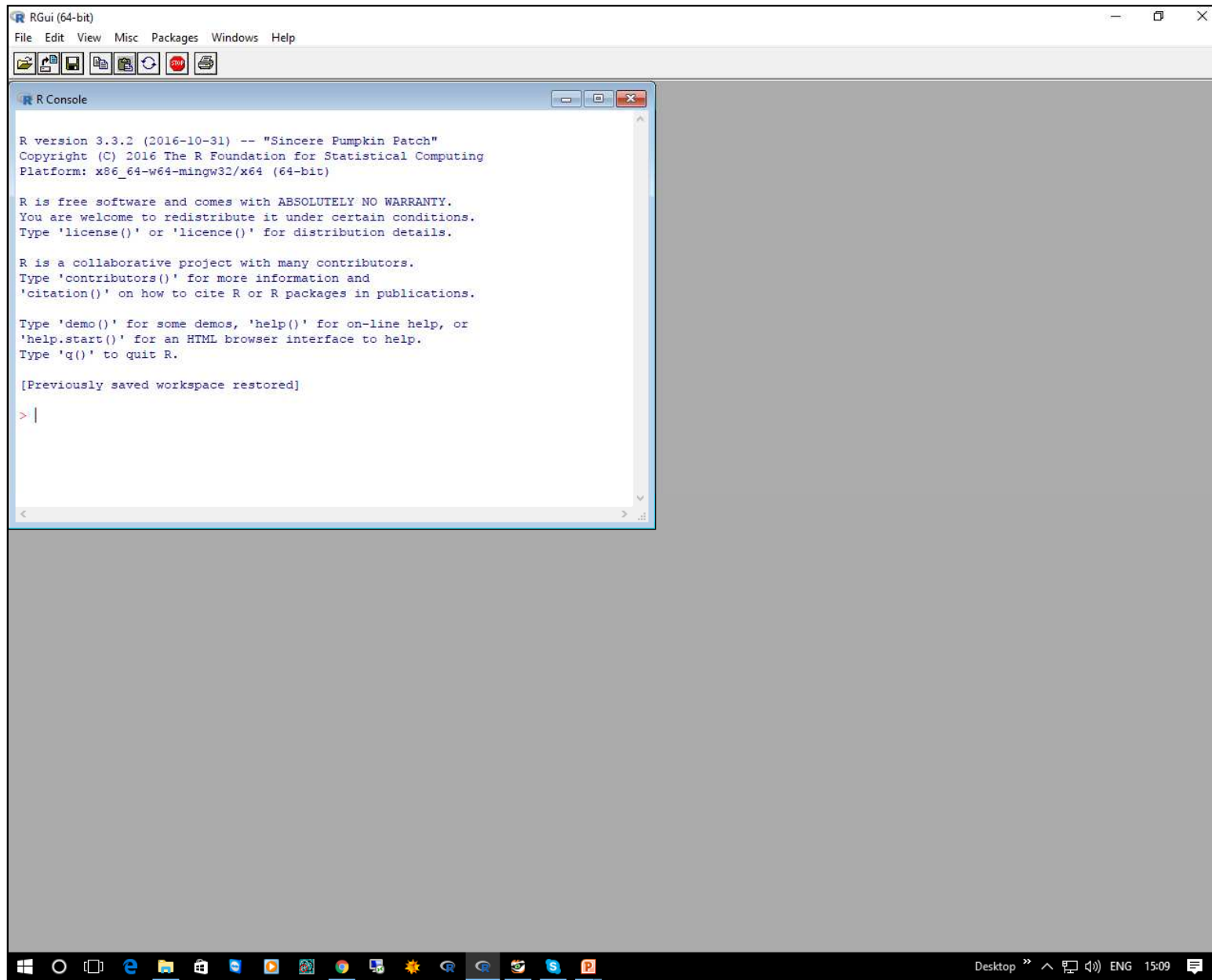


Statistica

Statistica descrittiva

GRAFICI e INDICI

Facciamo un salto in



Facciamo un salto in



RGui (64-bit)

File Edit View Misc Packages Windows Help

R Console

```
R version 3.3.2 (2016-10-31) -- "Sincere Pumpkin Patch"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> |
```

PER COMODITA',
CREIAMO SUL DESKTOP
UNA CARTELLA
"corsoR"

Desktop ^ ENG 15:09

Facciamo un salto in



The screenshot shows the RGui (64-bit) interface. The R Console window displays the following text:

```
R version 3.3.2 (2016-10-31) -- "Sincere Pumpkin Patch"  
Copyright (C) 2016 The R Foundation for Statistical Computing  
Platform: x86_64-w64-mingw32/x64 (64-bit)  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
  
[Previously saved workspace restored]  
  
> 2+2  
[1] 4  
> |
```

A callout box highlights the output of the command `2+2`, showing the result `[1] 4`.

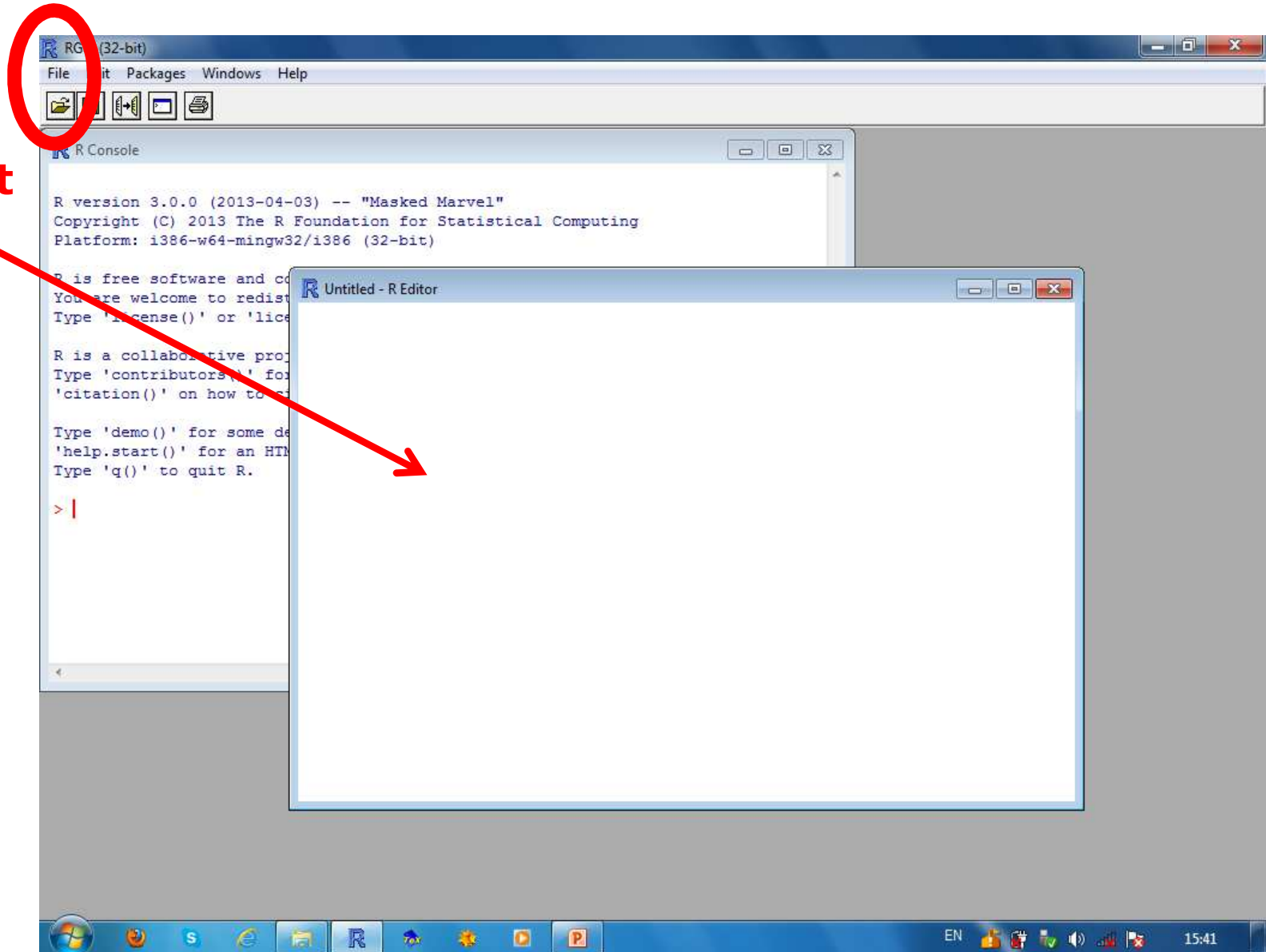
```
> 2+2  
[1] 4  
> |
```

The Windows taskbar at the bottom shows the system tray with the date and time: Desktop, ENG, 15:10.

Facciamo un salto in



File->New script



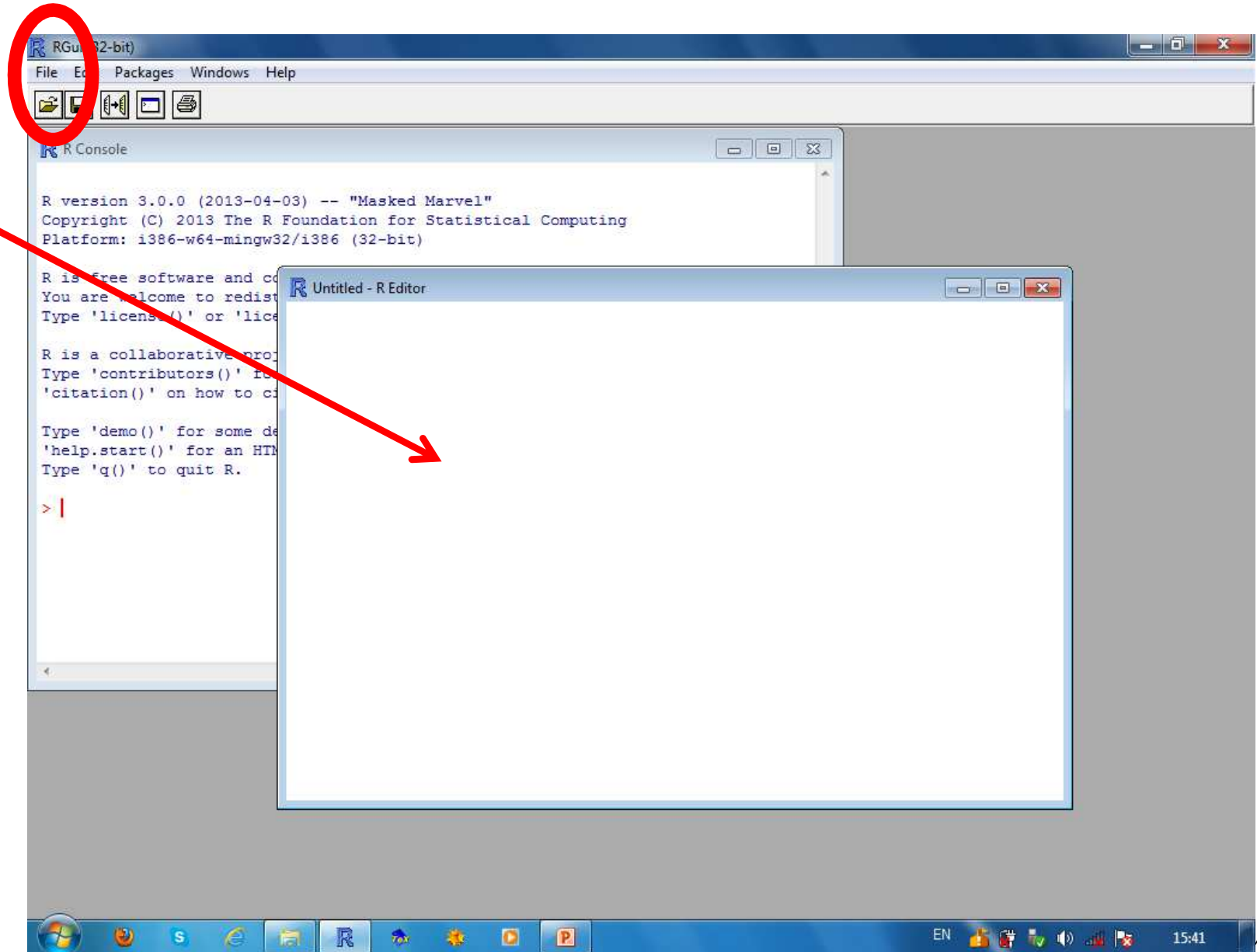
Facciamo un salto in



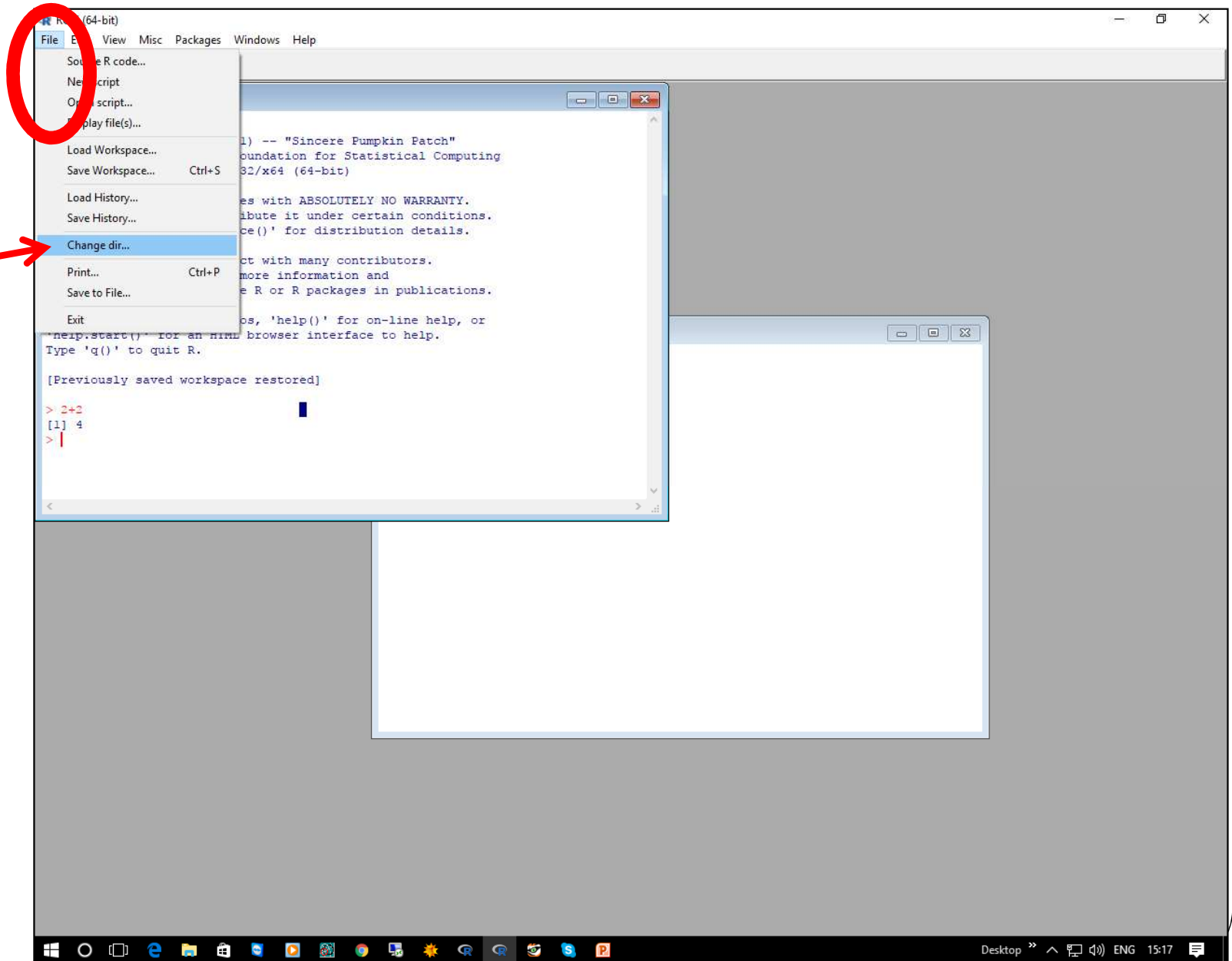
File->New script

File->Save

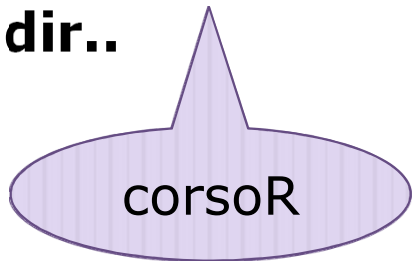
corsoR



Facciamo un salto in

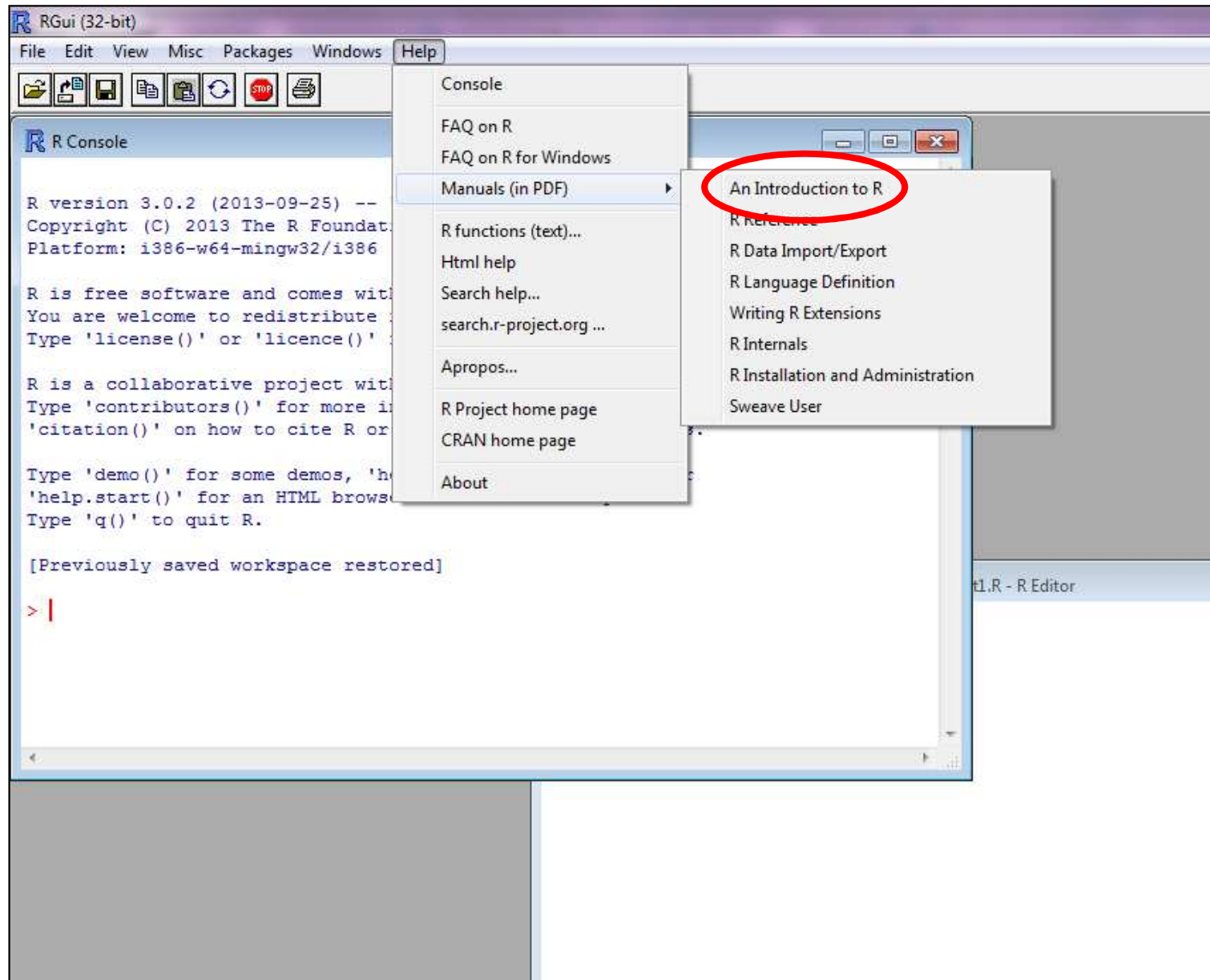


File-> Change dir..



corsoR

Facciamo un salto in



Facciamo un salto in



RGui (64-bit)

File Edit View Misc Packages Windows Help

R Console

```
R version 3.3.2 (2016-10-31) -- "Pumpkin Patch"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help,
'help.start()' for an HTML browser interface to help,
Type 'q()' to quit R.

[Previously saved workspace restored]

> |
```

The R Language

127.0.0.1:16837/doc/html/index.html

Statistical Data Analysis

Manuals

- [An Introduction to R](#)
- [Writing R Extensions](#)
- [R Data Import/Export](#)
- [The R Language Definition](#)
- [R Installation and Administration](#)
- [R Internals](#)

Reference

- [Search Engine & Keywords](#)

Miscellaneous Material

- [About R](#)
- [License](#)
- [NEWS](#)
- [Authors](#)
- [Frequently Asked Questions](#)
- [User Manuals](#)
- [Resources](#)
- [Thanks](#)
- [Technical papers](#)

Material specific to the Windows port

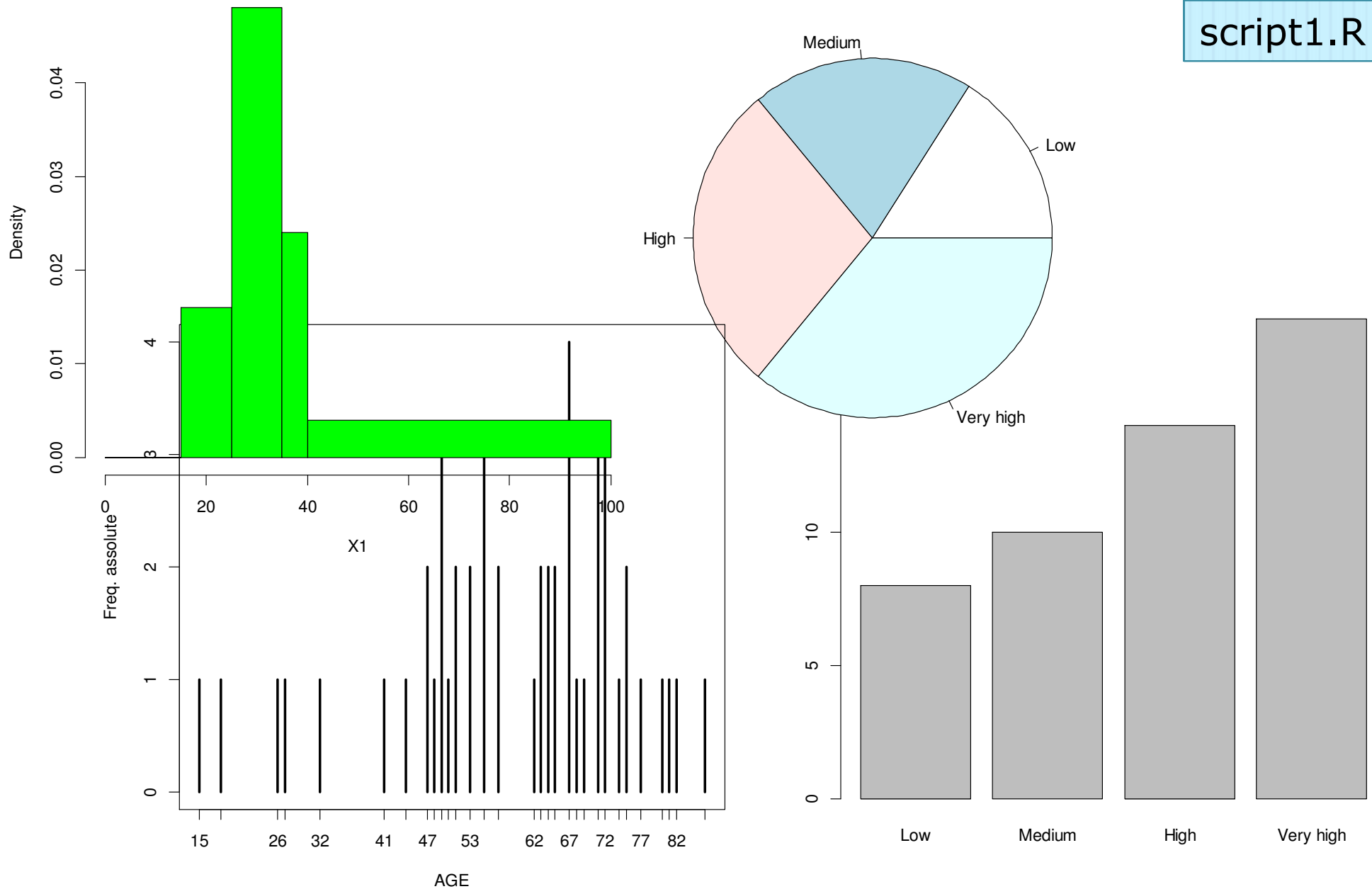
- [CHANGES up to R 2.15.0](#)
- [Windows FAQ](#)

da menu
Help:
Html help

Facciamo un salto in



script1.R



Facciamo un salto in



**Come diavolo faccio a sapere
il nome di un comando di R?**



Facciamo un salto in



“Istogramma con r” ecc.

<https://cran.r-project.org/doc/contrib/Frascati-FormularioStatisticaR.pdf>



Indici di centralità: **moda**

	Età (y)	Peso (kg)	Altezza (m)	Sesso	Malattia
1	19	50.2	1.65	F	Cardiov.
2	22	75.6	1.78	M	Diabete
3	21	80.1	1.91	M	Diabete
4	23	56.7	1.72	M	Cardiov.
5	22	75.0	1.81	M	Polmonare
6	20	58.3	1.68	F	Ipert.

Indici di centralità: moda

Il genere M compare più spesso di F

Delle 4 malattie considerate, due hanno frequenza maggiore (Cardiov. e Diabete)

Sesso	Malattia
F	Cardiov.
M	Diabete
M	Diabete
M	Cardiov.
M	Polmonare
F	Ipert.

Indici di centralità: moda

Nel Genere la **moda** è M

Nella Malattia ci sono **due mode**:
Cardiov. e Diabete (**bi-modalità**)

Sesso	Malattia
F	Cardiov.
M	Diabete
M	Diabete
M	Cardiov.
M	Polmonare
F	Ipert.

Indici di centralità: moda

Voto all'esame: da 18 a 30L

Rendimento:

18-22: suff.

23-26: buono

27-30L: ottimo

**Il voto BUONO è
la moda del
rendimento**

Rendimento	n_i	$f_i = n_i/n$	% = $100f_i$
S	35	0.233	23.3 %
B	62	0.413	41.3 %
O	53	0.354	35.4 %
tot.	150	1	100 %

Indici di centralità: moda

CARATTERE/
VARIABILE

FREQUENZE

num. figli	n_i		
0	5		
1	5		
2	3		
3	3		
4	4		
tot.	20		

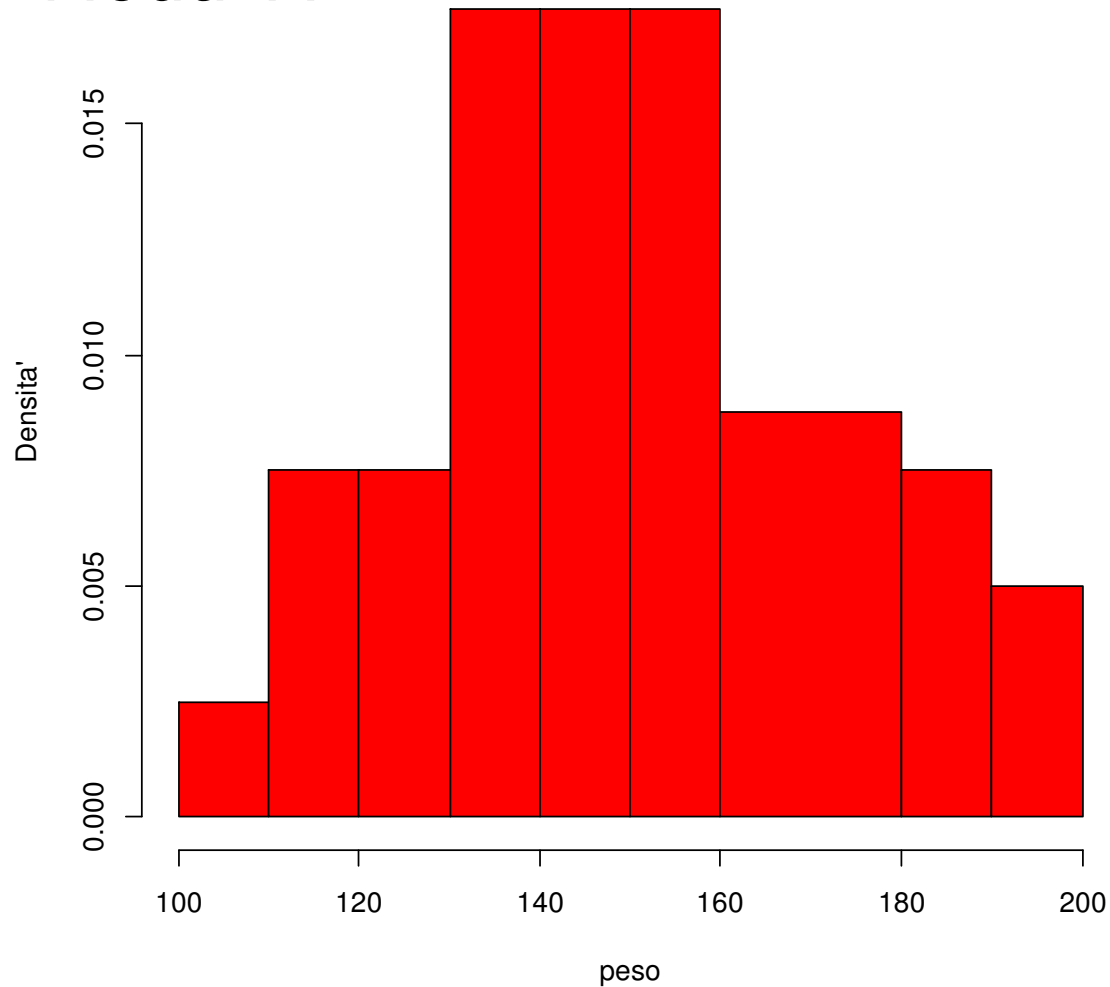
**Entrambi i
valori 0 e 1
sono una moda:
*bi-modalità***

Indici di centralità: mo

Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192

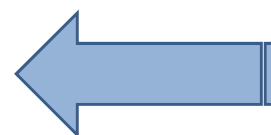
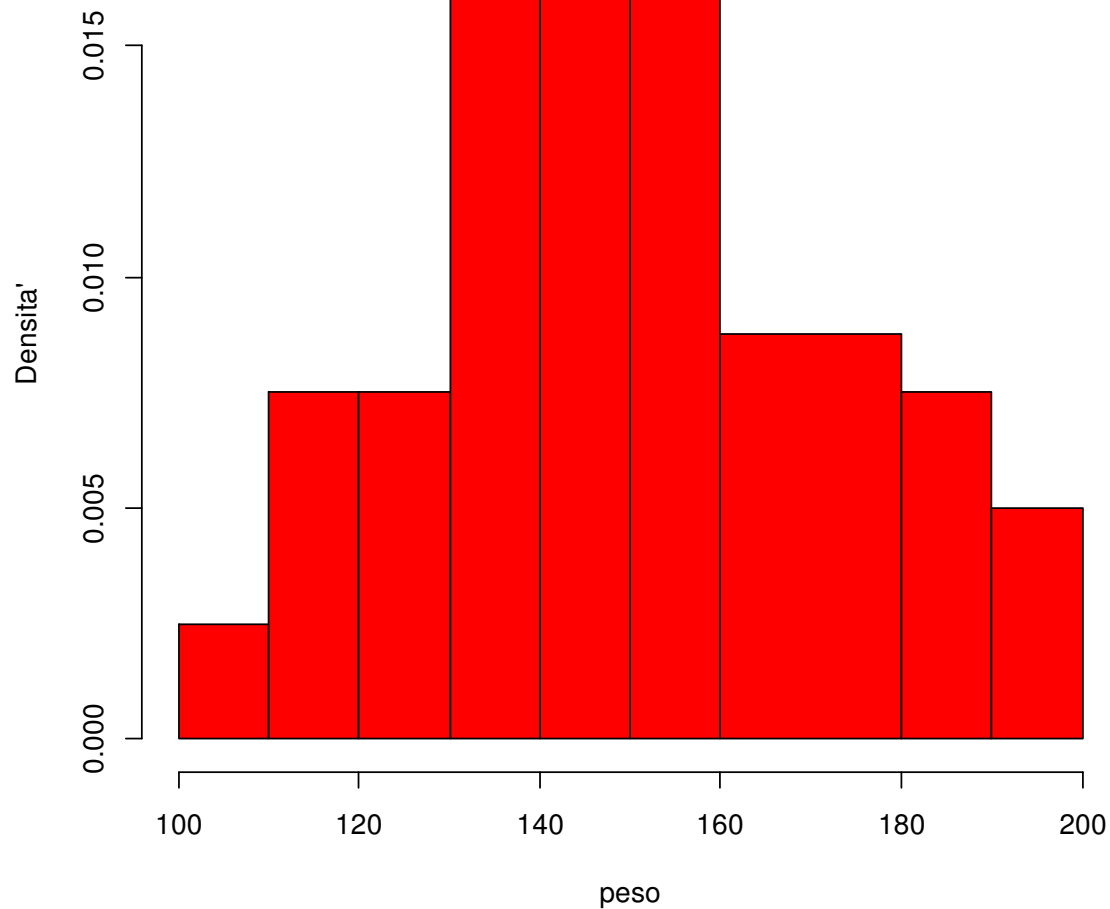
Moda ??

Moda ??



Indici di centralità: mo

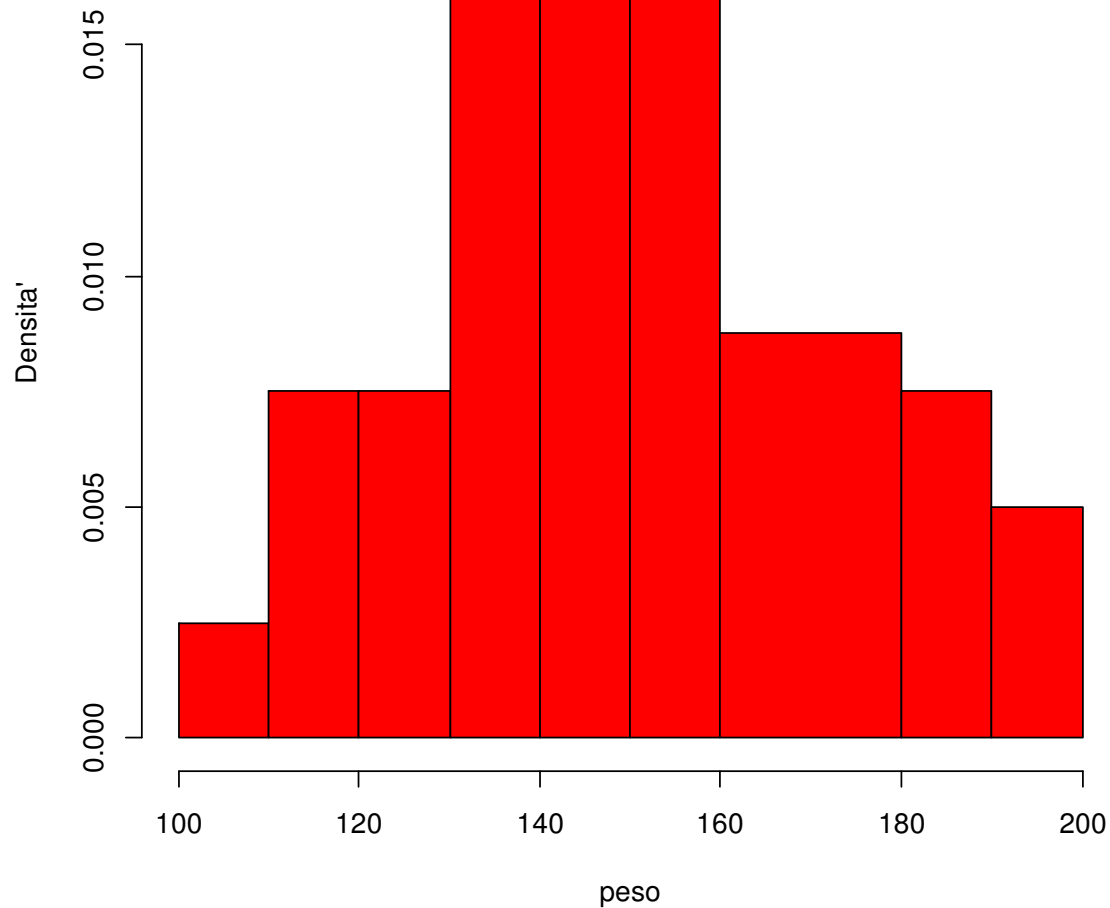
Moda



Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192

Indici di centralità: mo

Tre classi
modali



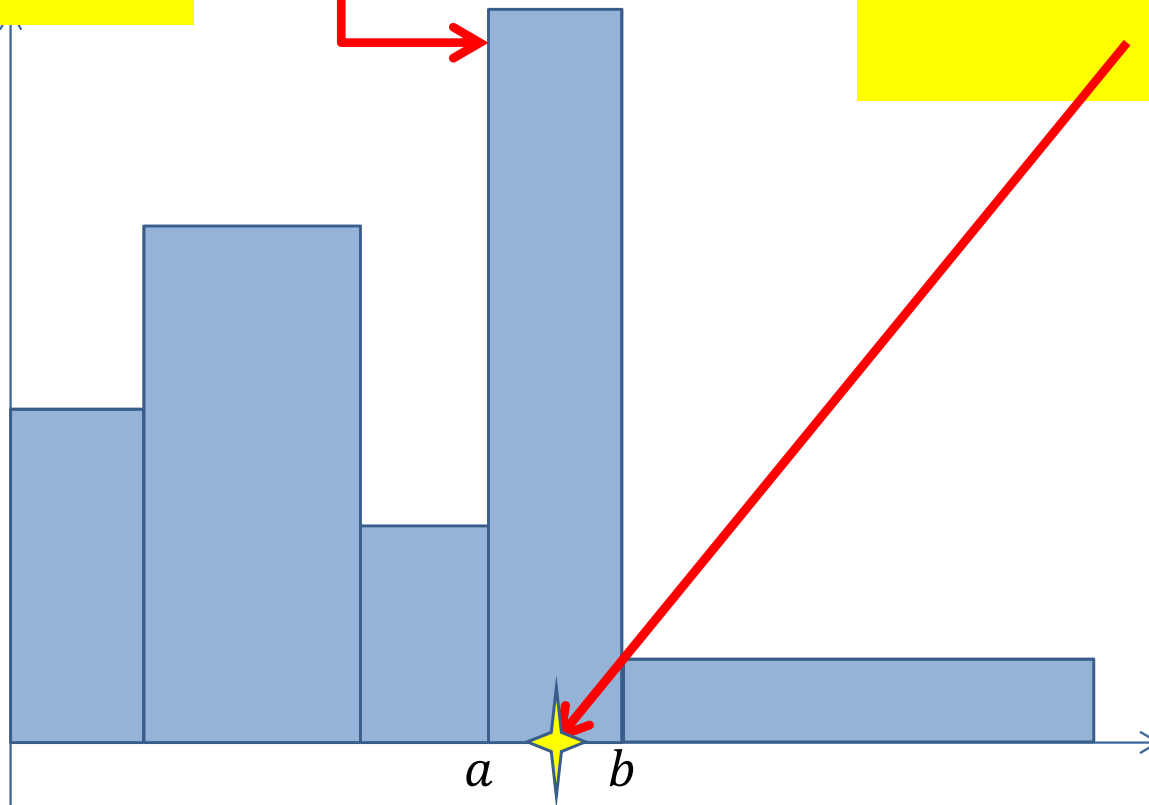
Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192

Indici di centralità: moda

**Classe/i
modale:** con
la **densità**
maggiore:
 $(a, b]$

Moda: valore
centrale della/e
classe modale:

$$\frac{a + b}{2}$$



Indici di centralità: **media**

	Età (y)	Peso (kg)	Altezza (m)	Sesso	Malattia
1	19	50.2	1.65	F	Cardiov.
2	22	75.6	1.78	M	Diabete
3	21	80.1	1.91	M	Diabete
4	23	56.7	1.72	M	Cardiov.
5	22	75.0	1.81	M	Polmonare
6	20	58.3	1.68	F	Ipert.

Indici di centralità: media

$$\frac{19 + 22 + 21 + 23 + 22 + 20}{6} = 21.17 \quad \text{anni}$$

	Età (y)	Peso (kg)	Altezza (m)	Sesso	Malattia
1	19	50.2	1.65	F	Cardiov.
2	22	75.6	1.78	M	Diabete
3	21	80.1	1.91	M	Diabete
4	23	56.7	1.72	M	Cardiov.
5	22	75.0	1.81	M	Polmonare
6	20	58.3	1.68	F	Ipert.

Indici di centralità: media

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

	Età (y)	Peso (kg)	Altezza (m)	Sesso	Malattia
1	19	50.2	1.65	F	Cardiov.
2	22	75.6	1.78	M	Diabete
3	21	80.1	1.91	M	Diabete
4	23	56.7	1.72	M	Cardiov.
5	22	75.0	1.81	M	Polmonare
6	20	58.3	1.68	F	Ipert.

La media aritmetica

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{x} = 151.05$$

Unità	Peso
1	118
2	151
3	143
4	172
5	147
6	146
7	138
8	175
9	134
10	172
11	118
12	151
13	155
14	155
15	146
16	135
17	127
18	178
19	136
20	180
21	151
22	186
23	122
24	132
25	114
26	171
27	140
28	187
29	106
30	159
31	127
32	191
33	192
34	181
35	143
36	153
37	144
38	139
39	148
40	179

X : n. figli

x_i	n_i
0	4
1	6
2	4
3	1

$$n = 15$$

Tabella di
frequenze

La media aritmetica

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{x} = 151.05$$

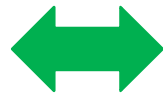
Unità	Peso
1	118
2	151
3	143
4	172
5	147
6	146
7	138
8	175
9	134
10	172
11	118
12	151
13	155
14	155
15	146
16	135
17	127
18	178
19	136
20	180
21	151
22	186
23	122
24	132
25	114
26	171
27	140
28	187
29	106
30	159
31	127
32	191
33	192
34	181
35	143
36	153
37	144
38	139
39	148
40	179

X : n. figli

x_i	n_i
0	4
1	6
2	4
3	1

$n = 15$

**Tabella di
frequenze**



Unità	n. figli
1	0
2	0
3	0
4	0
5	1
6	1
7	1
8	1
9	1
10	1
11	2
12	2
13	2
14	2
15	3

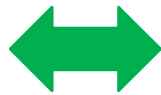
La media aritmetica

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

X : n. figli

x_i	n_i
0	4
1	6
2	4
3	1

$n = 15$



Unità	n. figli
1	0
2	0
3	0
4	0
5	1
6	1
7	1
8	1
9	1
10	1
11	2
12	2
13	2
14	2
15	3

$$\bar{x} = \frac{0 + 0 + 0 + 0 + 1 + 1 + \dots + 2 + 2 + 3}{15}$$

$$= \frac{0 \times 4 + 1 \times 6 + 2 \times 4 + 3 \times 1}{15} =$$

$$= \frac{1}{n} \sum_{i=1}^K x_i \times n_i = \sum_{i=1}^K x_i \times f_i$$

$$\left(f_i = \frac{n_i}{n} \right)$$

$$\bar{x} = 1.13$$

La media aritmetica

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

x_i	n_i
Diab	2
CV	2
Polm	1
Ipert	1

$$n = 6$$

Tabella di
frequenze

Diab.+Diab.+....

non senso!

La **moda** si può **sempre** determinare, perchè si calcola usando solo le frequenze;

la **media aritmetica** si calcola solo per variabili **quantitative.**

Indici di centralità: **mediana**

quale dato sta *in mezzo*?

	Età (y)	Peso (kg)	Altezza (m)	Sesso	Malattia
1	19	50.2	1.65	F	Cardiov.
2	22	75.6	1.78	M	Diabete
3	21	80.1	1.91	M	Diabete
4	23	56.7	1.72	M	Cardiov.
5	22	75.0	1.81	M	Polmonare
6	20	58.3	1.68	F	Ipert.

Indici di centralità: mediana

quale dato sta *in mezzo*?

	Età (y)	Età ordinata			
1	19	19			
2	22	20			
3	21	21			
4	23	22			
5	22	22			
6	20	23			

Indici di centralità: mediana

quale dato sta *in mezzo*?

	Età (y)	Età ordinata			
1	19	19	metà dei dati minore di 21.5 e metà maggiore di 21.5		
2	22	20			
3	21	21			
4	23	22			
5	22	22			
6	20	23			

Indici di centralità: mediana

quale dato sta *in mezzo*? **mediana: 21.5**

	Età (y)	Età ordinata			
1	19	19	metà dei dati minore di 21.5 e metà maggiore di 21.5		
2	22	20			
3	21	21			
4	23	22			
5	22	22			
6	20	23			

Indici di centralità: me

dati : x_1, x_2, \dots, x_n

$$n = 40$$

dati **riordinati**:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

$$x_{(20)} = 147$$

$$x_{(21)} = 148$$

Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192



Indici di centralità: me

dati : x_1, x_2, \dots, x_n

$$n = 40$$

dati **riordinati**:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

$$x_{(20)} = 147$$

$$x_{(21)} = 148$$

n pari :

La **posizione** della **mediana** è

$$\frac{n + 1}{2} \quad (20.5)$$

Il **valore** della **mediana** è
(**LA mediana**)

$$\frac{x_{(n/2)} + x_{(n/2+1)}}{2} \quad 147.5$$

Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192

147.5



Indici di centralità: mediana

dati : x_1, x_2, \dots, x_n

$$n = 11$$

dati **riordinati**:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

n dispari :

La **posizione** della **mediana** è

$$\frac{n + 1}{2} \quad (6)$$

Il **valore** della **mediana** è
(**LA mediana**)

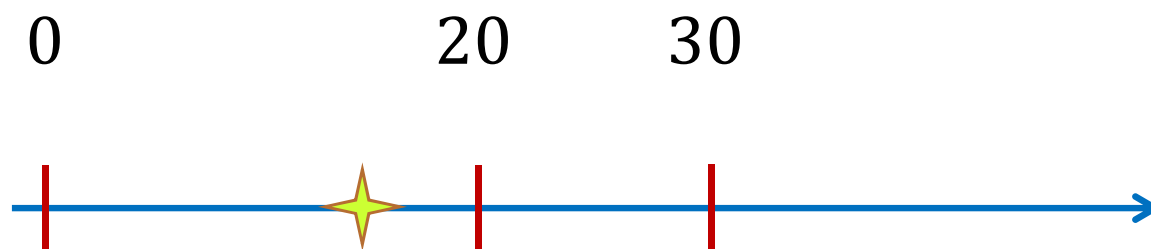
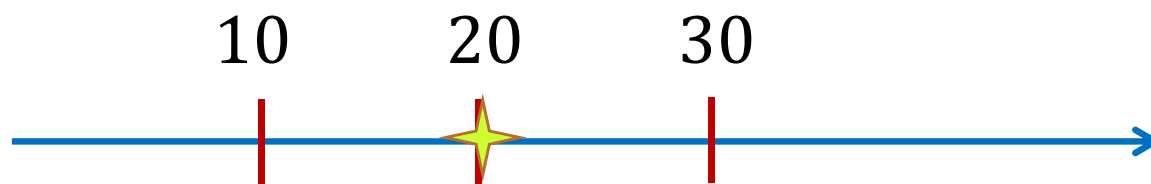
$$x_{\left(\frac{n+1}{2}\right)}$$

Unità	Peso	Peso o.
1	118	118
2	151	118
3	143	134
4	172	138
5	147	143
6	146	146
7	138	147
8	175	151
9	134	172
10	172	172
11	118	175

146

Media e mediana a confronto

$n = 3$



Indici di posizione

dati : x_1, x_2, \dots, x_n

dati **riordinati**:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

$$x_{(10)} = 135$$

$$x_{(11)} = 136$$

Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192

135.5

147.5

La **posizione** del **I° quartile** è

$$\frac{n + 1}{4} \quad (10.25)$$

Il **valore** della **I° quartile** è
(**IL I° quartile**)

$$\frac{x_{(n/4)} + x_{(n/4+1)}}{2}$$

Indici di posizione

dati : x_1, x_2, \dots, x_n

dati **riordinati**:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

La **posizione** del **III° quartile** è

$$\frac{3 \times (n + 1)}{4} \quad (30.75)$$

Il **valore** della **III° quartile** è
(**IL III° quartile**)

$$\frac{x_{(3(n+1)/4)} + x_{(3(n+1)/4)}}{2}$$

Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192

Vertical axis labels: 135.5 (between rows 10 and 11), 147.5 (between rows 20 and 21), 172 (between rows 30 and 31)

Indici di posizione

dati : x_1, x_2, \dots, x_n

$$n = 11$$

dati **riordinati**:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

Unità	Peso	Peso o.
1	118	118
2	151	118
3	143	134
4	172	138
5	147	143
6	146	146
7	138	147
8	175	151
9	134	172
10	172	172
11	118	175

La **posizione** del **I° quartile** è

$$\frac{n + 1}{4} \quad (3)$$

Il **valore** del **I° quartile** è
(**IL I° quartile**)

$$x_{\left(\frac{n+1}{4}\right)}$$

134

Indici di posizione

dati : x_1, x_2, \dots, x_n

$$n = 11$$

dati **riordinati**:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

Unità	Peso	Peso o.
1	118	118
2	151	118
3	143	134
4	172	138
5	147	143
6	146	146
7	138	147
8	175	151
9	134	172
10	172	172
11	118	175

La **posizione** del **III^o quartile** è $\frac{3(n+1)}{4}$ (9)

Il **valore** del **III^o quartile** è
(**IL III^o quartile**)

$$x_{\left(\frac{3(n+1)}{4}\right)} \quad 172$$

...e per dati raggruppati

Classi	n_i	N_i	l_i
100- 110	1	1	0,1
110- 120	3	4	0,3
120- 130	3	7	0,3
130- 140	7	14	0,7
140- 150	7	21	0,7
150- 160	7	28	0,7
160- 170	0	28	0,0
170- 180	7	35	0,7
180- 190	3	38	0,3
190- 200	2	40	0,2
Tot	40		

...e per dati raggruppati

Classi	n_i	N_i	l_i
100- 110	1	1	0,1
110- 120	3	4	0,3
120- 130	3	7	0,3
130- 140	7	14	0,7
140- 150	7	21	0,7
150- 160	7	28	0,7
160- 170	0	28	0,0
170- 180	7	35	0,7
180- 190	3	38	0,3
190- 200	2	40	0,2
Tot	40		

Posizione della mediana:

$$\frac{n + 1}{2} = 20.5$$

...e per dati raggruppati

Posizione della mediana:

$$\frac{n + 1}{2} = 20.5$$

Classi	n_i	N_i	l_i
100- 110	1	1	0,1
110- 120	3	4	0,3
120- 130	3	7	0,3
130- 140	7	14	0,7
140- 150	7	21	0,7
150- 160	7	28	0,7
160- 170	0	28	0,0
170- 180	7	35	0,7
180- 190	3	38	0,3
190- 200	2	40	0,2
Tot	40		

← classe mediana

...e per dati raggruppati

Classi	n_i	N_i	l_i
100- 110	1	1	0,1
110- 120	3	4	0,3
120- 130	3	7	0,3
130- 140	7	14	0,7
140- 150	7	21	0,7
150- 160	7	28	0,7
160- 170	0	28	0,0
170- 180	7	35	0,7
180- 190	3	38	0,3
190- 200	2	40	0,2
Tot	40		

Posizione della mediana:

$$\frac{n + 1}{2} = 20.5$$

Valore nella classe: quanti dati che stanno nella classe mediana vengono prima della mediana? **20 - 14!**

$$\frac{n}{2} - N_{i^*-1}$$

...e per dati raggruppati

Classi	n_i	N_i	l_i
100- 110	1	1	0,1
110- 120	3	4	0,3
120- 130	3	7	0,3
130- 140	7	14	0,7
140- 150	7	21	0,7
150- 160	7	28	0,7
160- 170	0	28	0,0
170- 180	7	35	0,7
180- 190	3	38	0,3
190- 200	2	40	0,2
Tot	40		

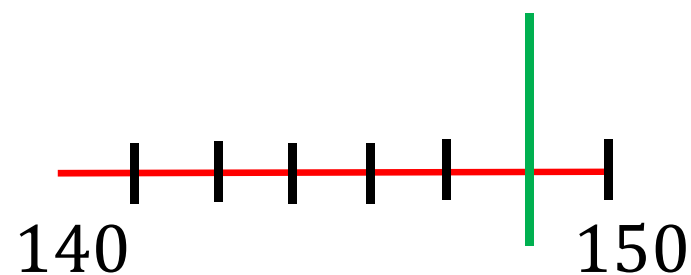
Posizione della mediana:

$$\frac{n + 1}{2} = 20.5$$

Valore nella classe: 20-14
dati *prima* della mediana.

Dei 7 dati nella classe, 6
sono prima della mediana
e 1 dopo:

$$\left(\frac{n}{2} - N_{i^*-1}\right) \times \frac{a_i}{n_i}$$



...e per dati raggruppati

Classi	n_i	N_i	l_i
100- 110	1	1	0,1
110- 120	3	4	0,3
120- 130	3	7	0,3
130- 140	7	14	0,7
140- 150	7	21	0,7
150- 160	7	28	0,7
160- 170	0	28	0,0
170- 180	7	35	0,7
180- 190	3	38	0,3
190- 200	2	40	0,2
Tot	40		

Posizione della mediana:

$$\frac{n + 1}{2} = 20.5$$

Valore della mediana:

$$x_{i^*} + \left(\frac{n}{2} - N_{i^*-1} \right) \times \frac{a_i}{n_i}$$

$$140 + (20 - 14) \times \frac{10}{7} = 148.57$$

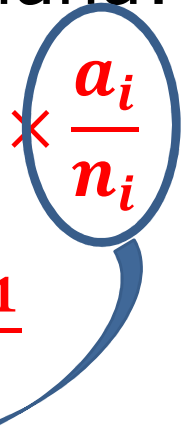
...e per dati raggruppati

Classi	n_i	N_i	l_i
100- 110	1	1	0,1
110- 120	3	4	0,3
120- 130	3	7	0,3
130- 140	7	14	0,7
140- 150	7	21	0,7
150- 160	7	28	0,7
160- 170	0	28	0,0
170- 180	7	35	0,7
180- 190	3	38	0,3
190- 200	2	40	0,2
Tot	40		

Posizione della mediana:

$$\frac{n + 1}{2} = 20.5$$

Valore della mediana:

$$x_{i^*} + \left(\frac{n}{2} - N_{i^*-1} \right) \times \frac{a_i}{n_i}$$


$$x_{i^*} + \frac{n/2 - N_{i^*-1}}{l_i}$$
$$140 + \frac{20 - 14}{0.7} = 148.57$$

...e per dati raggruppati

Classi	n_i	N_i	l_i
100- 110	1	1	0,1
110- 120	3	4	0,3
120- 130	3	7	0,3
130- 140	10	17	1,0
140- 150	3	20	0,3
150- 160	8	28	0,8
160- 170	0	28	0,0
170- 180	7	35	0,7
180- 190	3	38	0,3
190- 200	2	40	0,2
Tot	40		

Un caso particolare:

$$\frac{n+1}{2} = 20.5$$

$$x_{i^*} + \left(\frac{n}{2} - N_{i^*-1} \right) \times \frac{a_i}{n_i} = 20$$

20-17 = **3 dei 3** dati
nella classe vengono
prima della mediana

⇒

la mediana è = **150**

...e per dati raggruppati

Classi	n_i	N_i	l_i
100- 110	1	1	0,1
110- 120	3	4	0,3
120- 130	3	7	0,3
130- 140	7	14	0,7
140- 150	7	21	0,7
150- 160	7	28	0,7
160- 170	0	28	0,0
170- 180	7	35	0,7
180- 190	3	38	0,3
190- 200	2	40	0,2
Tot	40		

Posizione di Q1:

$$\frac{n + 1}{4} = 10.25$$

Valore di Q1:

$$x_{i^*} + \frac{n/4 - N_{i^*-1}}{l_i}$$

$$130 + \frac{10 - 7}{0.7} = 134.28$$

...e per dati raggruppati

Classi	n_i	N_i	l_i
100- 110	1	1	0,1
110- 120	3	4	0,3
120- 130	3	7	0,3
130- 140	7	14	0,7
140- 150	7	21	0,7
150- 160	7	28	0,7
160- 170	0	28	0,0
170- 180	7	35	0,7
180- 190	3	38	0,3
190- 200	2	40	0,2
Tot	40		

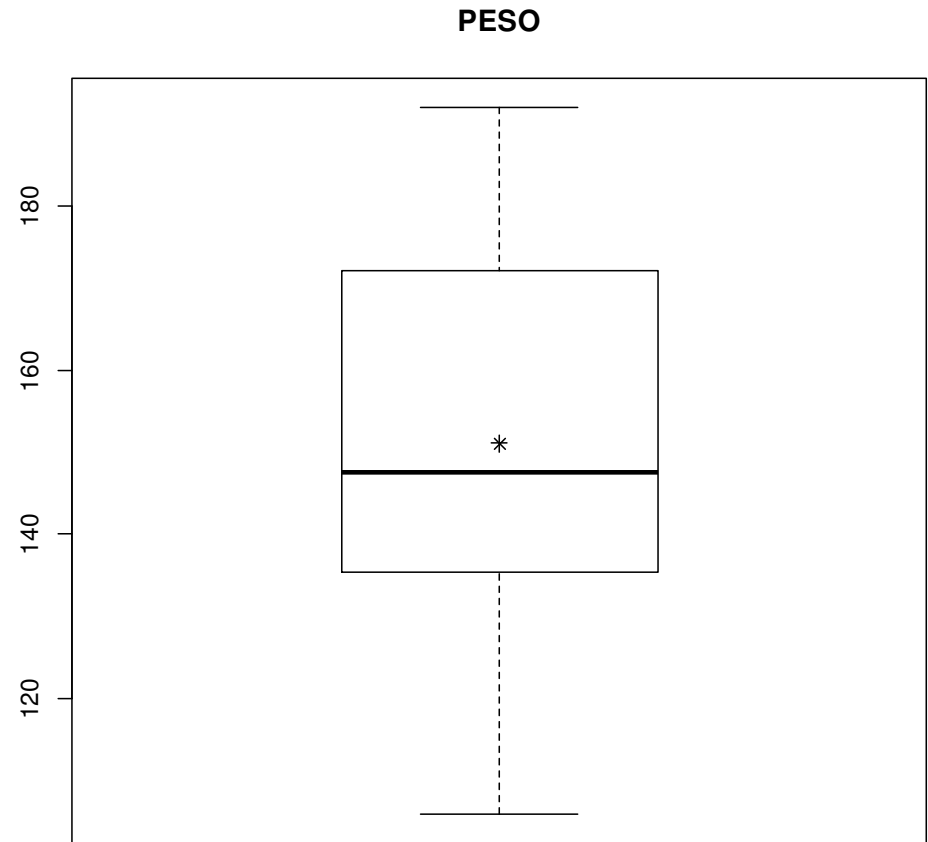
Posizione di Q3:

$$3 \times \frac{n + 1}{4} = 30.75$$

Valore di Q3:

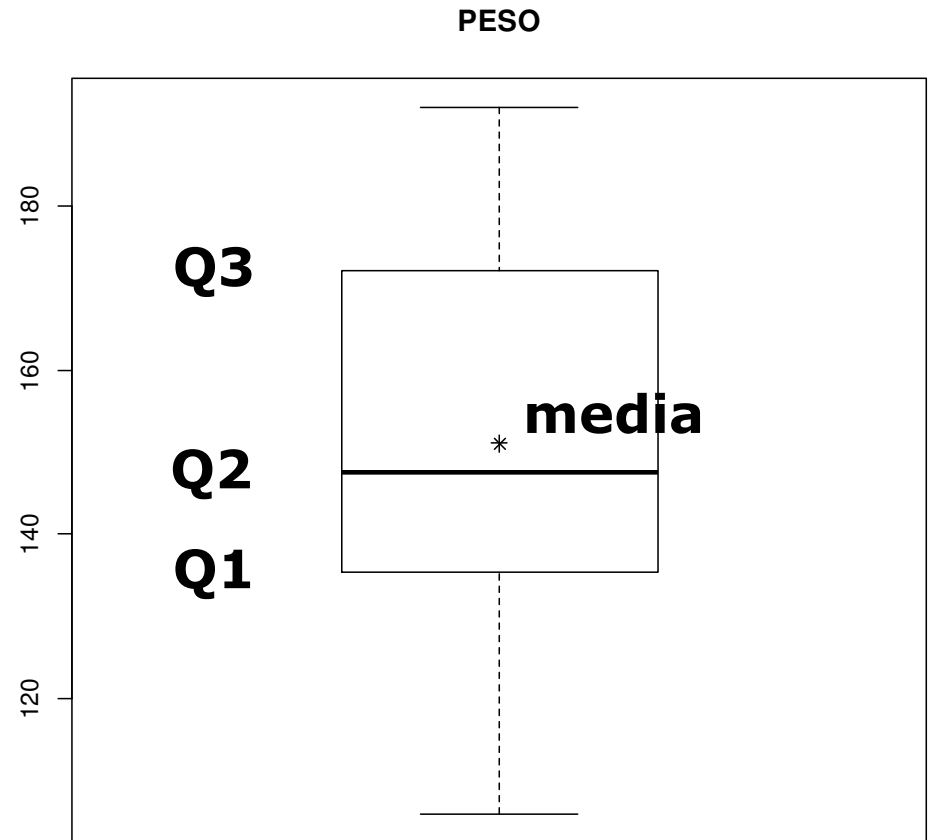
$$x_{i^*} + \frac{3n/4 - N_{i^*-1}}{l_i}$$
$$170 + \frac{30 - 28}{0.7} = 172.86$$

Boxplot



Boxplot

Differenza (*range*) interquartile: $Q3 - Q1$: IQR

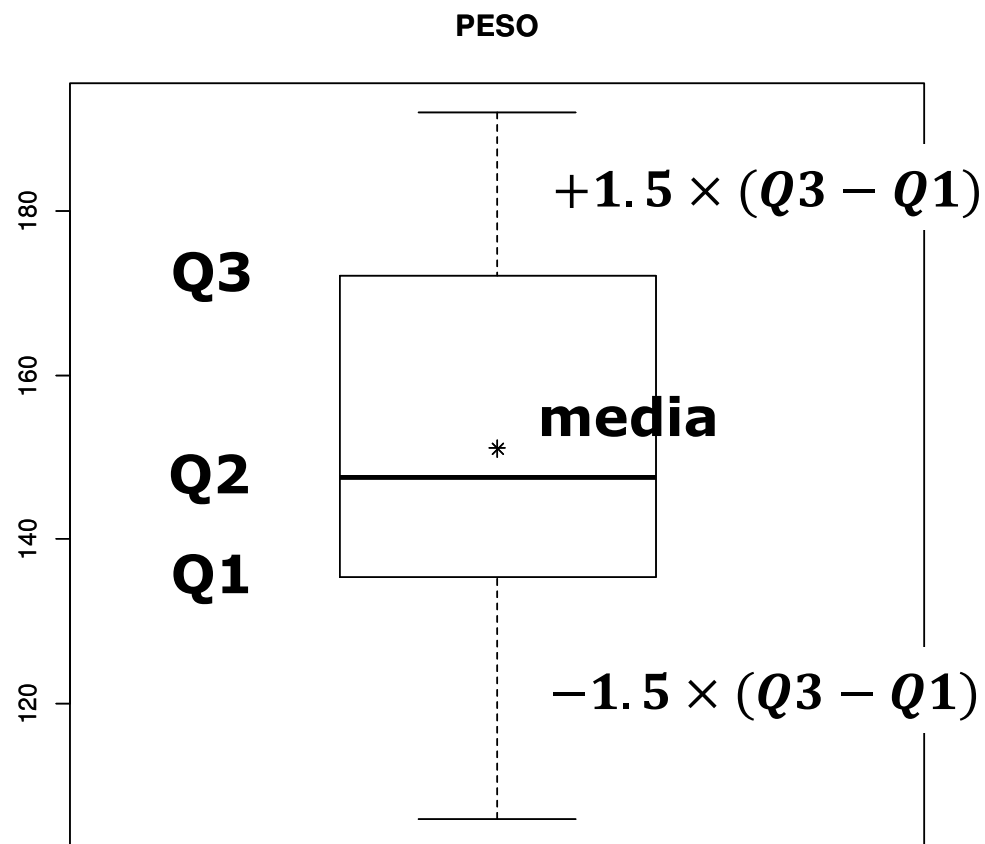


Boxplot

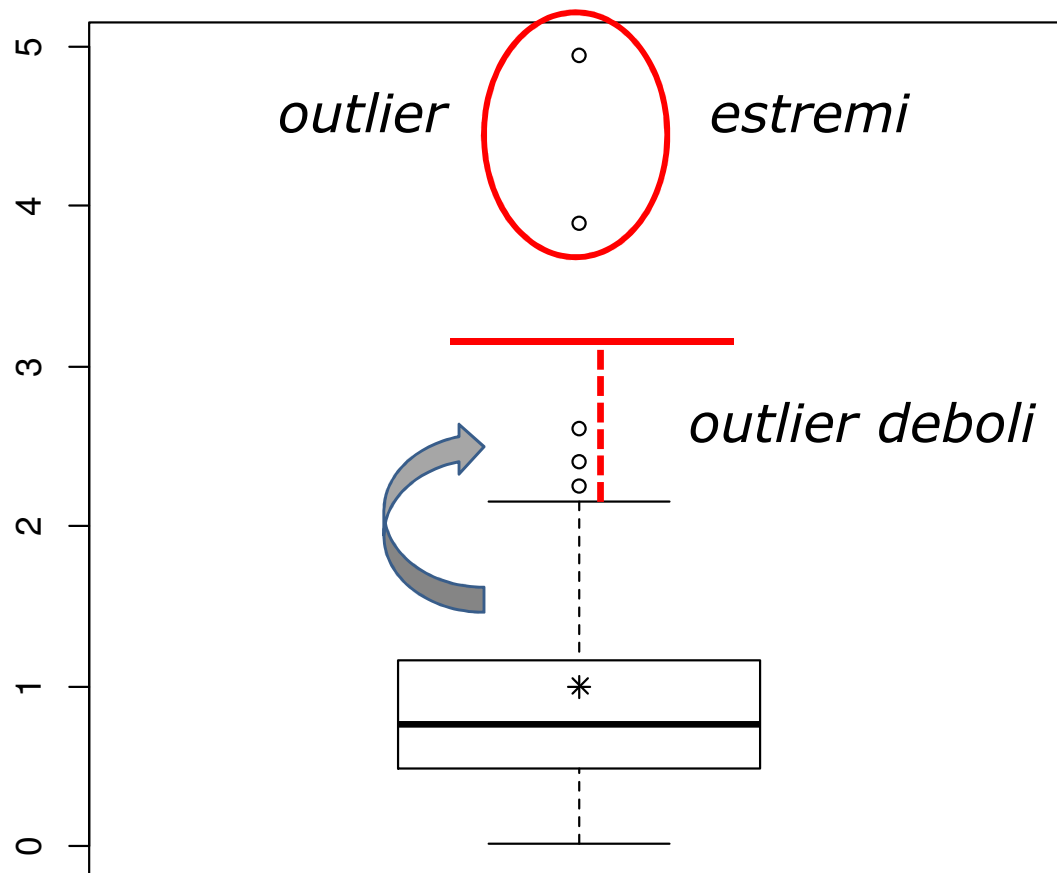
Differenza (*range*) interquartile: $Q3 - Q1$: IQR

$$1.5 \times (Q3 - Q1)$$

è la
lunghezza **massima**
dei baffi



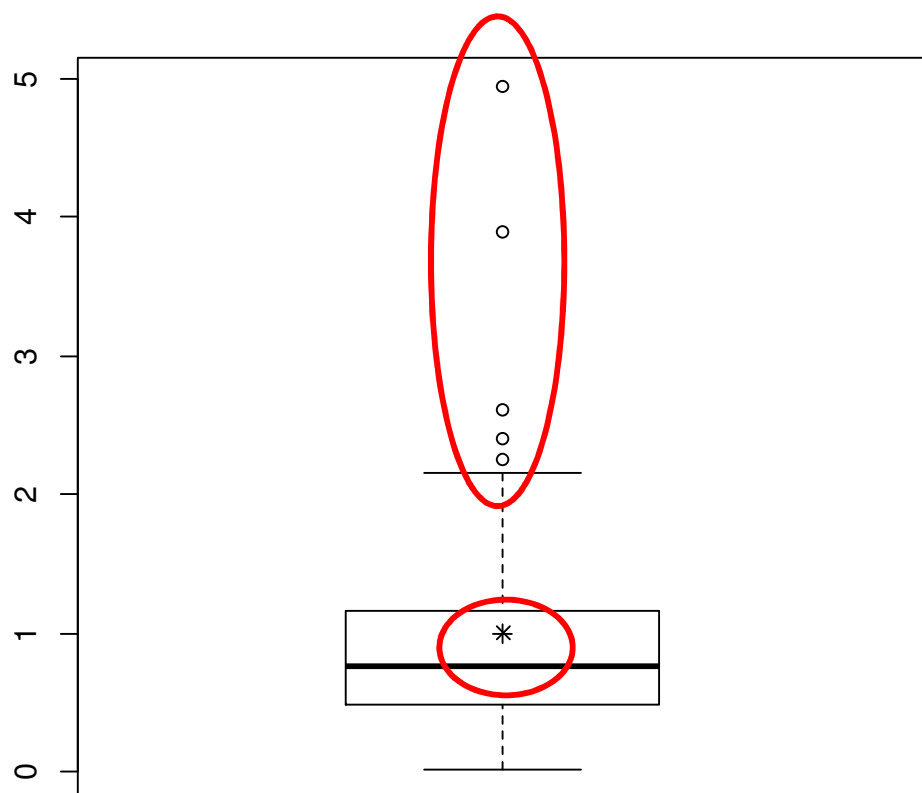
Boxplot



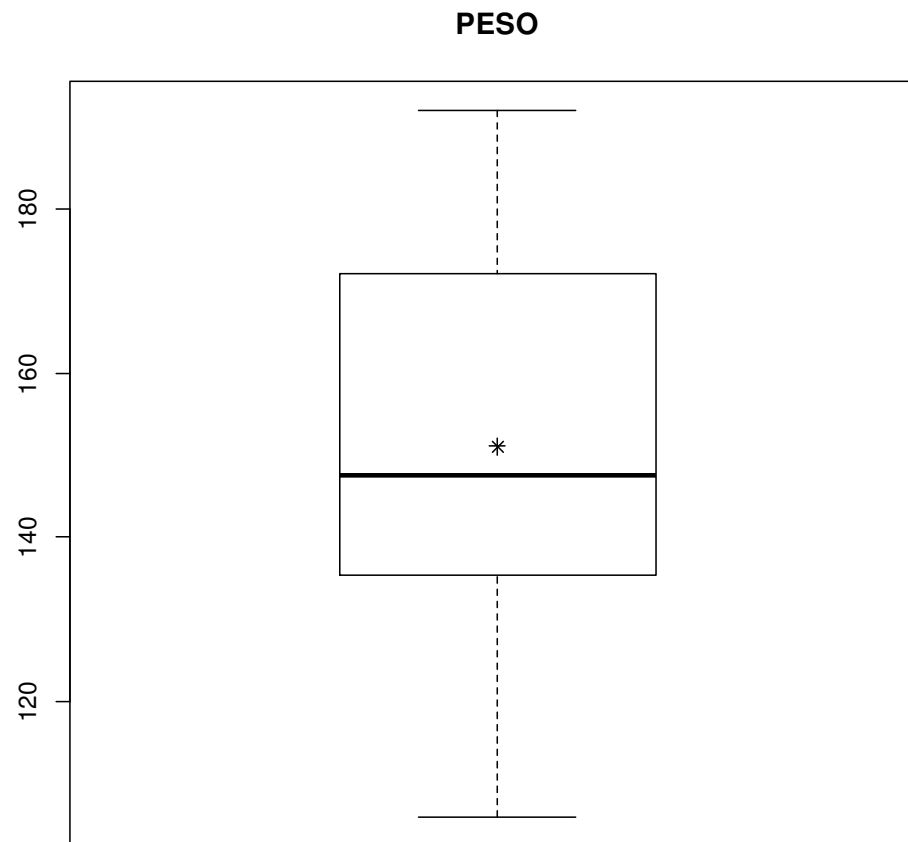
Dati fuori dal baffo
(superiore): *outlier*

Baffo più corto della
lunghezza massima

Boxplot



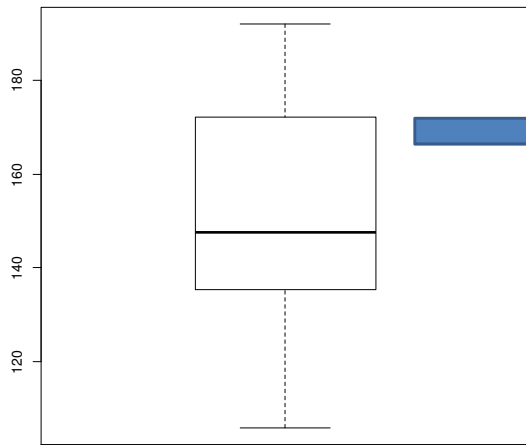
poca simmetria



simmetria

Boxplot vs istogramma

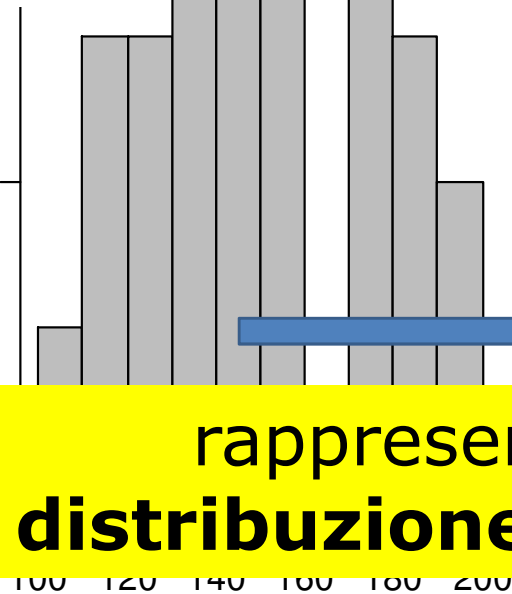
rappresentazione della
dispersione dei dati



\bar{x}

0.005

0.000



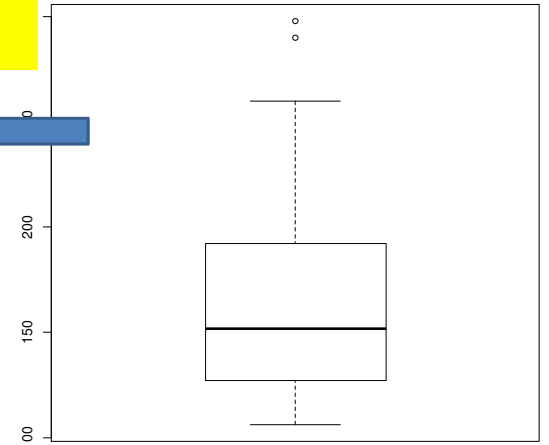
Density

0.010

0.005

rappresentazione della
distribuzione delle frequenze

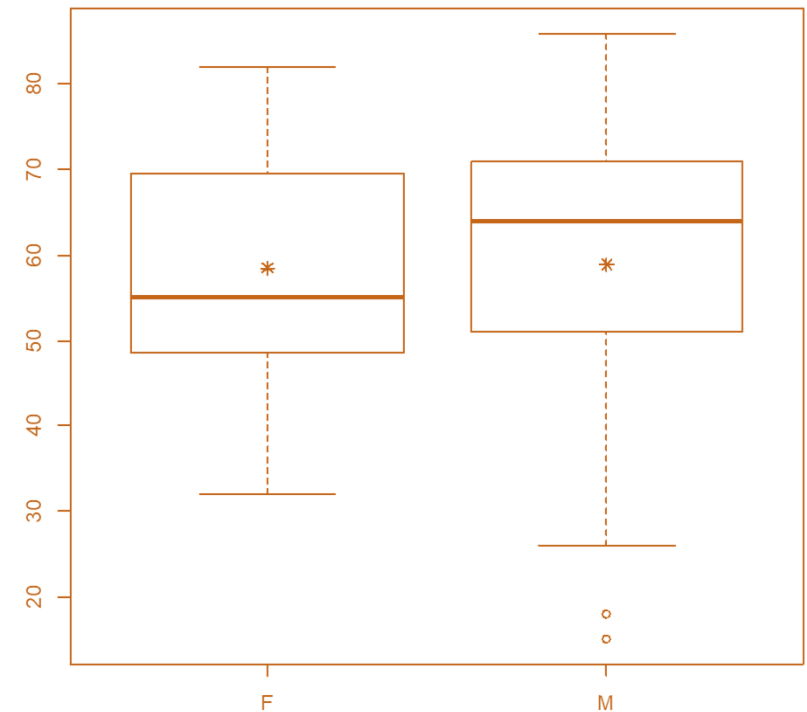
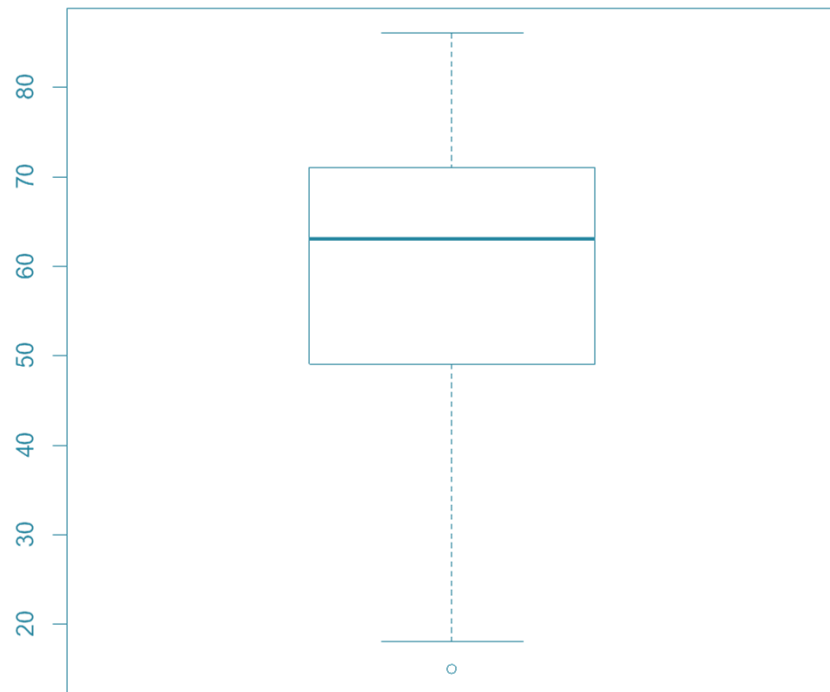
100 150 200 250 300



Ritorniamo in

```
> summary(AGE)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
15.00	49.25	63.00	58.78	71.00	86.00



```
> boxplot(AGE[GEN==0],AGE[GEN==1],names=c("F","M"))  
> points(c(mean(AGE[GEN==0]),mean(AGE[GEN==1])),pch=8)
```

Quantili

dati : x_1, x_2, \dots, x_n

$$0 < q < 1$$

dati **riordinati**:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

$$Q1: q = 0.25$$

$$Q2: q = 0.50$$

$$Q3: q = 0.75$$

Posizione del quantile di ordine q :

$$q \times (n + 1) \quad (q = 0.83 \Rightarrow 0.83 \times 41 = 34.03)$$

Valore del quantile:

$$\frac{x_{(i^*)} + x_{(i^*+1)}}{2}$$

$$i^* = \text{parte intera di } q(n + 1)$$

Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192

Quantili

dati : x_1, x_2, \dots, x_n

$$0 < q < 1$$

dati **riordinati**:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

$$Q1: q = 0.25$$

$$Q2: q = 0.50$$

$$Q3: q = 0.75$$

Posizione

percentili:

$$q = k \times 0.01$$

q :

$$q \times (n + 1)$$

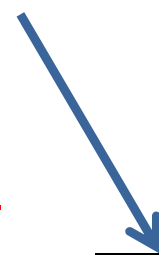
$$0.25 \times 139 = 34.03$$

Valore del quantile:

$$\frac{x_{(i^*)} + x_{(i^*+1)}}{2}$$

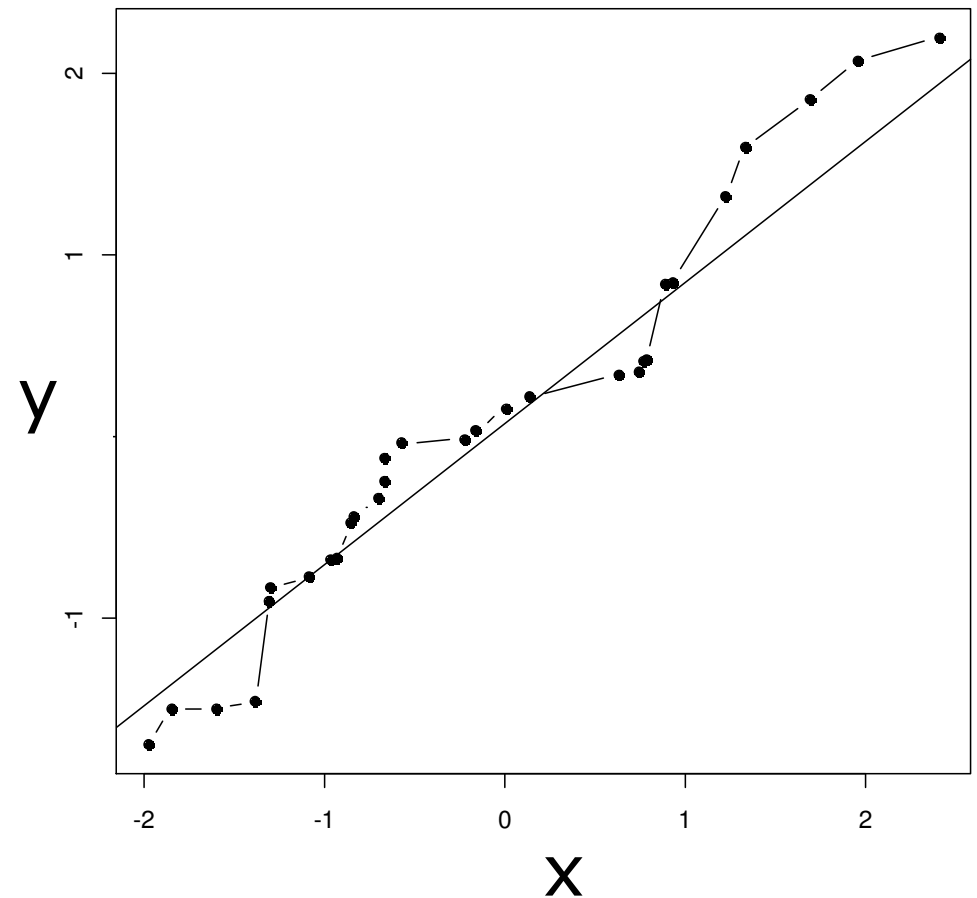
$$i^* = \text{parte intera di } q(n + 1)$$

Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192

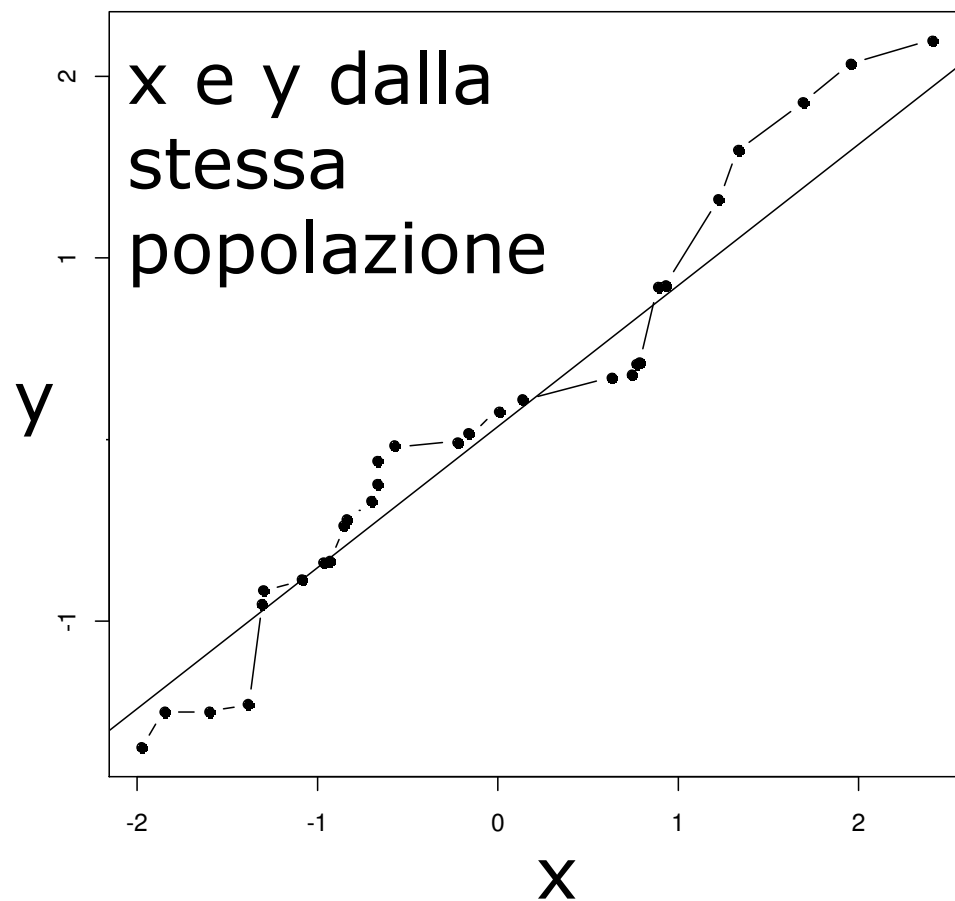
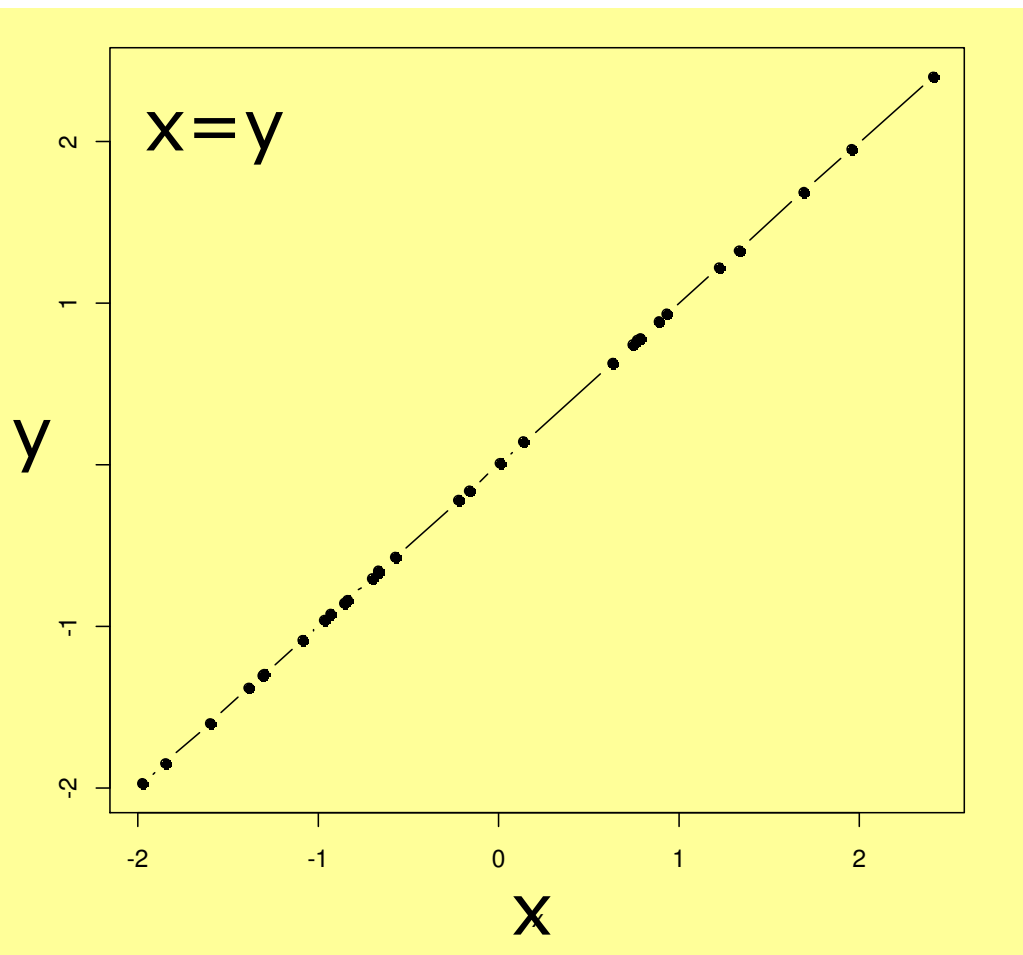


Q-Q plot

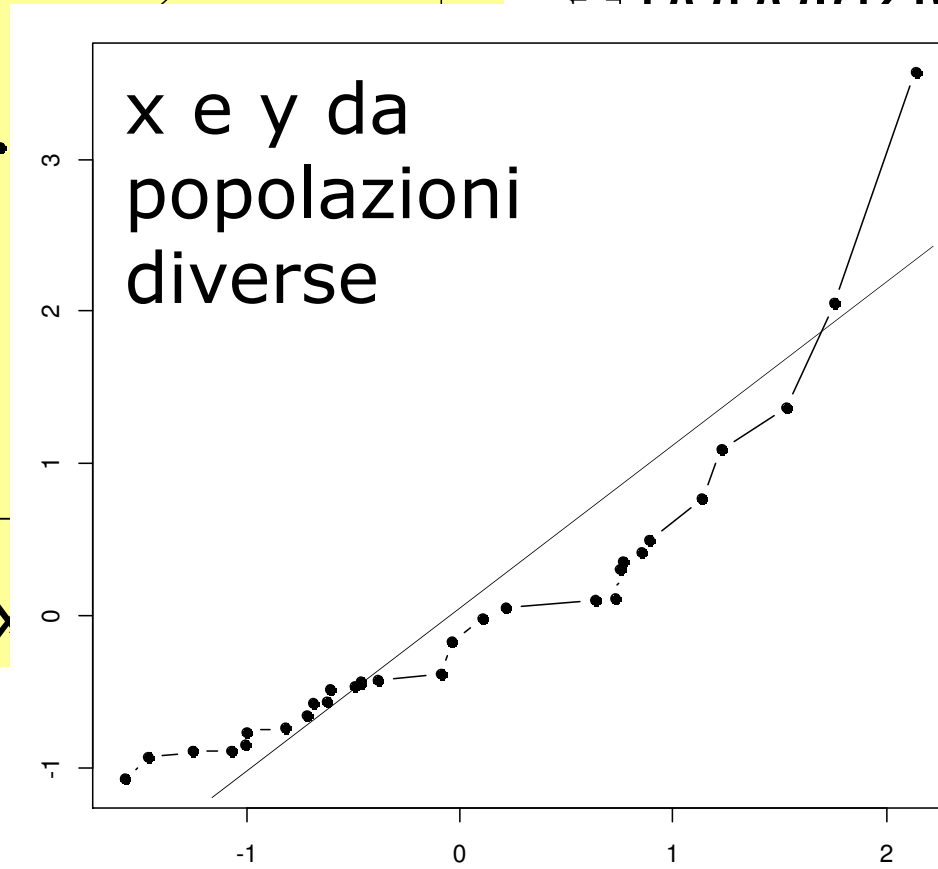
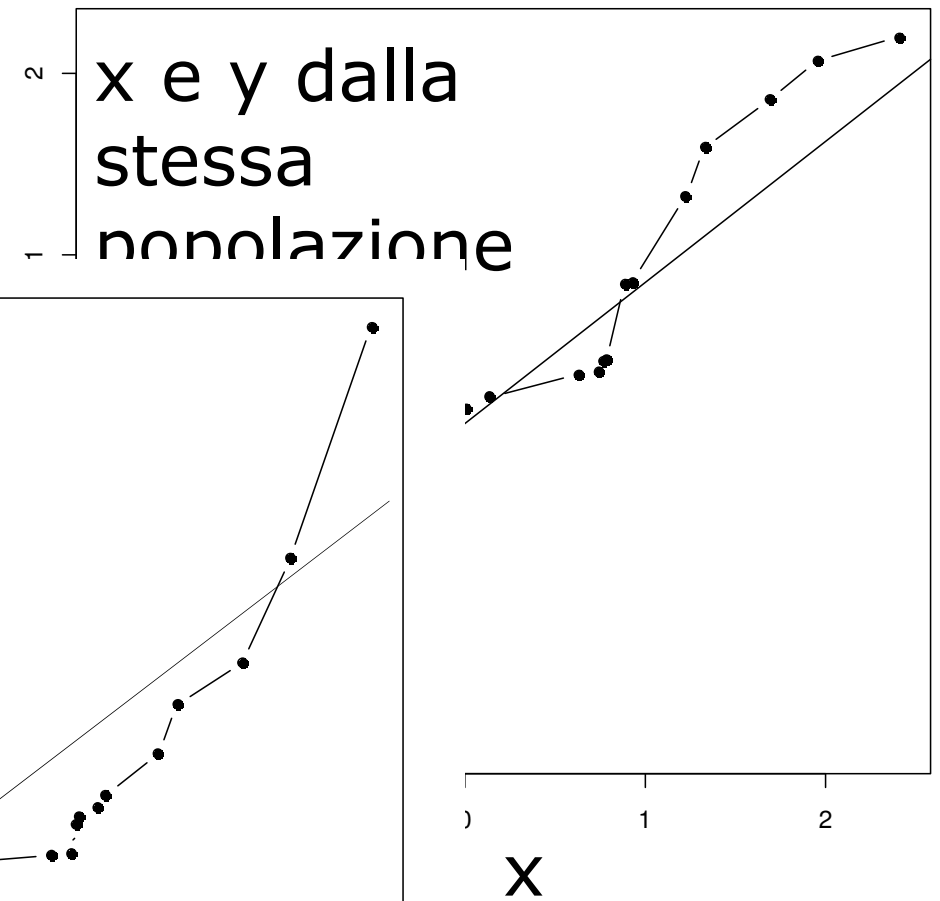
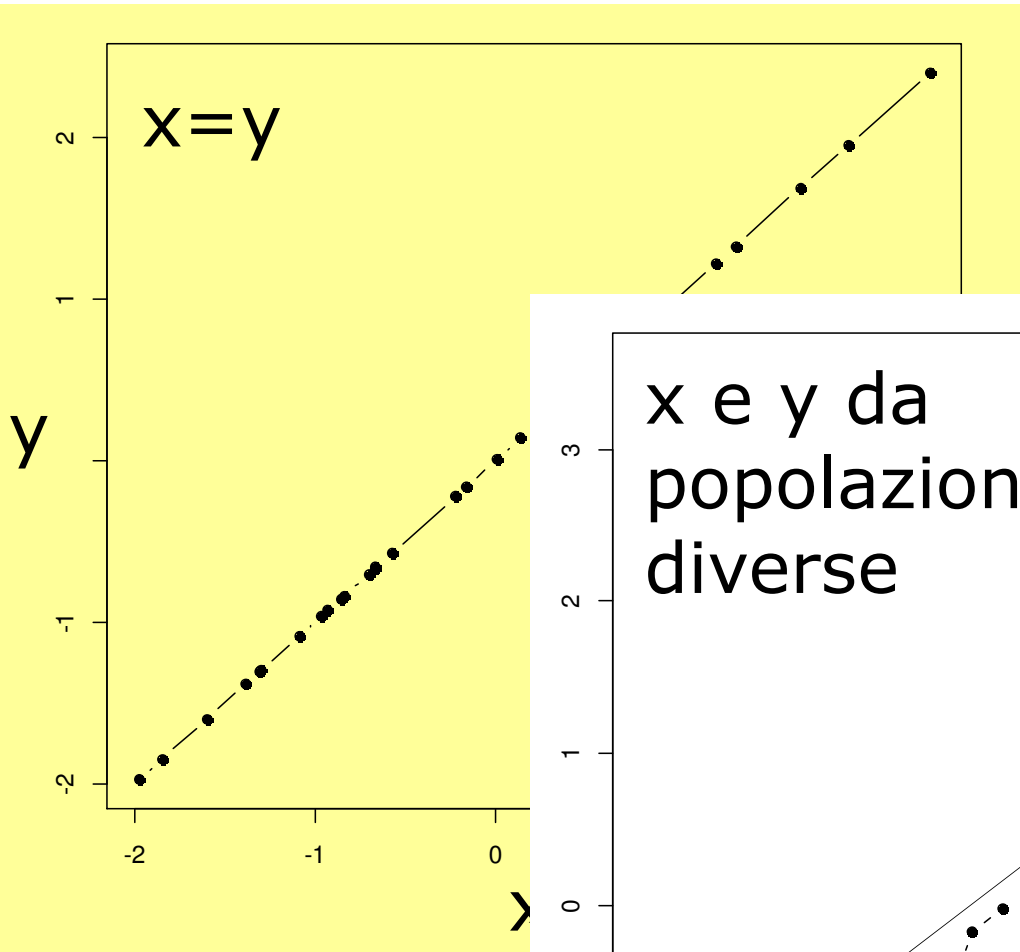
I due insiemi di dati
"si assomigliano"?



Q-Q plot



Q-Q plot



Media pesata

Media aritmetica:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \sum_{i=1}^n x_i \times \frac{1}{n}$$

↕ dato ↘ peso

Media pesata:

$$\overline{x_w} = \sum_{i=1}^n x_i \times w_i$$

pesi tali che: $w_i > 0$, $\sum_{i=1}^n w_i = 1$