Statistica

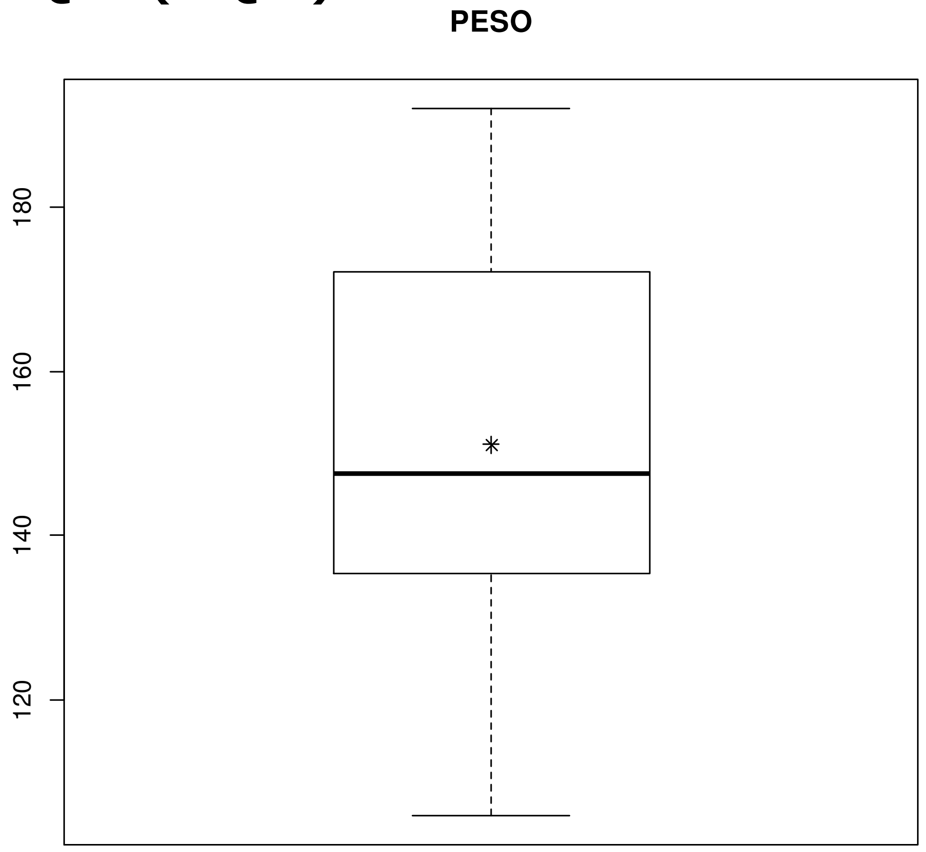
Statistica descrittiva

INDICI

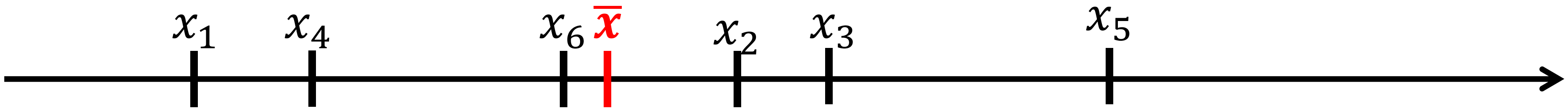
Indici di dispersione

Differenza interquartile: Q3-Q1 (IQR)

Range: Max-Min

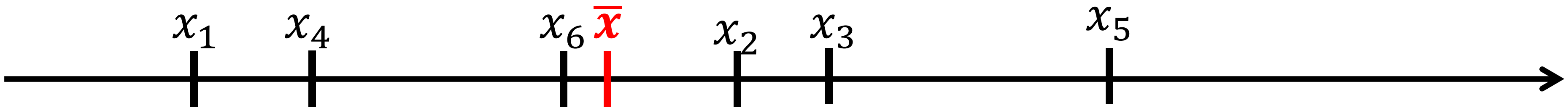


Indici di dispersione



Prendiamo un punto di riferimento, la **media**, e guardiamo a come si disperdono i dati attorno ad esso

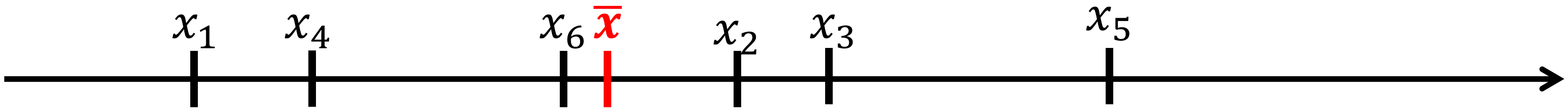
Indici di dispersione



Prendiamo un punto di riferimento, la **media**, e guardiamo a come si disperdono i dati attorno ad esso

$$\sum_{i=1}^{n}(x_i - \bar{x})$$

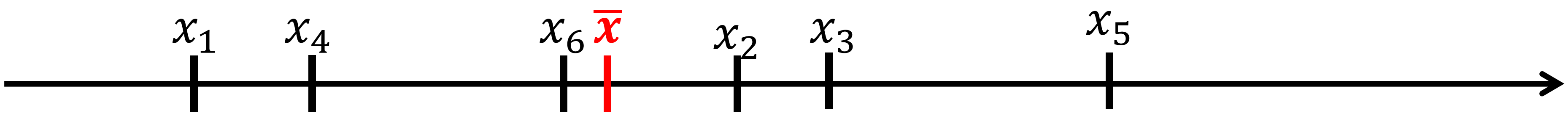
Indici di dispersione



Prendiamo un punto di riferimento, la **media**, e guardiamo a come si disperdono i dati attorno ad esso

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \left(\sum_{i=1}^{n} x_i\right) - n\bar{x} \quad = n\bar{x} - n\bar{x} = 0 \text{ !!!!!!!}$$

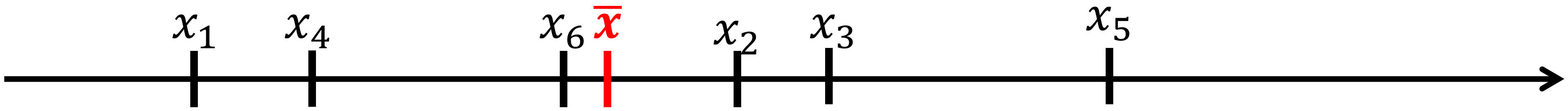
Indici di dispersione



Prendiamo un punto di riferimento, la **media**, e guardiamo a come si disperdono i dati attorno ad esso

$$\sum_{i=1}^{n} (x_i - \bar{x})^2$$

Indici di dispersione



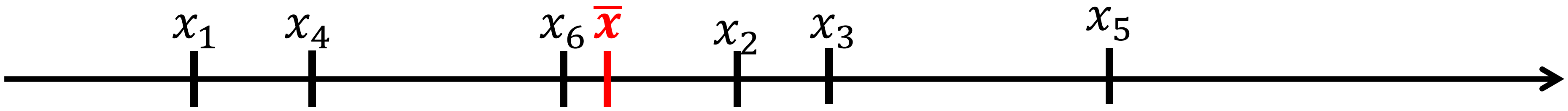
Prendiamo un punto di riferimento, la **media**, e guardiamo a come si disperdono i dati attorno ad esso

$$0 \leq \sum_{i=1}^{n} (x_i - \bar{x})^2$$

MA $= 0 \Leftrightarrow$ I DATI SONO TUTTI UGUALI!

Ex: 2, 2, 2, 2, 2 $\Rightarrow$ $\bar{x} = 2$

Indici di dispersione



Prendiamo un punto di riferimento, la **media**, e guardiamo a come si disperdono i dati attorno ad esso

$$\boxed{\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Varianza, o scarto quadratico medio

Varianza

$$\bar{x} = \frac{19 + 22 + 21 + 23 + 22 + 20}{6} = 21.17$$ anni

	Età (y)	Peso (kg)	Altezza (m)	Sesso	Causa di morte
1	19	50.2	1.65	F	Nat.
2	22	75.6	1.78	M	Inc.
3	21	80.1	1.91	M	Inc.
4	23	56.7	1.72	M	Nat.
5	22	75.0	1.81	M	M.C.
6	20	58.3	1.68	F	Tum.

$$\sigma^2 = \frac{(19-21.17)^2 + (22-21.17)^2 + (21-21.17)^2}{6} +$$

$$+ \frac{(23-21.17)^2 + (22-21.17)^2 + (20-21.17)^2}{6} = 1.81 \quad \text{anni}^2$$

$$\bar{x} = 21.17$$

	Età (y)	Peso (kg)	Altezza (m)	Sesso	Causa di morte
1	19	50.2	1.65	F	Nat.
2	22	75.6	1.78	M	Inc.
3	21	80.1	1.91	M	Inc.
4	23	56.7	1.72	M	Nat.
5	22	75.0	1.81	M	M.C.
6	20	58.3	1.68	F	Tum.

Varianza

X : n. figli

x_i	n_i
0	4
1	6
2	4
3	1

$\boldsymbol{n} = \mathbf{15}$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$\bar{x} = 1.13$

$\sigma^2 =$

$$\frac{4 \times (0 - 1.13)^2 + 6 \times (1 - 1.13)^2 + 4 \times (2 - 1.13)^2 + 1 \times (3 - 1.13)^2}{15} = 0.782$$

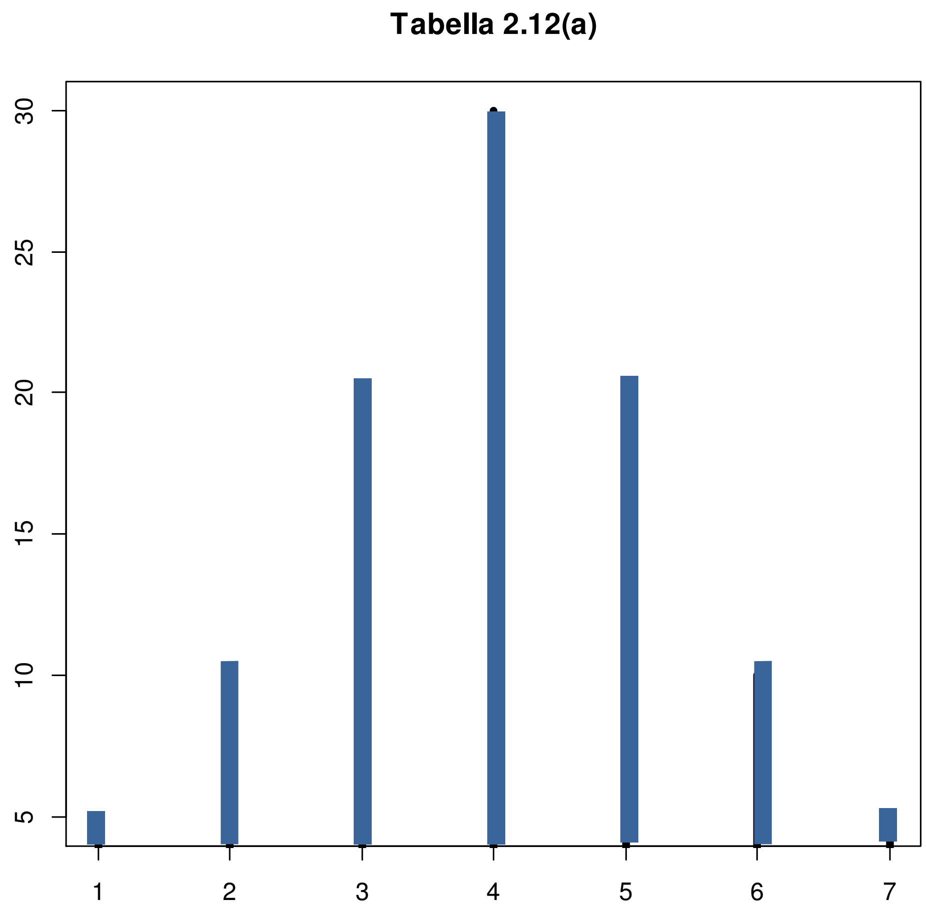
Varianza

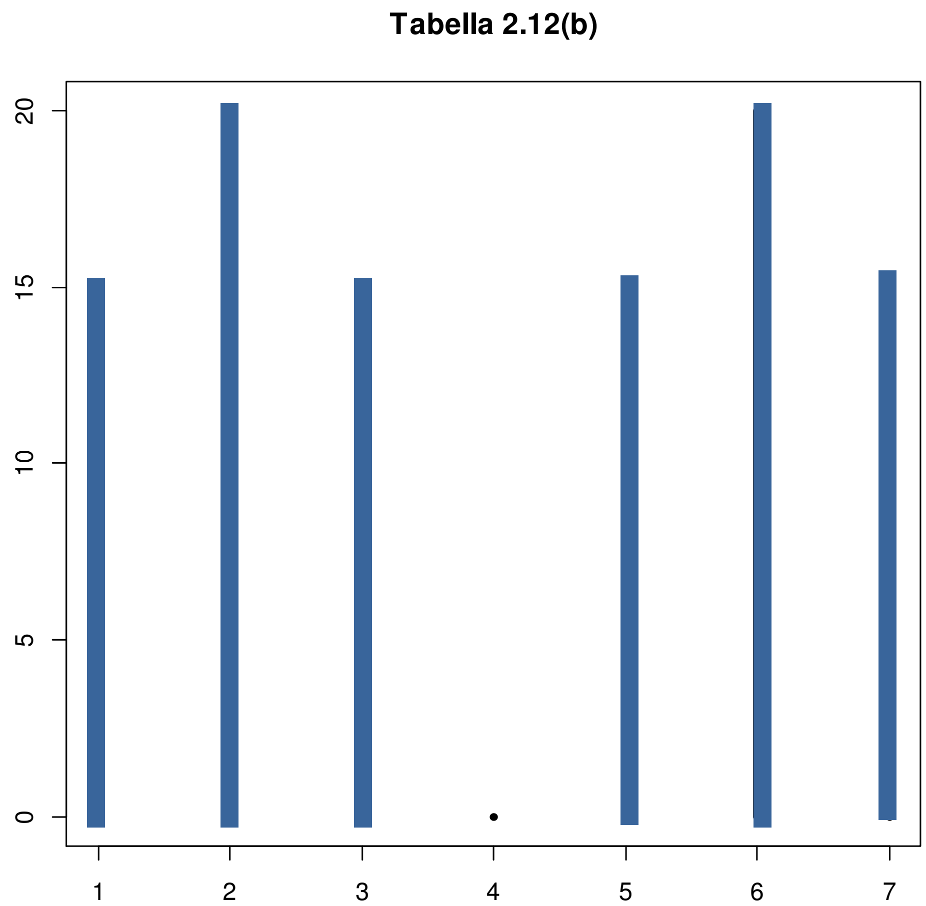
Prendiamo un punto di riferimento, la **media**, e guardiamo a come si disperdono i dati attorno ad esso

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i{}^2\right) - \bar{x}^2$$

(S.M. Iacus, *Statistica*)

Varianza

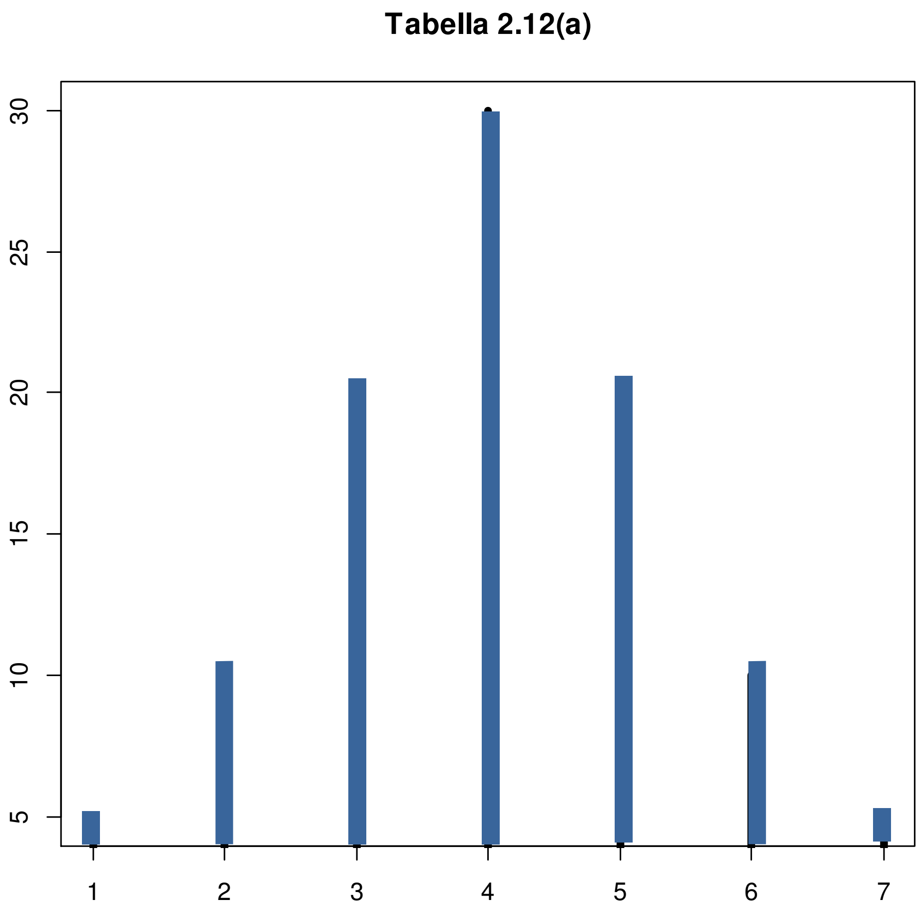


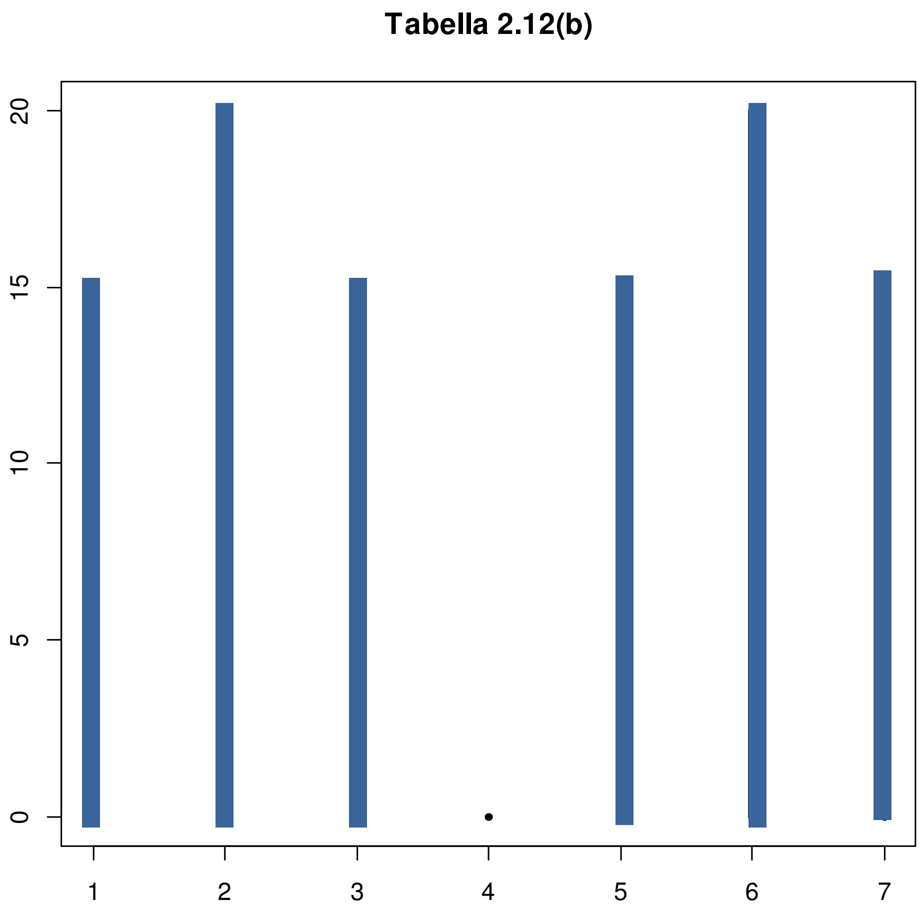


dati più frequenti **vicino a** 4

dati più frequenti **lontano da** 4

Varianza



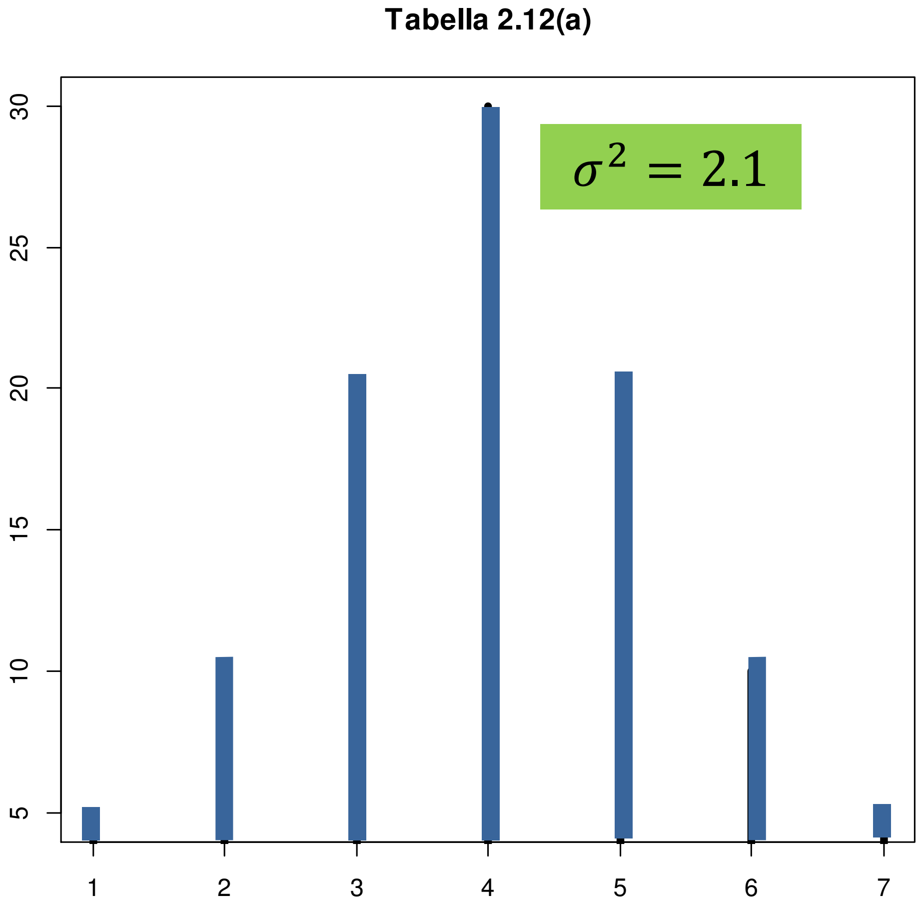


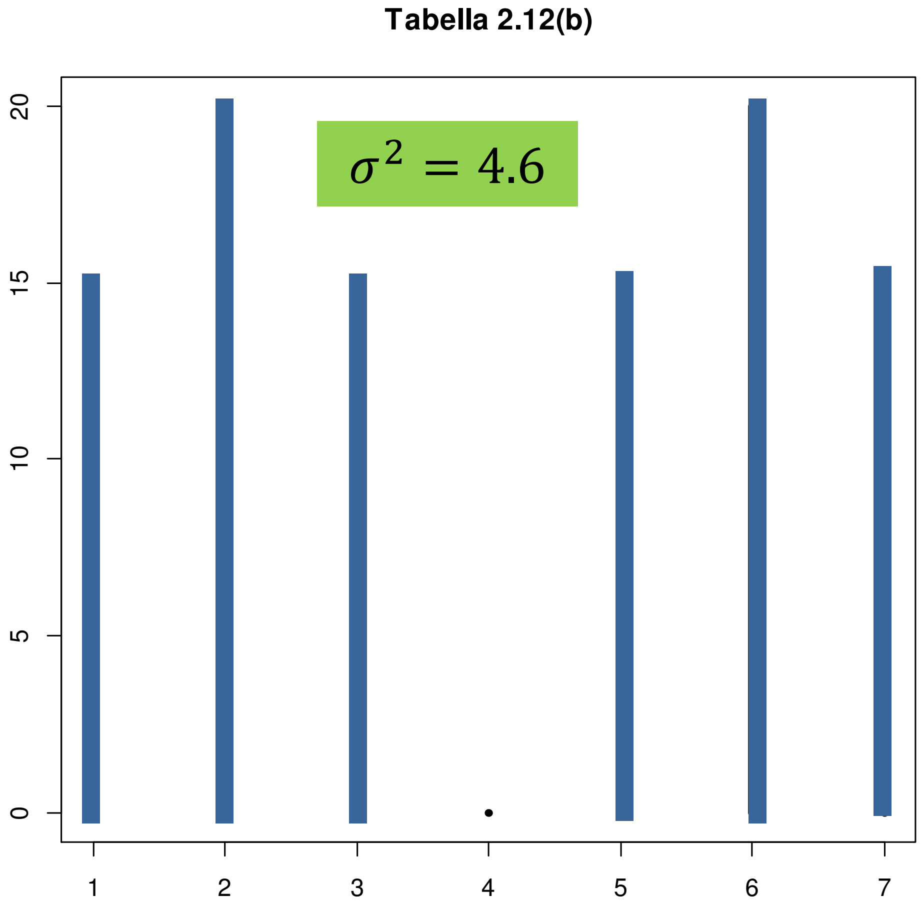
dati più frequenti **vicino a** 4

dati più frequenti **lontano da** 4

$\bar{x} = 4$ mediane pari a 4 $\bar{y} = 4$

Varianza





dati più frequenti **vicino a** 4

dati più frequenti **lontano da** 4

Proprietà della media

$$y_i = x_i + a \quad \Rightarrow \quad \boxed{\bar{y} = \bar{x} + a}$$

il ballo del mattooo neee

$$y_i = bx_i \quad \Rightarrow \quad \boxed{\bar{y} = b\bar{x}}$$

$$\boxed{y_i = bx_i + a \quad \Rightarrow \quad \bar{y} = b\bar{x} + a}$$

Proprietà della varianza

$$y_i = x_i + a \Rightarrow \boxed{\sigma^2{}_y = \sigma^2{}_x}$$

il ballo del mattooo neee

$$y_i = bx_i \Rightarrow \boxed{\sigma^2{}_y = b^2\sigma^2{}_x}$$

$$\boxed{y_i = bx_i + a \Rightarrow \sigma^2{}_y = b^2\sigma^2{}_x}$$

Varianza campionaria

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

varianza della popolazione

<u>campione</u> ***<u>casuale</u>***

$$\bar{s}_n{}^2 = \frac{1}{\boldsymbol{n-1}}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

varianza ***campionaria***

Varianza campionaria

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} {x_i}^2 - \bar{x}^2$$

varianza della popolazione

<u>campione</u> *<u>casuale</u>*

$$\bar{s}_n{}^2 = \frac{1}{n-1}\sum_{i=1}^{n} {x_i}^2 - \frac{\boldsymbol{n}}{\boldsymbol{n-1}}\,\bar{x}^2$$

varianza ***campionaria***

Deviazione standard

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2 - \bar{x}^2}$$

della popolazione

<u>campione</u> <u>*casuale*</u>

campionaria

$$\bar{s}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} {x_i}^2 - \frac{n}{n-1}\bar{x}^2}$$

Nella stessa unità di misura dei dati

Il nostro *test*:

Sezione 3

1. Sia $x_1, x_2, \cdots, x_n$ un campione di n dati riferiti ad una certa variabile quantitativa (numerica). Quale tra i seguenti indici **non** misura, secondo voi, la variabilità dei dati:

$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$	$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$	$\max_{1\le i\le n} x_i - \min_{1\le i\le n} x_i$

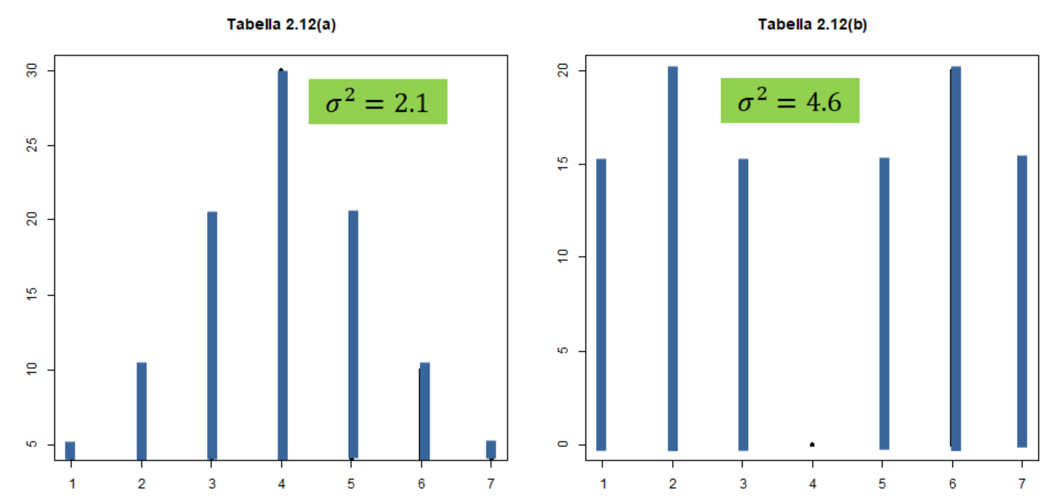
Il nostro *test*:

Sezione 3

1. Sia $x_1, x_2, \cdots, x_n$ un campione di n dati riferiti ad una certa variabile quantitativa (numerica). Quale tra i seguenti indici **non** misura, secondo voi, la variabilità dei dati:

$\boxed{\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i}$	$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$	$\max_{1\le i\le n} x_i - \min_{1\le i\le n} x_i$

$$\bar{x}_n = 4$$



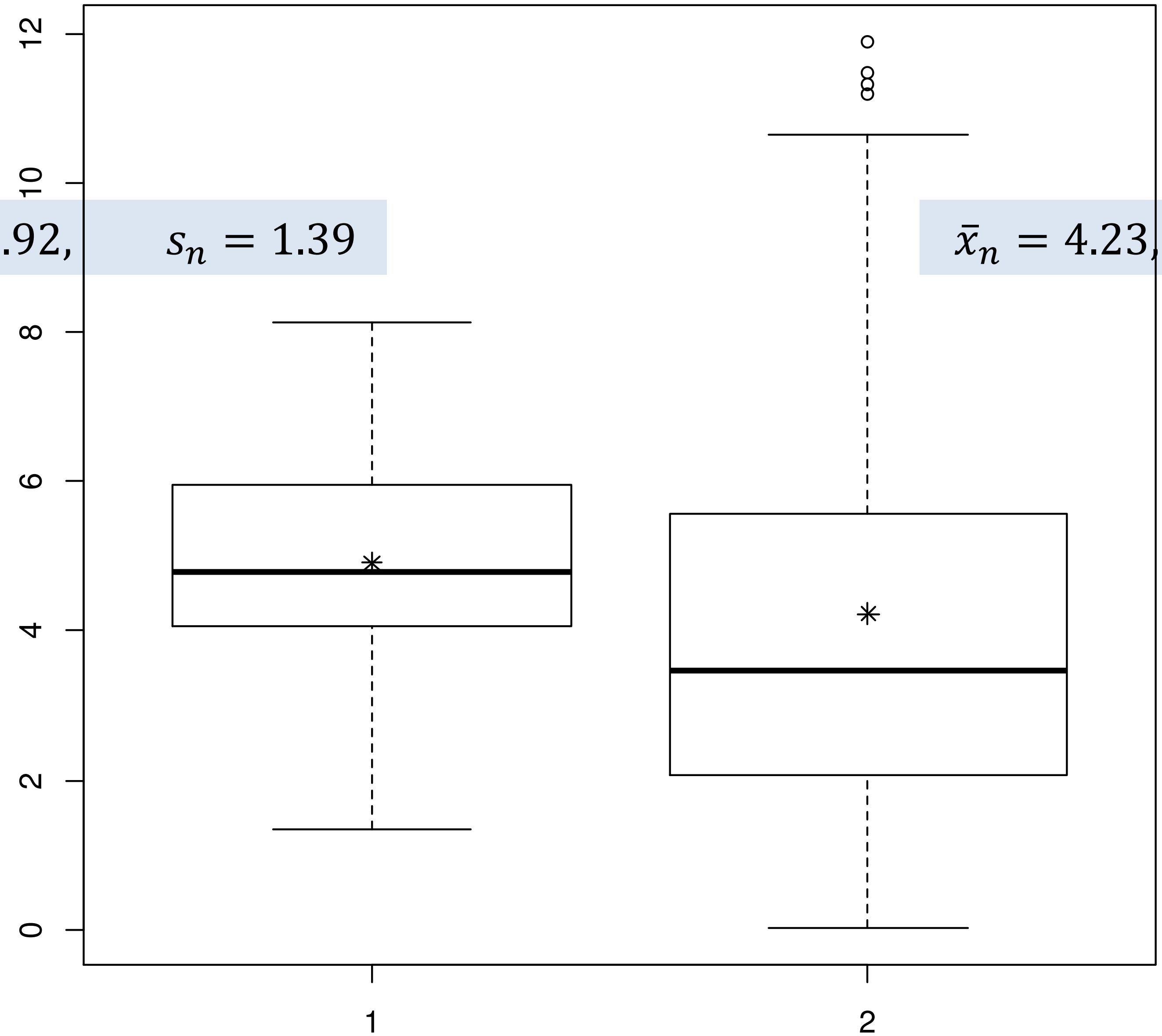
dati più frequenti **vicino a** 4

dati più frequenti **lontano da** 4

Sull'uso di $\bar{x}_n \mp s_n$

$\bar{x}_n = 4.92, \quad s_n = 1.39$

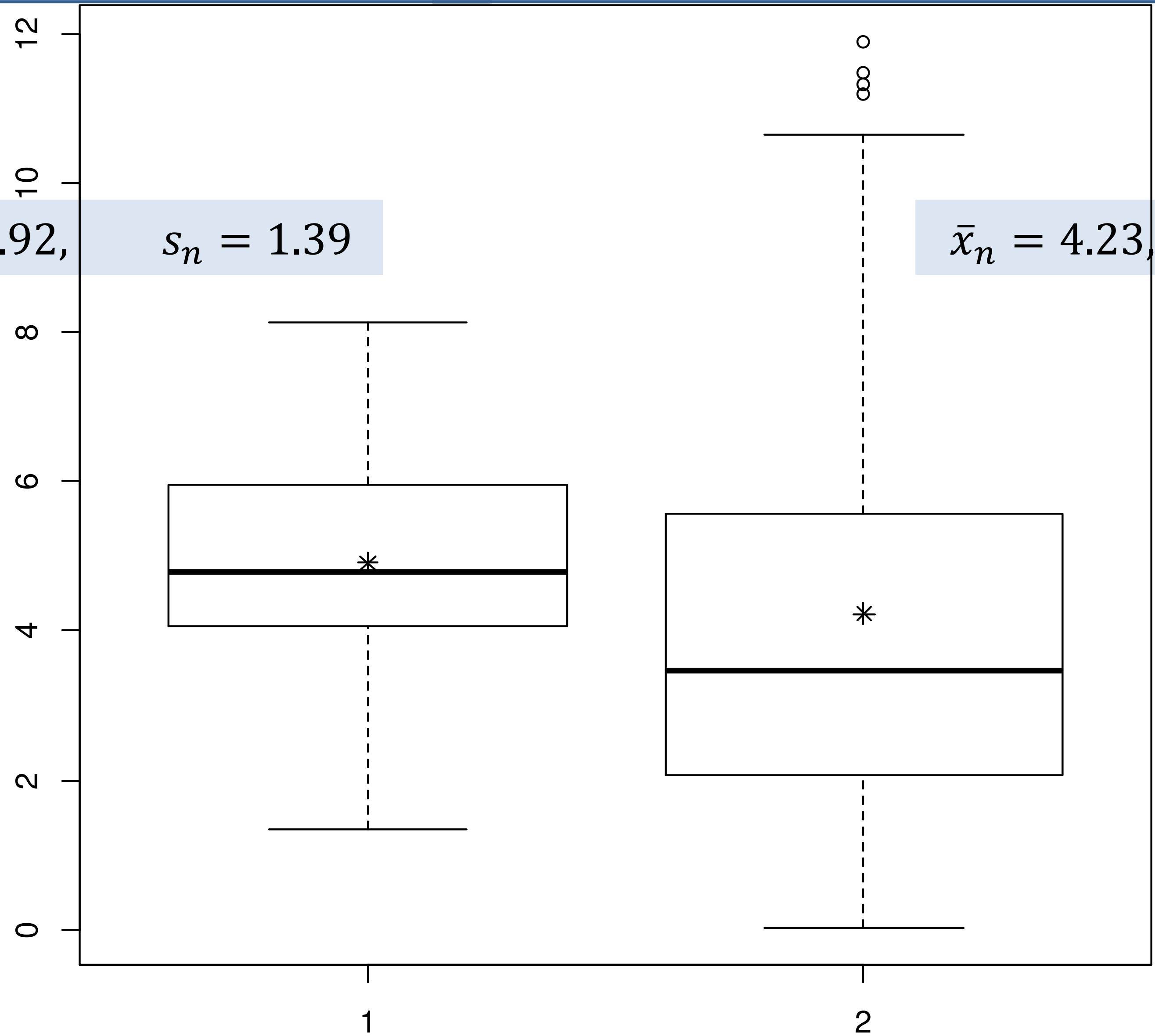
$\bar{x}_n = 4.23, \quad s_n = 2.87$



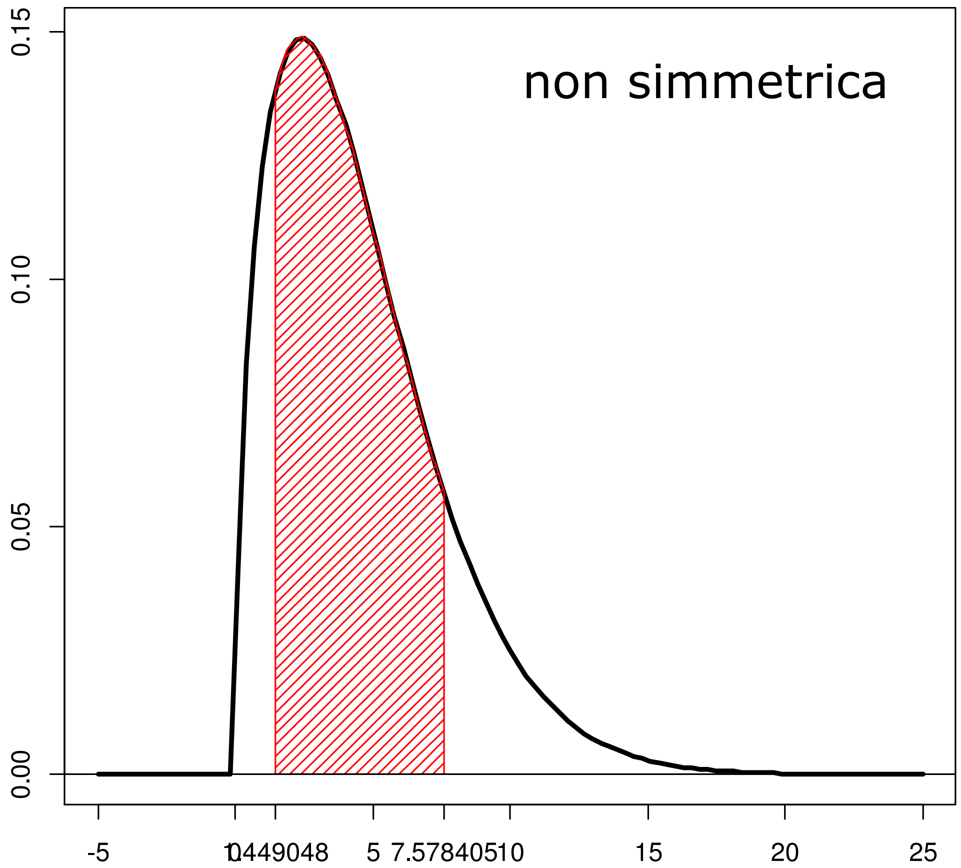
Per quale grafico $\bar{x}_n \mp s_n$ è rappresentativo della «posizione» dei dati?

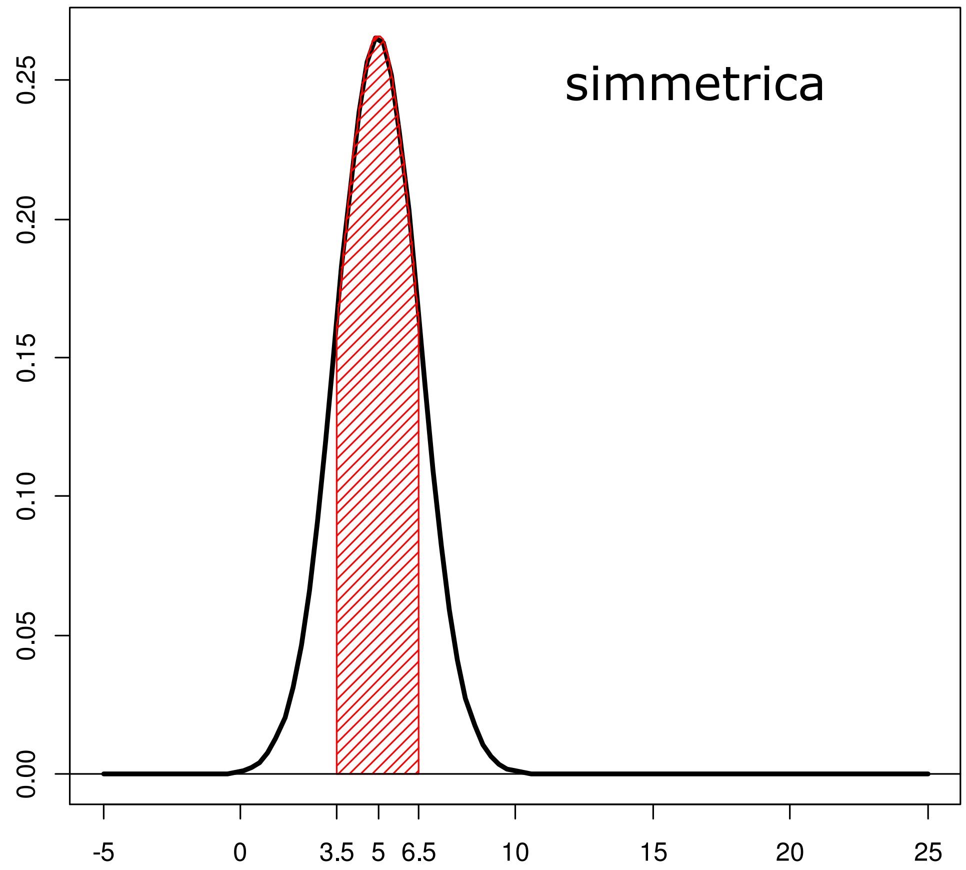
$\bar{x}_n = 4.92, \quad s_n = 1.39$

$\bar{x}_n = 4.23, \quad s_n = 2.87$



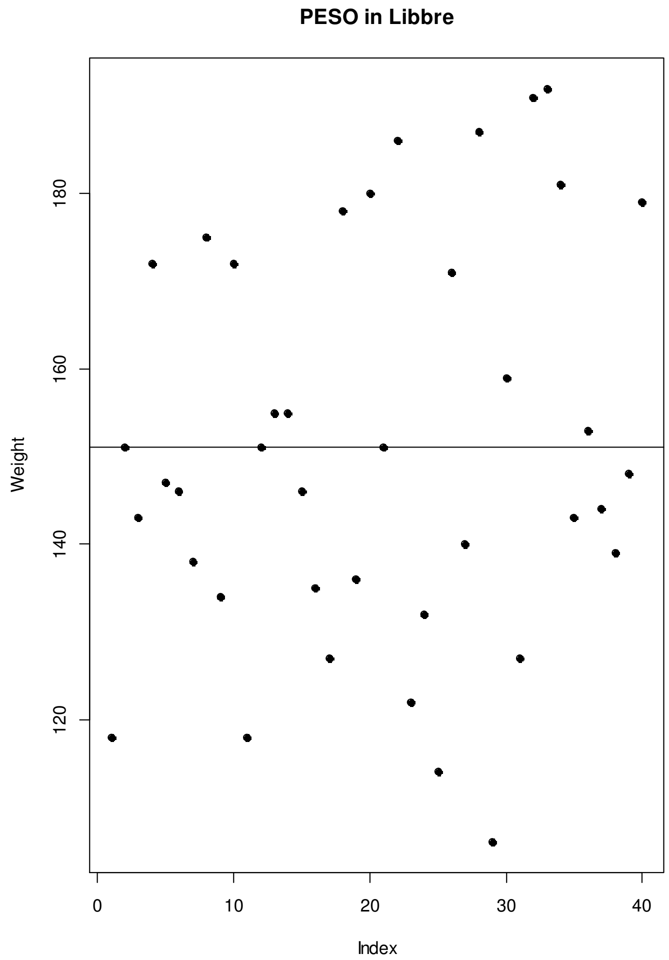
Sull’uso di $\bar{x}_n \mp s_n$





Il coefficiente di variazione

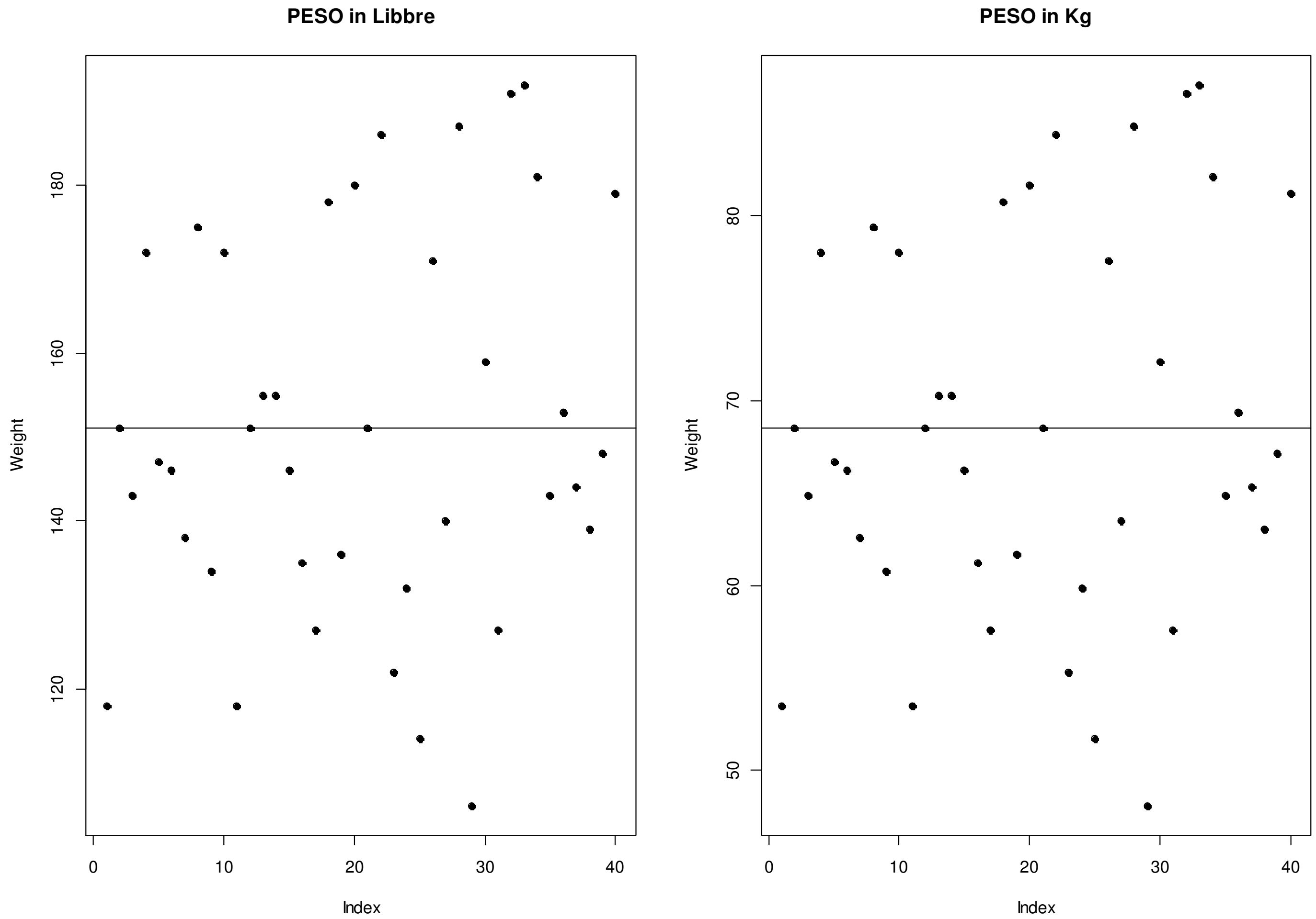
$s = 22.87$ libbre



Il coefficiente di variazione

$s = 22.87$ libbre

$s = 10.37$ Kg



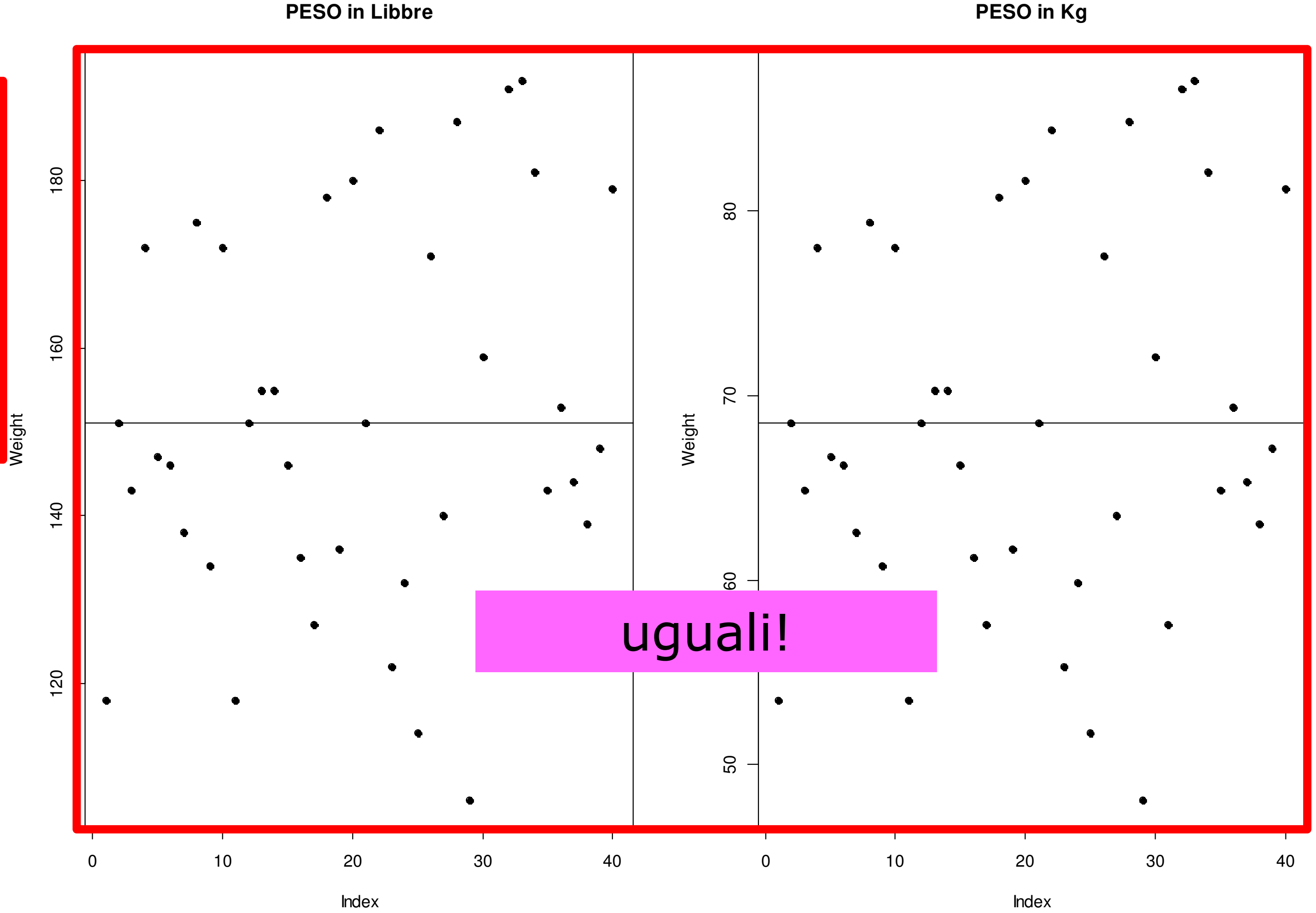
$$\boxed{W_{Kg} = 0.45359237 \times W_L}$$

Il coefficiente di variazione

$s = 22.87$ libbre

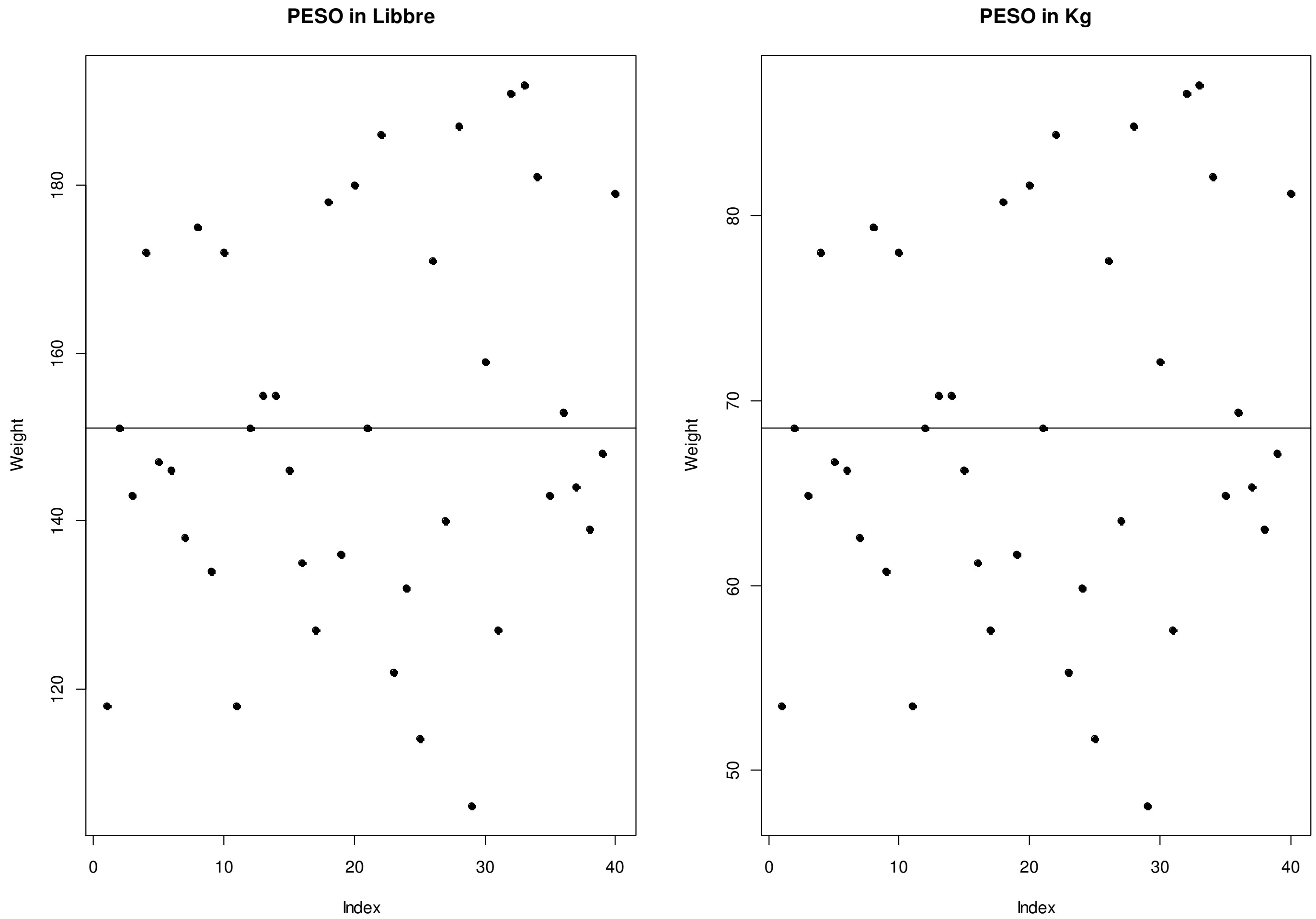
diversi!

$s = 10.37$ Kg



$$W_{Kg} = 0.45359237 \times W_L$$

Il coefficiente di variazione



$$\frac{\sigma}{|\mu|}$$

coeff. di variazione

$$W_{Kg} = 0.45359237 \times W_L$$

Il coefficiente di variazione

$\bar{w}_L = 151.05$ libbre

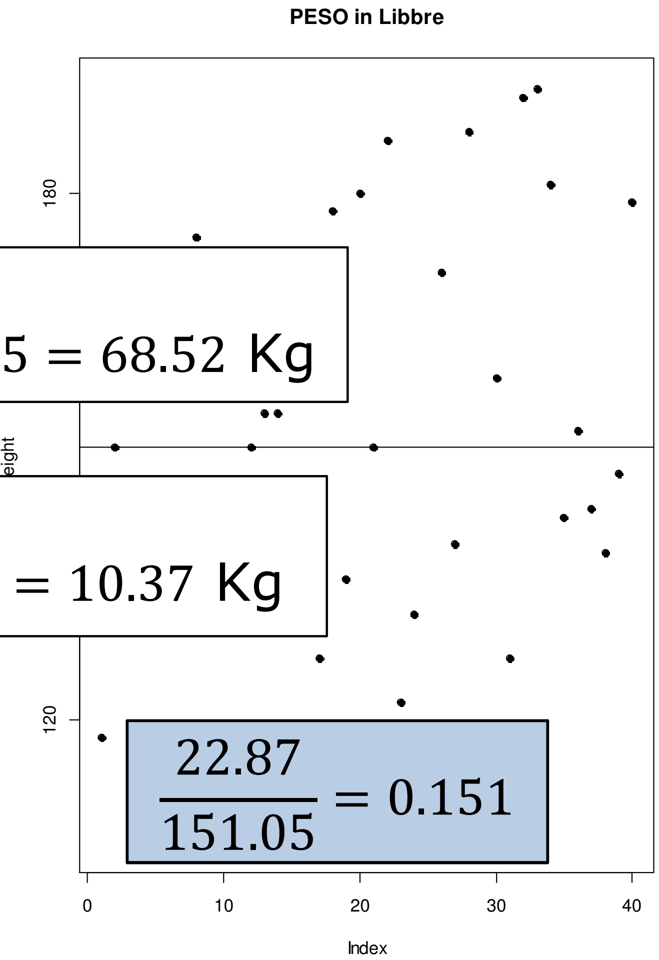
$\bar{w}_{Kg} = 0.45359237 \times 151.05 = 68.52$ Kg

$s_L = 22.87$ libbre

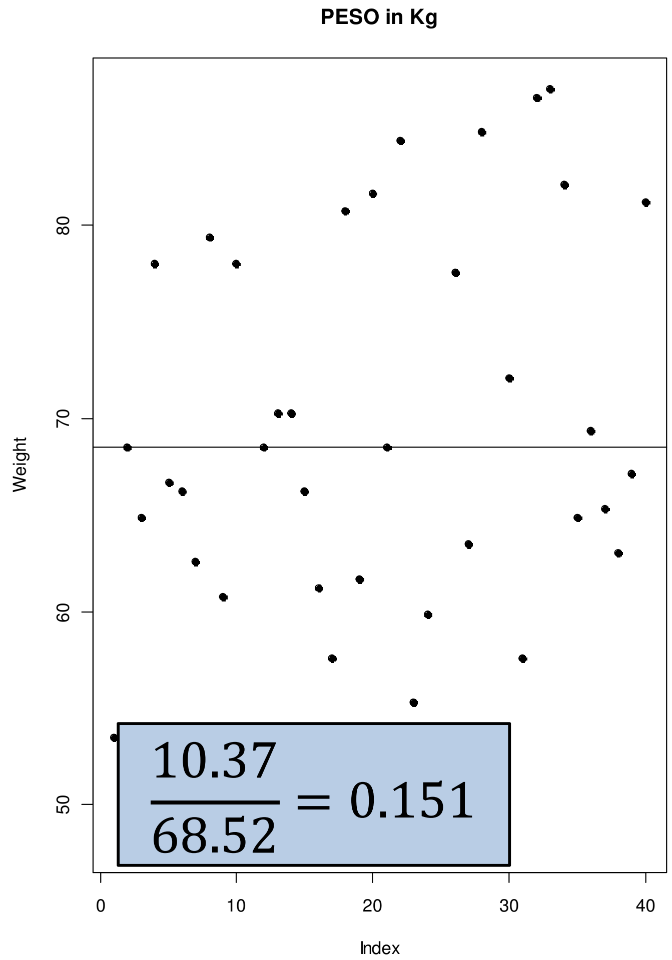
$s_{Kg} = 0.45359237 \times 22.87 = 10.37$ Kg

$$\frac{s}{|\bar{x}|}$$

coeff. di variazione campionario



$$\frac{22.87}{151.05} = 0.151$$



$$\frac{10.37}{68.52} = 0.151$$

$$W_{Kg} = 0.45359237 \times W_L$$

Il coefficiente di variazione

Es. 1 Livelli di colesterolo in due campioni, uno italiano (I) e uno tedesco (G):

$$\bar{x}_G = 210 \text{ mg/dl}, s_G = 30 \text{ mg/dl}$$

$$\bar{x}_I = 201 \text{ mg/dl}, s_I = 40 \text{ mg/dl}$$

Il coefficiente di variazione

Es. 1 Livelli di colesterolo in due campioni, uno italiano (I) e uno tedesco (G):

$$\bar{x}_G = 210 \text{ mg/dl}, s_G = 30 \text{ mg/dl} \Rightarrow \frac{30}{210} = 0.143$$

$$\bar{x}_I = 201 \text{ mg/dl}, s_I = 40 \text{ mg/dl} \Rightarrow \frac{40}{201} = 0.199$$

$\Rightarrow$ variabilità maggiore nella popolazione Italiana, una volta aggiustato rispetto alla media

Il coefficiente di variazione

Es. 1 Livelli di colesterolo in due campioni, uno italiano (I) e uno tedesco (G):

$$\bar{x}_G = 210 \text{ mg/dl}, s_G = 30 \text{ mg/dl} \quad \Rightarrow \quad \frac{30}{210} = 0.143$$

$$\bar{x}_I = 201 \text{ mg/dl}, s_I = 40 \text{ mg/dl} \quad \Rightarrow \quad \frac{40}{201} = 0.199$$

variabilità maggiore nella popolazione Italiana, una volta aggiustato rispetto alla media

Es. 2 Sulle stesse unità statistiche di cui ho misurato il peso ho misurato anche la dimensione del cervello (su una scala opportuna)

$$\bar{w} = 151.05 \text{ libbre}, \; s_w = 22.87 \text{ libbre}$$

$$\bar{b} = 908755 \; \ldots, \; s_b = 71372.8 \text{ n. pixel in una risonanza magnetica}$$

Il coefficiente di variazione

Es. 1 Livelli di colesterolo in due campioni, uno italiano (I) e uno tedesco (G):

$\bar{x}_G = 210 \text{ mg/dl}, s_G = 30 \text{ mg/dl}$ ⇨ $\frac{30}{210} = 0.143$

$\bar{x}_I = 201 \text{ mg/dl}, s_I = 40 \text{ mg/dl}$ ⇨ $\frac{40}{201} = 0.199$

⇨ variabilità maggiore nella popolazione Italiana, una volta aggiustato rispetto alla media

Es. 2 Sulle stesse unità statistiche di cui ho misurato il peso ho misurato anche la dimensione del cervello (su una scala opportuna)

$\bar{w} = 151.05 \text{ libbre}, \; s_w = 22.87 \text{ libbre}$ $\boxed{\frac{22.87}{151.05} = 0.151}$

$\bar{b} = 908755 \; \ldots, \; s_b = 71372.8 \text{ n. pixel in una risonanza magnetica}$ $\boxed{\frac{71372.8}{908755} = 0.078}$

Il coefficiente di variazione

Es. 1 Livelli di colesterolo in due campioni, uno italiano (I) e uno tedesco (G):

tà
nella
ne
una
stato
lla

Normativa europea per la valutazione sensoriale degli oli extravergine d'oliva prevede, perché la valutazione sia considerata valida, un coeff. variazione robusto (mediana al posto della media) inferiore al 20%.
(M. MAGLI, Ibimet)

Es. 2 Sulle stesse unità statistiche di cui ho misurato il peso ho misurato anche la dimensione del cervello (su una scala opportuna)

$\bar{w} = 151.05 \text{ libbre}, \ s_w = 22.87 \text{ libbre}$ $\boxed{\frac{22.87}{151.05} = 0.151}$

$\bar{b} = 908755 \ \ldots, \ s_b = 71372.8 \text{ n. pixel in una risonanza magnetica}$ $\boxed{\frac{71372.8}{908755} = 0.078}$

Rapporto di correlazione

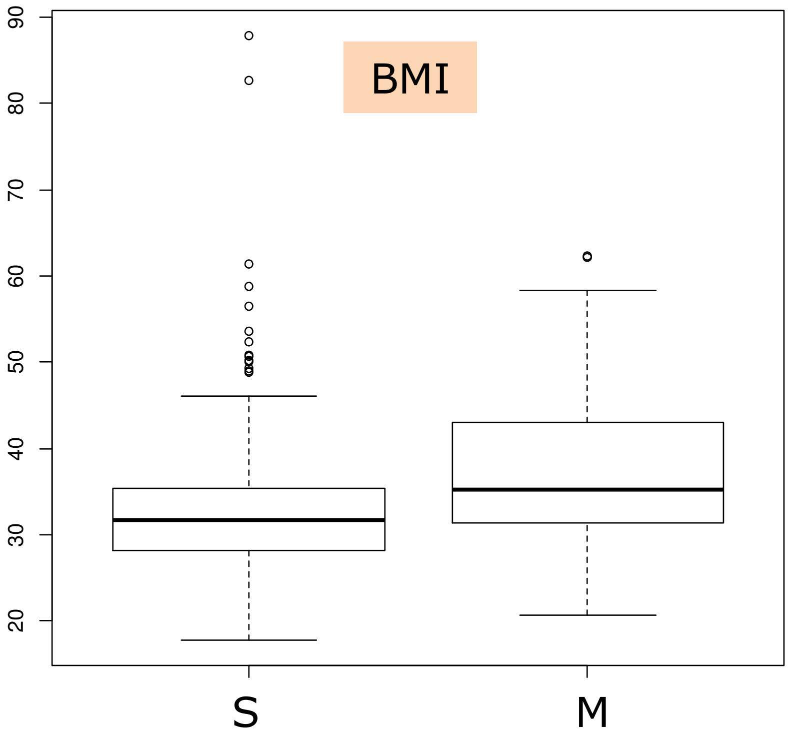
Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

$n_S = 392$
$\bar{x}_S = 32.76$
$\sigma_S = 7.38$

$n_M = 186$
$\bar{x}_M = 37.2$
$\sigma_M = 8.45$



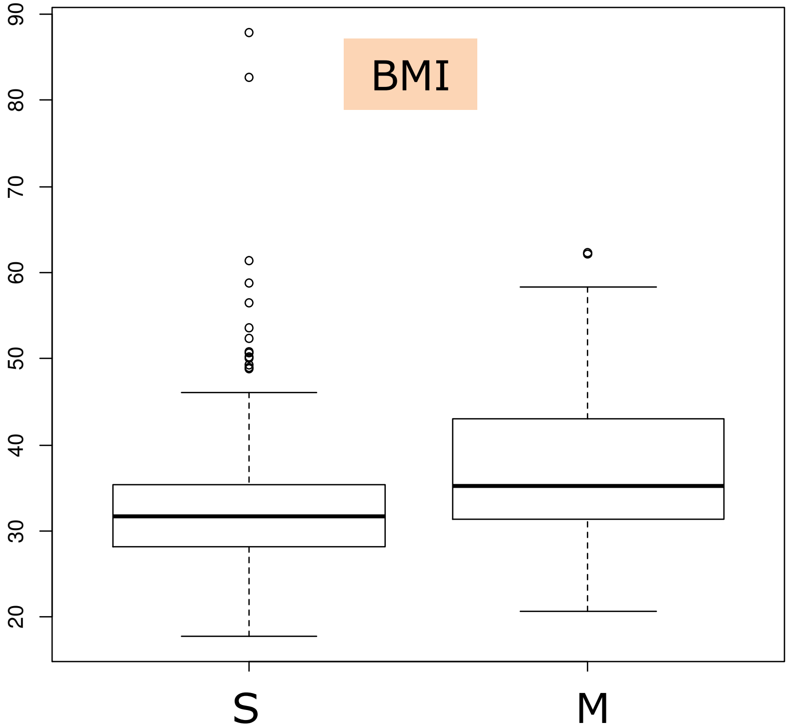
Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

$n_S = 392$
$\bar{x}_S = 32.76$
$\sigma_S = 7.38$

$n_M = 186$
$\bar{x}_M = 37.2$
$\sigma_M = 8.45$

Se BMI è ***mediamente*** maggiore tra i malati, possiamo dire che i dati sul BMI tendono ad essere **associati** con la malattia.

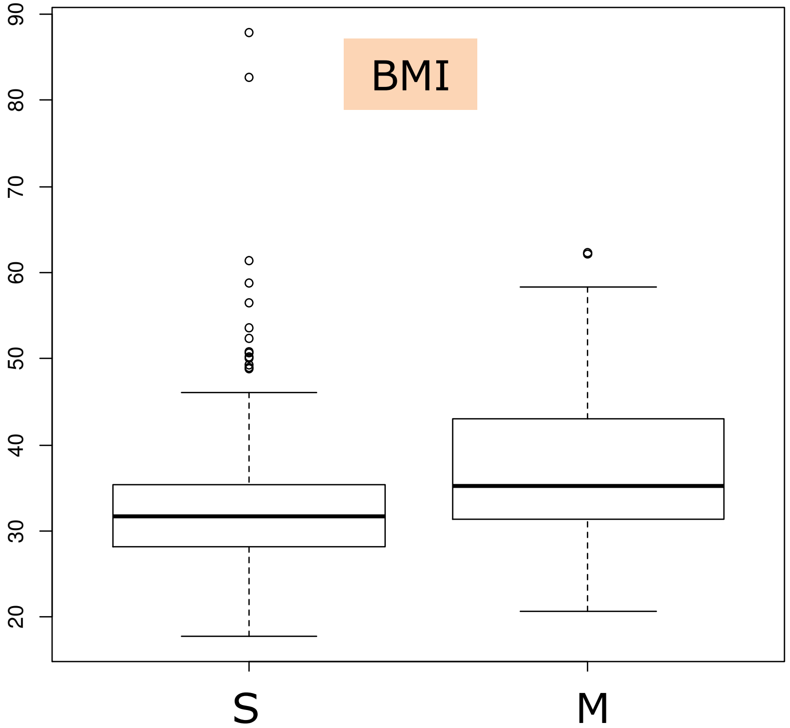


Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

$n_S = 392$
$\bar{x}_S = 32.76$
$\sigma_S = 7.38$

$n_M = 186$
$\bar{x}_M = 37.2$
$\sigma_M = 8.45$



Se BMI è ***mediamente*** maggiore tra i malati, possiamo dire che i dati sul BMI tendono ad essere **associati** con la malattia.

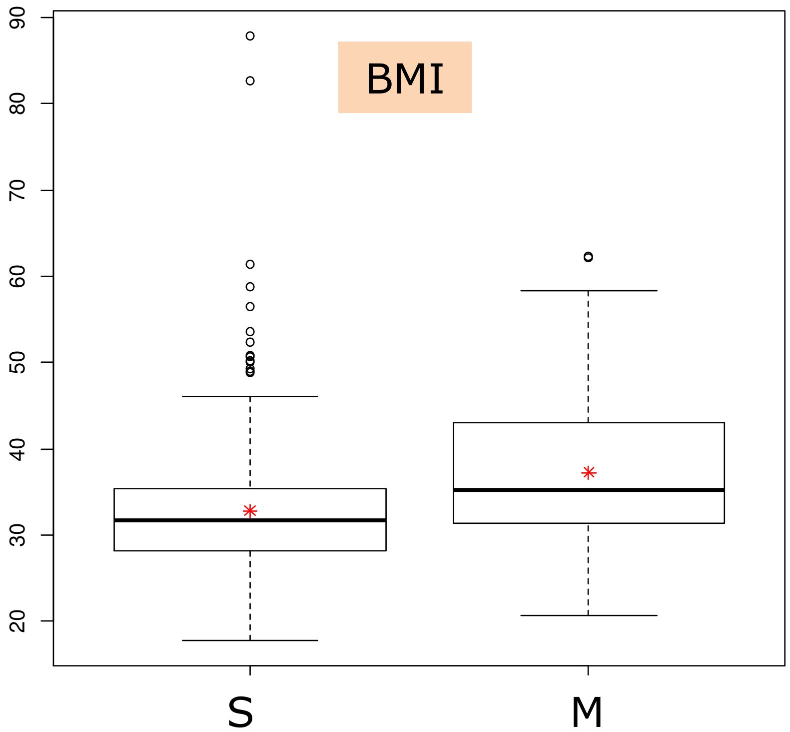
Confrontiamo le medie nei due gruppi (*medie condizionate*) tenendo conto della eventuale **diversa variabilità** (*varianze condizionate*) dei valori nei due gruppi.

Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

$n_S = 392$
$\bar{x}_S = 32.76$
$\sigma_S = 7.38$

$n_M = 186$
$\bar{x}_M = 37.2$
$\sigma_M = 8.45$



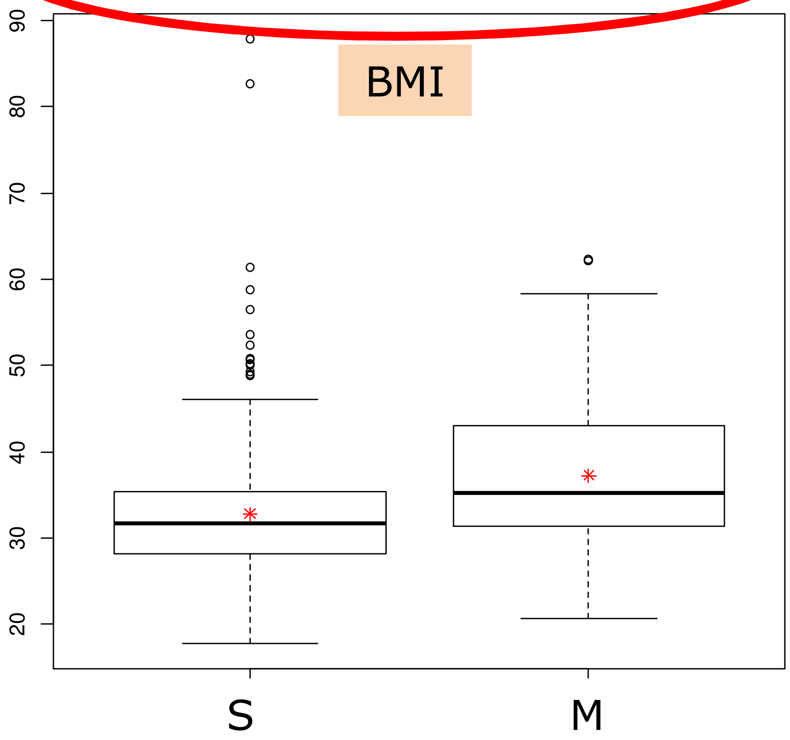
Confrontiamo le medie nei due gruppi (*medie condizionate*) tenendo conto della eventuale **diversa variabilità** (*varianze condizionate*) dei valori nei due gruppi.

Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

$n_S = 392$
$\bar{x}_S = 32.76$
$\sigma_S = 7.38$

$n_M = 186$
$\bar{x}_M = 37.2$
$\sigma_M = 8.45$



Confrontiamo le medie nei due gruppi (*medie condizionate*) tenendo conto della eventuale **diversa variabilità** (*varianze condizionate*) dei valori nei due gruppi.

Rapporto di correlazione

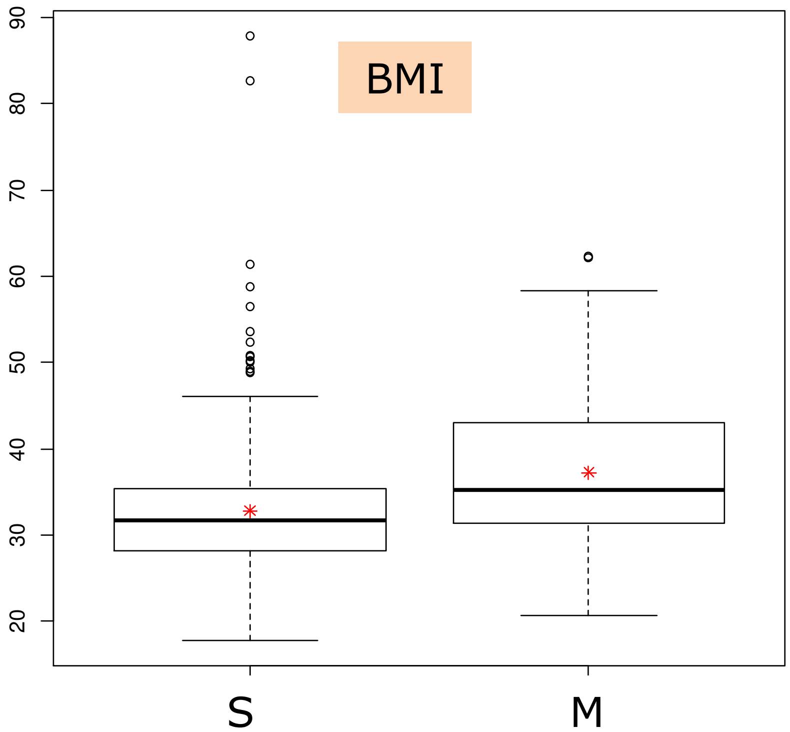
Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

$$\frac{\text{Variabilità delle medie nei gruppi rispetto alla media totale}}{\text{Variabilità totale}}$$

$n_S = 392$
$\bar{x}_S = 32.76$
$\sigma_S = 7.38$

$n_M = 186$
$\bar{x}_M = 37.2$
$\sigma_M = 8.45$

$$\frac{\text{Varianza } \textbf{between}}{\text{Varianza } \textbf{between} + \text{Varianza } \textbf{within}}$$



Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

Variabilità delle medie nei gruppi rispetto alla media totale

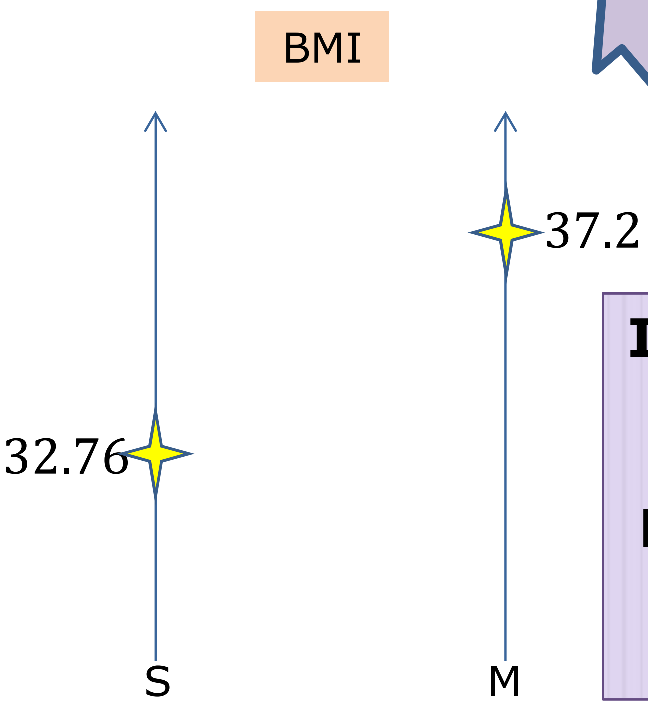
Variabilità totale

Varia

Varianza *betw* *thin*

$n_S = 392$
$\bar{x}_S = 32.76$
$\boldsymbol{\sigma_S = 0!}$

$n_M = 186$
$\bar{x}_M = 37.2$
$\boldsymbol{\sigma_M = 0!}$



IMMAGINIAMO CHE IN CIASCUN GRUPPO TUTTI I DATI COINCIDANO CON LA MEDIA IN QUEL GRUPPO

Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

Variabilità delle medie nei gruppi rispetto alla media totale

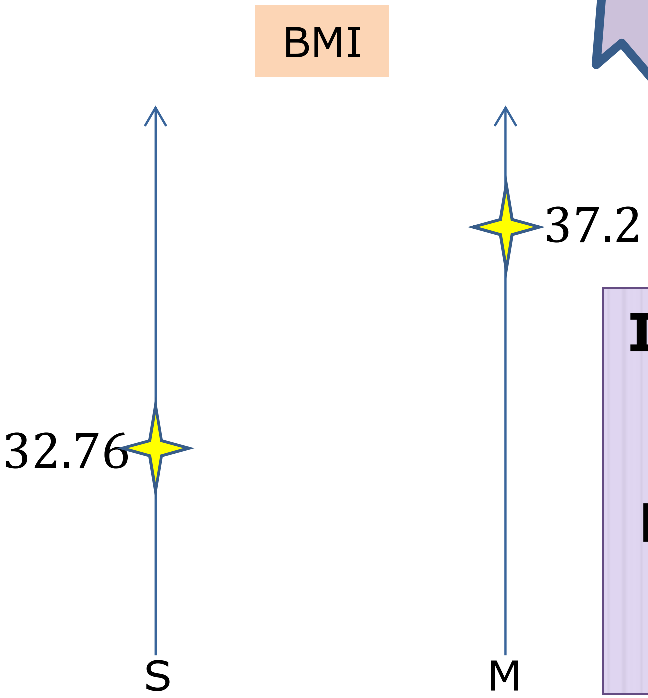
Variabilità totale

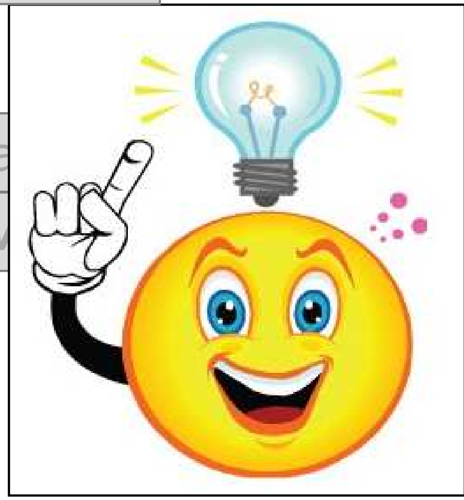
Varia

Varianza ***betw*** ***thin***

$n_S = 392$
$\bar{x}_S = 32.76$
$\boldsymbol{\sigma_S = 0}!$

$n_M = 186$
$\bar{x}_M = 37.2$
$\boldsymbol{\sigma_M = 0}!$





IMMAGINIAMO CHE IN CIASCUN GRUPPO TUTTI I DATI COINCIDANO CON LA MEDIA IN QUEL GRUPPO

LA SOLA VARIABILITA' NEI DATI E' DOVUTA ALLA DIVISIONE NEI DUE GRUPPI!

Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

Variabilità delle medie nei gruppi rispetto alla media totale

Variabilità totale

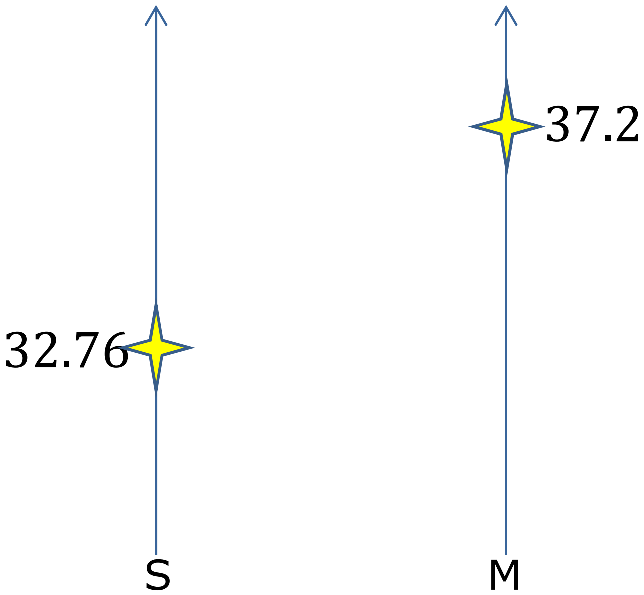
Varianza ***between***

Varianza ***between*** + Varianza ***within***

$n_S = 392$
$\bar{x}_S = 32.76$
$\boldsymbol{\sigma_S = 0!}$

$n_M = 186$
$\bar{x}_M = 37.2$
$\boldsymbol{\sigma_M = 0!}$

BMI



$$\bar{x} = \frac{392 \times 32.76 + 186 \times 37.2}{392 + 186} = 34.19$$

Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

Variabilità delle medie nei gruppi rispetto alla media totale

Variabilità totale

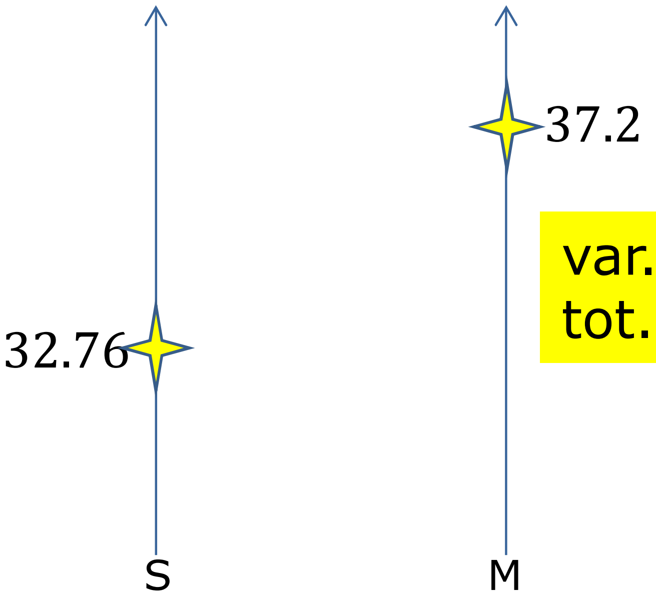
Varianza ***between***

Varianza ***between*** + Varianza ***within***

$n_S = 392$
$\bar{x}_S = 32.76$
$\boldsymbol{\sigma_S = 0}$!

$n_M = 186$
$\bar{x}_M = 37.2$
$\boldsymbol{\sigma_M = 0}$!

BMI



$$\bar{x} = \frac{392 \times 32.76 + 186 \times 37.2}{392 + 186} = 34.19$$

$$\text{var. tot.} = \frac{392 \times (32.76 - 34.19)^2 + 186 \times (37.2 - 34.19)^2}{392 + 186}$$

Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

Variabilità delle medie nei gruppi rispetto alla media totale

Variabilità totale

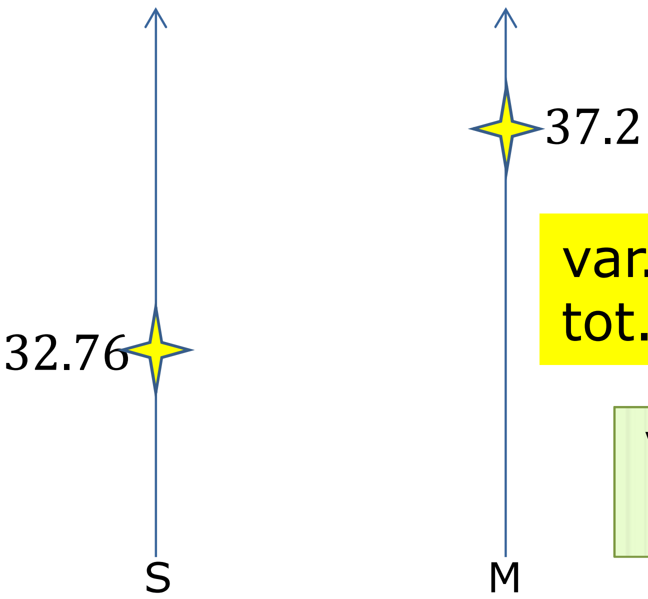
Varianza ***between***

Varianza ***between*** + Varianza ***within***

$n_S = 392$
$\bar{x}_S = 32.76$
$\boldsymbol{\sigma_S = 0}$!

$n_M = 186$
$\bar{x}_M = 37.2$
$\boldsymbol{\sigma_M = 0}$!

BMI



$$\bar{x} = \frac{392 \times 32.76 + 186 \times 37.2}{392 + 186} = 34.19$$

$$\text{var. tot.} = \frac{392 \times (32.76 - 34.19)^2 + 186 \times (37.2 - 34.19)^2}{392 + 186}$$

Variabilità totale = Varianza between !
più è grande e più sono diverse

Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

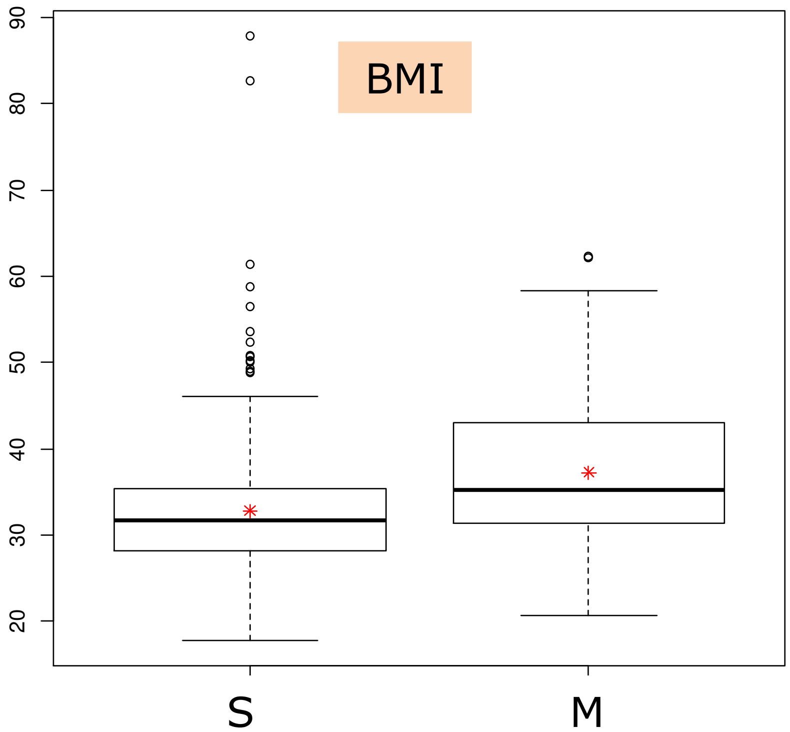
$$\frac{\text{Variabilità delle medie nei gruppi rispetto alla media totale}}{\text{Variabilità totale}}$$

$n_S = 392$ $\quad$ $n_M = 186$

$\bar{x}_S = 32.76$ $\quad$ $\bar{x}_M = 37.2$

$\sigma_S = 7.38$ $\quad$ $\sigma_M = 8.45$

$$\frac{\text{Varianza } \textbf{\textit{between}}}{\text{Varianza } \textbf{\textit{between}} + \text{Varianza } \textbf{\textit{within}}}$$



$$\frac{n_M}{n}(\bar{x}_M - \bar{x})^2 + \frac{n_F}{n}(\bar{x}_F - \bar{x})^2$$

$\bar{x}$ = media complessiva

Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

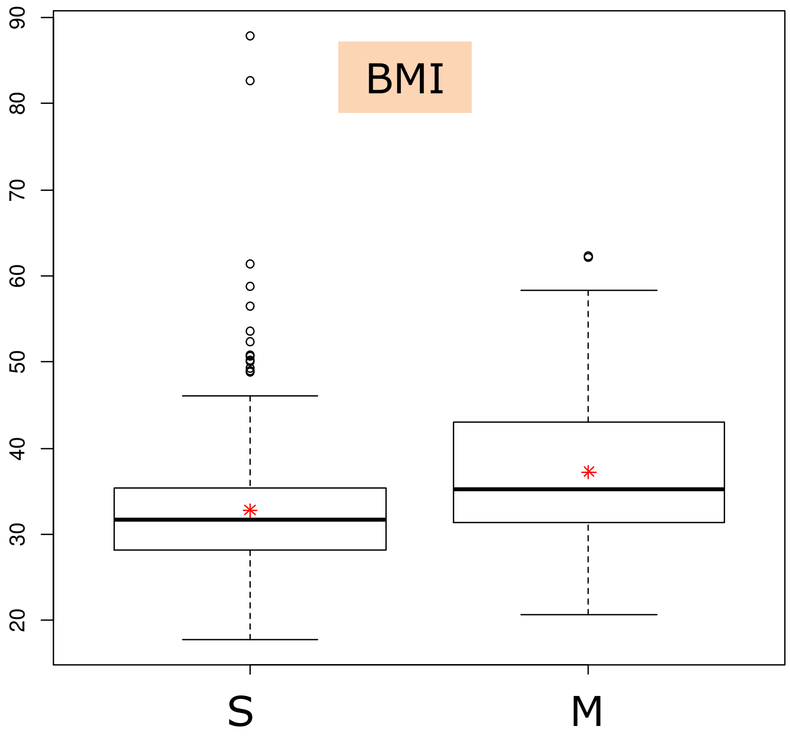
$$\frac{\text{Variabilità delle medie nei gruppi rispetto alla media totale}}{\text{Variabilità totale}}$$

$n_S = 392$ $\quad n_M = 186$

$\bar{x}_S = 32.76$ $\quad \bar{x}_M = 37.2$

$\sigma_S = 7.38$ $\quad \sigma_M = 8.45$

$$\frac{\text{Varianza } \textbf{\textit{between}}}{\text{Varianza } \textbf{\textit{between}} + \text{Varianza } \textbf{\textit{within}}}$$



$$\frac{n_M}{n}(\bar{x}_M - \bar{x})^2 + \frac{n_F}{n}(\bar{x}_F - \bar{x})^2$$

$$\bar{x} = \frac{392 \times 32.76 + 186 \times 37.2}{392 + 186} = 34.19$$

$$\frac{392}{392+186}(32.76 - 34.19)^2 + \frac{186}{392+186}(37.2 - 34.19)^2$$

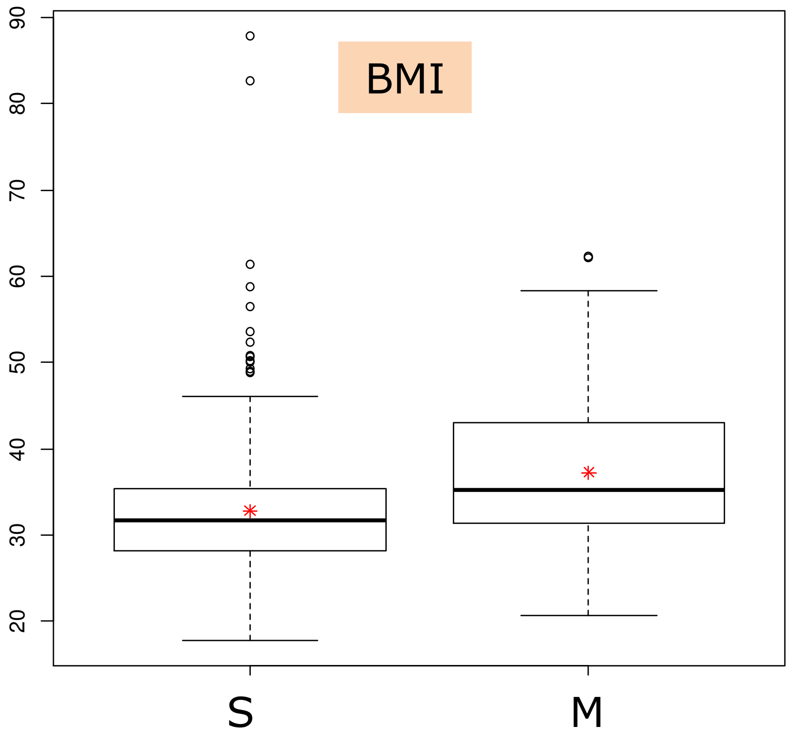
Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

$$\frac{\text{Variabilità delle medie nei gruppi rispetto alla media totale}}{\text{Variabilità totale}}$$

$n_S = 392$ | $n_M = 186$
$\bar{x}_S = 32.76$ | $\bar{x}_M = 37.2$
$\sigma_S = 7.38$ | $\sigma_M = 8.45$

$$\frac{\text{Varianza } \textbf{\textit{between}}}{\text{Varianza } \textbf{\textit{between}} + \text{Varianza } \textbf{\textit{within}}}$$



$$\frac{n_M}{n}(\bar{x}_M - \bar{x})^2 + \frac{n_F}{n}(\bar{x}_F - \bar{x})^2 = 4.30$$

$$\bar{x} = \frac{392 \times 32.76 + 186 \times 37.2}{392 + 186} = 34.19$$

$$\frac{392}{392 + 186}(32.76 - 34.19)^2 + \frac{186}{392 + 186}(37.2 - 34.19)^2$$

Rapporto di correlazione

Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

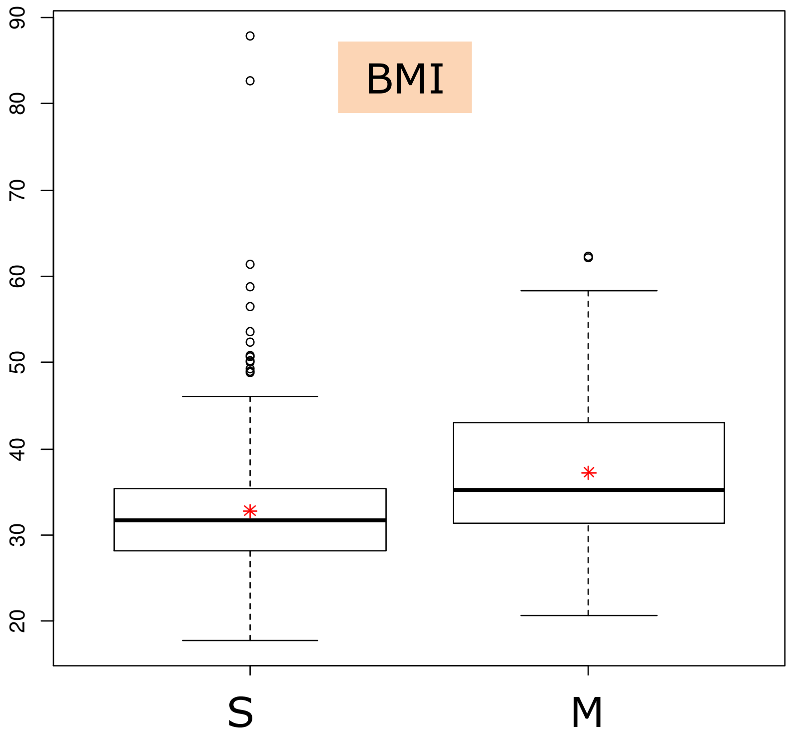
Variabilità delle medie nei gruppi rispetto alla media totale

Variabilità totale $= \frac{1}{n_M + n_F} \sum (x_i - \bar{x})^2$

$$n_S = 392 \qquad n_M = 186$$
$$\bar{x}_S = 32.76 \qquad \bar{x}_M = 37.2$$
$$\sigma_S = 7.38 \qquad \sigma_M = 8.45$$

Varianza ***between***

Varianza ***between*** **+** Varianza ***within***



$$\frac{n_M}{n}(\bar{x}_M - \bar{x})^2 + \frac{n_F}{n}(\bar{x}_F - \bar{x})^2$$

$$\frac{n_M}{n} \times \sigma^2{}_M + \frac{n_F}{n} \times \sigma^2{}_F$$

$$\frac{392}{392+186} 7.38^2 + \frac{186}{392+186} 8.45^2 = 59.91$$

Rapporto di correlazione

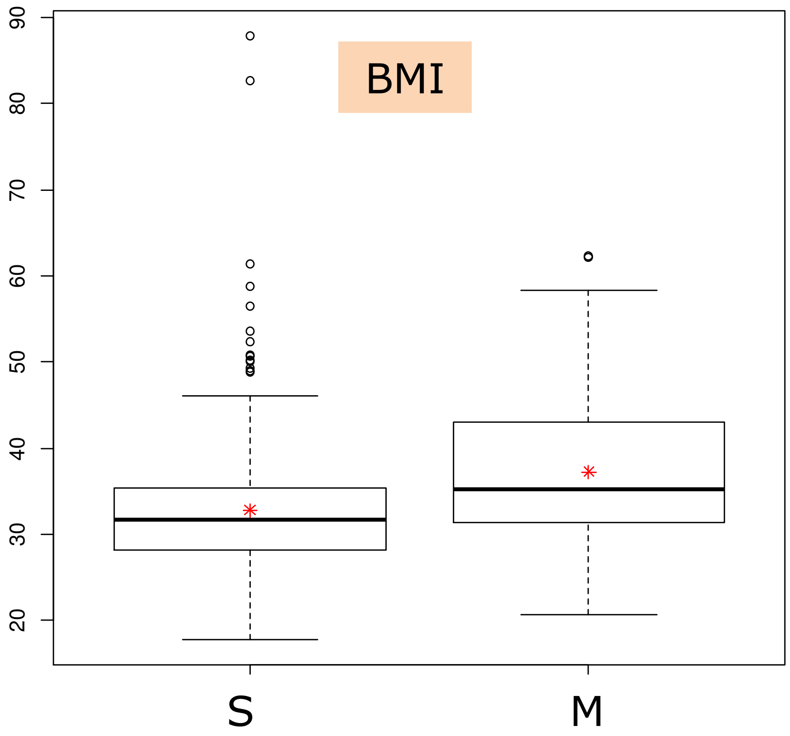
Il peso è lo stesso tra i malati ed i non malati di una certa patologia?

$$\frac{\text{Variabilità delle medie nei gruppi rispetto alla media totale}}{\text{Variabilità totale}}$$

$n_S = 392$ $\quad$ $n_M = 186$

$\bar{x}_S = 32.76$ $\quad$ $\bar{x}_M = 37.2$

$\sigma_S = 7.38$ $\quad$ $\sigma_M = 8.45$



$$\frac{\text{Varianza } \textbf{\textit{between}}}{\text{Varianza } \textbf{\textit{between}} + \text{Varianza } \textbf{\textit{within}}}$$

$$\eta^2 = \frac{4.30}{4.30+59.91} = 0.067$$

$$\boxed{0 \leq \eta^2 \leq 1}$$

Analisi della varianza

K gruppi di dati, anche di diverse numerosità: $n_1, \dots, n_K$

l'analisi della varianza riguarda le **medie**!

$\bar{x}_i$ e σ^2_i media e varianza distorta(!) nel gruppo *i*-mo

Variabilità delle medie nei **gruppi** rispetto alla **media totale**

media totale: $\bar{x} = \dfrac{n_1\bar{x}_1 + \cdots + n_K\bar{x}_K}{n_1 + \cdots + n_K}$

varianza totale: $\sigma^2 = \sigma_W{}^2 + \sigma_B{}^2$

Within

Between

media pesata *delle varianze* nei gruppi

normale formula della varianza, ottenuta *come se* tutti i dati in ciascun gruppo coincidessero con la media nel gruppo: *varianza delle medie*

$$\sigma_W{}^2 = \frac{1}{n}\sum_{i=1,\dots,K} n_i\sigma^2{}_i$$

(per K=2, è un punto interno, cioè *within*)

$$\sigma_B{}^2 = \frac{1}{n}\sum_{i=1,\dots,K} n_i(\bar{x}_i - \bar{x})^2$$

è piccola se le $\bar{x}_i$ **non** sono molto diverse tra di loro

Analisi della varianza

K gruppi di dati, anche di diverse numerosità: $n_1, \ldots, n_K$

l'analisi della varianza riguarda le **medie**!

$\bar{x}_i$ e σ^2_i media e varianza distorta(!) nel gruppo *i*-mo

Variabilità delle medie nei **gruppi** rispetto alla **media totale**

media totale: $\bar{x} = \dfrac{n_1\bar{x}_1 + \cdots + n_K\bar{x}_K}{n_1 + \cdots + n_K}$

varianza totale: $\sigma^2 = \sigma_W{}^2 + \sigma_B{}^2$

Within

Between

$$\frac{\sigma_B{}^2}{\sigma_W{}^2 + \sigma_B{}^2}$$

media pesata *delle varianze* nei gruppi

(per *K*=2, è un punto interno, cioè *within*)

$$\sigma_W{}^2 = \frac{1}{n}\sum_{i=1,\ldots,K} n_i\sigma^2_i$$

normale formula della varianza, ottenuta *come se* tutti i dati in ciascun gruppo coincidessero con la media nel gruppo: *varianza delle medie*

$$\sigma_B{}^2 = \frac{1}{n}\sum_{i=1,\ldots,K} n_i(\bar{x}_i - \bar{x})^2$$

è piccola se le $\bar{x}_i$ **non** sono molto diverse tra di loro