 48 45 - 49 22 10 11425536000 55 55 - 59 23 10 11553235200 50 50 - 54 24 10 11367302400 56 55 - 59 25 10 11585635200 49 45 - 49 26 10 11572156800 50 50 - 54 27 10 11590992000 49 45 - 49 28 10 11293516800 58 55 - 59 29 10 11215238400 61 60 - 64 30 10 11405664000 55 55 - 59 31 10 11633155200 48 45 - 49 32 10 11287728000 59 55 - 59 34 10 10992758400 68 65 - 69 35 10 1043868800 48 45<	20 10 11332483200 57 55 59 64.0 149.0 21 10 11631945600 48 45 49 70.3 160.0 22 10 11425536000 55 55 59 74.4 160.0 23 10 11553235200 50 50 54 55.5 154.5 24 10 11367302400 56 55 59 89.0 166.0 25 10 11585635200 49 45 49 50.6 157.0 26 10 11572156800 50 50 54 71.4 152.0 27 10 11590992000 49 45 49 78.0 157.0 28 10 11293516800 58 55 59 72.0 162.0 29 10 11215238400 61 60 64 68.0 155.5 30 10 11405664000 55 55 59 75.0 161.0 31 10 11633155200 48 <td>20 10 11332483200 57 55 - 59 64.0 149.0 61 21 10 11631945600 48 45 - 49 70.3 160.0 67 22 10 11425536000 55 55 - 59 74.4 160.0 68 23 10 11553235200 50 50 - 54 55.5 154.5 73 24 10 11367302400 56 55 - 59 89.0 166.0 61 25 10 11585635200 49 45 - 49 50.6 157.0 68 26 10 11572156800 50 50 - 54 71.4 152.0 74 27 10 11590992000 49 45 - 49 78.0 157.0 62 28 10 11215238400 61 60 - 64 68.0 155.5 65 30 10 11405664000 55 55 -</td> <td>20 10 11332483200 57 55 - 59 64.0 149.0 61 28.82753 21 10 11631945600 48 45 - 49 70.3 160.0 67 27.46094 22 10 11425536000 55 55 - 59 74.4 160.0 68 29.06250 23 10 11553235200 50 - 54 55.5 154.5 73 23.25070 24 10 11367302400 56 55 - 59 89.0 166.0 61 32.29787 25 10 11585635200 49 45 - 49 50.6 157.0 68 20.52822 26 10 11570156800 50 50 - 54 71.4 152.0 74 30.90374 27 10 11590992000 49 45 - 49 78.0 157.0 62 31.64429 28 10 11293516800 58 55 - 59 72.0</td> <td>20 10 11332483200 57 55 - 59 64.0 149.0 61 28.82753 OSTEOPENIA 21 10 11631945600 48 45 - 49 70.3 160.0 67 27.46094 OSTEOPENIA 22 10 11425536000 55 55 - 59 74.4 160.0 68 29.06250 OSTEOPENIA 23 10 11553235200 50 50 - 54 55.5 154.5 73 23.25070 OSTEOPENIA 24 10 11367302400 56 55 - 59 89.0 166.0 61 32.29787 OSTEOPENIA 25 10 11585635200 49 45 - 49 50.6 157.0 68 20.52822 OSTEOPENIA 26 10 11572156800 50 50 - 54 71.4 152.0 74 30.90374 NORMAL 27 10 11590992000 49 45 - 49 78.0 157.0 62 31.64429 OSTEOPENIA 28 10 11293516800 58 55 - 59 72.0 162.0 65 27.43484 OSTEOPENIA 29 10 11215238400 61 60 - 64 68.0 155.5 65 28.12212 OSTEOPENIA 30 10 11405664000 55 55 - 59 75.0 161.0 92 28.93407 NORMAL 31 10 11633155200 48 45 - 49 66.5 153.0 11 28.40788 OSTEOPOROSIS 32 10 11287728000 59 55 - 59 101.0 156.0 82 41.50230 NORMAL 34 10 10992758400 68 65 - 69 66.5 145.0 57 31.62901 OSTEOPENIA 35 10 10999382400 69 65 - 69 70.0 168.0 48 24.80159 OSTEOPOROSIS 36 10 11643868800 48 45 - 49 60.1 153.0 86 25.67389 NORMAL 37 10 11551420800 50 50 - 54 67.0 159.0 105 26.50212 NORMAL 39 10 10948089600 69 65 - 69 67.0 144.0 79 32.31096 NORMAL 39 10 10948089600 69 65 - 69 67.0 144.0 79 32.31096 NORMAL 39 10 10948089600 69 65 - 69 66.5 147.0 48 30.77421 OSTEOPOROSIS 40 10 11051251200 66 65 - 69 66.5 147.0 48 30.77421 OSTEOPOROSIS</td>	20 10 11332483200 57 55 - 59 64.0 149.0 61 21 10 11631945600 48 45 - 49 70.3 160.0 67 22 10 11425536000 55 55 - 59 74.4 160.0 68 23 10 11553235200 50 50 - 54 55.5 154.5 73 24 10 11367302400 56 55 - 59 89.0 166.0 61 25 10 11585635200 49 45 - 49 50.6 157.0 68 26 10 11572156800 50 50 - 54 71.4 152.0 74 27 10 11590992000 49 45 - 49 78.0 157.0 62 28 10 11215238400 61 60 - 64 68.0 155.5 65 30 10 11405664000 55 55 -	20 10 11332483200 57 55 - 59 64.0 149.0 61 28.82753 21 10 11631945600 48 45 - 49 70.3 160.0 67 27.46094 22 10 11425536000 55 55 - 59 74.4 160.0 68 29.06250 23 10 11553235200 50 - 54 55.5 154.5 73 23.25070 24 10 11367302400 56 55 - 59 89.0 166.0 61 32.29787 25 10 11585635200 49 45 - 49 50.6 157.0 68 20.52822 26 10 11570156800 50 50 - 54 71.4 152.0 74 30.90374 27 10 11590992000 49 45 - 49 78.0 157.0 62 31.64429 28 10 11293516800 58 55 - 59 72.0	20 10 11332483200 57 55 - 59 64.0 149.0 61 28.82753 OSTEOPENIA 21 10 11631945600 48 45 - 49 70.3 160.0 67 27.46094 OSTEOPENIA 22 10 11425536000 55 55 - 59 74.4 160.0 68 29.06250 OSTEOPENIA 23 10 11553235200 50 50 - 54 55.5 154.5 73 23.25070 OSTEOPENIA 24 10 11367302400 56 55 - 59 89.0 166.0 61 32.29787 OSTEOPENIA 25 10 11585635200 49 45 - 49 50.6 157.0 68 20.52822 OSTEOPENIA 26 10 11572156800 50 50 - 54 71.4 152.0 74 30.90374 NORMAL 27 10 11590992000 49 45 - 49 78.0 157.0 62 31.64429 OSTEOPENIA 28 10 11293516800 58 55 - 59 72.0 162.0 65 27.43484 OSTEOPENIA 29 10 11215238400 61 60 - 64 68.0 155.5 65 28.12212 OSTEOPENIA 30 10 11405664000 55 55 - 59 75.0 161.0 92 28.93407 NORMAL 31 10 11633155200 48 45 - 49 66.5 153.0 11 28.40788 OSTEOPOROSIS 32 10 11287728000 59 55 - 59 101.0 156.0 82 41.50230 NORMAL 34 10 10992758400 68 65 - 69 66.5 145.0 57 31.62901 OSTEOPENIA 35 10 10999382400 69 65 - 69 70.0 168.0 48 24.80159 OSTEOPOROSIS 36 10 11643868800 48 45 - 49 60.1 153.0 86 25.67389 NORMAL 37 10 11551420800 50 50 - 54 67.0 159.0 105 26.50212 NORMAL 39 10 10948089600 69 65 - 69 67.0 144.0 79 32.31096 NORMAL 39 10 10948089600 69 65 - 69 67.0 144.0 79 32.31096 NORMAL 39 10 10948089600 69 65 - 69 66.5 147.0 48 30.77421 OSTEOPOROSIS 40 10 11051251200 66 65 - 69 66.5 147.0 48 30.77421 OSTEOPOROSIS



The way to study the relation will depend on the variable types:

• One qualitative variable and one quantitative variable: Table of statistics

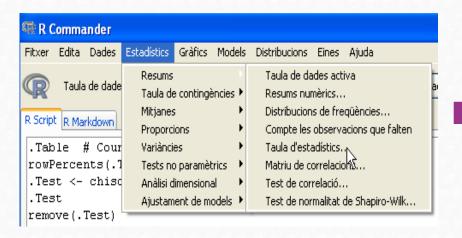


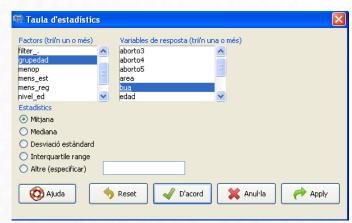
Mean value of the variable in each category for each individual



Let's do in R Commander: Osteoporosis dataset

Study if bone density (bua) how change in each group of age

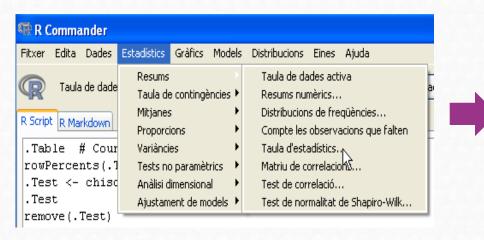






Let's do in R Commander: Osteoporosis dataset

Study if bone density (bua) how change in each group of age



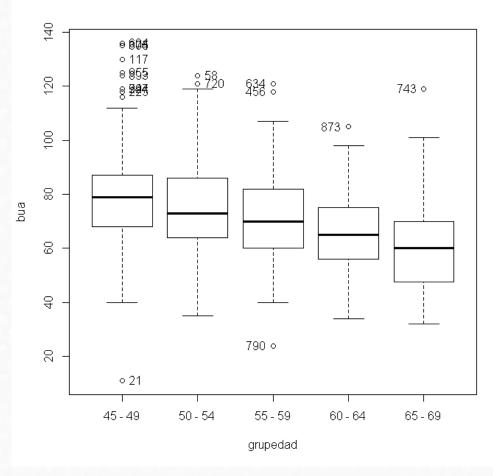


```
> with(osteo, tapply(bua, list(grupedad), mean, na.rm=TRUE))
45-49 50-54 55-59 60-64 65-69
78.75926 75.05150 71.43182 64.89147 60.66667
```



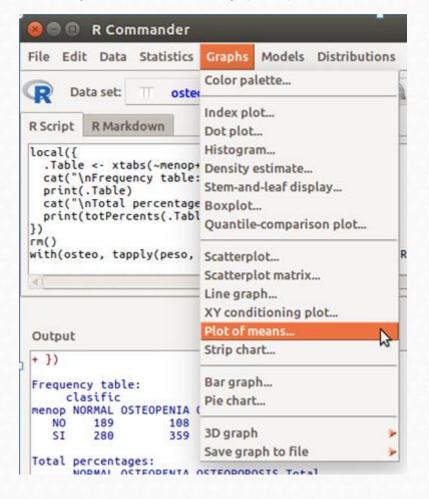
Study if bone density (bua) is different in each group of age

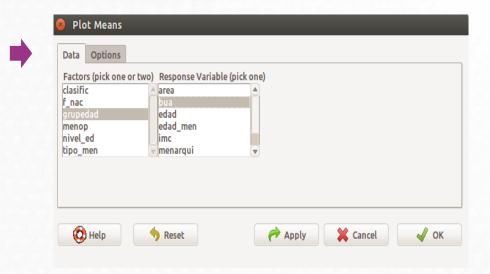






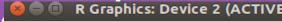
Study if bone density (bua) is different in each group of age





2. Bivariate analysis

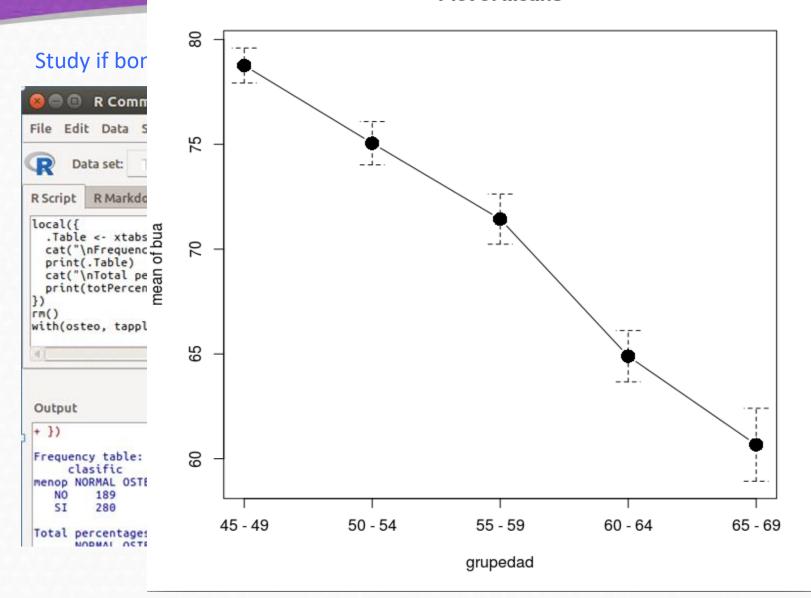
R Graphics: Device 2 (ACTIVE)





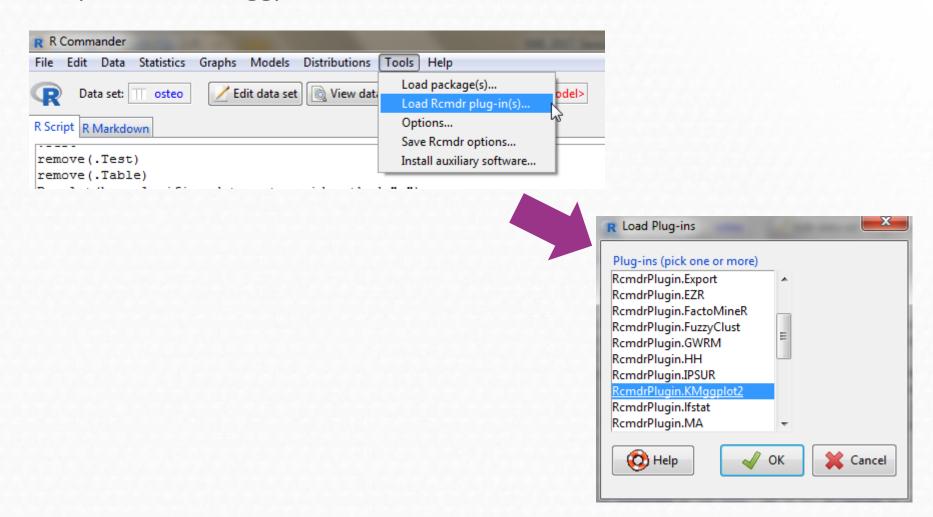
√ ok

Plot of Means



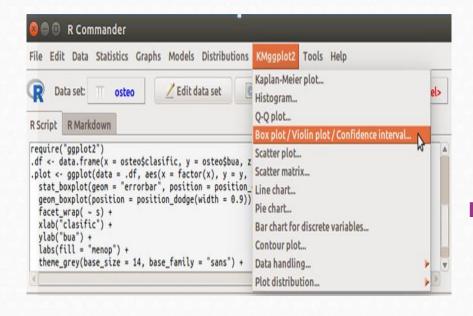


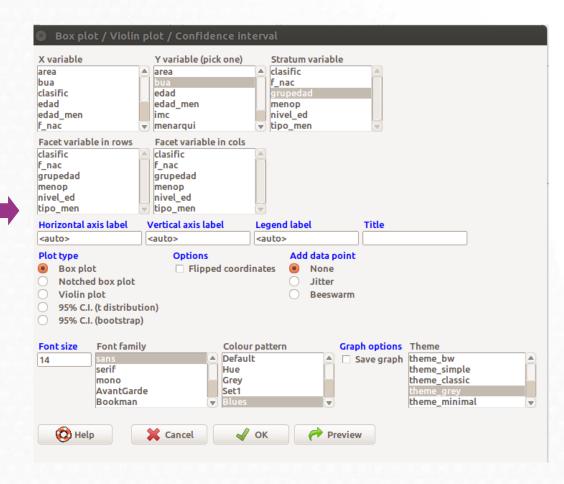
Graphics with KMggplots2:

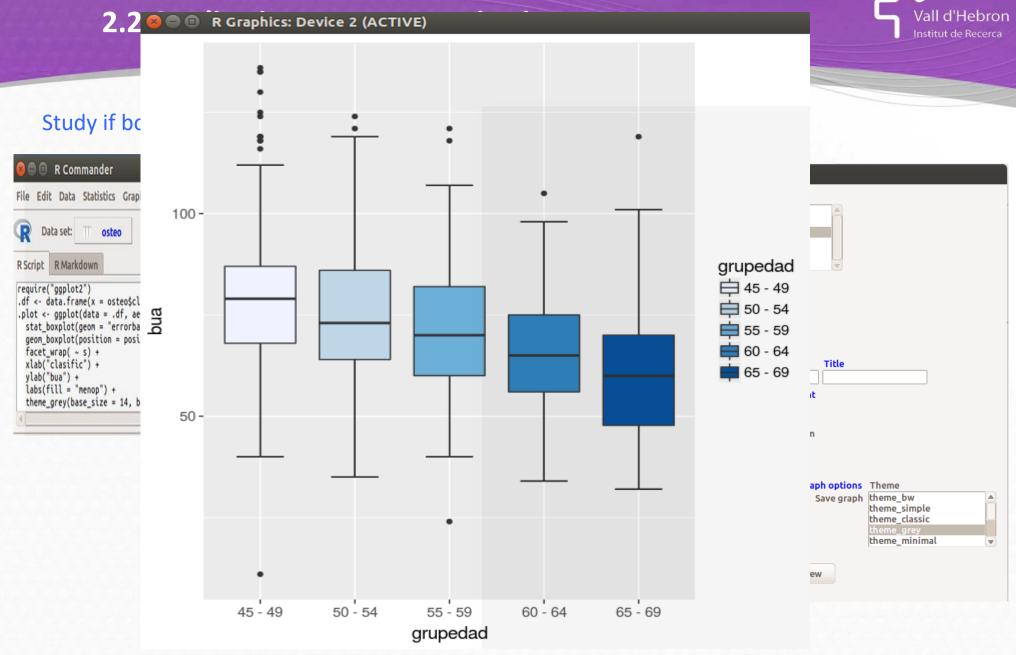




Study if bone density (bua) is different in each group of age (with KMggplot2)

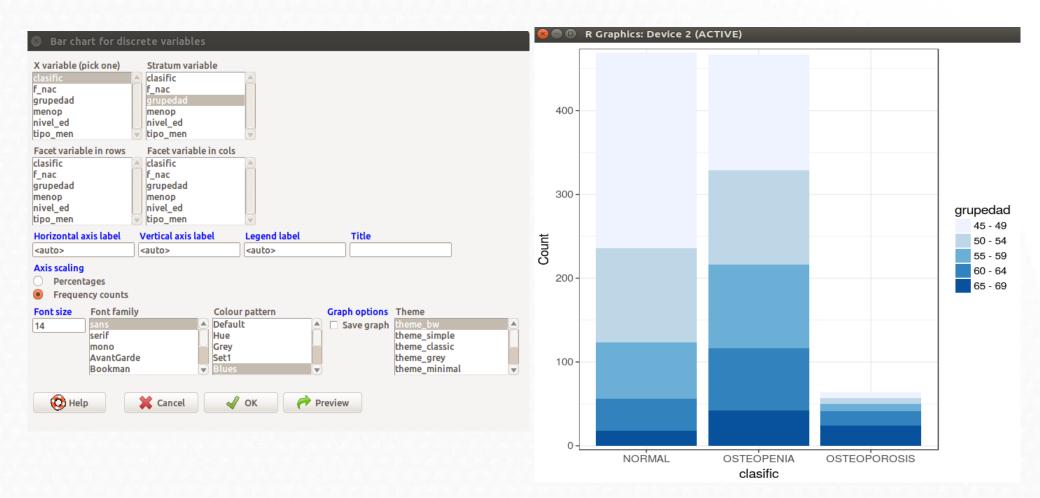








Study if grupedad, influence in the illness type (classific) (with KMgglplots2)



2. Bivariate analysis



Exercise

Study if the relationship between menop and group of illness (classific)

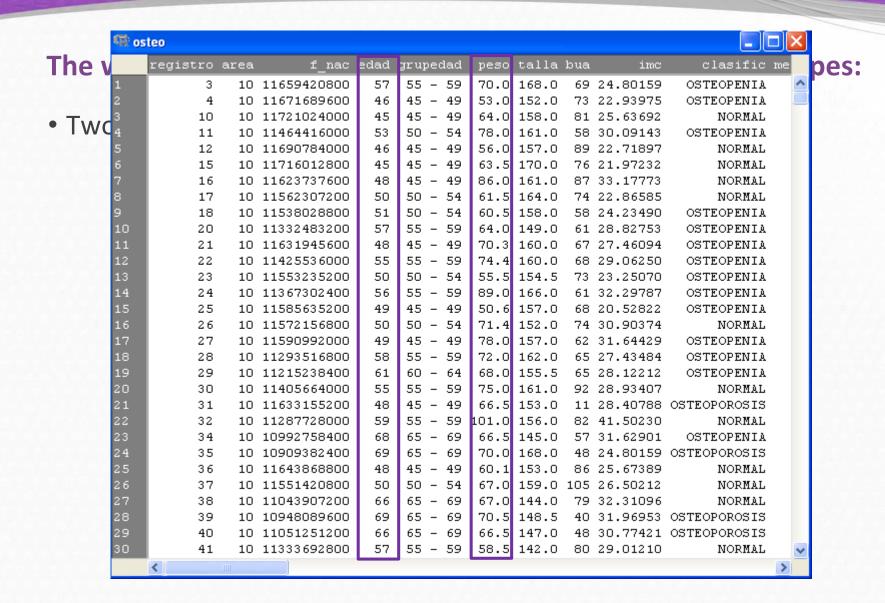
Study if peso is different in each group of illness (classific).



The way to study the relation will depend on the variable types:

• Two quantitatives variables:

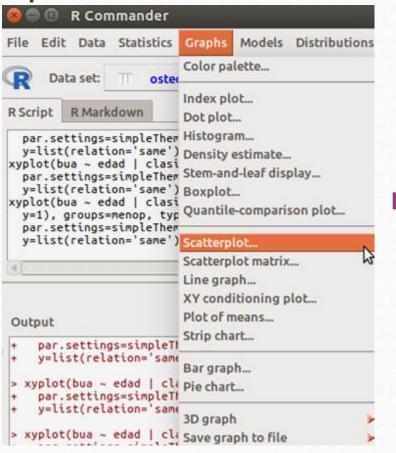


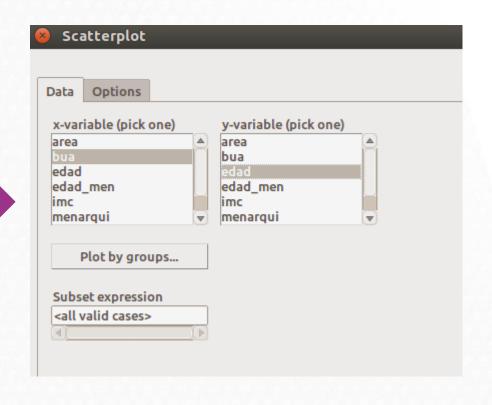




The way to study the relation will depend on the variable types:

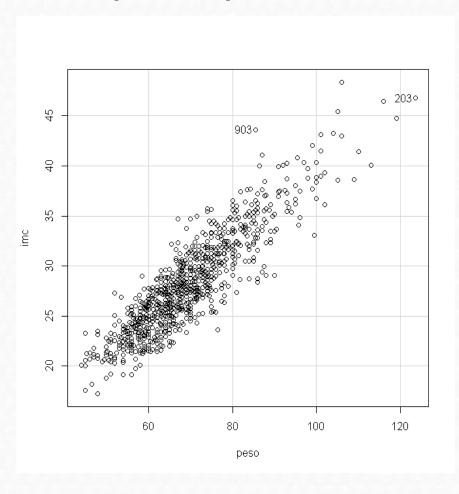
• Two quantitatives variables





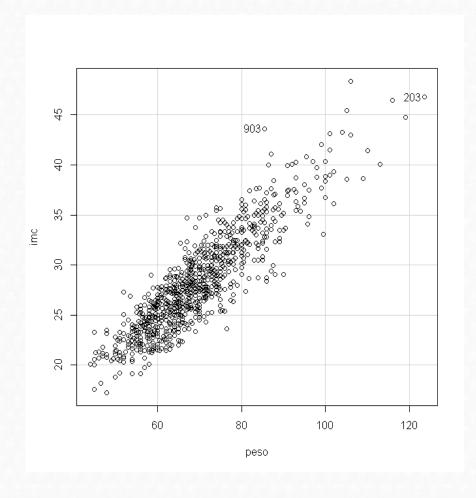


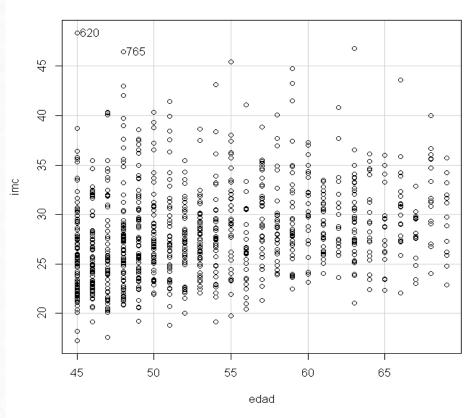
The way to study the relation will depend on the variable types:





The way to study the relation will depend on the variable types:

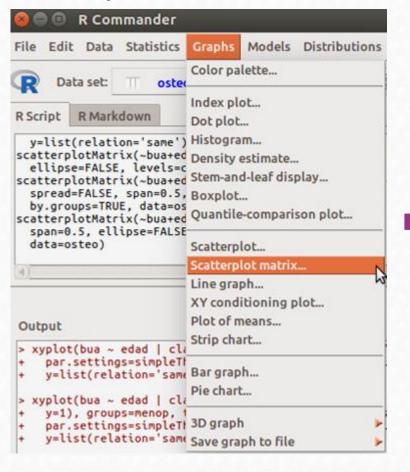


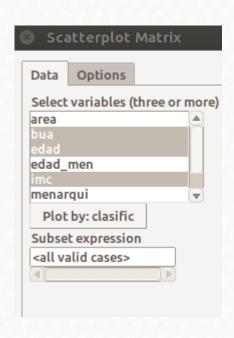




The way to study the relation will depend on the variable types:

• Two quantitatives variables:





2.3

Vall d'Hebron Institut de Recerca

The war

• Two q

File Edit Data

P Data set:

R Script R Mar

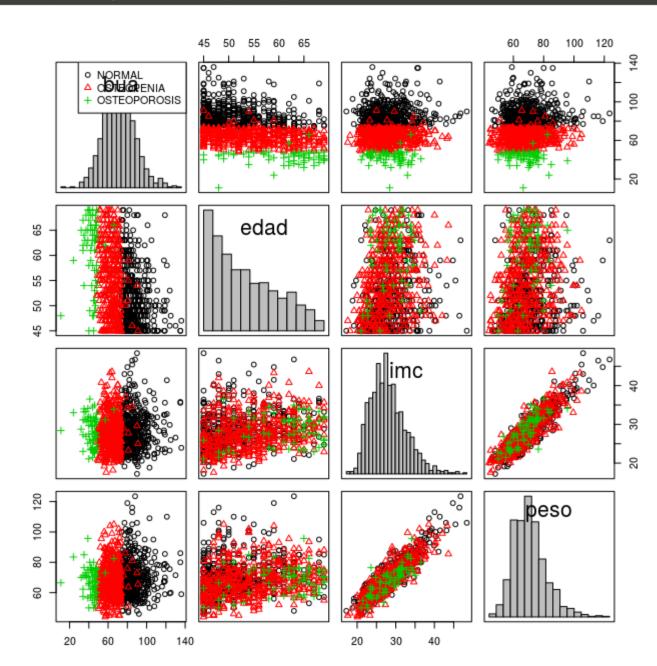
y=list(relat scatterplotMat ellipse=FALS scatterplotMat spread=FALSE by.groups=TR scatterplotMat span=0.5, el data=osteo)

Output

> xyplot(bua + par.settir
+ y=list(rel
> xyplot(bua -

> xyplot(bua -+ y=1), grou + par.settir

y=list(rel



pes:



TABLE OF CONTENTS

1. From univariate to bivariate analysis

2. Bivariate analysis

- 1. Qualitative vs Qualitative
- 2. Qualitative vs Quantitative
- 3. Quantitative vs Quantitative

3. Correlation

- 1. Definition
- 2. Types of correlation (Pearson, Spearman)

3. Correlation1. Definition



Main characteristics of correlation analysis:

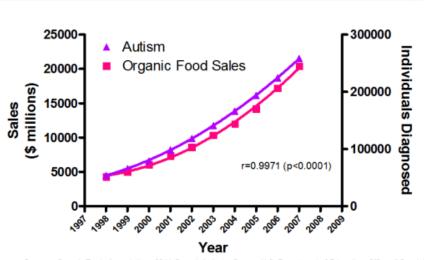
- Correlation analysis allow:
 - Study the way of relation between the two variables
 - Quantify the intensity of relation
- Correlation is not causation one thing does not causes the other
- In the correlation analysis, the two variables have the same weigh
- The correlation coefficient measures the strength of the relation

3. Correlation1. Definition

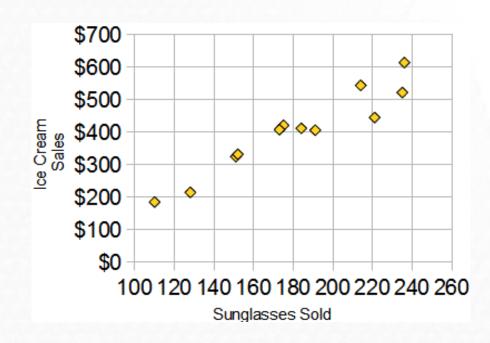


Main characteristics of correlation analysis:

Correlation is not causation



Sources: Organic Trade Association, 2011 Organic Industry Survey, U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS), OM B# 1820-0043: "Children with Disabilities Receiving Special Education Under Part B of the Individuals with Disabilities Education Act



2. Types of correlation



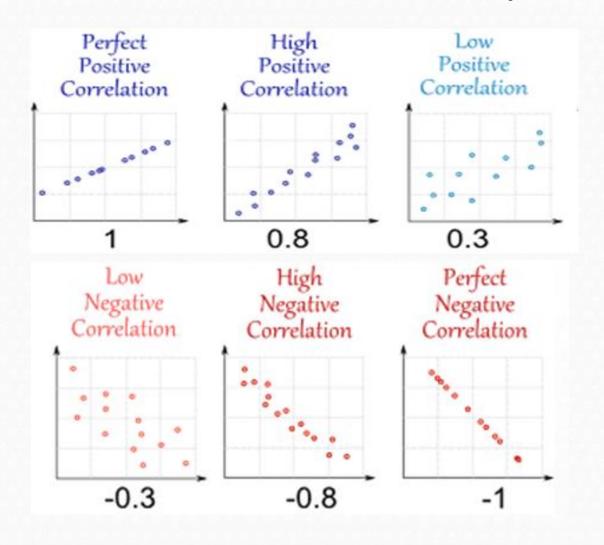
Pearson correlation coefficient

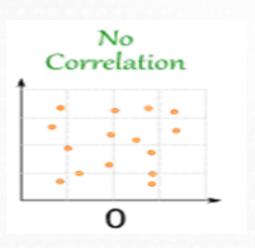
- It is represented by letter "r". It has no dimensions (no units)
- Values go from -1 to +1
 - >r=0 indicates no linear relation between the variables
 - >r>0 indicates direct relation between the variables
 - >r<0 indicates indirect relation between the variables
 - >r=1/-1 indicates a perfect relation between the variables

2. Types of correlation



Pearson correlation coefficient. Examples



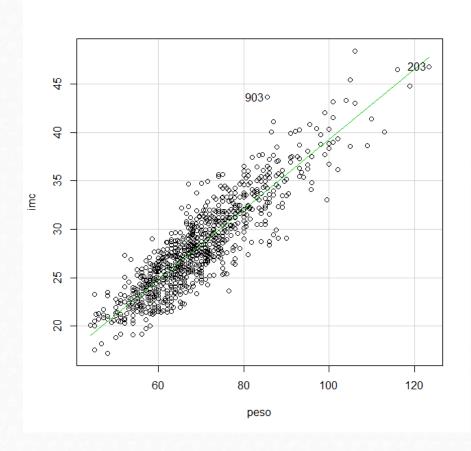


3. Correlation2. Types of correlation



Study the relationship between peso and body mass index (imc):

imc peso imc 1.0000000 0.8927967 peso 0.8927967 1.0000000

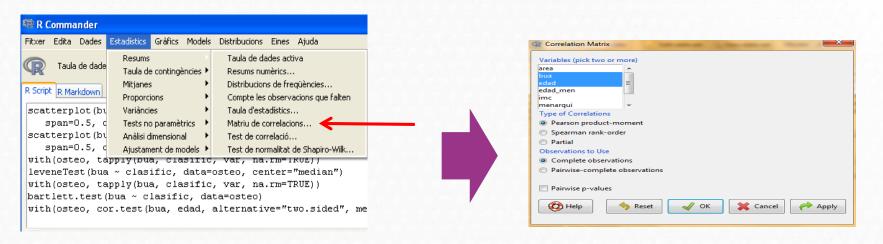


2. Types of correlation



Pearson correlation coefficient. How to in R-commander?

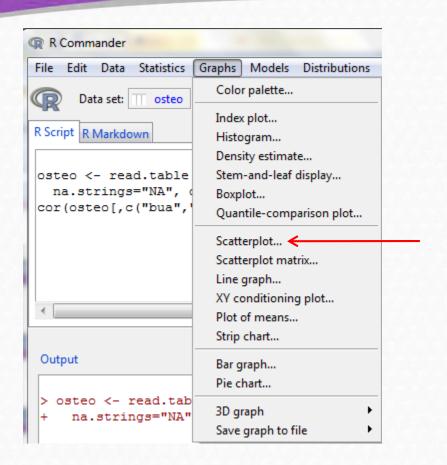
Bone density and **age** are correlated?



Don't forget to look the graphic!!

2. Types of correlation





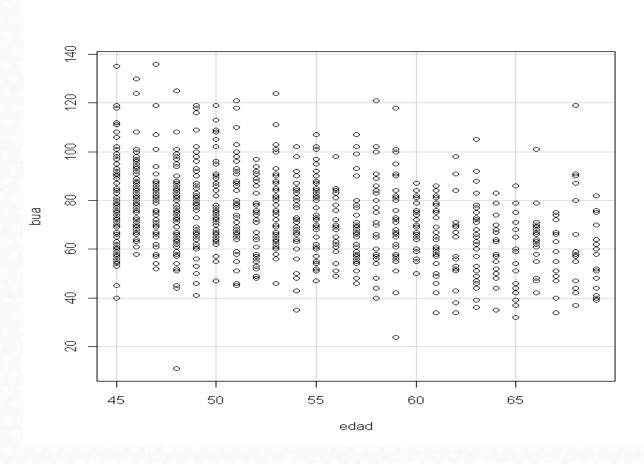


© Scatterplot	K	J
Data Options x-variable (pick one) aborto3 aborto4 aborto5 area bua edad edad Plot by groups Data Options y-variable (pick one) aborto4 aborto5 area bua edad edad edad Plot by groups	X	
Subset expression <all cases="" valid=""> Help Reset OK Cancel Apply</all>		

2. Types of correlation



Pearson correlation coefficient. How to in R-commander?



3. Correlation2. Types of correlation



Pearson correlation coefficient. How to in R-commander?

Exercise 1. Do you think that exists a relationship between *peso* and *talla*? What type of relationship? Show a scatterplot of the values.

2. Types of correlation



Non Parametric correlation: Spearman correlation coefficient

 Pearson correlation coefficient is severely affected by outliers and if the relation is not lineal



Better to use **Spearman** correlation coefficient (use the ranks between the numbers instead the values) to calculate the correlation coefficient

• Evaluates the monotonic relationship between the variables (not the linear relationship as Pearson does).

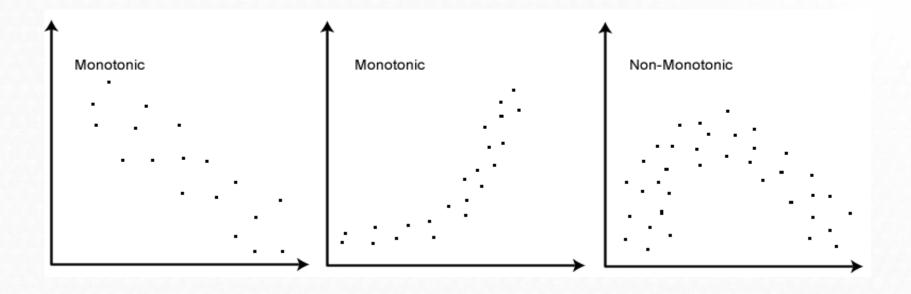


The variables tend to change together but not necessarily at a constant rate

3. Correlation2. Types of correlation



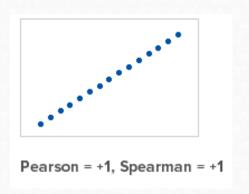
Non Parametric correlation: Spearman correlation coefficient

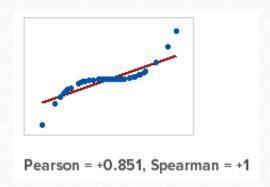


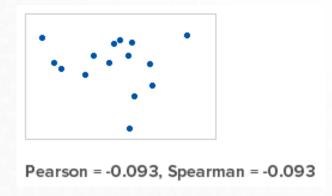
2. Types of correlation

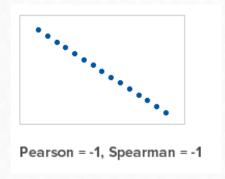


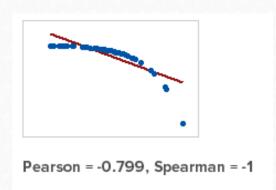
Comparison of Pearson and Spearman coefficients.











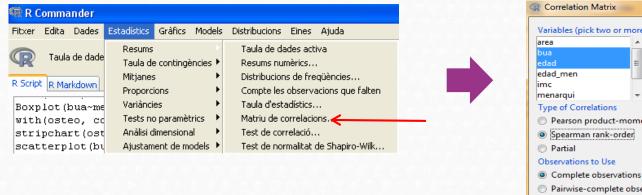


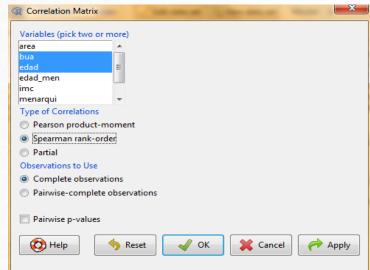
Always examine a scatterplot to determine the form of the relationship

3. Correlation2. Types of correlation



Spearman correlation coefficient. How to in R-commander?





3. Correlation. Exercises



Exercise 2. An hypothetic study, published last year that exists a relation between *age* and *systolic blood pressure (sbp)*? Do you think is it true? Show a scatterplot of the values? If not, find another variable in the dataset that has a good correlation with *systolic blood pressure*.

Use dataset Framingham250.csv