# STATISTICA

Esercizi
regressione

# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, $X$ ed $Y$:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

1. Supposto $Y$ gaussiana, si verifichi l'ipotesi che la media $\mu_Y$ della popolazione di riferimento sia pari a 100 al livello di significatività dell'1%.

# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

1. Supposto $Y$ gaussiana, si verifichi l'ipotesi che la media $\mu_Y$ della popolazione di riferimento sia pari a 100 al livello di significatività dell'1%.

$$H_0\colon \mu_Y = 100 \;\; ; \quad H_1 : \mu_Y \neq 100$$

$$\bar{y} = 95.82, \qquad s_Y{}^2 = \frac{1}{8}\sum_{i=1}^{9}(y_i - \bar{y})^2 = 674.507 \quad \Rightarrow \quad \left|\frac{95.82 - 100}{\sqrt{\frac{674.507}{9}}}\right| = 0.48$$

non possiamo rifiutare $H_0$ al livello dell'1%. $\Leftarrow$ $t(8)_{0.01/2} = t(8)_{0.005} = 3.355387$

# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, $X$ ed $Y$:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

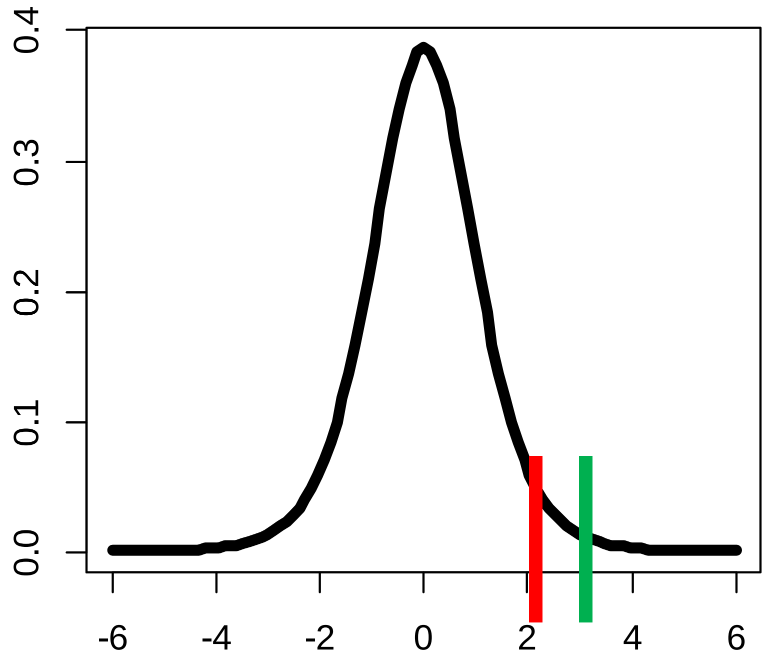
2. Se non possiamo rifiutare $H_0$ al livello di significatività dell'1% cosa potremo dire al livello di significatività del 5%?

# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

2. Se non possiamo rifiutare $H_0$ al livello di significatività dell'1% cosa potremo dire al livello di significatività del 5%?



# Esercizio 1

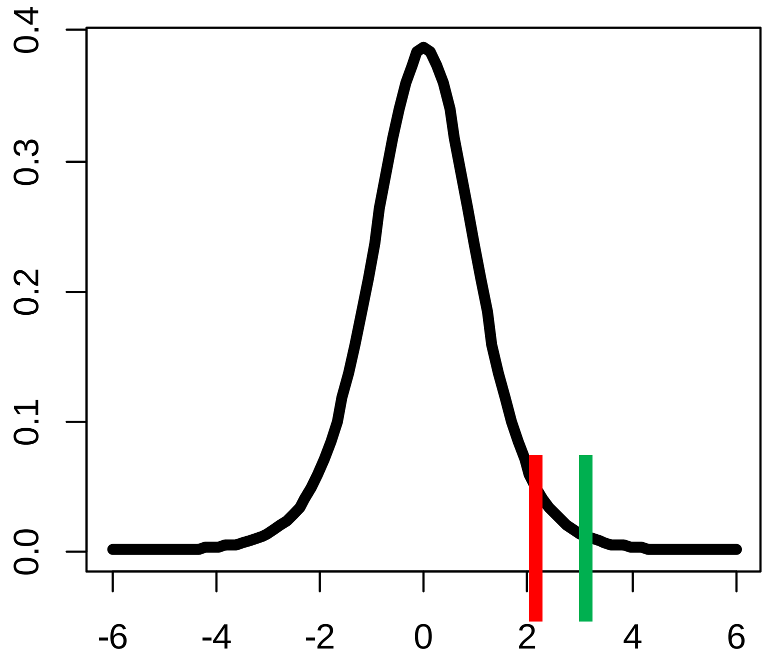
Su un campione di $n = 9$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

2. Se non possiamo rifiutare $H_0$ al livello di significatività dell'1% cosa potremo dire al livello di significatività del 5%?

Potrebbe essere che al livello del 5% si possa rifiutare! Basta che la statistica test cada tra

$t(8)_{0.025} = 2.306004$ e $t(8)_{0.005} = 3.355387$



# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

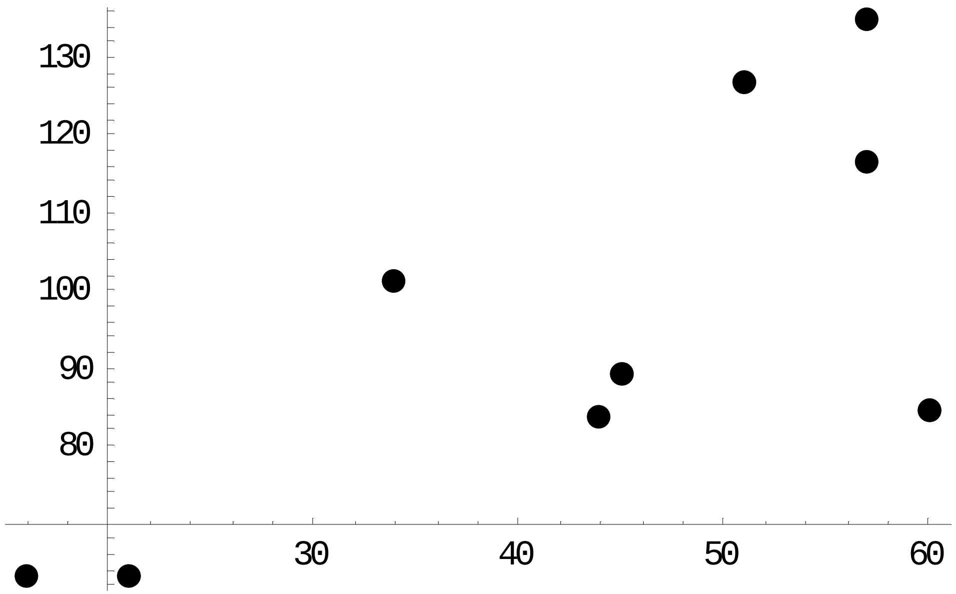
3. Si ritiene che la variabilità di *Y* possa essere spiegata in funzione di *X*. Rappresentare la distribuzione congiunta di *X* ed *Y* mediante un opportuno grafico: le due variabili sono correlate? Come e quanto?

# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, $X$ ed $Y$:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

3. Si ritiene che la variabilità di $Y$ possa essere spiegata in funzione di $X$. Rappresentare la distribuzione congiunta di $X$ ed $Y$ mediante un opportuno grafico: le due variabili sono correlate? Come e quanto?



# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

3. Si ritiene che la variabilità di *Y* possa essere spiegata in funzione di *X*. Rappresentare la distribuzione congiunta di *X* ed *Y* mediante un opportuno grafico: le due variabili sono correlate? Come e quanto?

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| XY | 1006.4 | 1329.3 | 3669.6 | 6640.5 | 6461.7 | 5070.0 | 7683.6 | 3437.4 | 4014.0 |

$\bar{y} = 95.82$

$\bar{x} = 42.78$

$$cov(x, y) = \frac{1}{9}\sum_{i=1}^{9} x_i y_i - \bar{x}\bar{y} = 4368.06 - 4099.18 = 268.88$$

# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

3. Si ritiene che la variabilità di *Y* possa essere spiegata in funzione di *X*. Rappresentare la distribuzione congiunta di *X* ed *Y* mediante un opportuno grafico: le due variabili sono correlate? Come e quanto?

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| XY | 1006.4 | 1329.3 | 3669.6 | 6640.5 | 6461.7 | 5070.0 | 7683.6 | 3437.4 | 4014.0 |

$$\bar{y} = 95.82$$

$$\bar{x} = 42.78$$

$$cov(x, y) = \frac{1}{9}\sum_{i=1}^{9} x_i y_i - \bar{x}\bar{y} = 4368.06 - 4099.18 = 268.88$$

$$\sigma_x{}^2 = \frac{1}{9}\sum_{i=1}^{9} (x_i - \bar{x})^2 = 227.06$$

$$\sigma_y{}^2 = 599.56$$

$$\rho = \frac{268.88}{\sqrt{227.06 \times 599.56}} = \mathbf{0.73}$$

# Esercizio 1

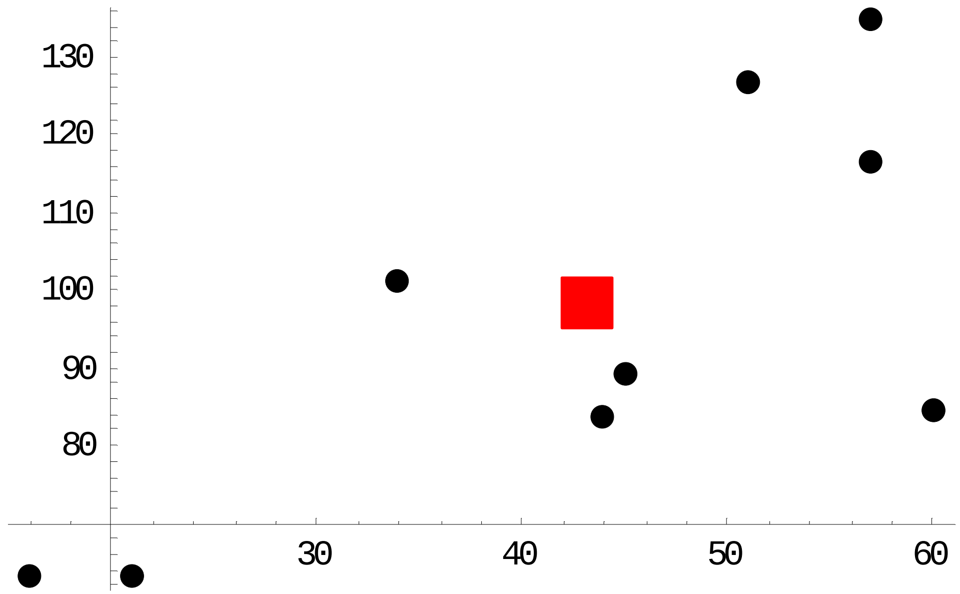
Su un campione di $n = 9$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

3. Si ritiene che la variabilità di *Y* possa essere spiegata in funzione di *X*. Rappresentare la distribuzione congiunta di *X* ed *Y* mediante un opportuno grafico: le due variabili sono correlate? Come e quanto?

$\bar{y} = 95.82$

$\bar{x} = 42.78$




$\rho = \mathbf{0.73}$

# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

4. Calcolare le stime dei minimi quadrati dei parametri $a$ e $b$ del modello di regressione lineare $y_i = a + bx_i + \varepsilon_i$

# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

4. Calcolare le stime dei minimi quadrati dei parametri $a$ e $b$ del modello di regressione lineare $y_i = a + bx_i + \varepsilon_i$

$$\bar{y} = 95.82$$

$$\bar{x} = 42.78$$

$$\sigma_x{}^2 = \frac{1}{9}\sum_{i=1}^{9}(x_i - \bar{x})^2 = 227.06$$

$$cov(x,y) = \frac{1}{9}\sum_{i=1}^{9} x_i y_i - \bar{x}\bar{y} = 268.88$$

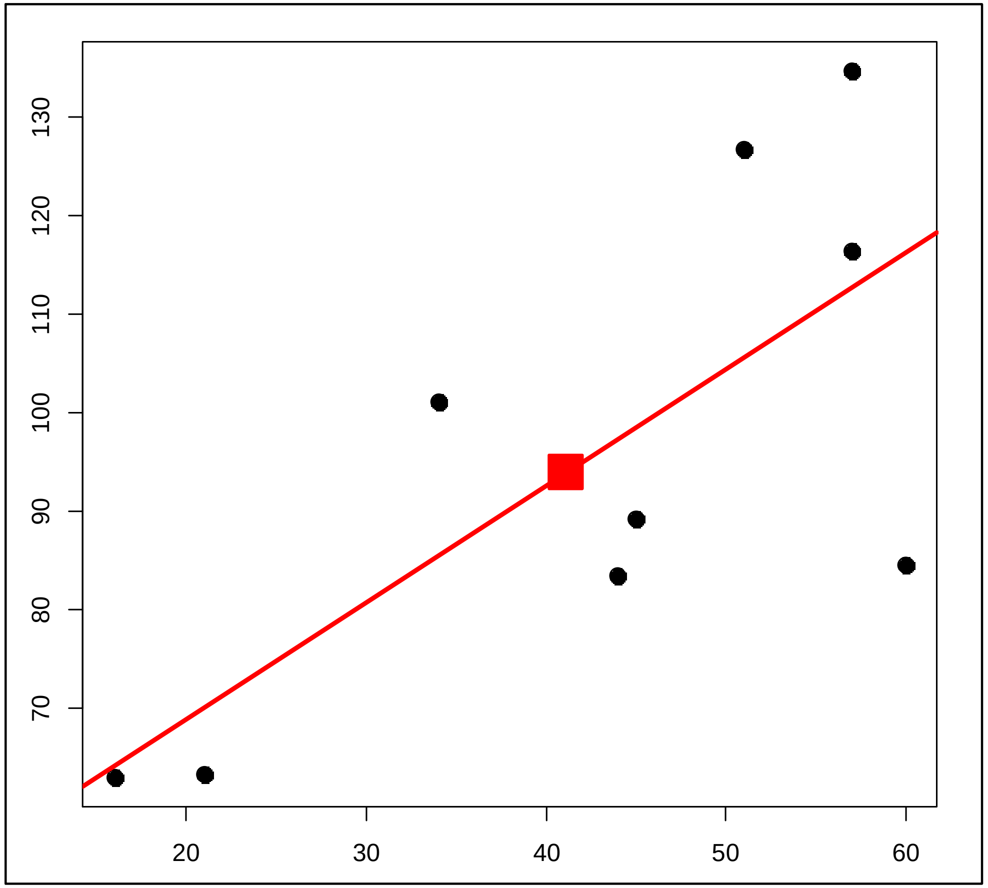
$$\hat{b} = \frac{268.88}{227.06} = 1.184$$

$$\hat{a} = 95.82 - 1.184 \times 42.78 = 45.17$$

# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, $X$ ed $Y$:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

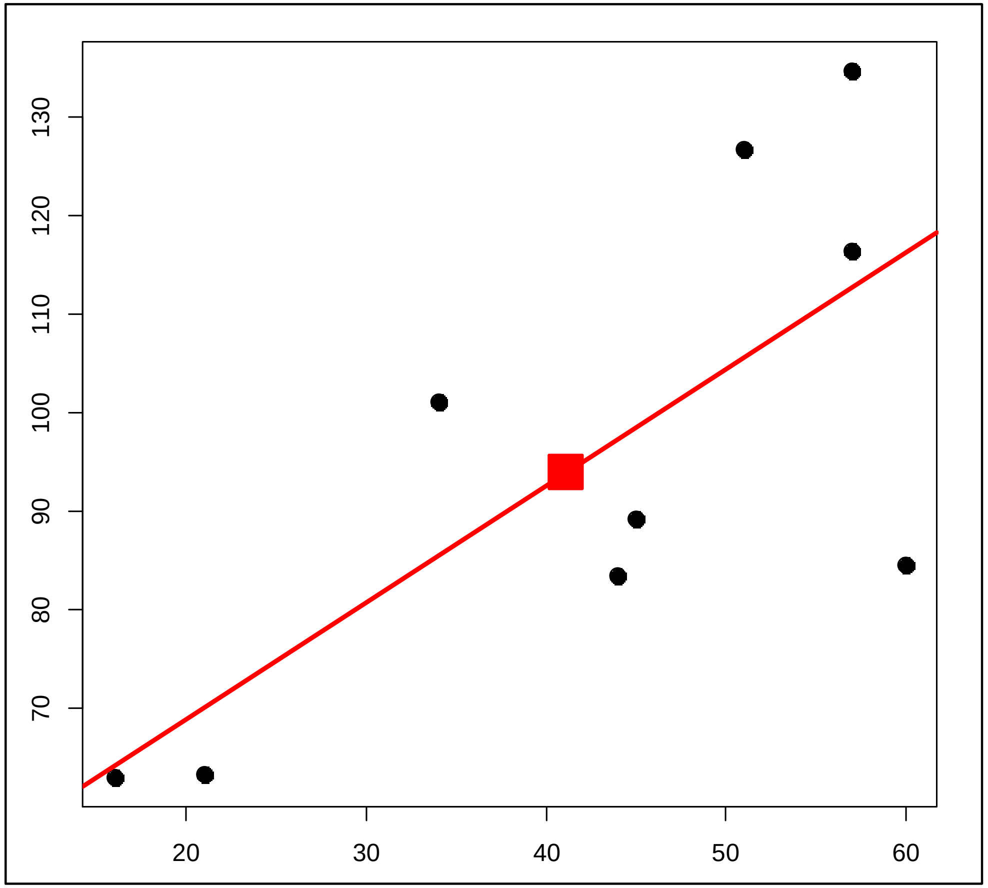


# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, $X$ ed $Y$:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

5. Calcolare una misura della bontà di adattamento del modello ai dati




$$R^2 = 0.73^2 = 0.53$$

Il valore è medio: **non** è indicazione forte di bontà di adattamento del modello ai dati

# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

6. Testare la significatività della regressione al livello si significatività del 5%

# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

$$\left| \frac{\hat{\boldsymbol{b}}}{\sqrt{\frac{s^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} \right| > t(n-2)_{\alpha/2}$$

$$s^2 = \frac{1}{n-2}\sum {e_i}^2$$

# E' molto utile sapere che

$$\sum_i (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i {e_i}^2$$

**SST = SSR + SSE**

$$\boxed{R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}}}$$

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{(n-2)s^2}{n{\sigma_y}^2}$$

$$\boxed{s^2 = \frac{n}{n-2} \times (1 - R^2) \times {\sigma_y}^2}$$

# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

$$\left| \frac{\hat{\boldsymbol{b}}}{\sqrt{\frac{s^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} \right| > t(7)_{0.025}$$

$$\hat{b} = 1.184$$

$$t(7)_{0.025} = 2.3646$$

$$\sigma_x{}^2 = 227.06 \Longrightarrow \sum_{i=1}^{9}(x_i - \bar{x})^2 = 9 \times 227.06 = 2043.54$$

$$\sigma_y{}^2 = 599.56, \quad R^2 = 0.53 \quad \Rightarrow \quad s^2 = \frac{n}{n-2} \times (1 - R^2) \times \sigma_y{}^2 =$$

$$= \frac{9}{7}(1 - 0.53) \times 599.56 = 361.31$$

# Esercizio 1

Su un campione di $n = 9$ unità sono state osservate due variabili, $X$ ed $Y$:

| $x_i$ | 16 | 21 | 44 | 57 | 51 | 60 | 57 | 34 | 45 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 62.9 | 63.3 | 83.4 | 116.5 | 126.7 | 84.5 | 134.8 | 101.1 | 89.2 |

$$\left|\frac{\hat{\boldsymbol{b}}}{\sqrt{\frac{s^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}}\right| > t(7)_{0.025}$$

$$\hat{b} = 1.184$$

$$t(7)_{0.025} = 2.3646$$

$$\left|\frac{1.184}{\sqrt{\frac{361.31}{2043.54}}}\right| = 2.815423$$

rifiutiamo $H_0$ al 5%

$$\sigma_x{}^2 = 227.06 \Longrightarrow \sum_{i=1}^{9}(x_i - \bar{x})^2 = 9 \times 227.06 = 2043.54$$

$$\sigma_y{}^2 = 599.56, \quad R^2 = 0.53$$

$$s^2 = \frac{n}{n-2} \times (1 - R^2) \times \sigma_y{}^2 =$$

$$= \frac{9}{7}(1 - 0.53) \times 599.56 = 361.31$$

# Esercizio 2

$n = 50$

$\bar{x} = 52.48$

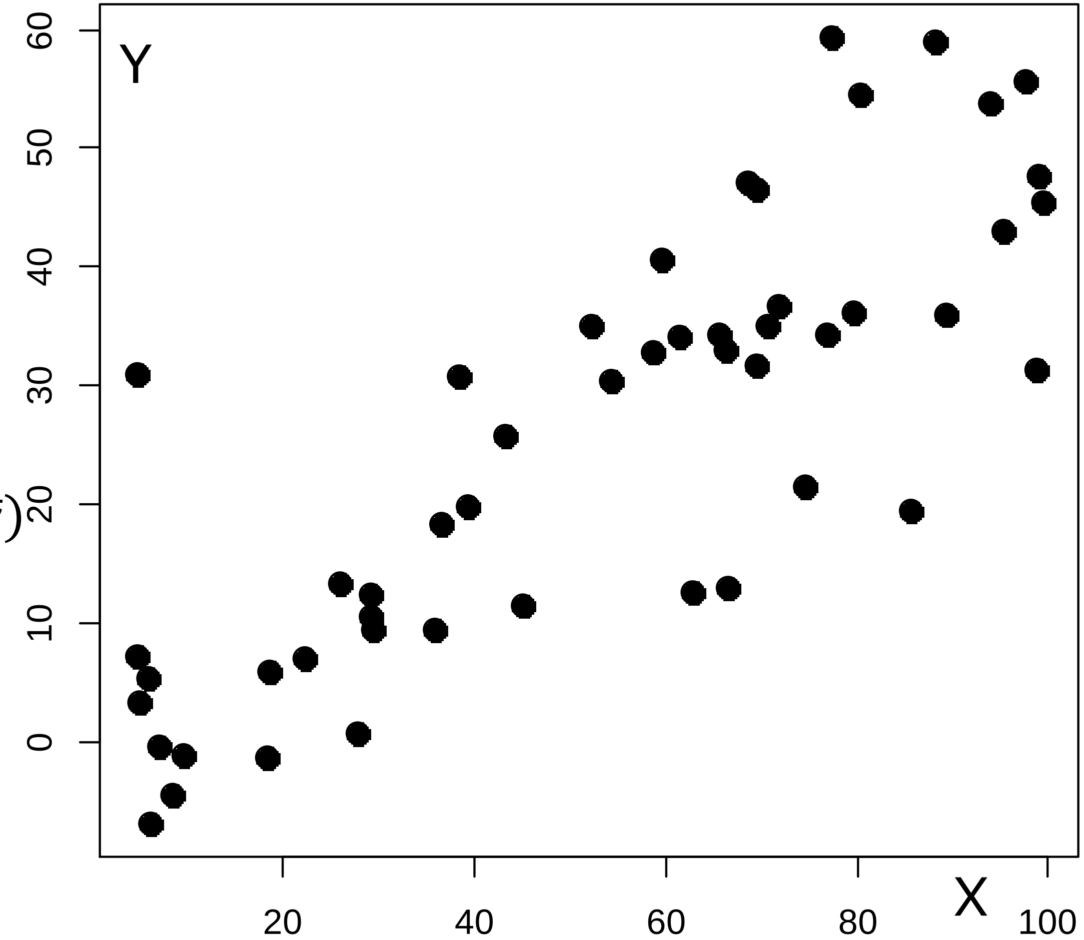
$\bar{y} = 25.36$

$${\sigma_x}^2 = 912.45 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$${\sigma_y}^2 = 329.97$$

$$cov_{xy} = 452.49 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

**1**. Le variabili X e Y sono correlate? Come e quanto?



# Esercizio 2

$n = 50$

$\bar{x} = 52.48$

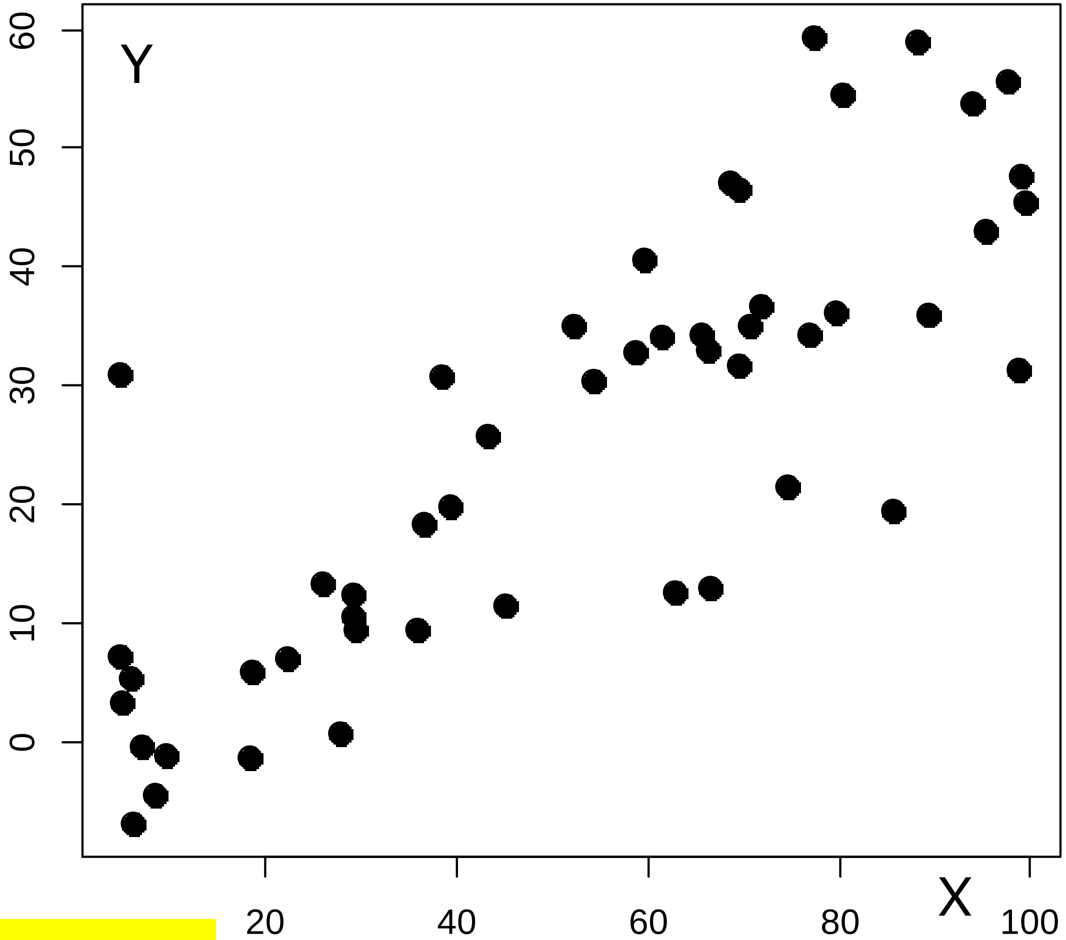
$\bar{y} = 25.36$

$\sigma_x{}^2 = 912.45$

$\sigma_y{}^2 = 329.97$

$cov_{xy} = 452.49$

$$\rho_{xy} = \frac{452.49}{\sqrt{912.45 \times 329.97}} = \mathbf{0.82}$$




Le due var. sono *linearmente* correlate, con correlazione positiva.

# Esercizio 2

$n = 50$

$\bar{x} = 52.48$

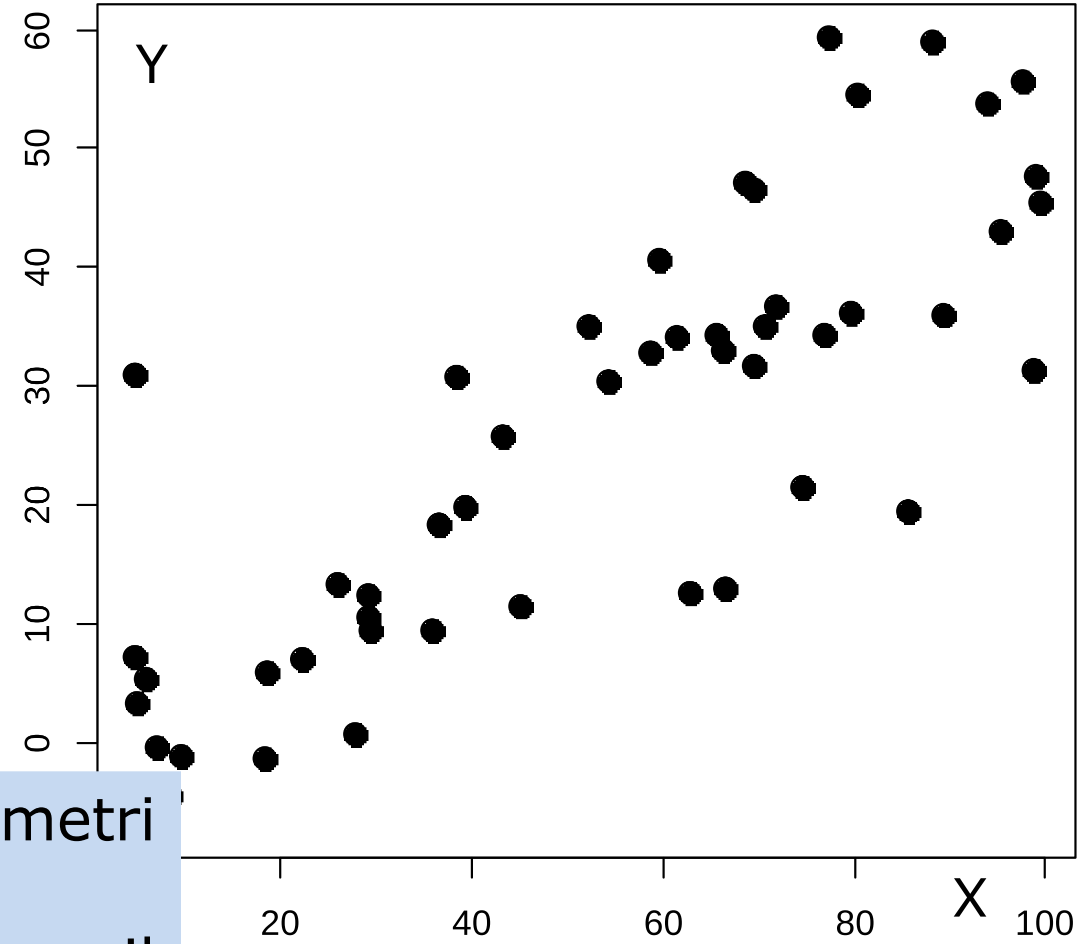
$\bar{y} = 25.36$

${\sigma_x}^2 = 912.45$

${\sigma_y}^2 = 329.97$

$cov_{xy} = 452.49$

$\rho_{xy} = 0.82$




**2**. Fornire una stima dei parametri di un modello di regressione lineare semplice e rappresentare il modello stimato.

# Esercizio 2

$$Y = a + bX$$

$n = 50$

$\bar{x} = 52.48$

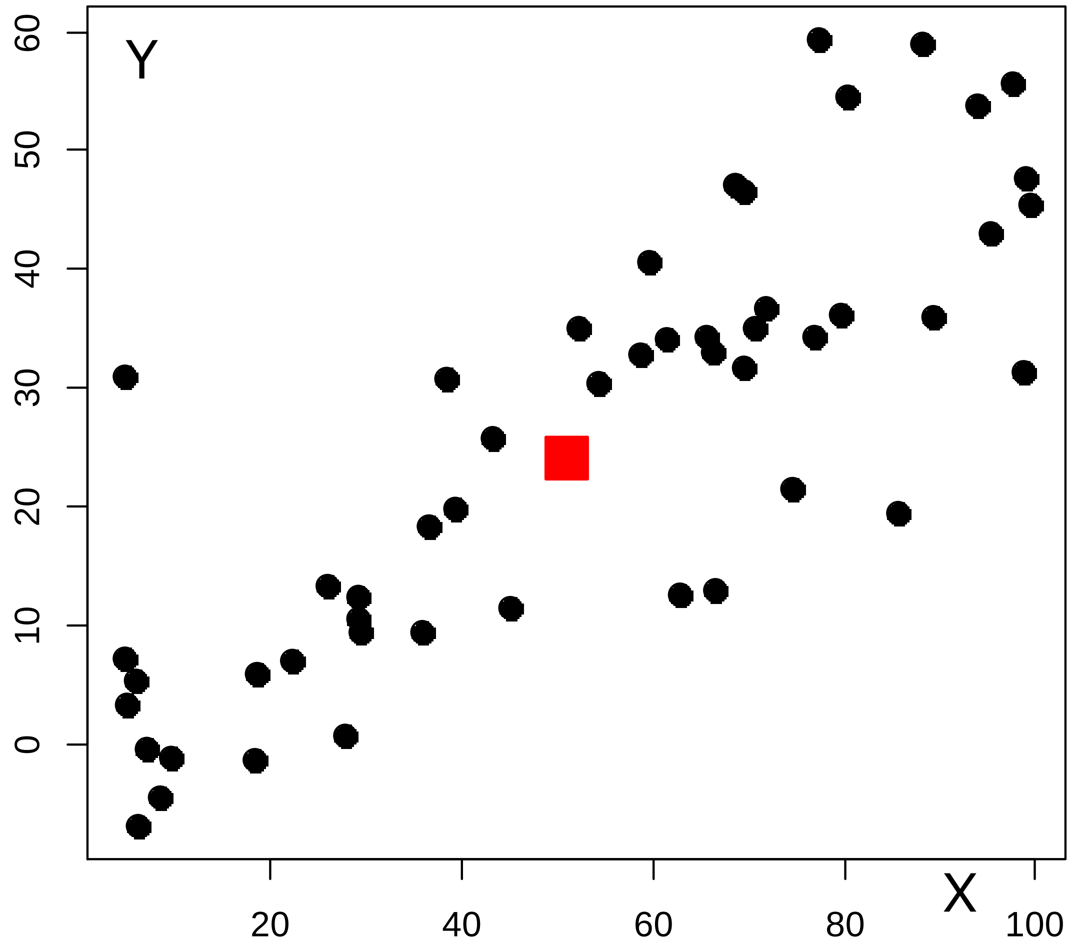
$\bar{y} = 25.36$

${\sigma_x}^2 = 912.45$

${\sigma_y}^2 = 329.97$

$cov_{xy} = 452.49$

$\rho_{xy} = 0.82$




$$\hat{b} = \frac{cov_{xy}}{{\sigma_x}^2} = \frac{452.49}{912.45} = 0.496$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 25.36 - 0.496 \times 52.48 = -0.67$$

# Esercizio 2

$$Y = a + bX$$

$n = 50$

$\bar{x} = 52.48$

$\bar{y} = 25.36$

${\sigma_x}^2 = 912.45$

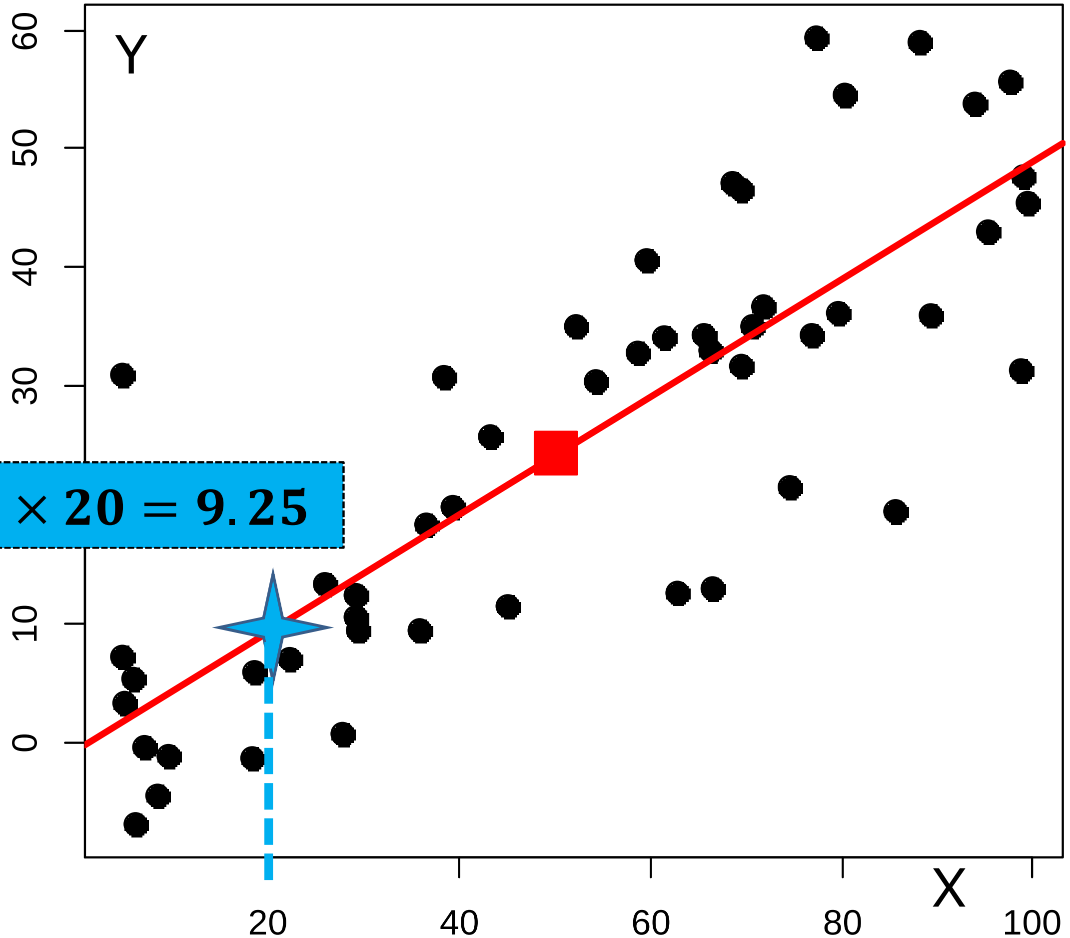
${\sigma_y}^2 = 329.97$

$cov_{xy} = 452.49$

$\rho_{xy} = 0.82$

$$\hat{b} = \frac{cov_{xy}}{{\sigma_x}^2} = \frac{452.49}{912.45} = 0.496$$

$$\mathbf{-0.67 + 0.496 \times 20 = 9.25}$$




$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 25.36 - 0.496 \times 52.48 = -0.67$$

# Esercizio 2

$n = 50$

$\bar{x} = 52.48$

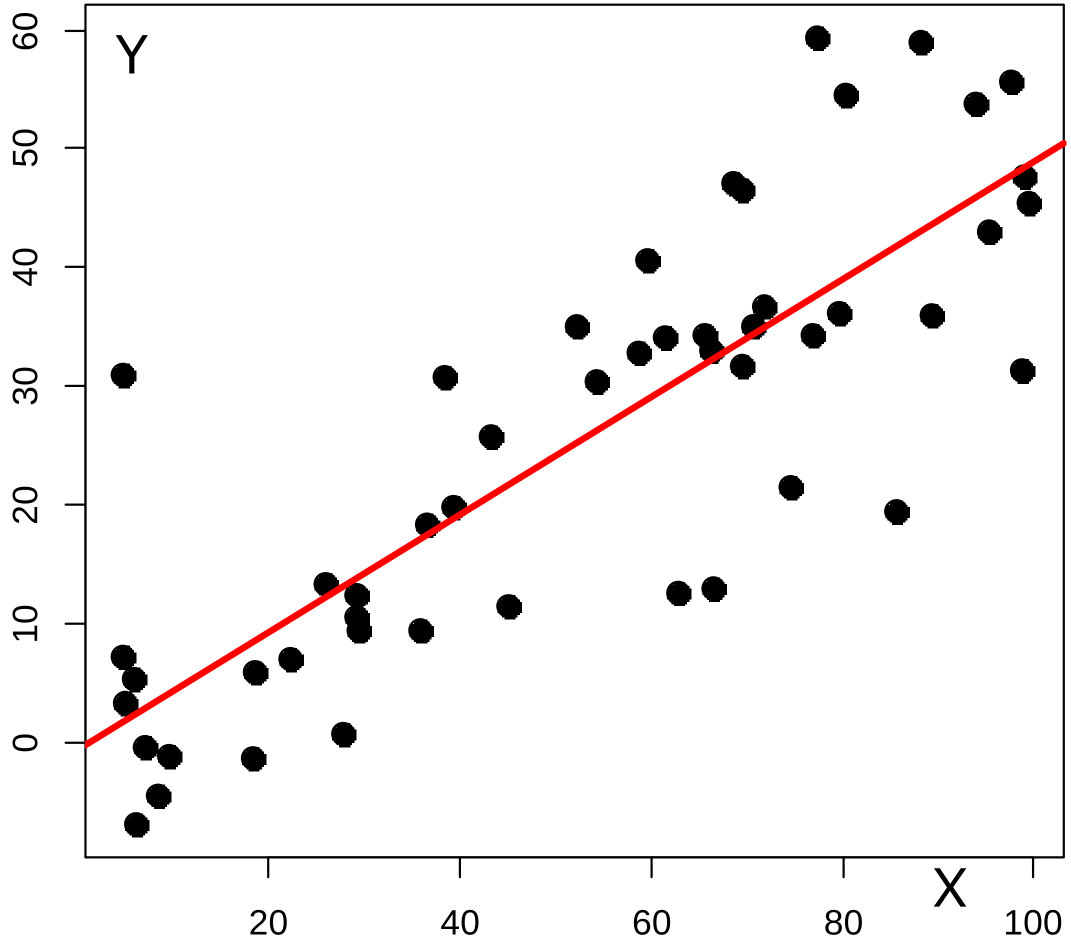
$\bar{y} = 25.36$

$\sigma_x{}^2 = 912.45$

$\sigma_y{}^2 = 329.97$

$cov_{xy} = 452.49$

$\rho_{xy} = 0.82$




**3**. Fornire una stima della varianza dell'errore e una valutazione della bontà del modello stimato.

# E' molto utile ricordare che:

$$\sum_i (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i {e_i}^2$$

**SST = SSR + SSE**

$$\boxed{R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}}}$$

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{(n-2)s^2}{n{\sigma_y}^2}$$

$$\boxed{s^2 = \frac{n}{n-2} \times (1 - R^2) \times {\sigma_y}^2}$$

# Esercizio 2

$n = 50$

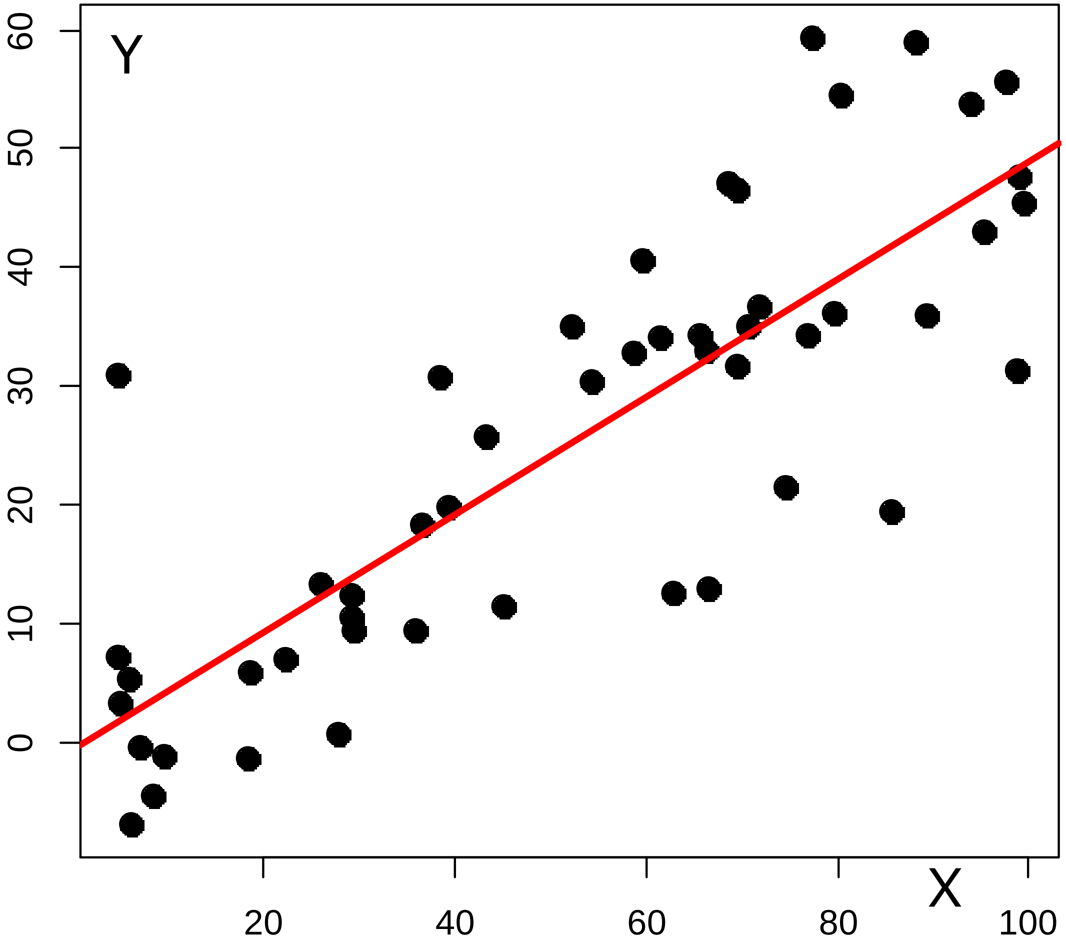
$\boldsymbol{\rho_{xy} = 0.82} \implies \boldsymbol{R^2 = 0.67}$

$\sigma_y{}^2 = 329.97$

$$\boxed{s^2 = \frac{n}{n-2} \times (1 - R^2) \times \sigma_y{}^2}$$

$$\boldsymbol{s^2} = \frac{50}{48} \times (1 - 0.67) \times 329.97$$

$$\boldsymbol{= 113.47}$$




$\hat{b} = 0.496 \qquad \hat{a} = -0.67$

# Esercizio 2

$n = 50$

$\bar{x} = 52.48$

$\bar{y} = 25.36$

${\sigma_x}^2 = 912.45$

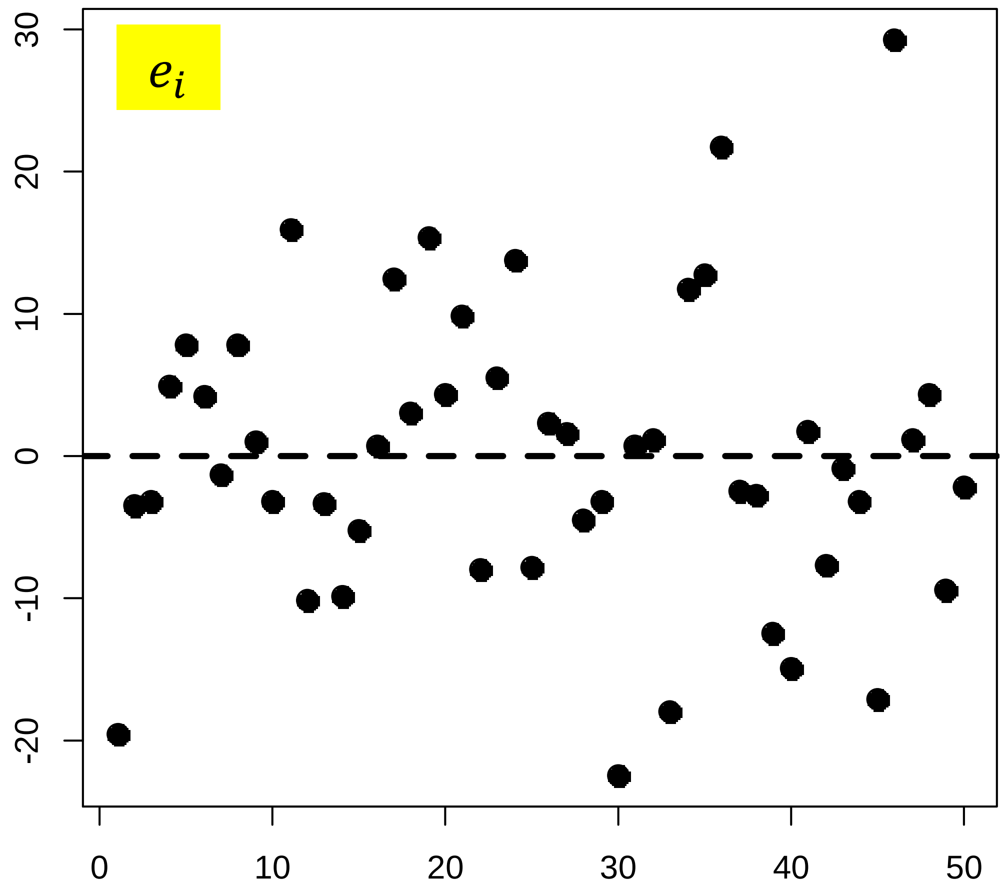
${\sigma_y}^2 = 329.97$

$cov_{xy} = 452.49$

$\rho_{xy} = 0.82$

**4**. Commentare il grafico dei residui qui riportato. Quale altro grafico si potrebbe valutare, e perchè?

grafico dei residui




$\hat{b} = 0.496$ $\hat{a} = -0.67$

# Esercizio 2

$n = 50$

$\bar{x} = 52.48$

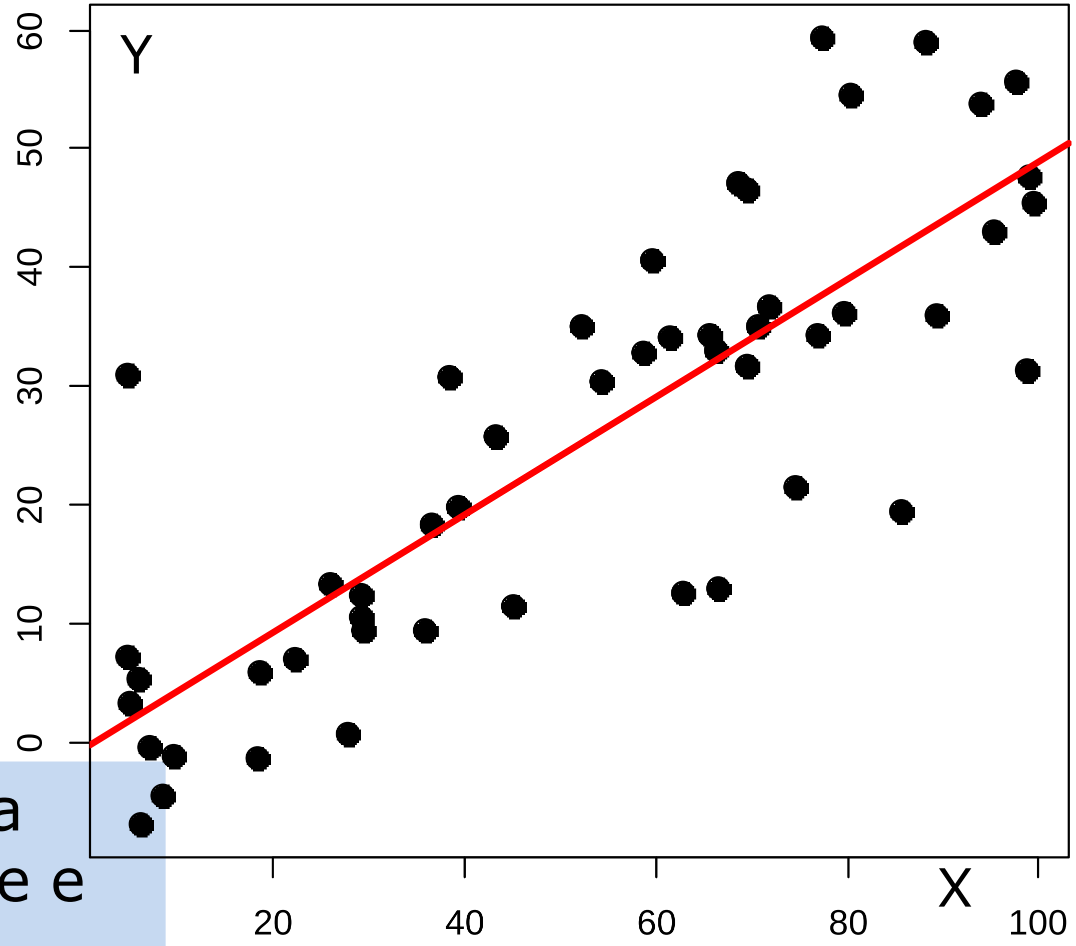
$\bar{y} = 25.36$

${\sigma_x}^2 = 912.45$

${\sigma_y}^2 = 329.97$

$cov_{xy} = 452.49$

$\rho_{xy} = 0.82$




$\hat{b} = 0.496$ $\hat{a} = -0.67$

**5**. Valutare al livello dell’1% la significatività della regressione e determinare l’IC al livello corrispondente per la pendenza.

# Esercizio 2

$n = 50$

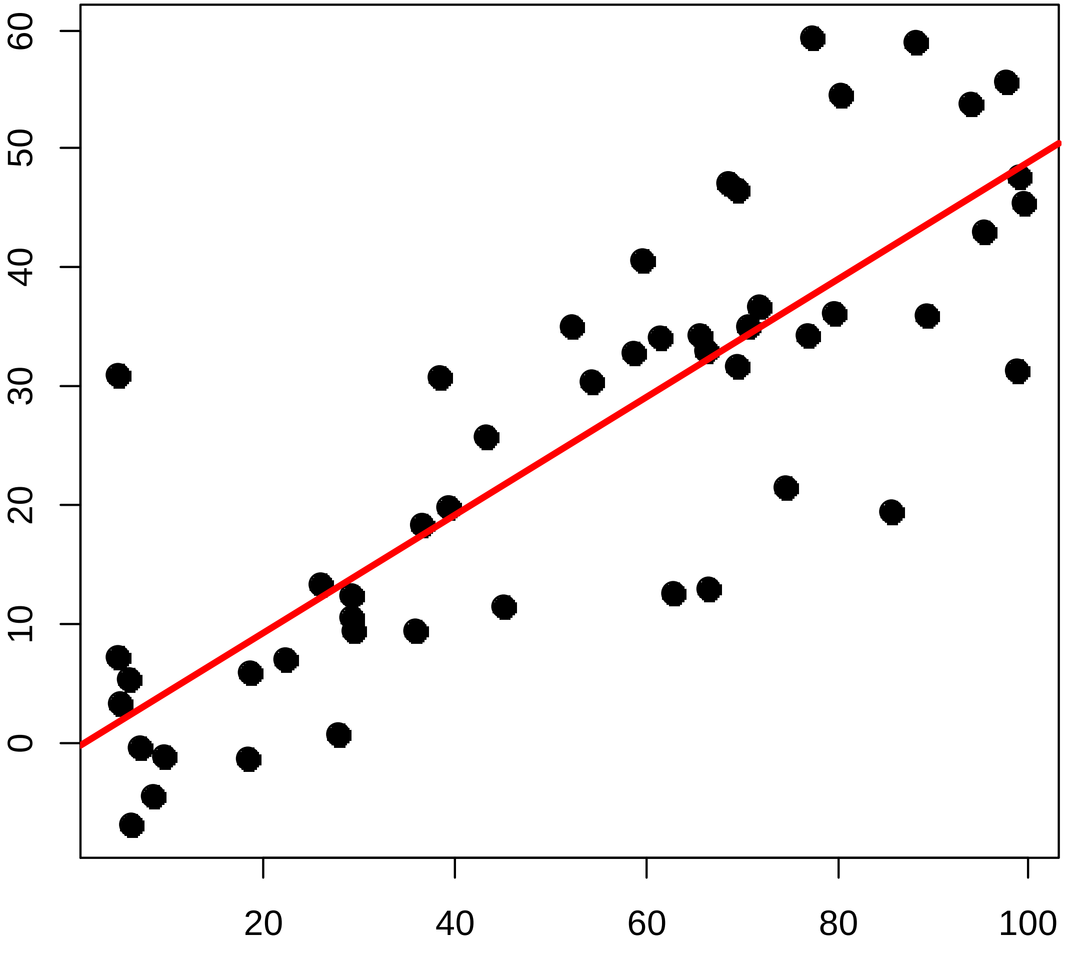
$s^2 = 113.47 \qquad \sigma_x{}^2 = 912.45$

$$\left| \frac{\hat{b}}{\sqrt{\frac{s^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} \right| =$$

$$\left| \frac{0.496}{\sqrt{\frac{113.47}{50 \times 912.45}}} \right| = 9.945$$

$$\alpha = 0.01 \Rightarrow t(48)_{\frac{0.01}{2}} = 2.6822$$

**rifiutiamo $\boldsymbol{H_0}$: $\boldsymbol{b = 0}$**
al livello dell'1%




$\hat{b} = 0.496 \qquad \hat{a} = -0.67$

# Esercizio 2

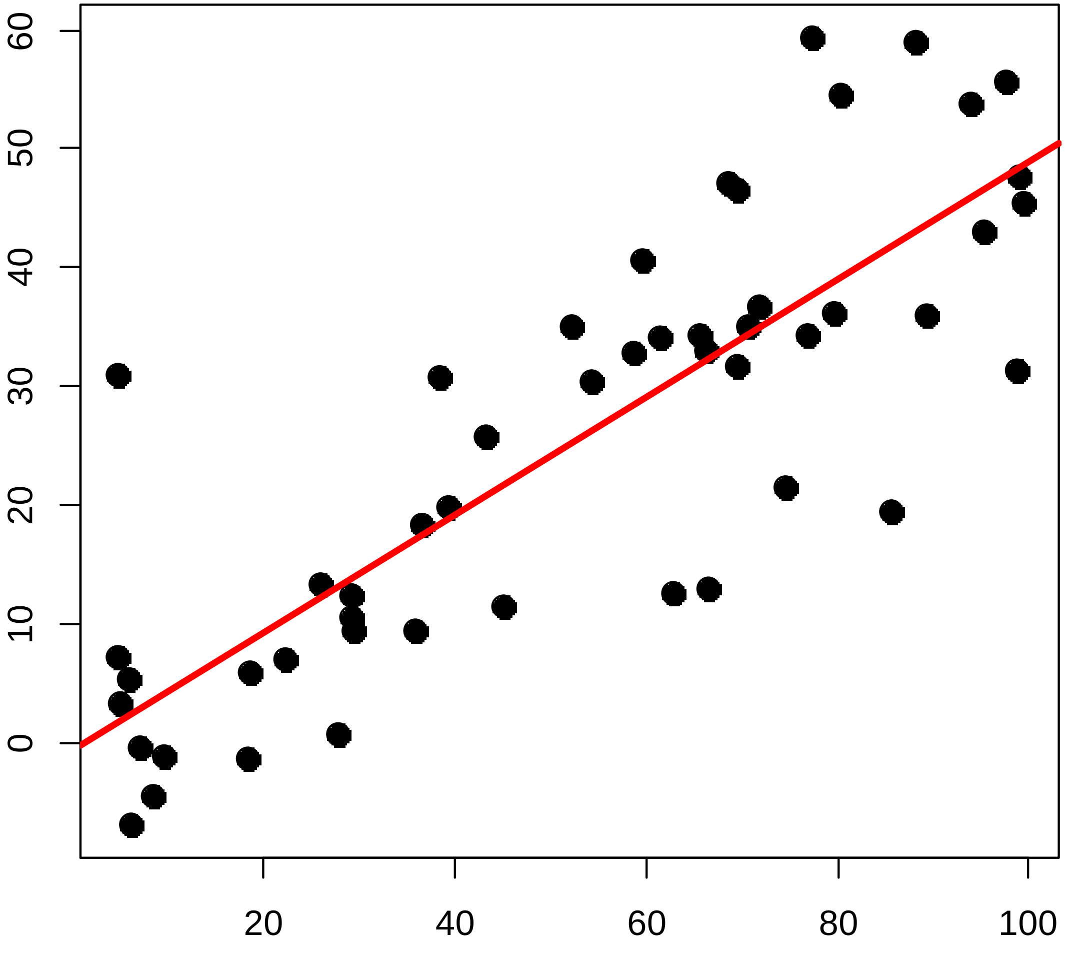
$n = 50$

$s^2 = 113.47 \qquad \sigma_x{}^2 = 912.45$

$\alpha = 0.01 \Rightarrow \; t(48)_{0.005} = 2.6822$

$$0.496 \mp t(48)_{0.005} \times \sqrt{\frac{113.47}{50 \times 912.45}} =$$

$$(0.362, 0.630)$$




$\hat{b} = 0.496 \qquad \hat{a} = -0.67$

# Esercizio 2

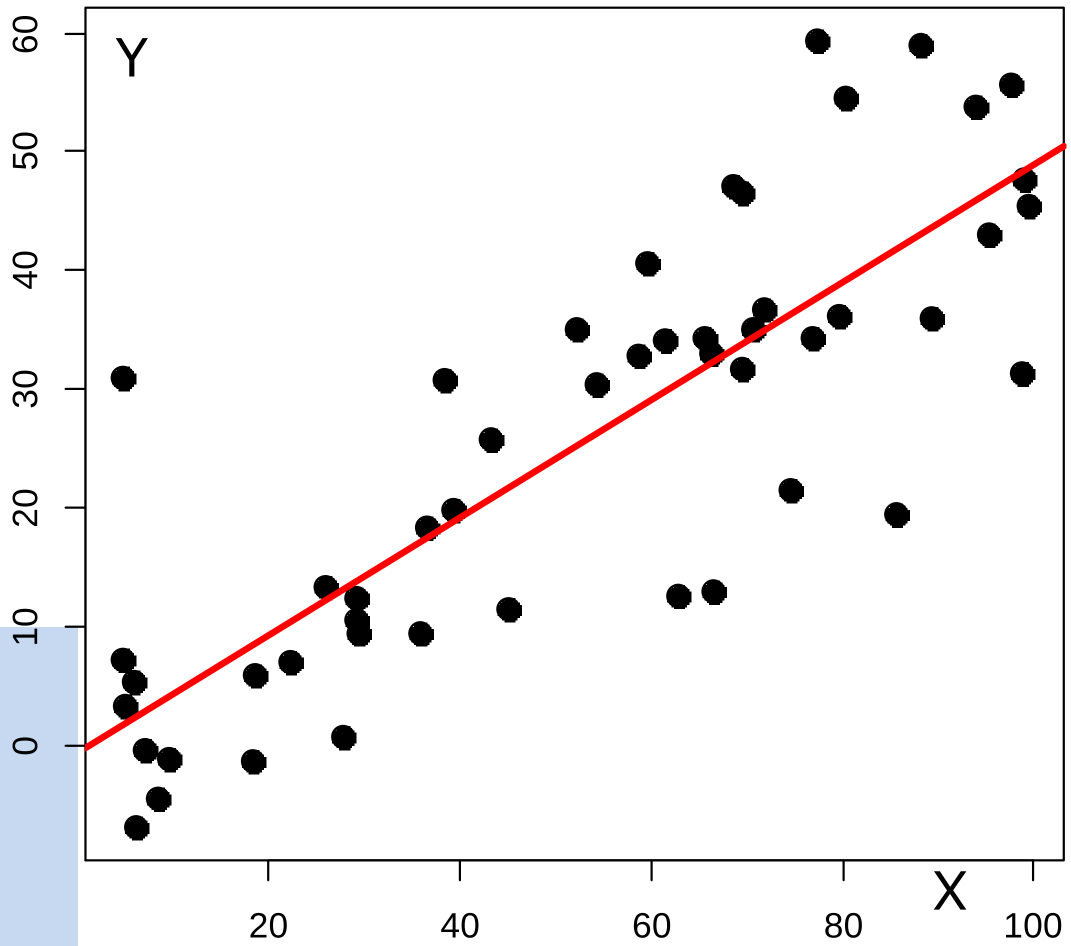
$n = 50$

$\bar{x} = 52.48$

$\bar{y} = 25.36$

$\sigma_x{}^2 = 912.45$

$\sigma_y{}^2 = 329.97$




$\hat{b} = 0.496$ $\hat{a} = -0.67$

**6**. Per quali dei seguenti valori di X la previsione risulta più attendibile?
$X = 20, X = 40, X = 100, X = 150.$
Giustificare la risposta.

# Esercizio 2

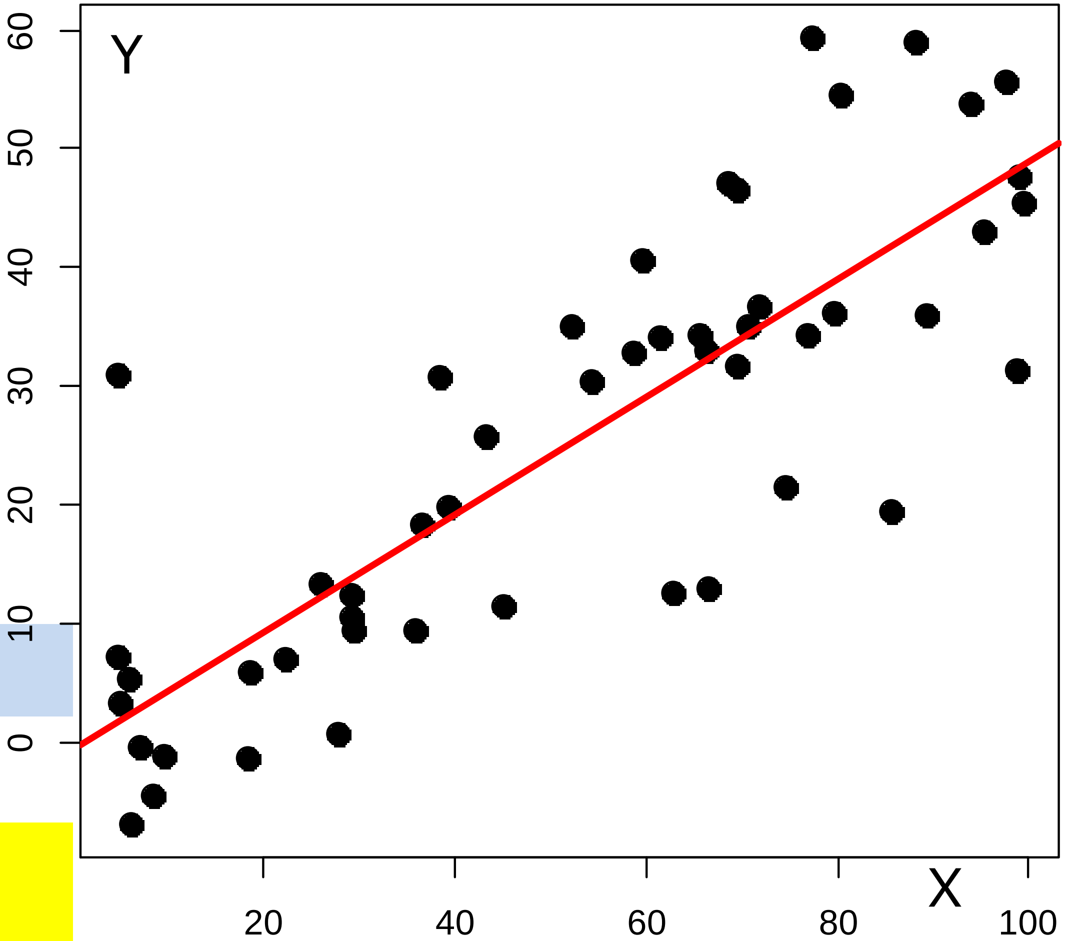
$n = 50$

$\bar{x} = 52.48$

$\bar{y} = 25.36$

${\sigma_x}^2 = 912.45$

${\sigma_y}^2 = 329.97$




$X = 20, \boldsymbol{X = 40}, X = 100, X = 150.$

$\hat{b} = 0.496$  $\hat{a} = -0.67$

Siccome l’ampiezza dell’IC è direttamente proporzionale a $\boldsymbol{(\bar{x} - x_0)^2}$…

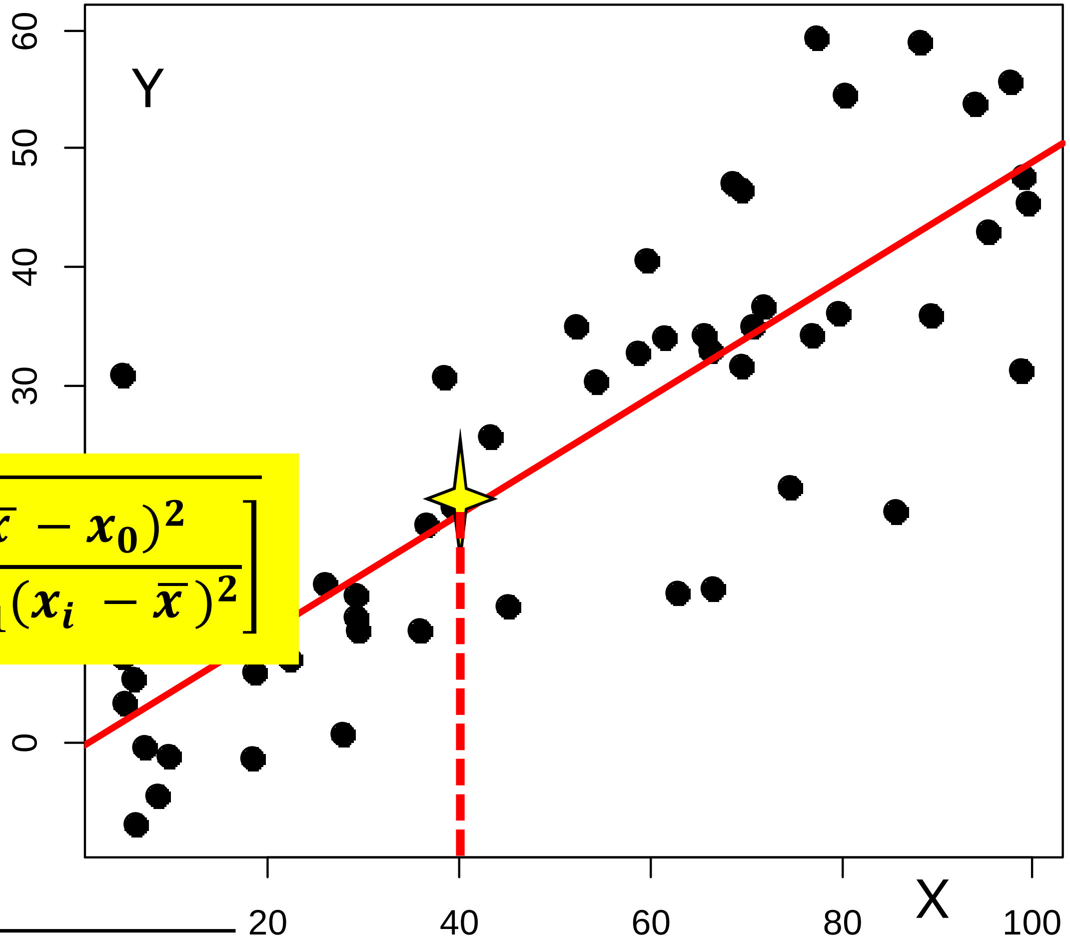
# Esercizio 2

$n = 50$

$s^2 = 113.47 \qquad {\sigma_x}^2 = 912.45$

$\bar{x} = 52.48$

Previsione per $x = 40$:
$\hat{y} = -0.67 + 0.496 \times 40 = 19.17$




$$\boldsymbol{\hat{y}_0 \mp t(n-2)_{\alpha/2} \times \sqrt{s^2\left[1 + n^{-1} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]}}$$

$\alpha = 0.01$

$$19.17 \pm 2.6822\sqrt{113.47\left[1 + \frac{1}{50} + \frac{(52.48-40)^2}{50 \times 912.45}\right]} : (-9.73\,, 48.07)$$

# Formulario

$$\bar{x} = n^{-1}\sum x_i$$

$$\bar{y} = n^{-1}\sum y_i$$

$$\sigma_x{}^2 = n^{-1}\sum x_i{}^2 - \bar{x}^{\,2}$$

$$\sigma_y{}^2 = n^{-1}\sum y_i{}^2 - \bar{y}^{\,2}$$

$$cov_{xy} = n^{-1}\sum x_i y_i - \bar{x}\,\bar{y}$$

$$\boldsymbol{\rho_{xy} = \frac{n^{-1}\sum x_i y_i - \bar{x} \times \bar{y}}{\sqrt{\sigma_x{}^2 \sigma_y{}^2}}}$$

$$\boldsymbol{R^2 = \rho_{xy}{}^2}$$

$$\boldsymbol{\hat{b} = \frac{n^{-1}\sum x_i y_i - \bar{x} \times \bar{y}}{n^{-1}\sum x_i{}^2 - \bar{x}^{\,2}}}$$

$$\boldsymbol{\hat{a} = \bar{y} - \hat{b}\bar{x}}$$

$$\boldsymbol{s^2 = \frac{n}{n-2} \times (1 - R^2) \times \sigma_y{}^2}$$

$$\boldsymbol{\sum_{i=1}^{n} x_i \qquad \sum_{i=1}^{n} y_i}$$

$$\boldsymbol{\sum_{i=1}^{n} x_i{}^2 \qquad \sum_{i=1}^{n} y_i{}^2}$$

$$\boldsymbol{\sum_{i=1}^{n} x_i\, y_i}$$

# Esercizio di compito

$$\sum_{i=1}^{n} x_i = 1416.11$$

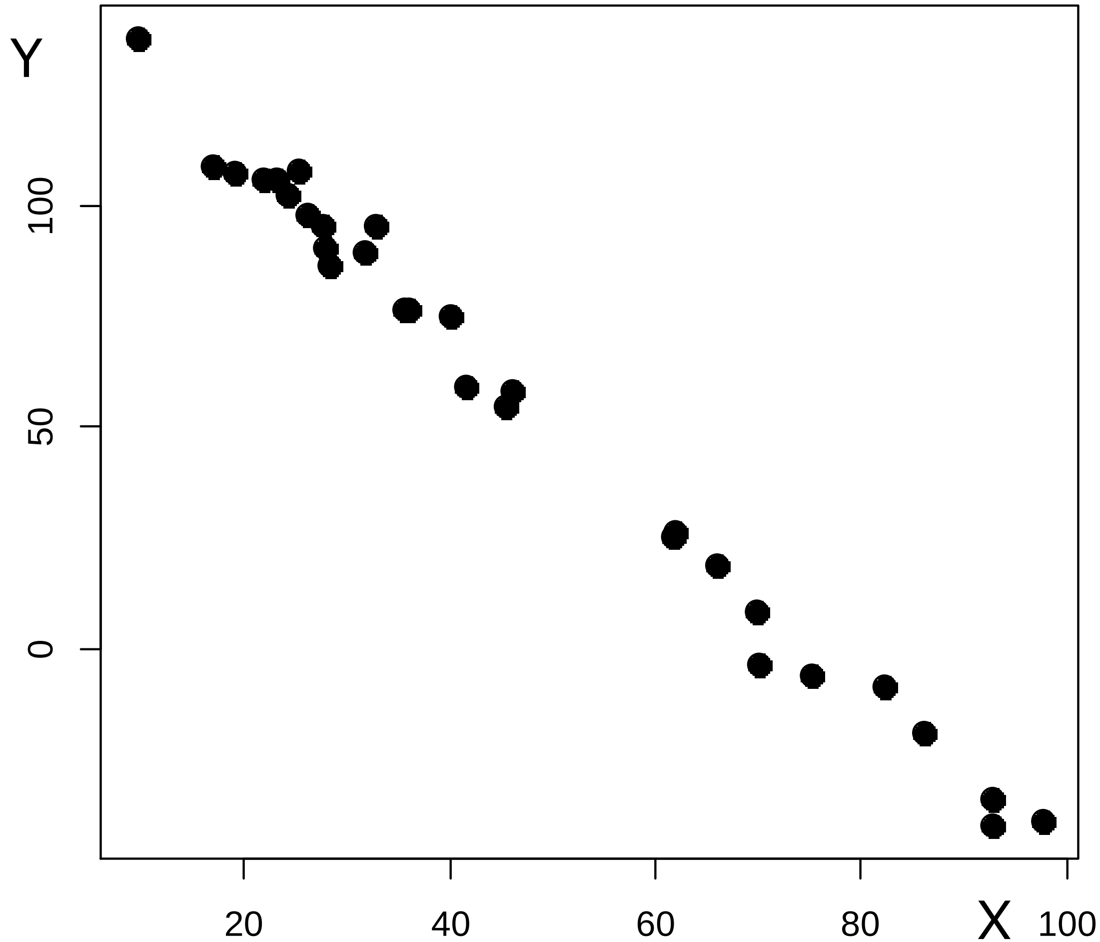
$$\sum_{i=1}^{n} y_i = 1662.26$$

$$\sum_{i=1}^{n} {x_i}^2 = 86627.69$$

$$\sum_{i=1}^{n} {y_i}^2 = 172424.8$$

$$\sum_{i=1}^{n} x_i\, y_i = 38811.94$$

$n = 30$



# Esercizio di compito

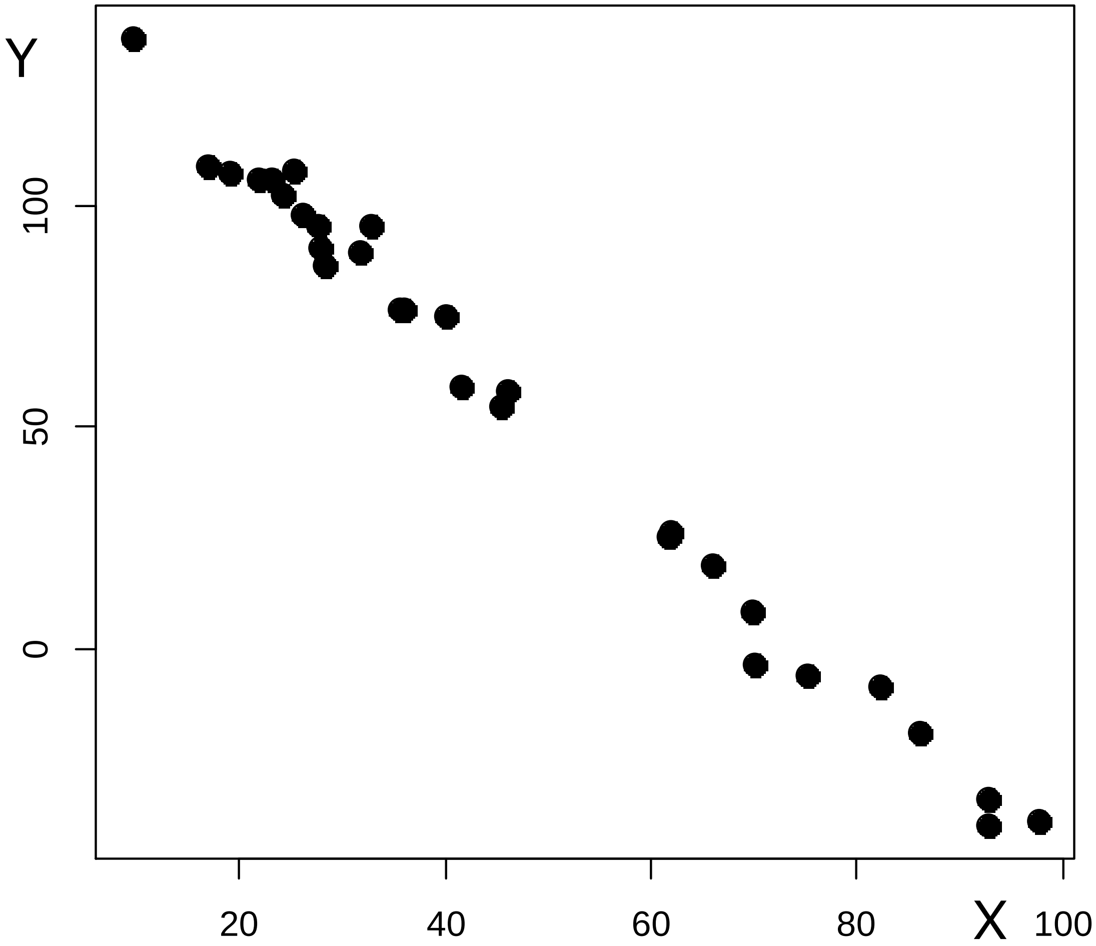
1. Le due variabili sono correlate? Come e quanto?

2. Fornire la stima dei parametri di un modello di regressione lineare e una valutazione della bontà del modello.

3. Valutare la significatività del modello di regressione.

4. Fornire una previsione del valore di Y in corrispondenza ai valori X= 57 e X= 120, con il relativo intervallo di confidenza al 95%.

$n = 30$



# Esercizio 6 (di compito)

Nella tabella sottostante sono riportati i valori delle variabili $X$ e $Y$

| $X$ | -3 | 0 | 4.3 | 10.7 | -1.2 | 8.1 | 1.5 | 15.8 | 15.2 | 5.6 | 2.1 | 11.2 | 12.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 5.0 | 9.3 | 7.7 | -1.8 | 11.5 | 3.4 | 9.2 | -5.5 | 0 | 7.9 | 5.6 | 2.5 | 1.2 |

1. Valutare la possibilità di utilizzare modello lineare per descrivere la dipendenza di $Y$ da $X$, giustificando la risposta.
2. Se possibile, stimare i parametri del modello lineare e discutere la significatività statistica della pendenza stimata.
3. Sulla base delle considerazioni precedenti, proporre una stima per il valore di $Y$ in corrispondenza a $x_0 = 6$
4. Modificare il valore di $Y$ in corrispondenza ad un solo valore di $x$ in modo da aumentare la correlazione lineare tra i dati. Si giustifichi la scelta fatta.

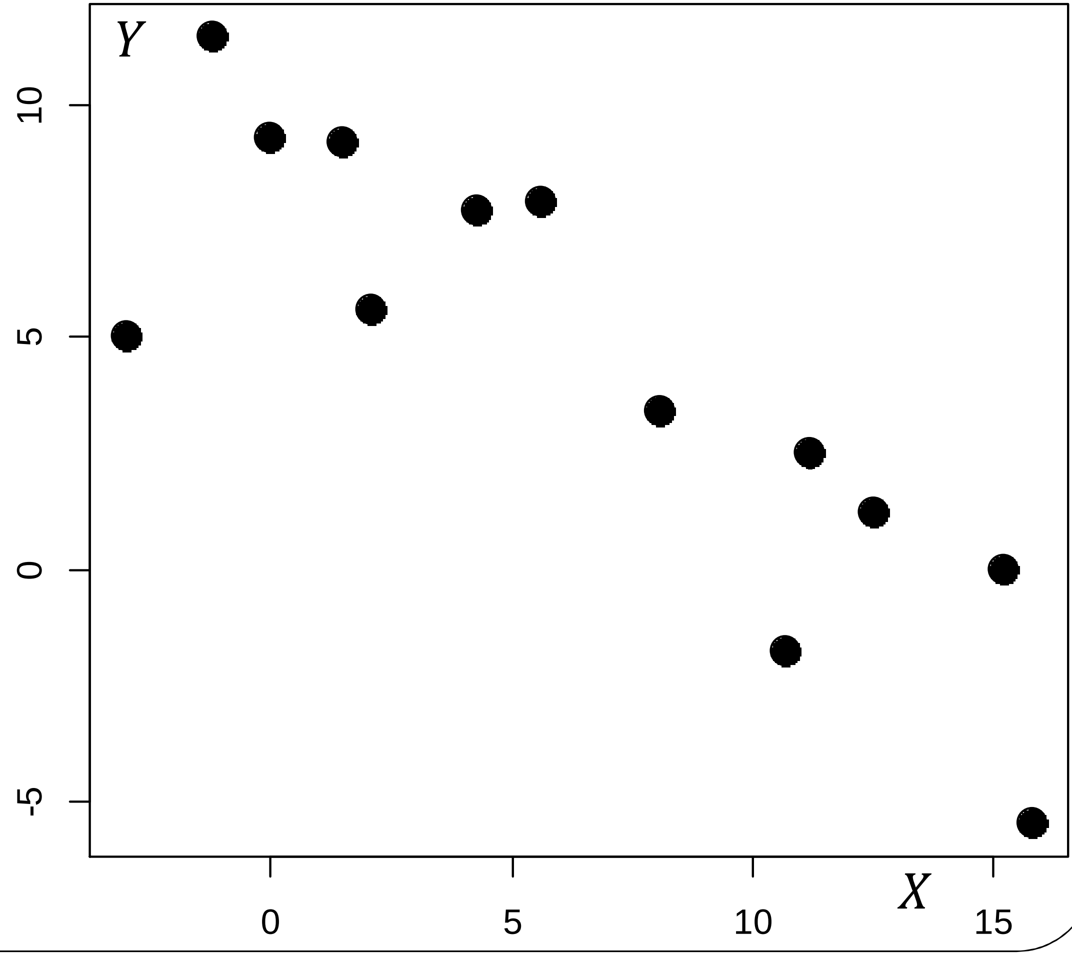
# Esercizio 6 (di compito)

Nella tabella sottostante sono riportati i valori delle variabili $X$ e $Y$

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X$ | -3 | 0 | 4.3 | 10.7 | -1.2 | 8.1 | 1.5 | 15.8 | 15.2 | 5.6 | 2.1 | 11.2 | 12.5 |
| $Y$ | 5.0 | 9.3 | 7.7 | -1.8 | 11.5 | 3.4 | 9.2 | -5.5 | 0 | 7.9 | 5.6 | 2.5 | 1.2 |

$n = 13$

$$\left.\begin{aligned} \sum x_i &= 82.8 \\ \sum y_i &= 56 \\ \sum {x_i}^2 &= 1009.42 \\ \sum {y_i}^2 &= 534.18 \\ \sum x_i y_i &= 38.49 \end{aligned}\right\} \quad \begin{aligned} \bar{x} &= 6.369231 \\ \bar{y} &= 4.307692 \\ {\sigma_x}^2 &= 37.08059 \\ {\sigma_y}^2 &= 22.53456 \\ \sigma_{xy} &= -24.47592 \end{aligned}$$



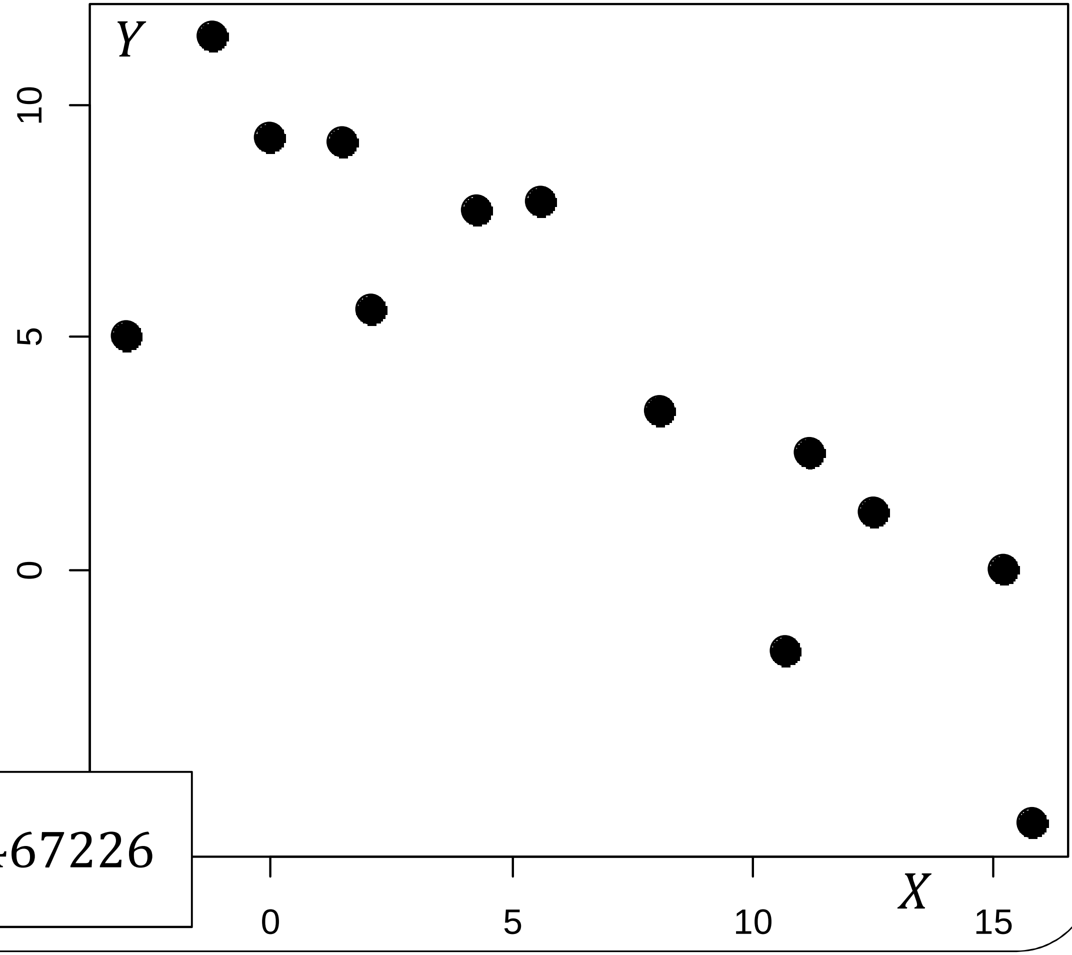
# Esercizio 6 (di compito)

Nella tabella sottostante sono riportati i valori delle variabili $X$ e $Y$

| $X$ | -3 | 0 | 4.3 | 10.7 | -1.2 | 8.1 | 1.5 | 15.8 | 15.2 | 5.6 | 2.1 | 11.2 | 12.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 5.0 | 9.3 | 7.7 | -1.8 | 11.5 | 3.4 | 9.2 | -5.5 | 0 | 7.9 | 5.6 | 2.5 | 1.2 |

$n = 13$

$$\left.\begin{aligned} \sum x_i &= 82.8 \\ \sum y_i &= 56 \\ \sum x_i^{\,2} &= 1009.42 \\ \sum y_i^{\,2} &= 534.18 \\ \sum x_i y_i &= 38.49 \end{aligned}\right\} \quad \begin{aligned} \bar{x} &= 6.369231 \\ \bar{y} &= 4.307692 \\ \sigma_x^{\,2} &= 37.08059 \\ \sigma_y^{\,2} &= 22.53456 \\ \sigma_{xy} &= -24.47592 \end{aligned}$$




$$\boxed{\rho_{xy} = \frac{-24.47592}{\sqrt{37.08059 \ \times 22.53456}} = -0.8467226}$$

# Esercizio 6 (di compito)

Nella tabella sottostante sono riportati i valori delle variabili $X$ e $Y$

| $X$ | -3 | 0 | 4.3 | 10.7 | -1.2 | 8.1 | 1.5 | 15.8 | 15.2 | 5.6 | 2.1 | 11.2 | 12.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 5.0 | 9.3 | 7.7 | -1.8 | 11.5 | 3.4 | 9.2 | -5.5 | 0 | 7.9 | 5.6 | 2.5 | 1.2 |

$n = 13$

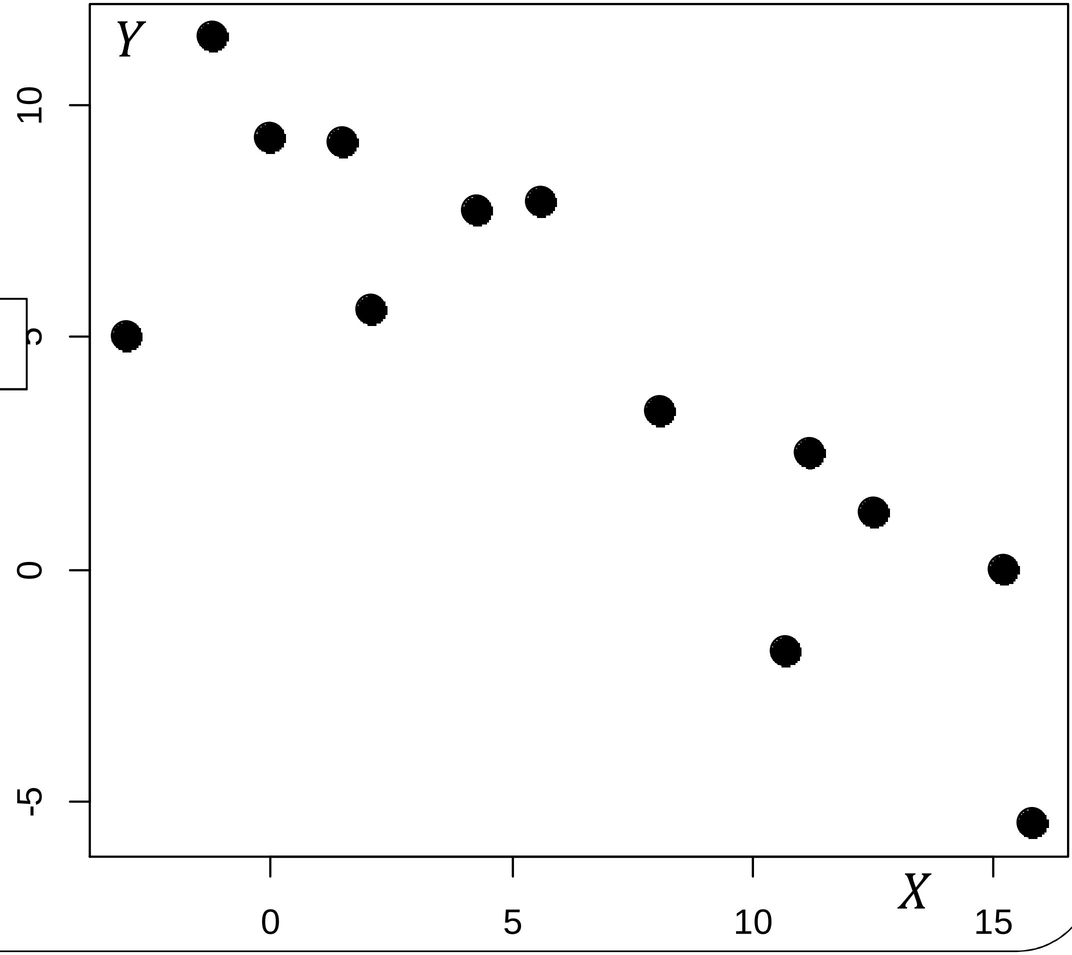
$\bar{x} = 6.369231$ $\quad$ $\bar{y} = 4.307692$

${\sigma_x}^2 = 37.08059$ $\quad$ ${\sigma_y}^2 = 22.53456$

$\sigma_{xy} = -24.47592$ $\quad$ $\boxed{\rho_{xy} = -0.8467226}$

$R^2 = {\rho_{xy}}^2 = 0.7169391$

La retta dei minimi quadrati spiega il 72% della varianza totale di $Y$.



# Esercizio 6 (di compito)

Nella tabella sottostante sono riportati i valori delle variabili $X$ e $Y$

| $X$ | -3 | 0 | 4.3 | 10.7 | -1.2 | 8.1 | 1.5 | 15.8 | 15.2 | 5.6 | 2.1 | 11.2 | 12.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 5.0 | 9.3 | 7.7 | -1.8 | 11.5 | 3.4 | 9.2 | -5.5 | 0 | 7.9 | 5.6 | 2.5 | 1.2 |

$n = 13$

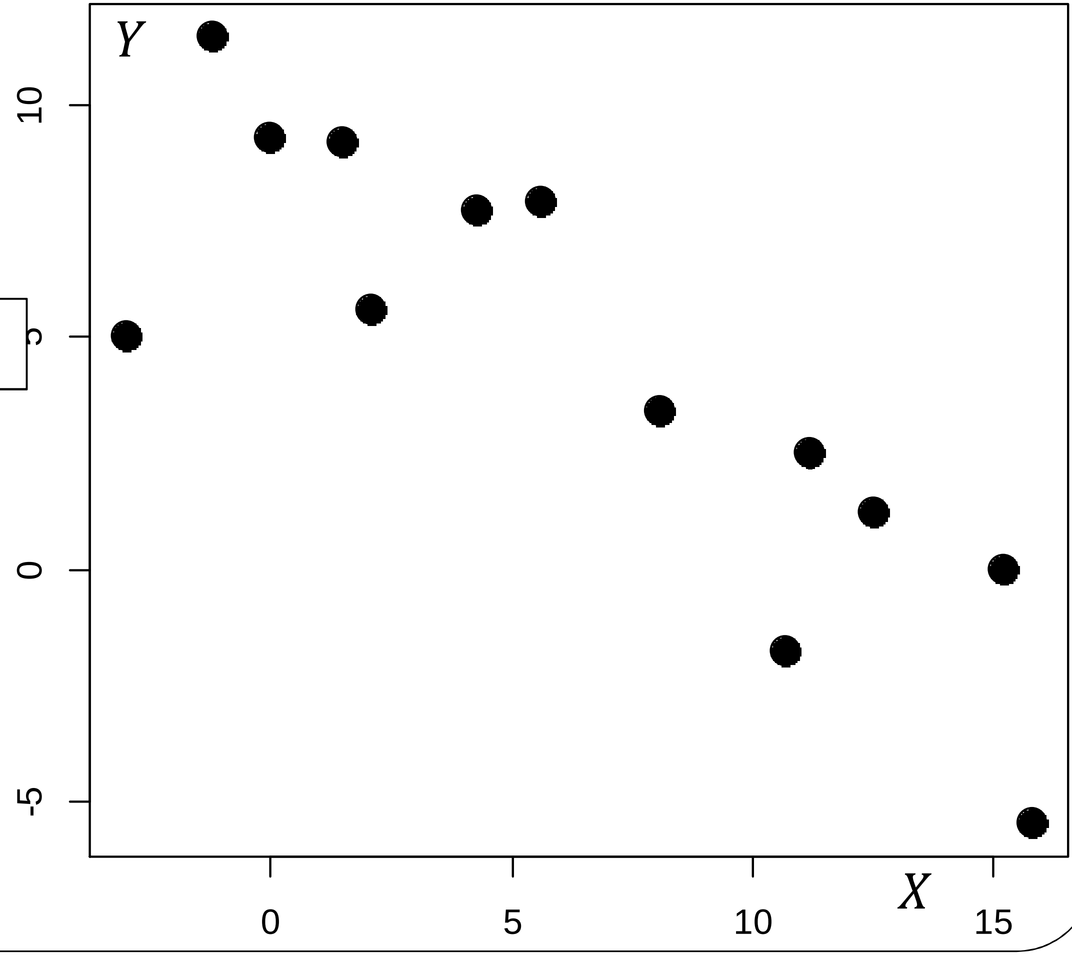
$\bar{x} = 6.369231$ $\qquad$ $\bar{y} = 4.307692$

${\sigma_x}^2 = 37.08059$ $\qquad$ ${\sigma_y}^2 = 22.53456$

$\sigma_{xy} = -24.47592$ $\qquad$ $\boxed{\rho_{xy} = -0.8467226}$

$$\hat{b} = \frac{-24.47592}{37.08059} = -0.6600736$$

$$\hat{a} = 4.307692 + 0.6600736 \times 6.369231$$
$$= 8.511853$$



# Esercizio 6 (di compito)

Nella tabella sottostante sono riportati i valori delle variabili $X$ e $Y$

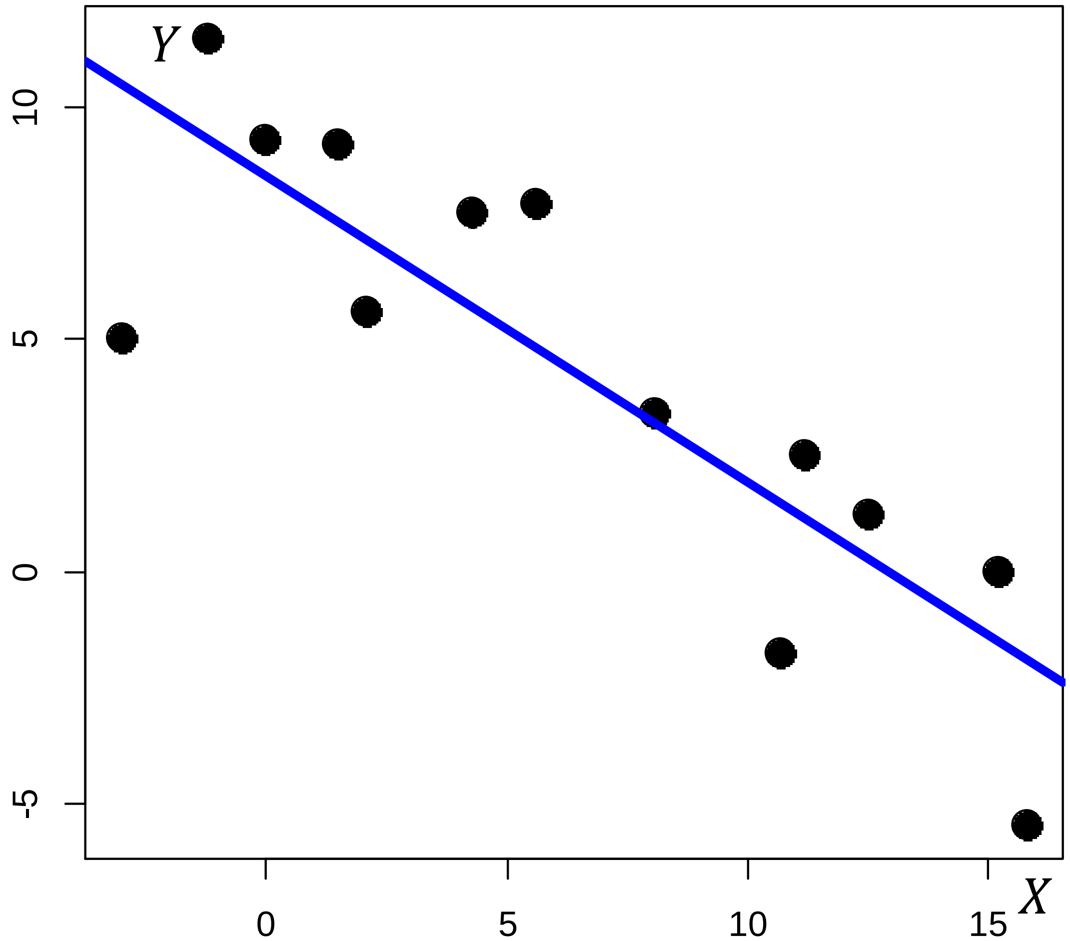
| $X$ | -3 | 0 | 4.3 | 10.7 | -1.2 | 8.1 | 1.5 | 15.8 | 15.2 | 5.6 | 2.1 | 11.2 | 12.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 5.0 | 9.3 | 7.7 | -1.8 | 11.5 | 3.4 | 9.2 | -5.5 | 0 | 7.9 | 5.6 | 2.5 | 1.2 |

$n = 13$

$$\boxed{\rho_{xy} = -0.8467226}$$

$\hat{b} = -0.6600736$

$\hat{a} = 8.511853$



# Esercizio 6 (di compito)

Nella tabella sottostante sono riportati i valori delle variabili $X$ e $Y$

| $X$ | -3 | 0 | 4.3 | 10.7 | -1.2 | 8.1 | 1.5 | 15.8 | 15.2 | 5.6 | 2.1 | 11.2 | 12.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 5.0 | 9.3 | 7.7 | -1.8 | 11.5 | 3.4 | 9.2 | -5.5 | 0 | 7.9 | 5.6 | 2.5 | 1.2 |

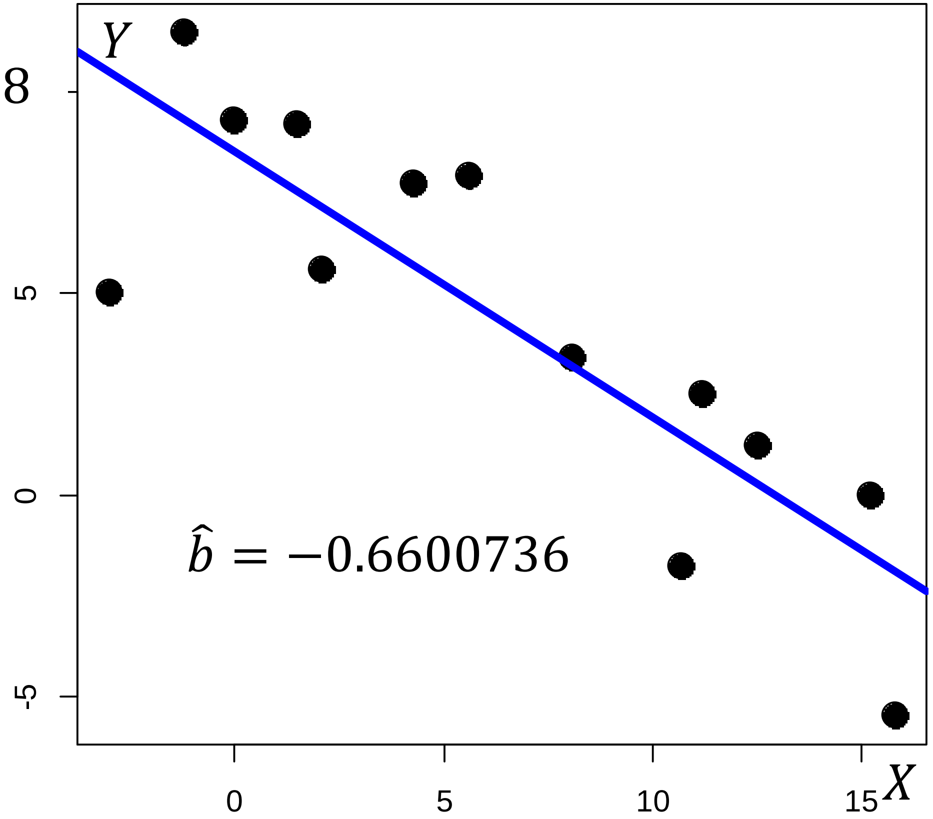
$n = 13 \quad {\sigma_x}^2 = 37.08059 \quad {\sigma_y}^2 = 22.53456$

$$s^2 = \frac{13}{11} \times (1 - 0.7169391) \times 22.53456 = 7.538$$

$$\left| \frac{-0.6600736}{\sqrt{\frac{7.538}{13 \times 37.08059}}} \right| = 5.278$$

$t(11)_{0.025} = 2.201$

La regressione è significativa al livello del 5%.



# Esercizio 6 (di compito)

Nella tabella sottostante sono riportati i valori delle variabili $X$ e $Y$

| $X$ | -3 | 0 | 4.3 | 10.7 | -1.2 | 8.1 | 1.5 | 15.8 | 15.2 | 5.6 | 2.1 | 11.2 | 12.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 5.0 | 9.3 | 7.7 | -1.8 | 11.5 | 3.4 | 9.2 | -5.5 | 0 | 7.9 | 5.6 | 2.5 | 1.2 |

$\hat{b} = -0.6600736$

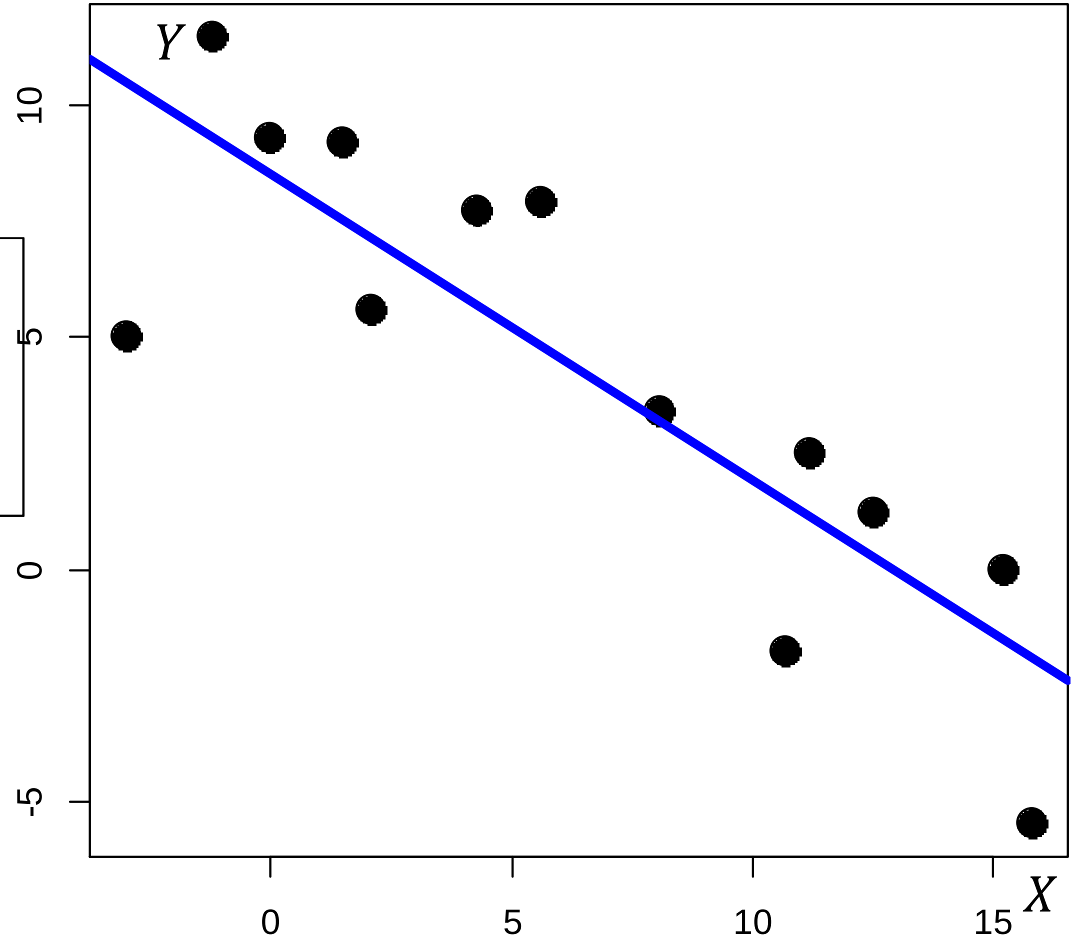
$\hat{a} = 8.511853$

3. Sulla base delle considerazioni precedenti, si proponga una stima per il valore di $Y$ in corrispondenza a $x_0 = 6$

$\hat{y} = 8.511853 - 0.6600736 \times 6 =$

$= 4.551411$

calcolate IC(95%)!



# Esercizio 6

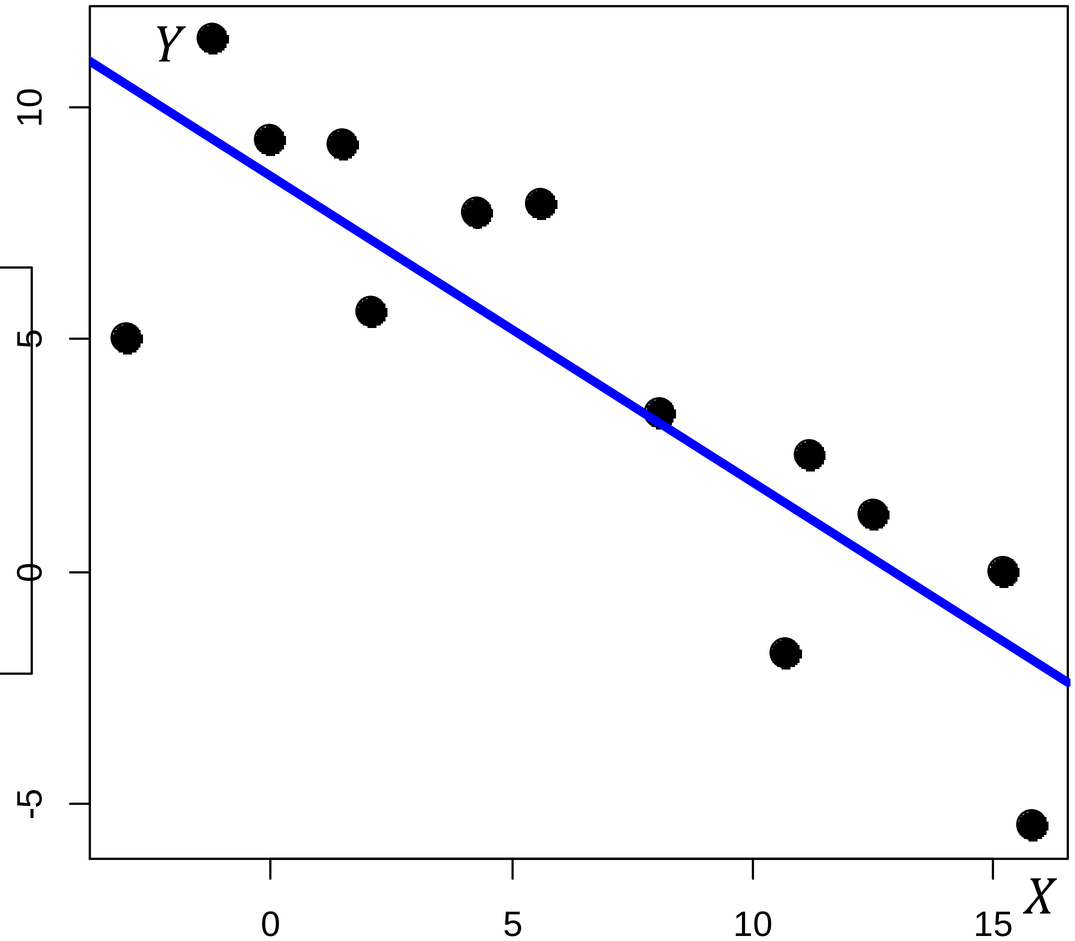
Nella tabella sottostante sono riportati i valori delle variabili $X$ e $Y$

| $X$ | -3 | 0 | 4.3 | 10.7 | -1.2 | 8.1 | 1.5 | 15.8 | 15.2 | 5.6 | 2.1 | 11.2 | 12.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 5.0 | 9.3 | 7.7 | -1.8 | 11.5 | 3.4 | 9.2 | -5.5 | 0 | 7.9 | 5.6 | 2.5 | 1.2 |

$\hat{b} = -0.6600736$

$\hat{a} = 8.511853$

4. Si modifichi il valore di $Y$ in corrispondenza ad un solo valore di $x$ in modo da aumentare la correlazione lineare tra i dati. Si giustifichi la scelta fatta.



# Esercizio 6

Nella tabella sottostante sono riportati i valori delle variabili $X$ e $Y$

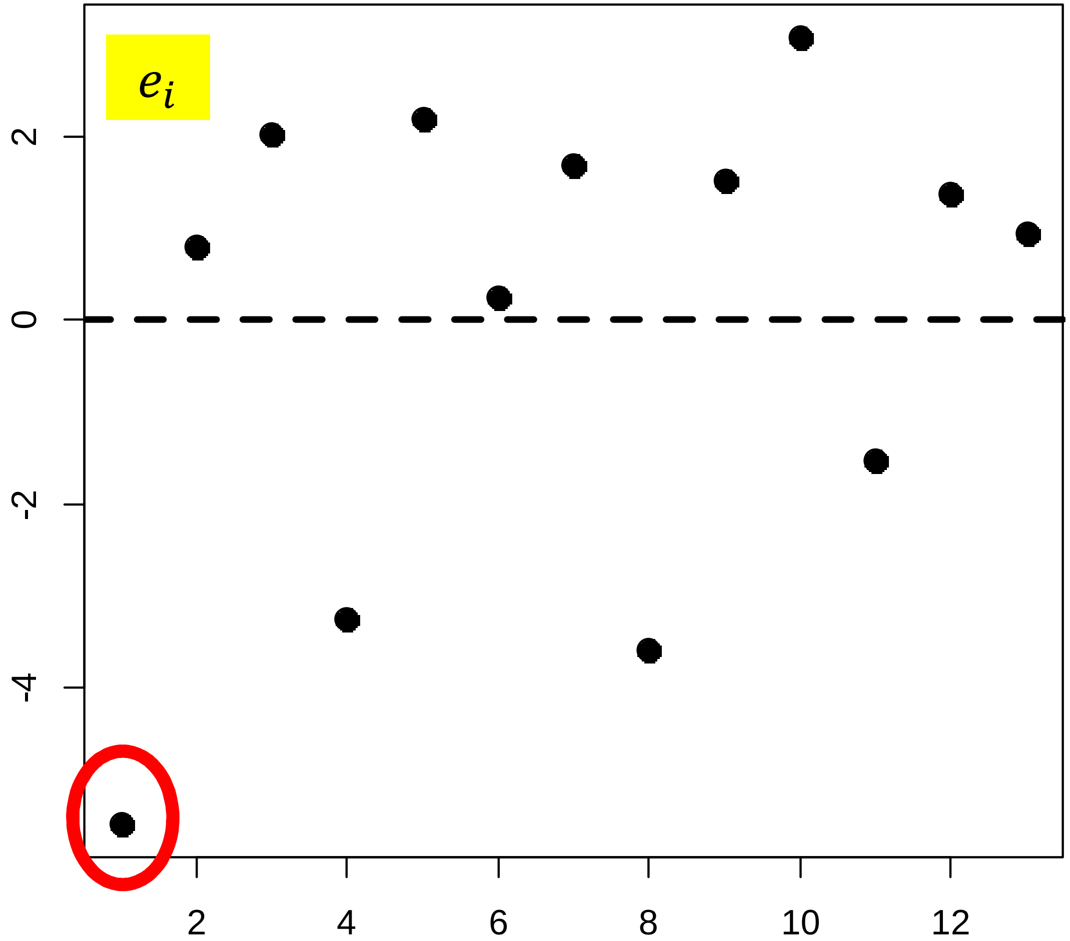
| $X$ | -3 | 0 | 4.3 | 10.7 | -1.2 | 8.1 | 1.5 | 15.8 | 15.2 | 5.6 | 2.1 | 11.2 | 12.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 5.0 | 9.3 | 7.7 | -1.8 | 11.5 | 3.4 | 9.2 | -5.5 | 0 | 7.9 | 5.6 | 2.5 | 1.2 |

$n = 13$

$y = 8.511853 - 0.6600736x$

$\hat{y}_1 = 8.511853 - 0.6600736 \times (-3)$

$= 10.4920736$



# Esercizio 6

Nella tabella sottostante sono riportati i valori delle variabili $X$ e $Y$

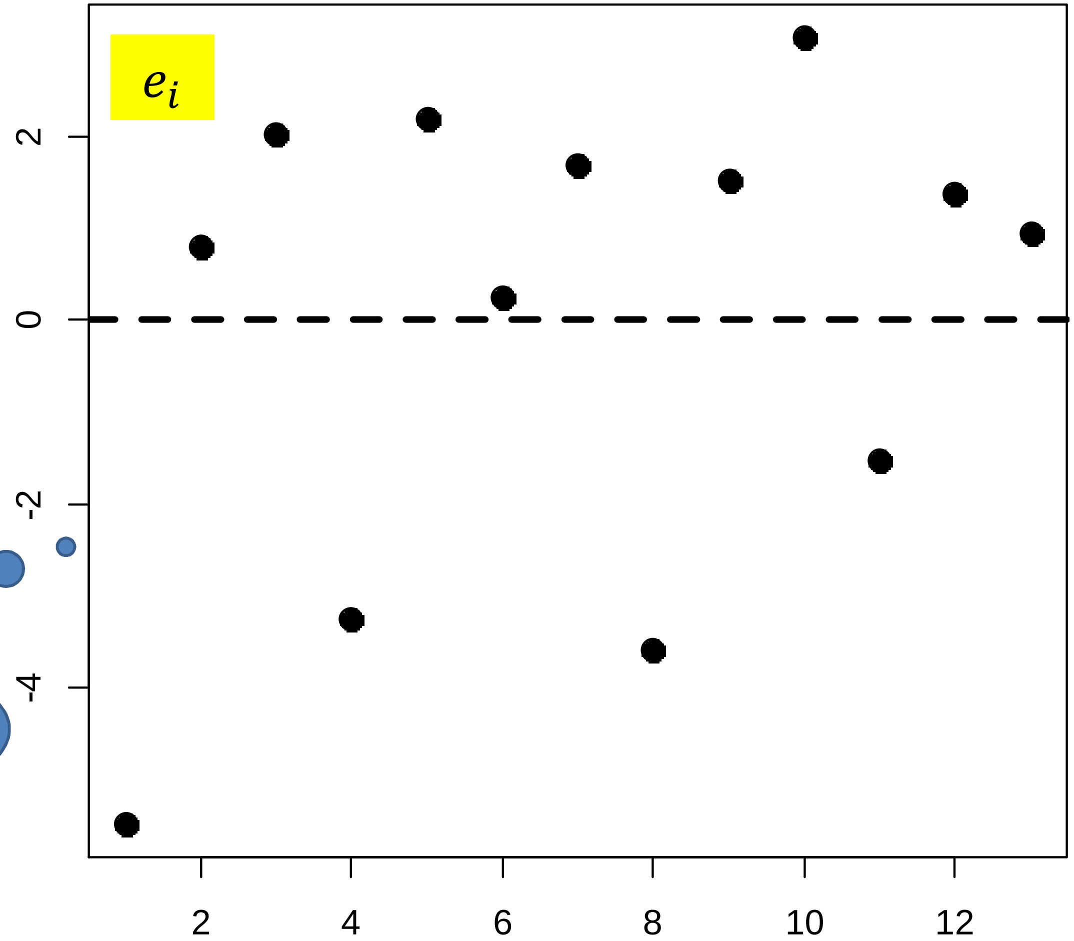
| $X$ | -3 | 0 | 4.3 | 10.7 | -1.2 | 8.1 | 1.5 | 15.8 | 15.2 | 5.6 | 2.1 | 11.2 | 12.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 5.0 | 9.3 | 7.7 | -1.8 | 11.5 | 3.4 | 9.2 | -5.5 | 0 | 7.9 | 5.6 | 2.5 | 1.2 |

$n = 13$

$y = 8.511853 - 0.6600736x$

$\hat{y}_1 = 8.511853 - 0.6600736 \times (-3)$

$= 10.4920736$




**NON E' UN BEL GRAFICO PER I RESIDUI…**

# Esercizio 6

Nella tabella sottostante sono riportati i valori delle variabili $X$ e $Y$

| $X$ | -3 | 0 | 4.3 | 10.7 | -1.2 | 8.1 | 1.5 | 15.8 | 15.2 | 5.6 | 2.1 | 11.2 | 12.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | $\hat{y}_1$ | 9.3 | 7.7 | -1.8 | 11.5 | 3.4 | 9.2 | -5.5 | 0 | 7.9 | 5.6 | 2.5 | 1.2 |

$n = 13$

$y = 8.511853 - 0.6600736x$

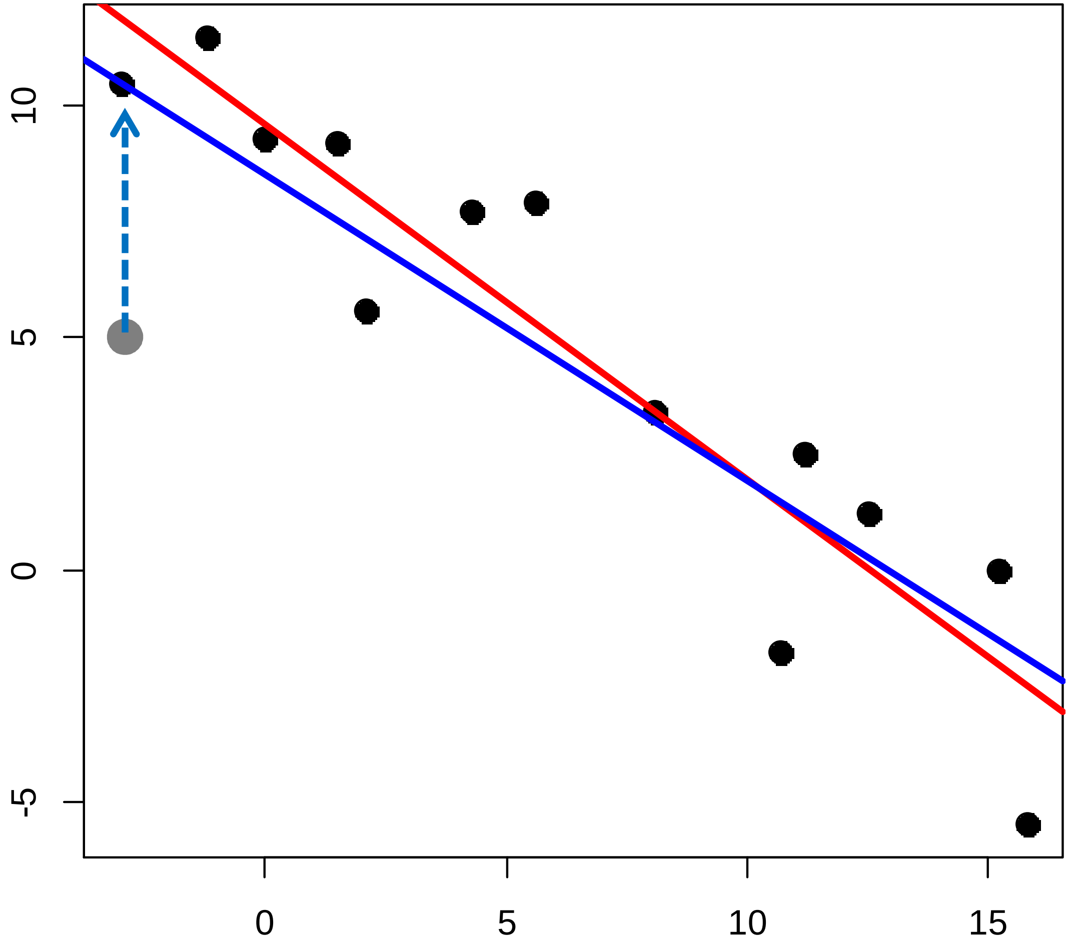
$\hat{y}_1 = 8.511853 - 0.6600736 \times (-3)$

$= 10.4920736$

$\rho_{xy} = -0.8467226$

$\rho_{xy'} = -0.9290495$

$R^2 = 0.863133$



# Esercizio 6

Nella tabella sottostante sono riportati i valori delle variabili $X$ e $Y$

| $X$ | -3 | 0 | 4.3 | 10.7 | -1.2 | 8.1 | 1.5 | 15.8 | 15.2 | 5.6 | 2.1 | 11.2 | 12.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | $\hat{y}_1$ | 9.3 | 7.7 | -1.8 | 11.5 | 3.4 | 9.2 | -5.5 | 0 | 7.9 | 5.6 | 2.5 | 1.2 |

$n = 13$

$y = 8.511853 - 0.6600736x$

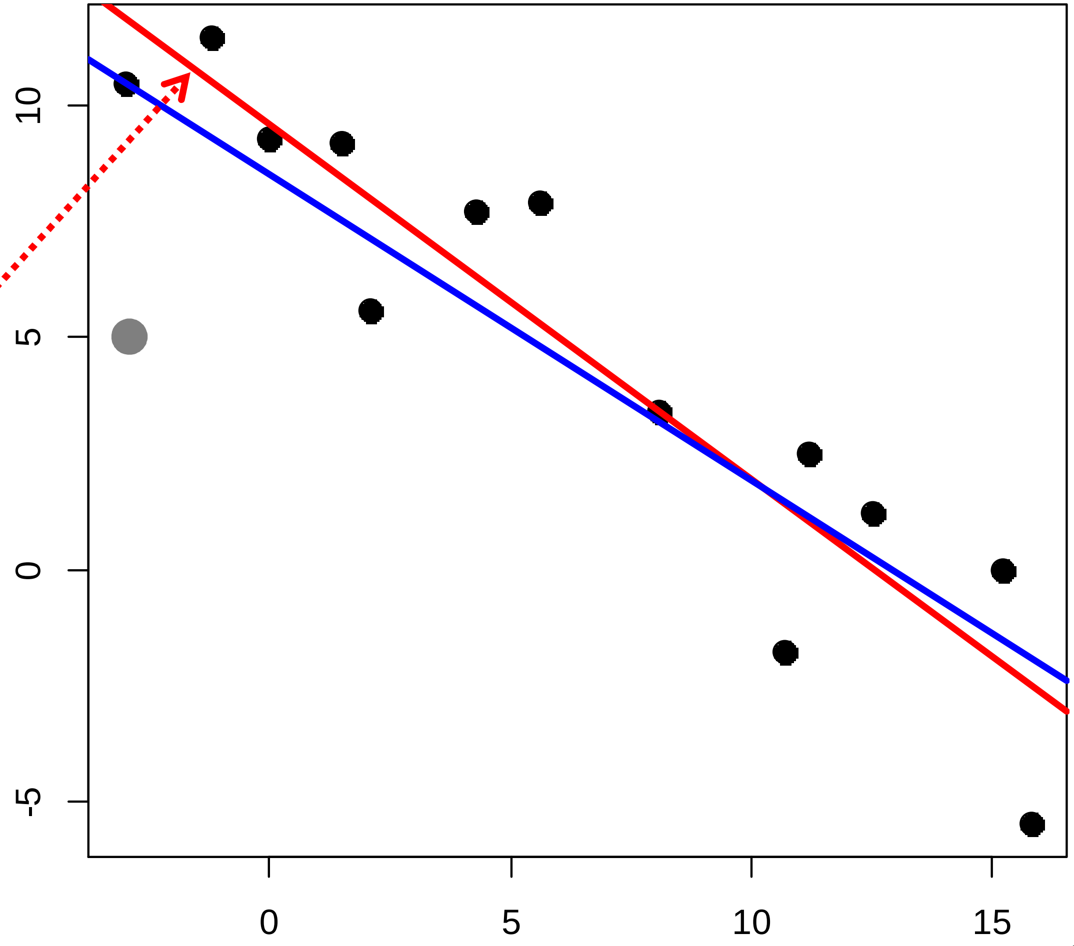
$\hat{y}_1 = 8.511853 - 0.6600736 \times (-3)$

$= 10.4920736$

$\rho_{xy} = -0.8467226$

$\rho_{xy'} = -0.9290495$

$R^2 = 0.863133$



# Esercizio 7

Il seguente grafico si riferisce al numero di passaggi tra Camera e Senato necessari per approvare una legge nella precedente legislatura (da *La Stampa* del 10/11/2016)

1. Quanti sono i dati in totale?
2. Di che tipo di variabile si tratta?
3. Cosa rappresenta questo grafico?
4. Quali indici di posizione possiamo calcolare?