# STATISTICA

Regressione-3

L’inferenza per il modello lineare semplice

# Regressione lineare:

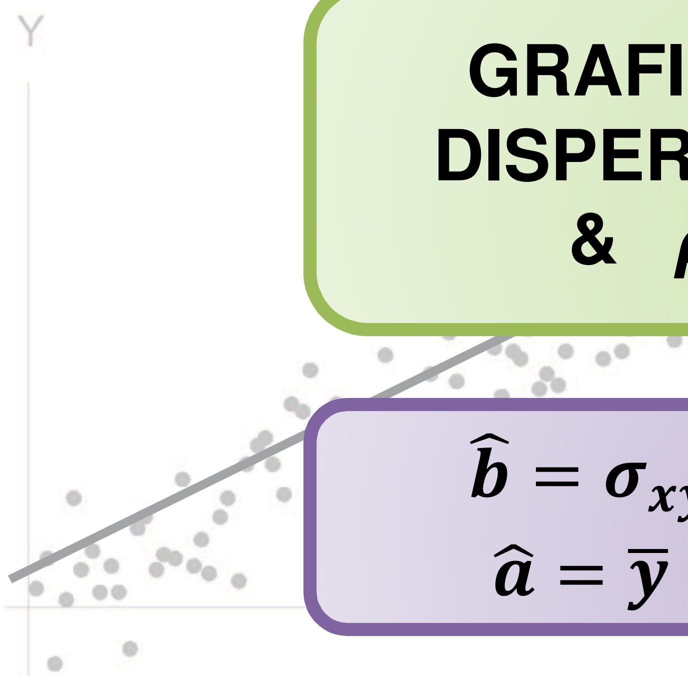



**GRAFICO DI DISPERSIONE**
**&** $\rho_{xy}$

**A.** Valutazione preliminare se una retta possa essere una buona approssimazione

$$\hat{b} = \sigma_{xy}/{\sigma_x}^2$$
$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

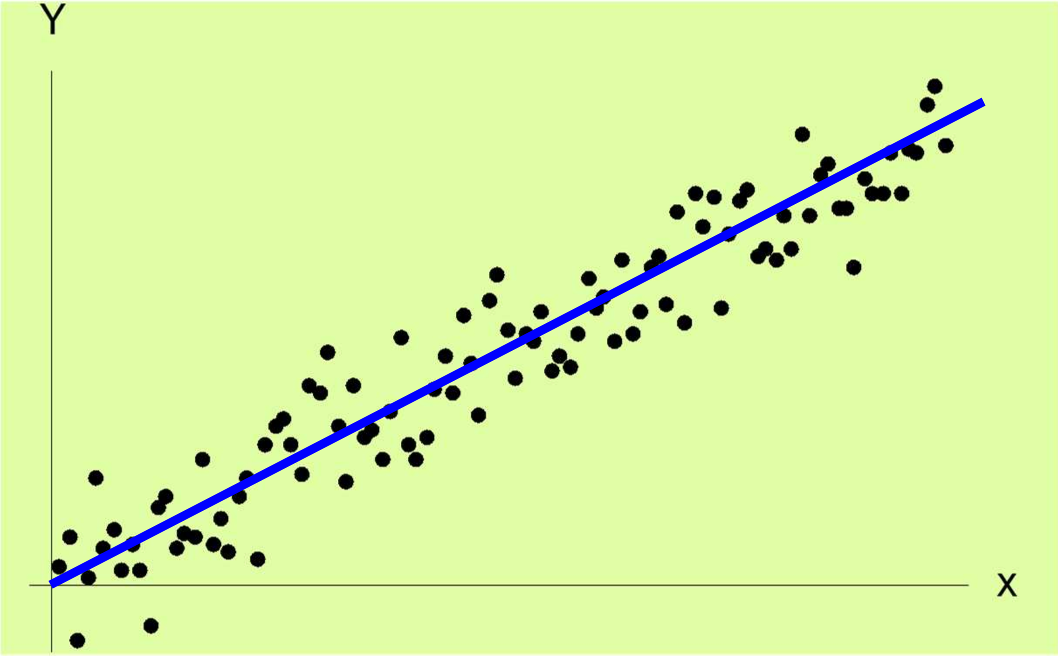
**B.** Stima dei parametri della retta.

$$R^2 = {\rho_{xy}}^2$$
**& analisi residui**

**C.** Valutazione della bontà di adattamento del **modello** ai dati

**D. Significatività della regressione**

# Inferenza




Il modello della **regressione lineare semplice**:

$$\boldsymbol{Y_i = a + b x_i + \varepsilon_i}$$

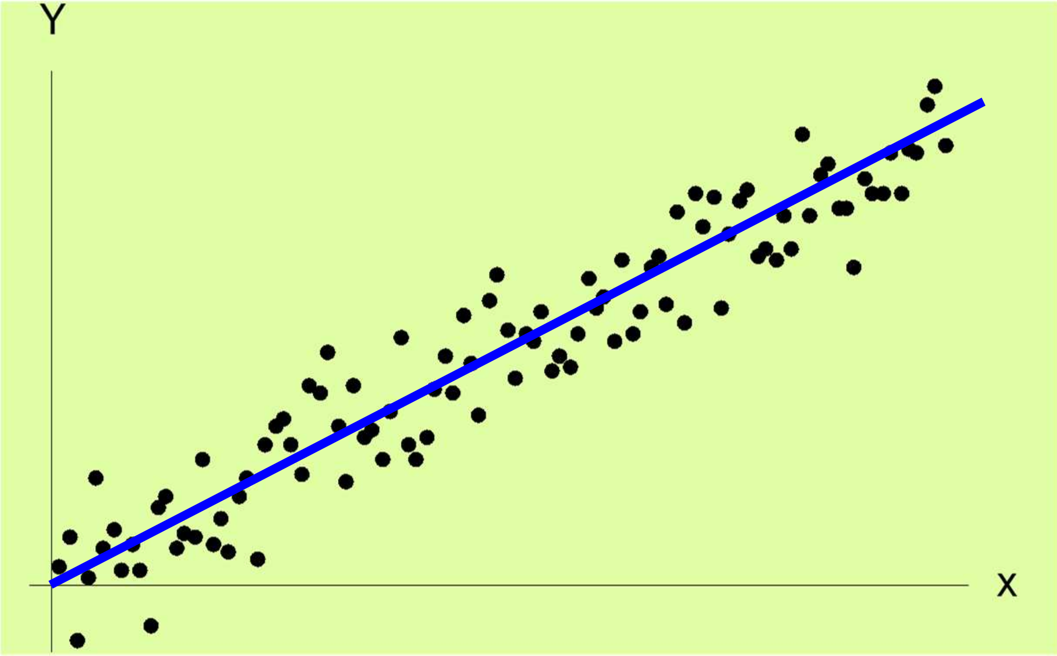
$$\boldsymbol{\varepsilon_i \sim N(0, \sigma^2)}$$

$$\boldsymbol{\varepsilon_i \text{ indipendenti}}$$

Verificare se il vero valore della pendenza nella popolazione di riferimento è davvero diverso da zero (⇔ previsioni!) oppure no:

$$\boldsymbol{H_0 : b = 0, \quad H_1 : b \neq 0}$$

# Inferenza




Il modello della
**regressione lineare semplice**:

$\boldsymbol{\varepsilon_1, \dots, \varepsilon_n}$ **i.i.d.** $\boldsymbol{\varepsilon_i \sim N(0, \sigma^2)}$

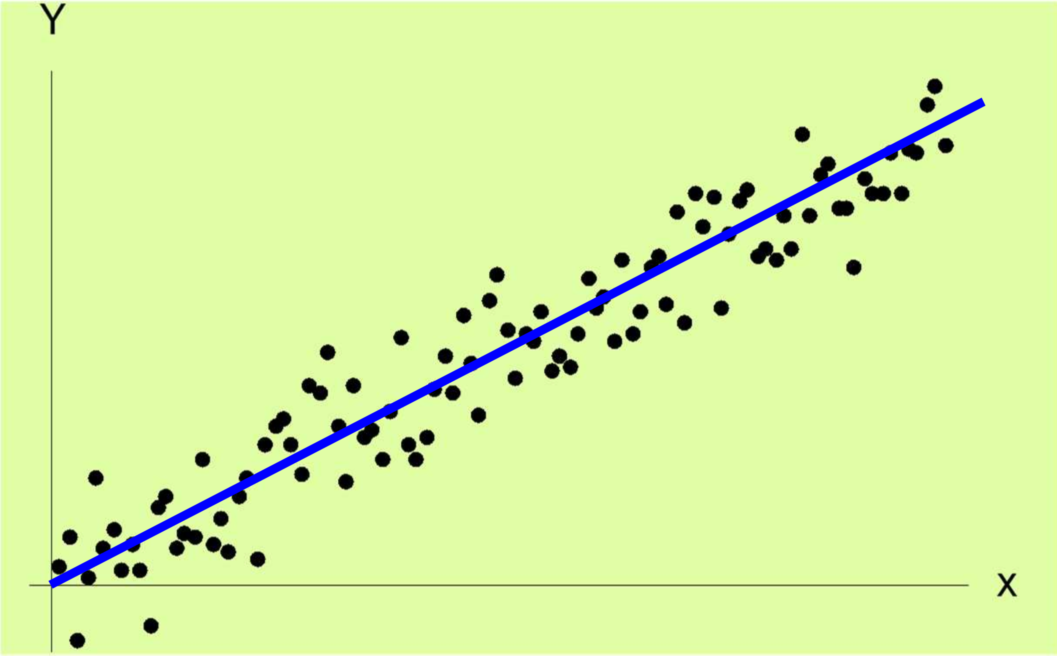
Osserviamo $\boldsymbol{e_i = y_i - \hat{y}_i}$

errori o ***residui***

**stima di $\sigma^2$ :**

**varianza**
**degli errori**

# Inferenza




Il modello della **regressione lineare semplice**:

$\boldsymbol{\varepsilon_1, \dots, \varepsilon_n}$ **i.i.d.** $\boldsymbol{\varepsilon_i \sim N(0, \sigma^2)}$

Osserviamo $\boldsymbol{e_i = y_i - \hat{y}_i}$

errori o ***residui***

$$\frac{1}{n}\sum y_i = \frac{1}{n}\sum \hat{y}_i$$

$$\frac{1}{n}\sum_{i=1}^{n} e_i = 0$$

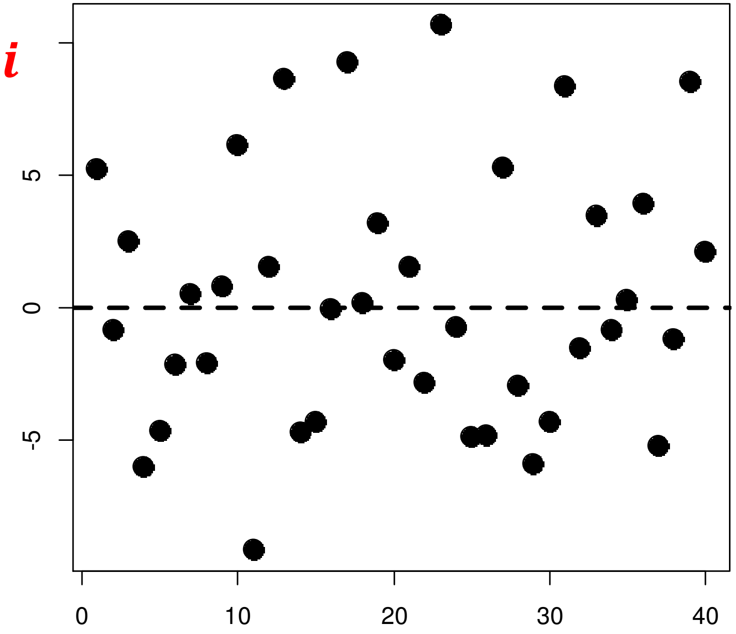
$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n} {e_i}^2$$

**stima di $\sigma^2$ :**

**varianza degli errori**

# Inferenza

$e_i$




Il modello della **regressione lineare semplice**:

$$\boldsymbol{\varepsilon_1, \dots, \varepsilon_n \ \text{i.i.d.}\ \varepsilon_i \sim N(0, \sigma^2)}$$

Osserviamo $\boldsymbol{e_i = y_i - \hat{y}_i}$

errori o ***residui***

$$\frac{1}{n}\sum y_i = \frac{1}{n}\sum \hat{y}_i$$

$$\boldsymbol{\frac{1}{n}\sum_{i=1}^{n} e_i = 0}$$

$$\boldsymbol{s^2 = \frac{1}{n-2}\sum_{i=1}^{n} {e_i}^2}$$

**stima di $\boldsymbol{\sigma^2}$ :**

**varianza degli errori**

# Inferenza

dalle stime agli stimatori:

$$B_n = \frac{\sum(Y_i - \bar{Y}_n)(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}$$

$$A_n = \bar{Y}_n - B_n\bar{x}$$

$$\boxed{\varepsilon_i \sim N(0, \sigma^2)}$$

$$\boxed{Y_i \sim N(a + bx_i, \sigma^2)}$$

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

$A_n$ e $B_n$ v.c. gaussiane

$$H_0 : b = 0 \qquad H_1 : b \neq 0$$

**rifiutiamo $H_0$ se:**

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

(rifiutiamo la casualità di una pendenza $\neq 0$)

# Inferenza

dalle stime agli stimatori:

$$B_n = \frac{\sum(Y_i - \bar{Y}_n)(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}$$

$$A_n = \bar{Y}_n - B_n\bar{x}$$

$$\boxed{\varepsilon_i \sim N(0, \sigma^2)}$$

$$\boxed{Y_i \sim N(a + bx_i, \sigma^2)}$$

$$\boxed{s^2 = \frac{1}{n-2}\sum_{i=1}^{n} {e_i}^2}$$

$A_n$ e $B_n$ v.c. gaussiane

$$\boldsymbol{H_0 : a = 0 \qquad H_1 : a \neq 0}$$

rifiutiamo $H_0$ se:

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

# Regressione lineare:

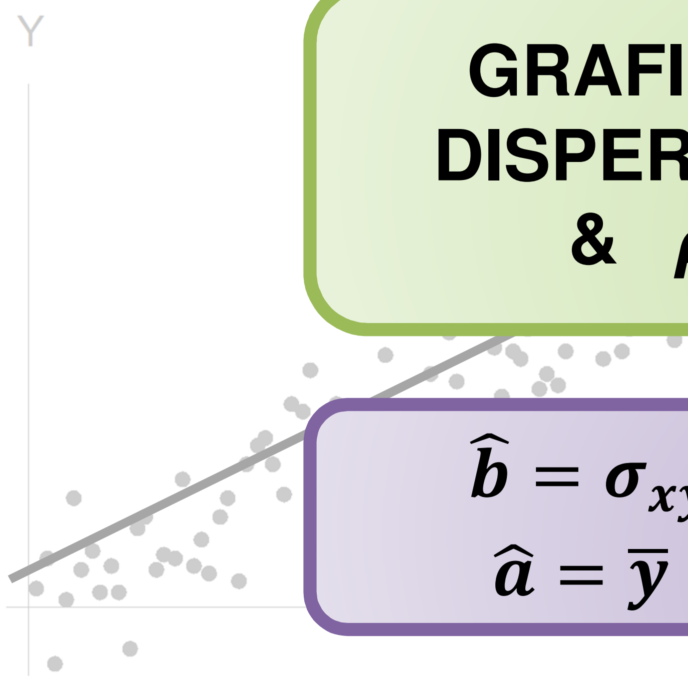



**GRAFICO DI DISPERSIONE & $\rho_{xy}$**

**A.** Valutazione preliminare se una retta possa essere una buona approssimazione

$$\hat{b} = \sigma_{xy}/\sigma_x^{\,2}$$

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

**B.** Stima dei parametri della retta.

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

**& analisi residui**

**C.** Valutazione della bontà di adattamento del **modello** ai dati

$$H_0 : b = 0$$

$$H_1 : b \neq 0$$

**D. Significatività della regressione**

# Esempio, Cont.

| $x_i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $y_i$ | 1.5 | 2.5 | 3 | 2.5 | 3.5 |
| $\hat{y}_i$ | 1.8 | 2.2 | 2.6 | 3.0 | 3.4 |
| $e_i$ | -0.3 | 0.3 | 0.4 | -0.5 | 0.1 |

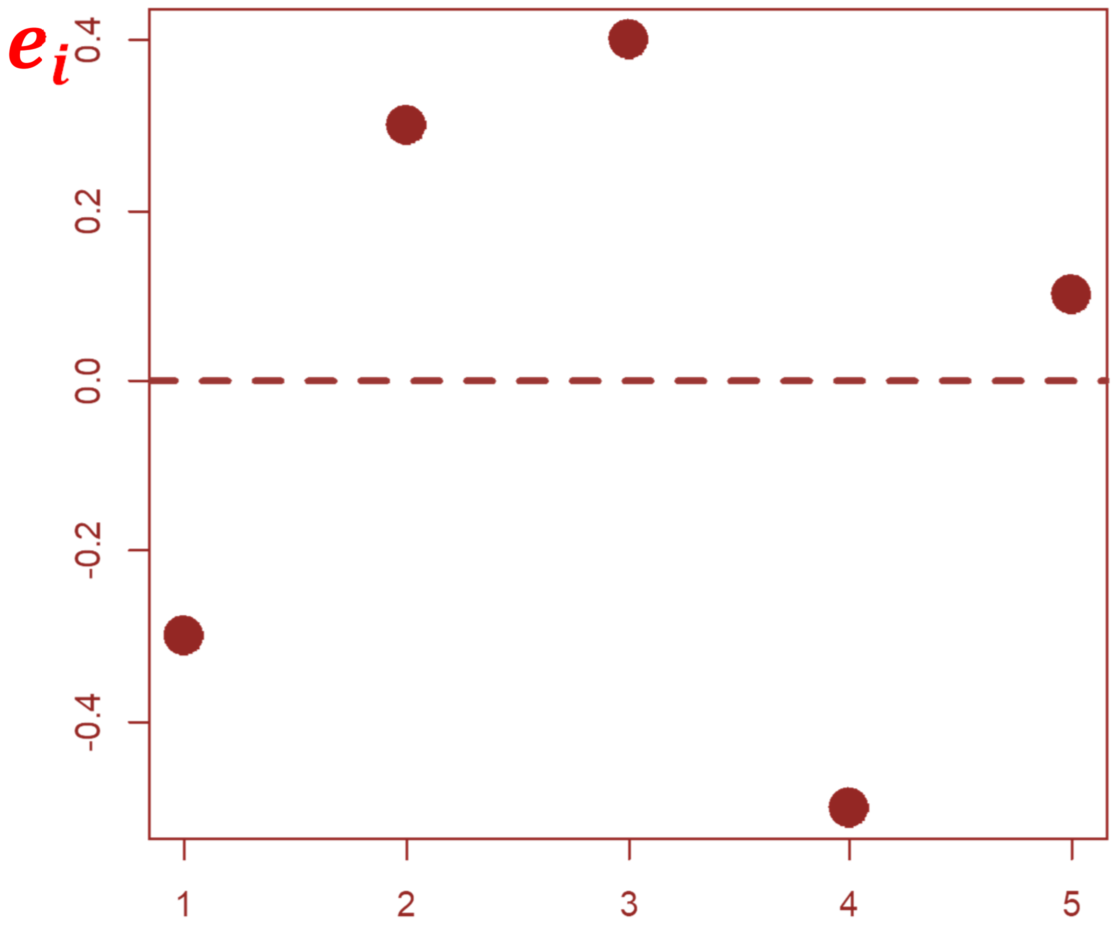
$$\hat{y}_i = 1.4 + 0.4x_i$$

$$\rho_{xy} = 0.85$$

$$R^2 = \rho^2 = 0.72$$

$$e_i = y_i - \hat{y}_i$$

$$s^2 = \frac{1}{3}\sum_{i=1}^{5} {e_i}^2 = 0.2$$



# Esempio, Cont.

| $x_i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $y_i$ | 1.5 | 2.5 | 3 | 2.5 | 3.5 |
| $\hat{y}_i$ | 1.8 | 2.2 | 2.6 | 3.0 | 3.4 |
| $e_i$ | -0.3 | 0.3 | 0.4 | -0.5 | 0.1 |

$$\hat{y}_i = 1.4 + \mathbf{0.4} x_i$$

$$H_0 : b = 0 \qquad H_1 : b \neq 0$$

$$\rho_{xy} = 0.85$$

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

$$R^2 = \rho^2 = 0.72$$

$$e_i = y_i - \hat{y}_i$$

$$s^2 = \frac{1}{3}\sum_{i=1}^{5} {e_i}^2 = \mathbf{0.2}$$

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = n{\sigma_x}^2 = 5 \times 2 = 10$$

# Esempio, Cont.

| $x_i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $y_i$ | 1.5 | 2.5 | 3 | 2.5 | 3.5 |
| $\hat{y}_i$ | 1.8 | 2.2 | 2.6 | 3.0 | 3.4 |
| $e_i$ | -0.3 | 0.3 | 0.4 | -0.5 | 0.1 |

$$\hat{y}_i = 1.4 + \mathbf{0.4}x_i$$

$$H_0 : b = 0 \qquad H_1 : b \neq 0$$

$$\rho_{xy} = 0.85$$

$$\frac{\mathbf{0.4}}{\sqrt{\frac{\mathbf{0.2}}{\mathbf{10}}}} = \mathbf{2.83} > \boldsymbol{t(5-2)}_{\mathbf{1-\frac{0.05}{2}}} ???$$

$$= \mathbf{3.182446}$$

$$R^2 = \rho^2 = 0.72$$

$$e_i = y_i - \hat{y}_i$$

**NON** POSSIAMO RIFIUTARE $H_0$, LA REGR. **NON** E' SIGNIFICATIVA , AL 5%!

$$\boldsymbol{s^2} = \frac{1}{3}\sum_{i=1}^{5} {e_i}^2 = \mathbf{0.2} \qquad \boldsymbol{\sum_{i=1}^{n}(x_i - \bar{x})^2 = n{\sigma_x}^2 = 5 \times 2 = 10}$$

# Esempio Cont.

| $x_i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $y_i$ | 1.5 | 2.5 | 3 | 2.5 | 3.5 |

$\bar{x} = 3, \bar{y} = 2.6$

$\hat{b} = 0.4$

$\hat{a} = 1.4$

$y = 1.4 + 0.4x$

**La previsione è attendibile?**

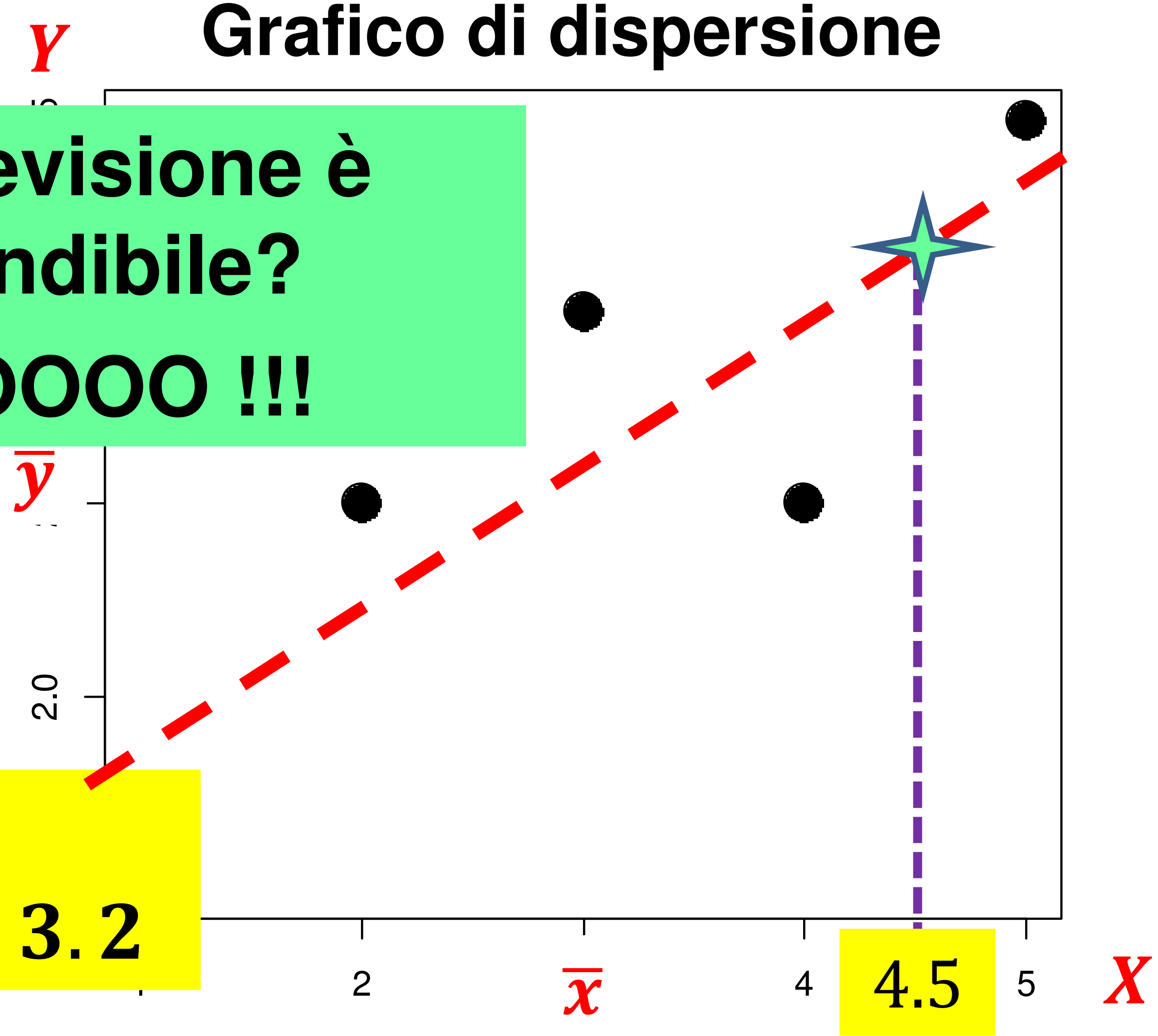
**NOOOOO !!!**

**Previsione:**

$\boldsymbol{x = 4.5,}$

$\boldsymbol{y = 1.4 + 0.4 \times 4.5 = 3.2}$

**Grafico di dispersione**



# Esempio Cont.

| $x_i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $y_i$ | 1.5 | 2.5 | 3 | 2.5 | 3.5 |

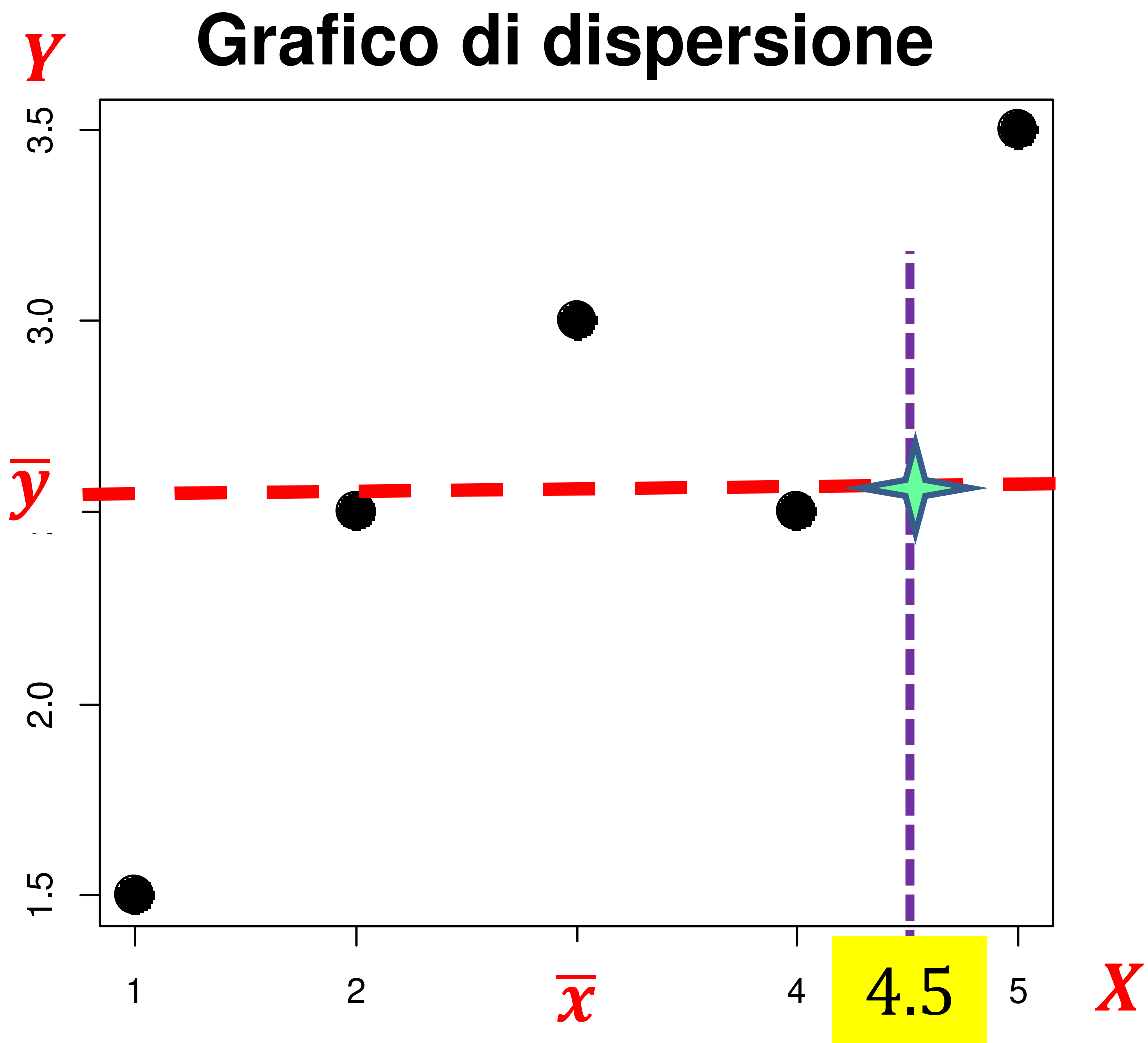
$\bar{x} = 3, \bar{y} = 2.6$

$\hat{b} = 0.4$

$\hat{a} = 1.4$

$y = 1.4 + 0.4x$

La pendenza è positiva «per caso»



# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |

1. Le due variabili sono correlate? Come e quanto?
2. Stimare i parametri della retta dei minimi quadrati per i dati $(x_i, y_i)$.
3. Valutare la bontà di adattamento ai dati della retta
4. Determinare se la regressione sia statisticamente significativa al livello del 5%
5. Fornire la previsione del modello per i valori 4 e 15 della variabile indipendente

# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |

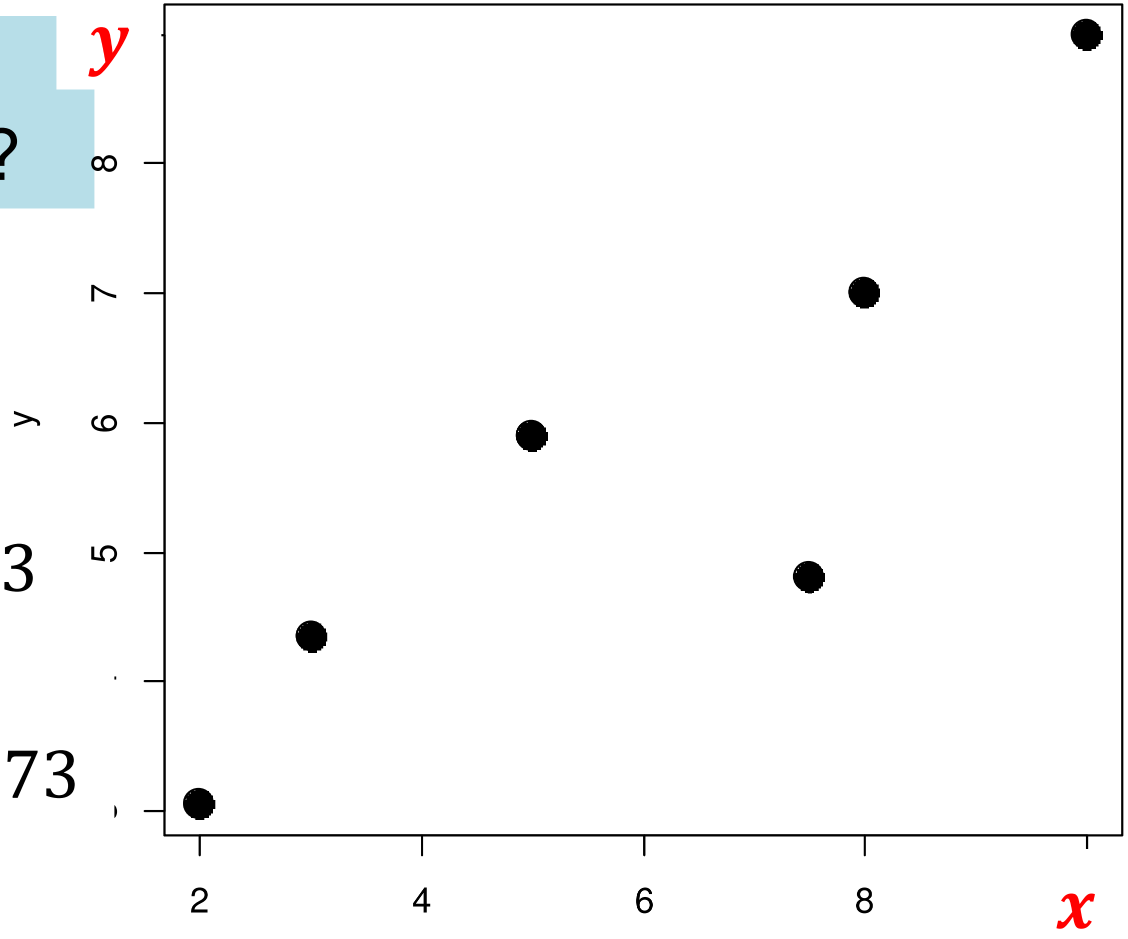
1. Le due variabili sono correlate? Come e quanto?

$\bar{x} = 5.9$

$\bar{y} = 5.68$

$${\sigma_x}^2 = \frac{1}{6}\sum_i (x_i - 5.9)^2 = 8.23$$

$${\sigma_y}^2 = \frac{1}{6}\sum_i (y_i - 5.68)^2 = 3.73$$



# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |
| $x_i y_i$ | 6.2 | 12.9 | 29.5 | 36 | 56 | 90 |

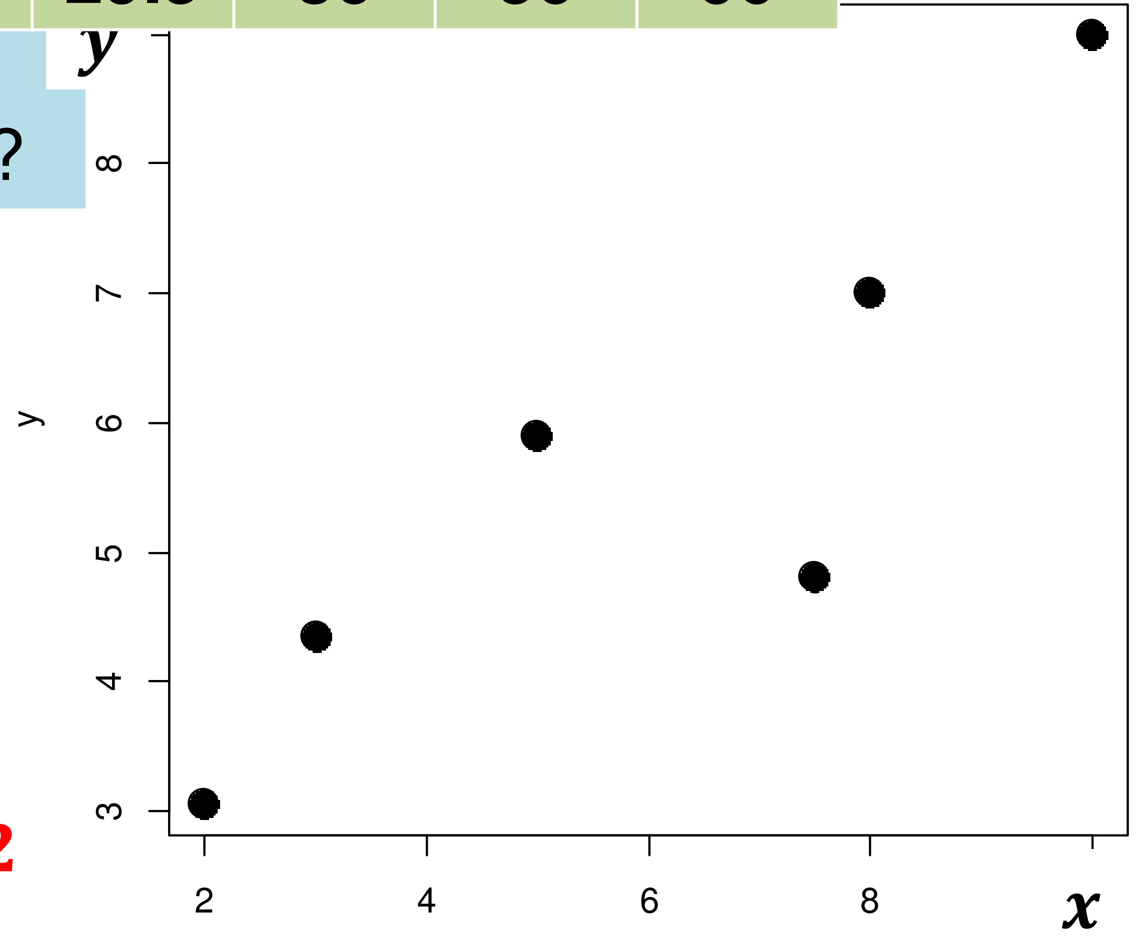
1. Le due variabili sono correlate? Come e quanto?

$\bar{x} = 5.9 \qquad \sigma_x{}^2 = 8.23$

$\bar{y} = 5.68 \qquad \sigma_y{}^2 = 3.73$

$$\boldsymbol{\sigma_{xy}} = \frac{\mathbf{1}}{\mathbf{6}} \sum_{\boldsymbol{i}} \boldsymbol{x_i y_i} - \boldsymbol{\bar{x}} \times \boldsymbol{\bar{y}} =$$

$$= 38.43 - 33.51 = \mathbf{4.92}$$



# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |
| $x_i y_i$ | 6.2 | 12.9 | 29.5 | 36 | 56 | 90 |

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

$\bar{x} = 5.9 \qquad {\sigma_x}^2 = 8.23$

$\bar{y} = 5.68 \qquad {\sigma_y}^2 = 3.73$

$$\boldsymbol{\sigma_{xy}} = \frac{1}{6}\sum_i x_i y_i - \bar{x} \times \bar{y} =$$

$$= 38.43 - 33.51 = \mathbf{4.92}$$

$$\boldsymbol{\rho_{xy}} = \frac{\mathbf{4.92}}{\sqrt{\mathbf{8.23} \times \mathbf{3.73}}} = \mathbf{0.89}$$

Indicazione di correlazione lineare, positiva

# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | **3.1** | **4.3** | **5.9** | **4.8** | **7** | **9** |

1. Le due variabili sono correlate? Come e quanto?
2. Stimare i parametri della retta dei minimi quadrati per i dati $(x_i, y_i)$.
3. Valutare la bontà di adattamento ai dati della retta

$$\boldsymbol{R^2 = \rho^2 = 0.89^2 = 0.79 > 0.70}\ ,\ \textbf{accettabile}$$

(*risposta parziale*)

# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |

2. Stimare i parametri della retta dei minimi quadrati per i dati $(x_i, y_i)$.

$\bar{x} = 5.9$

$\bar{y} = 5.68$

${\sigma_x}^2 = 8.23$

${\sigma_y}^2 = 3.73$

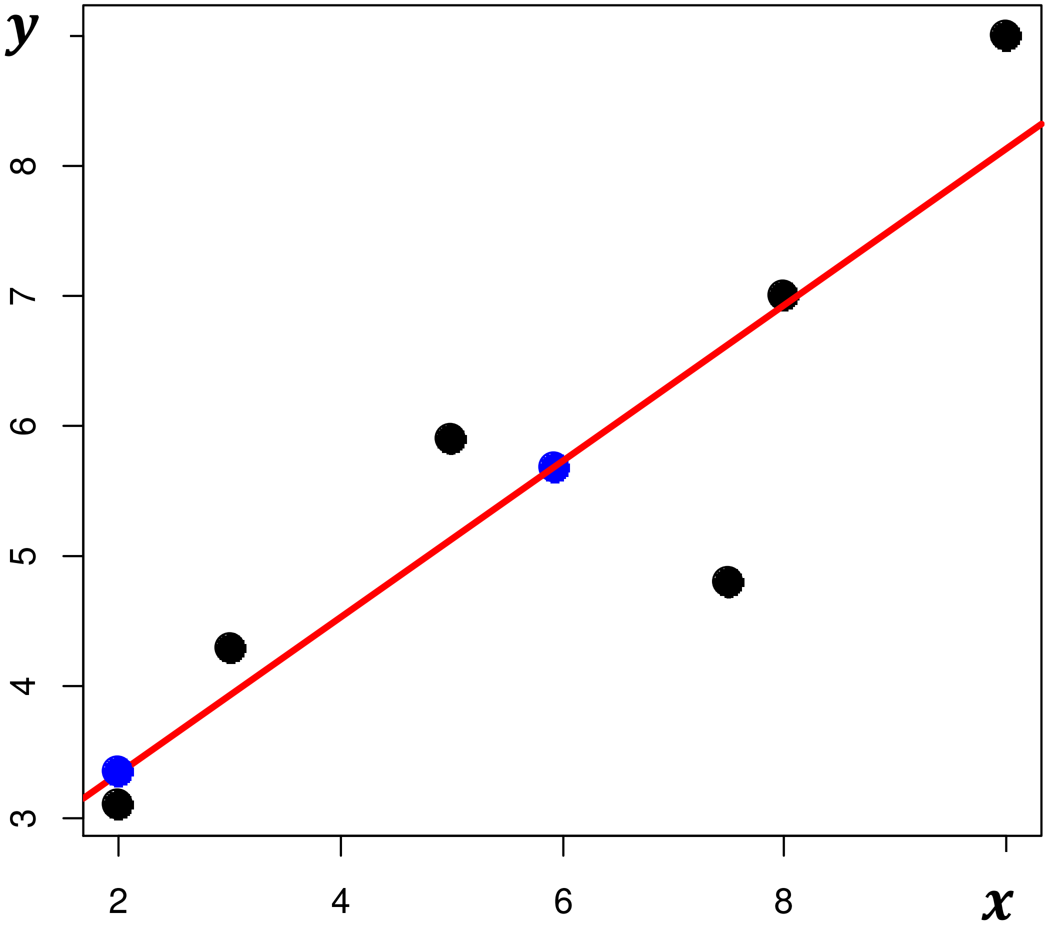
$\sigma_{xy} = 4.92$

$$\hat{b} = \frac{\sigma_{xy}}{{\sigma_x}^2} = \frac{4.92}{8.23} = 0.60$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} =$$
$$= 5.68 - 0.60 \times 5.9 = 2.14$$

# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |




$$\bar{x} = 5.9$$
$$\bar{y} = 5.68$$

$$x = 2,$$
$$\hat{y} = 2.14 + 0.6 \times = 3.34$$

$$\hat{b} = 0.60$$

$$\hat{a} = 2.14$$

# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |
| $\hat{y}_i$ | 3.34 | 3.94 | 5.14 | 6.64 | 6.94 | 8.14 |
| $e_i$ | - 0.24 | 0.36 | 0.76 | - 1.84 | 0.06 | 0.86 |

$$\hat{b} = 0.60 \qquad \hat{a} = 2.14$$

$$\hat{y}_i = 2.14 + 0.6x_i$$

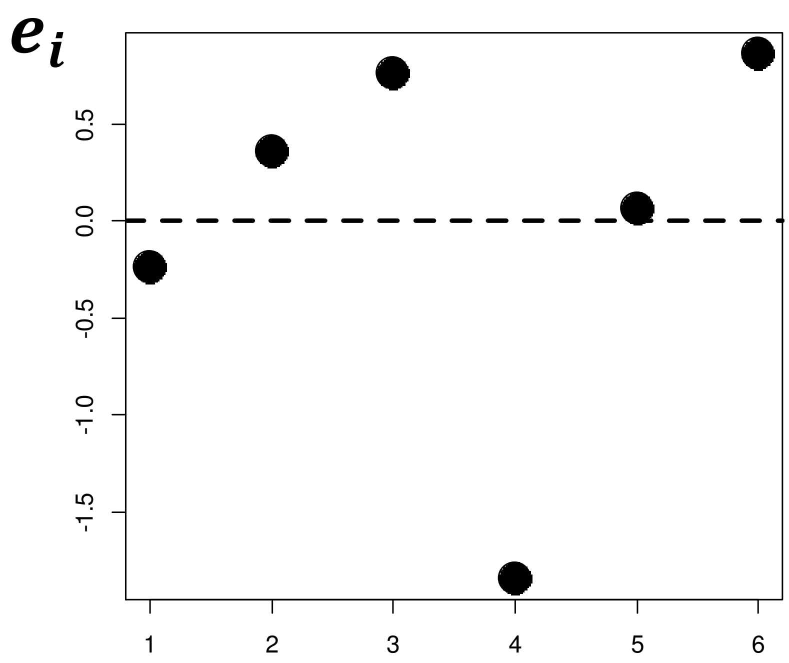
$$e_i = y_i - (2.14 + 0.6x_i)$$

$$s^2 = \frac{1}{4}\sum_i {e_i}^2 = 1.22$$

# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |
| $\hat{y}_i$ | 3.34 | 3.94 | 5.14 | 6.64 | 6.94 | 8.14 |
| $e_i$ | - 0.24 | 0.36 | 0.76 | - 1.84 | 0.06 | 0.86 |

3. Valutare la bontà di adattamento ai dati della retta (se avessimo più dati…)



# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |

4. Determinare se la regressione sia statisticamente significativa al livello del 5%

$\bar{x} = 5.9$

$\bar{y} = 5.68$

${\sigma_x}^2 = 8.23$

${\sigma_y}^2 = 3.73$

$\sigma_{xy} = 4.92$

$\boldsymbol{s^2 = 1.22}$

$\boldsymbol{\hat{b} = 0.60}$

$\boldsymbol{H_0 : b = 0 \quad vs \quad H_1 : b \neq 0}$

# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |

4. Determinare se la regressione sia statisticamente significativa al livello del 5%

$\bar{x} = 5.9$

$\bar{y} = 5.68$

${\sigma_x}^2 = 8.23$

${\sigma_y}^2 = 3.73$

$\sigma_{xy} = 4.92$

$\boldsymbol{s^2 = 1.22}$

$\boldsymbol{\hat{b} = 0.60}$

$\boldsymbol{H_0 : b = 0 \quad vs \quad H_1 : b \neq 0}$

$$\frac{|\hat{b}|}{\sqrt{\dfrac{s^2}{n{\sigma_x}^2}}} =$$

# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |

4. Determinare se la regressione sia statisticamente significativa al livello del 5%

$\bar{x} = 5.9$

$\bar{y} = 5.68$

${\sigma_x}^2 = 8.23$

${\sigma_y}^2 = 3.73$

$\sigma_{xy} = 4.92$

$\boldsymbol{s^2 = 1.22}$

$\boldsymbol{\hat{b} = 0.60}$

$\boldsymbol{H_0 : b = 0 \quad vs \quad H_1 : b \neq 0}$

$$\frac{|\hat{b}|}{\sqrt{\dfrac{s^2}{n{\sigma_x}^2}}} = \frac{0.60}{\sqrt{\dfrac{1.22}{6 \times 8.23}}} = 3.82$$

# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |

4. Determinare se la regressione sia statisticamente significativa al livello del 5%

Rifiutiamo l'ipotesi nulla, e quindi **la regressione é significativa** al livello del 5%

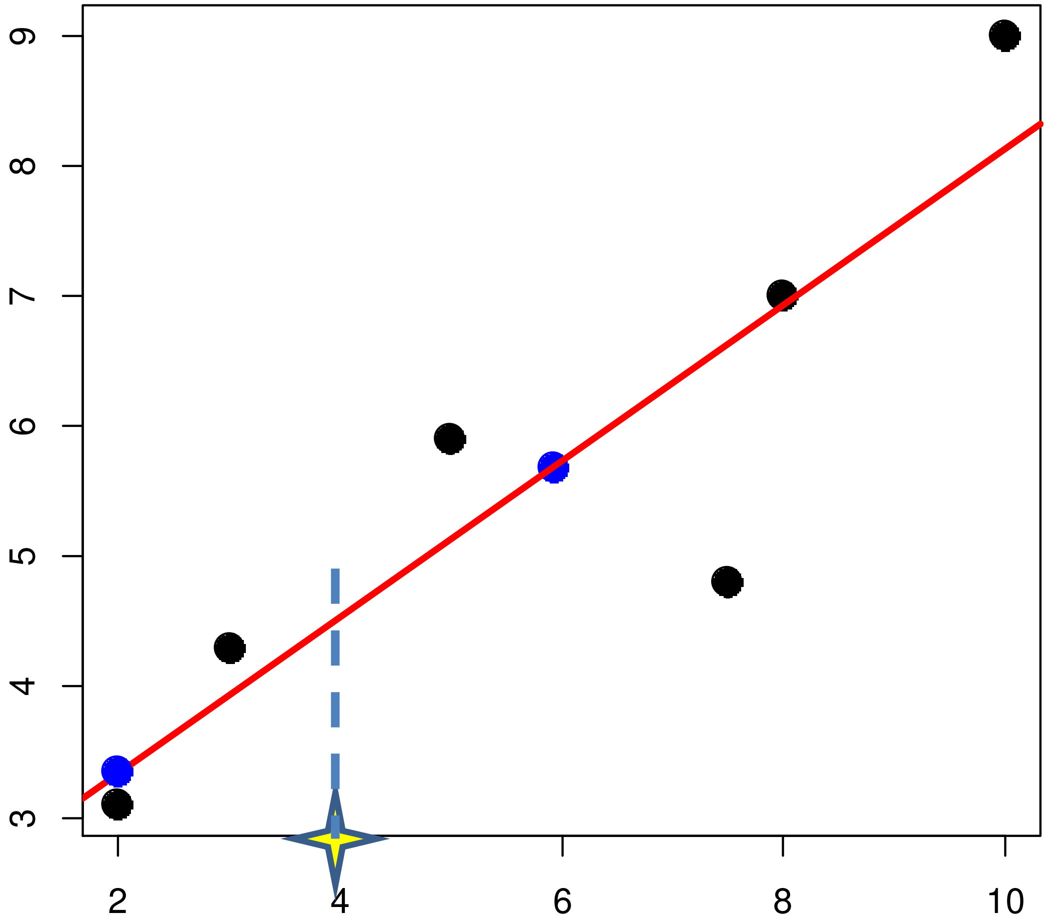
$$H_0 : b = 0 \quad vs \quad H_1 : b \neq 0$$

$$\frac{|\hat{b}|}{\sqrt{\frac{s^2}{n\sigma_x{}^2}}} = \frac{0.60}{\sqrt{\frac{1.22}{6 \times 8.23}}} = 3.82$$

$$t(4)_{1-\frac{0.05}{2}} = t(4)_{0.975} = 2.7764$$

# Esercizio 2

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |



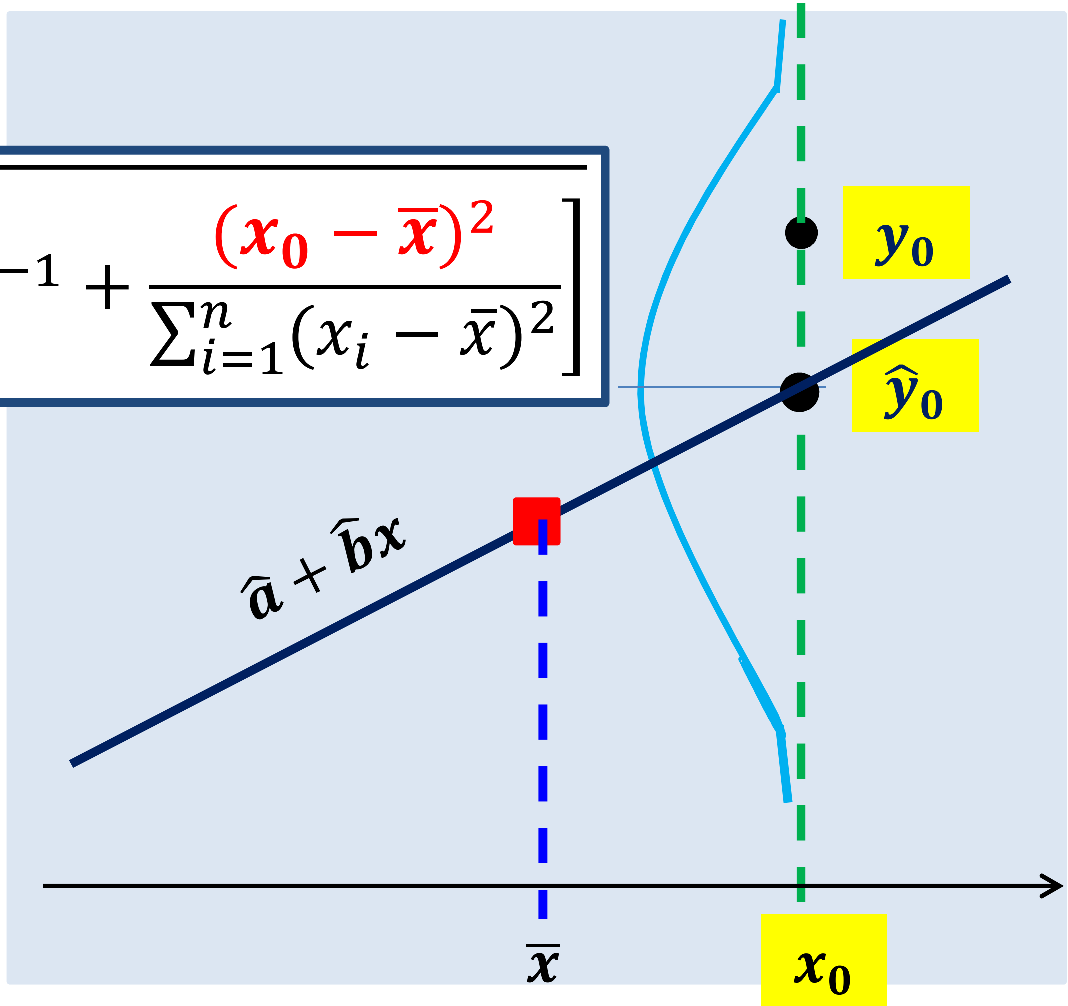

5. Fornire la previsione del modello per i valori 4 e 15 della variabile indipendente

$$x = 4, \hat{y} = 2.14 + 0.6 \times 4 = 4.54$$

$x = 15$ : non è possibile fare una previsione, perché fuori dal campo delle osservazioni

# Inferenza per la previsione

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




IC della **risposta** di un nuovo "individuo" con covariata pari a $\boldsymbol{x_0}$

# Esercizio 2, cont.

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |

5. Fornire la previsione del
er i valori 4 e
ariabile
nte

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

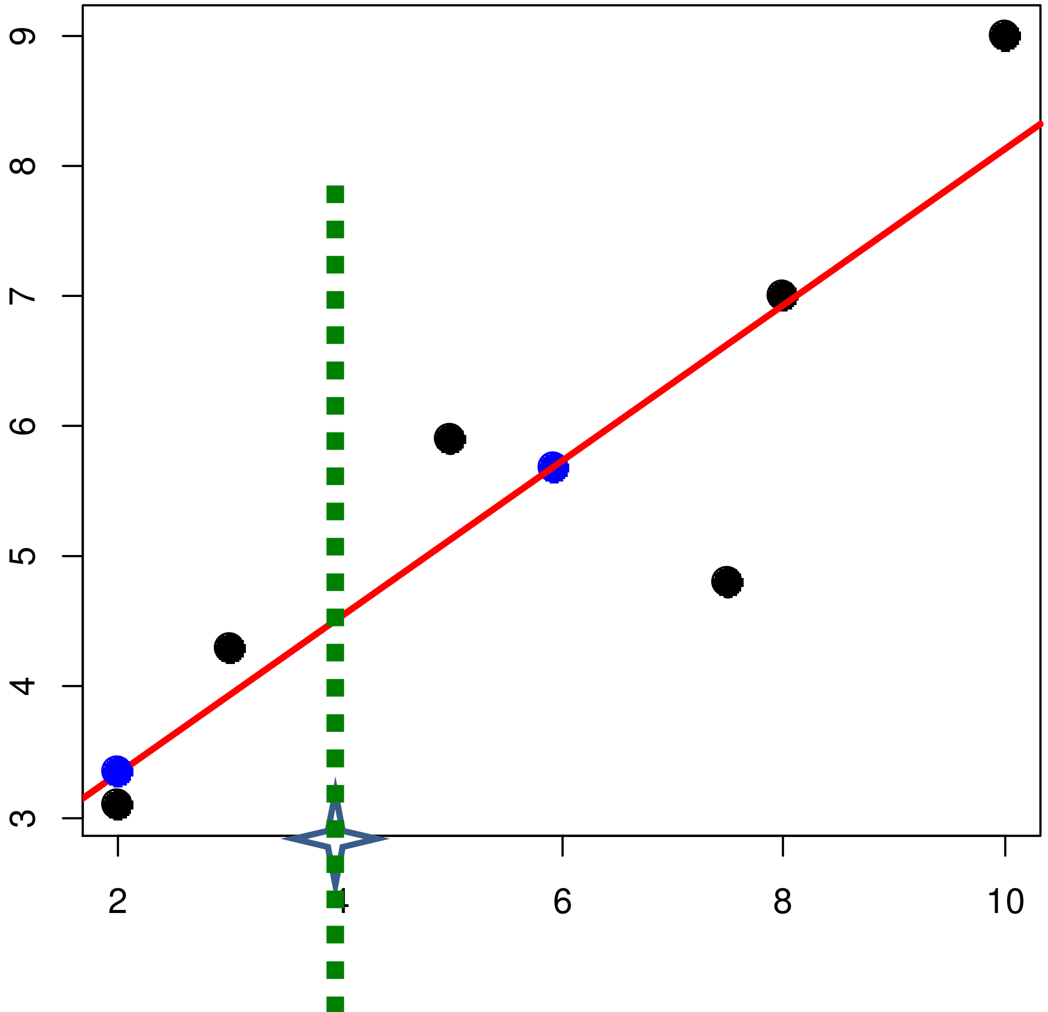
$$\boldsymbol{x = 4, \hat{y} = 2.14 + 0.6 \times 4 = 4.54}$$

$$\alpha = 0.05$$

$$\boxed{4.54 \mp 2.7764 \times \sqrt{1.22\left[1+\frac{1}{6}+\frac{(4-5.9)^2}{6\times 8.23}\right]} \Rightarrow (1.12, 7.95)}$$

# Esercizio 2, cont.

| $x_i$ | 2 | 3 | 5 | 7.5 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 3.1 | 4.3 | 5.9 | 4.8 | 7 | 9 |




5. Fornire la previsione del modello per i valori 4 e 15 della variabile indipendente

$$x = 4, \hat{y} = 2.14 + 0.6 \times 4 = 4.54$$

$$\boxed{(1.12, 7.95)}$$

# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, $X$ ed $Y$:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |

1. Rappresentare l'andamento congiunto di $Y$ in funzione di $X$ mediante un opportuno grafico: le due variabili sono correlate? Come e quanto?
2. Determinare le stime dei minimi quadrati dei parametri della retta interpolatrice dei dati
3. Determinare la significatività del modello di regressione lineare basato sulla retta al punto 2.
4. Valutare con un opportuno indice la bontà di adattamento del modello ai dati
5. Indicare entro quale intervallo sia possibile fare previsione sui valori di $Y$

# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

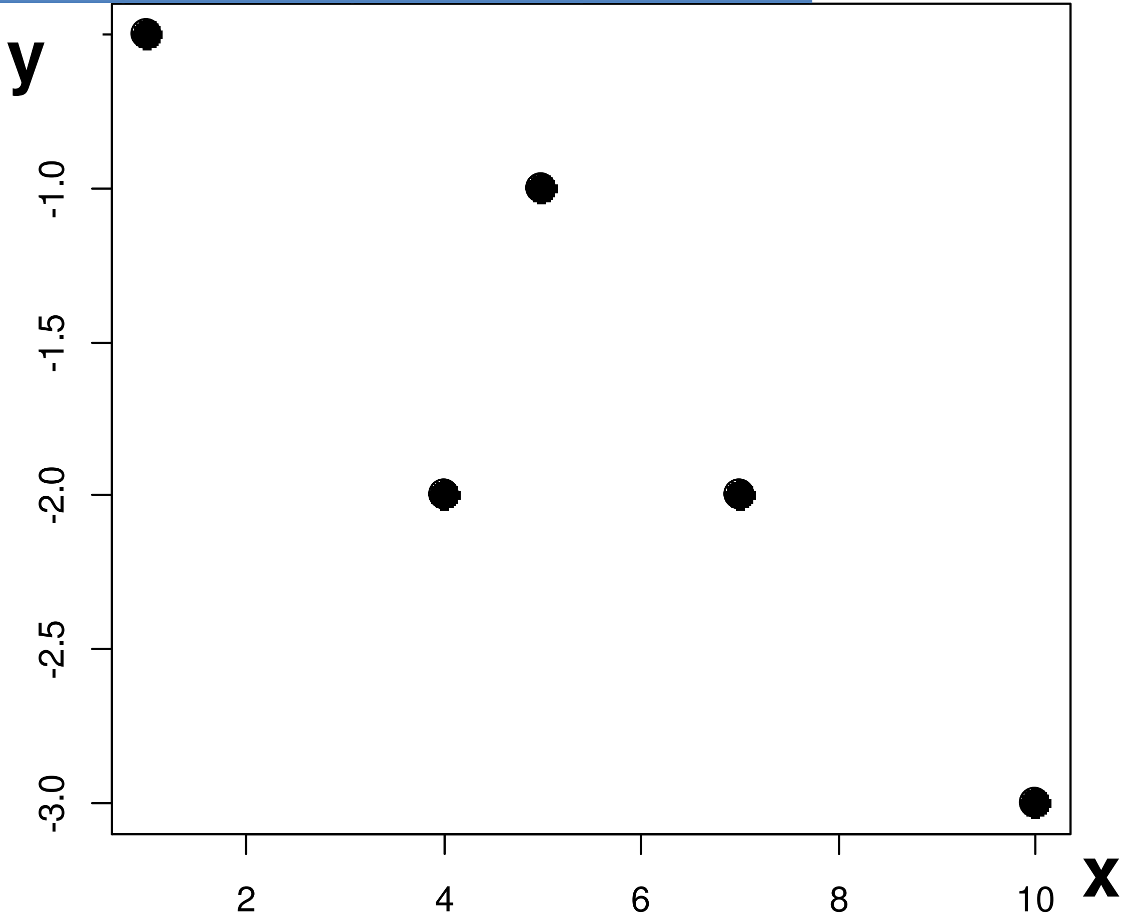
| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |

$$\bar{x} = 5.4$$

$$\bar{y} = -1.7$$

$${\sigma_x}^2 = 9.04$$

$${\sigma_y}^2 = 0.76$$



# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |
| xy | -0.5 | -8 | -5 | -14 | -30 |

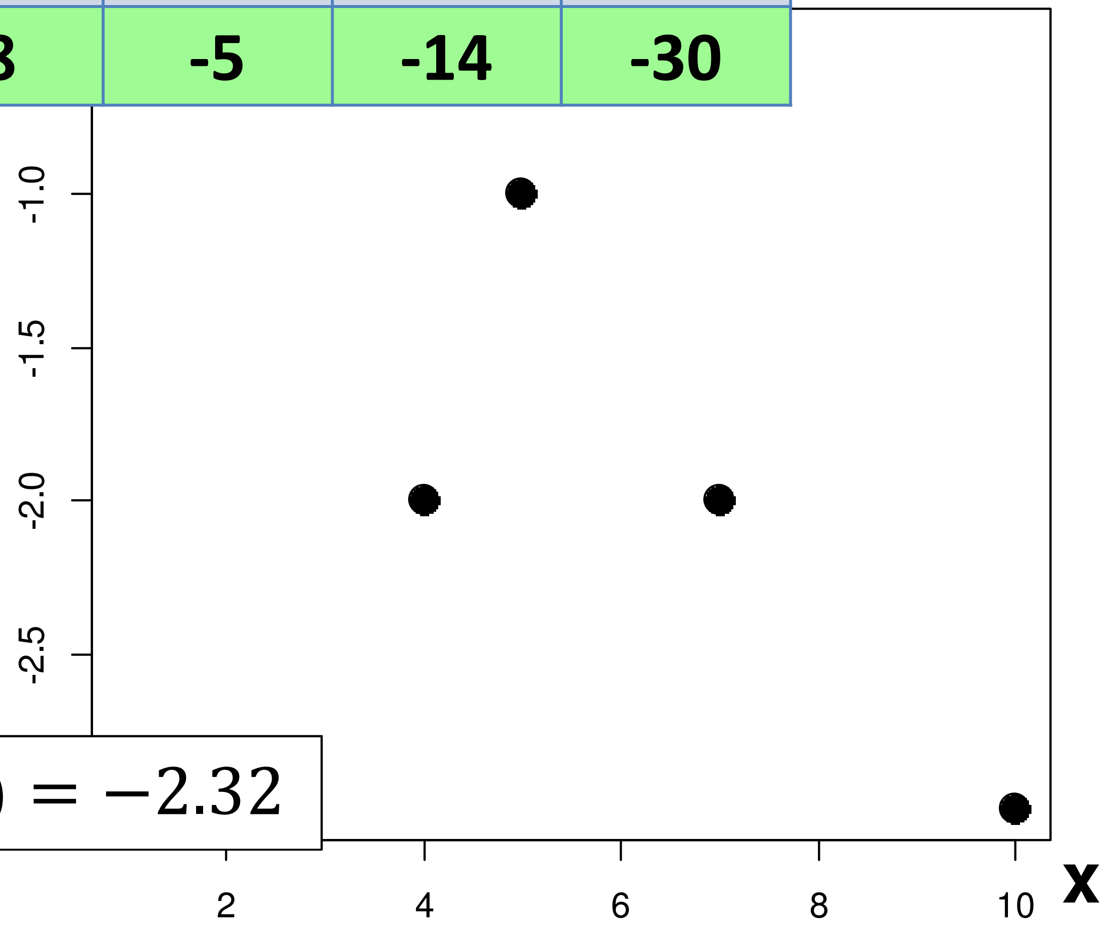
$\bar{x} = 5.4$

$\bar{y} = -1.7$

${\sigma_x}^2 = 9.04$

${\sigma_y}^2 = 0.76$

$$\boxed{\sigma_{xy} = -11.5 - 5.4 \times (-1.7) = -2.32}$$



# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

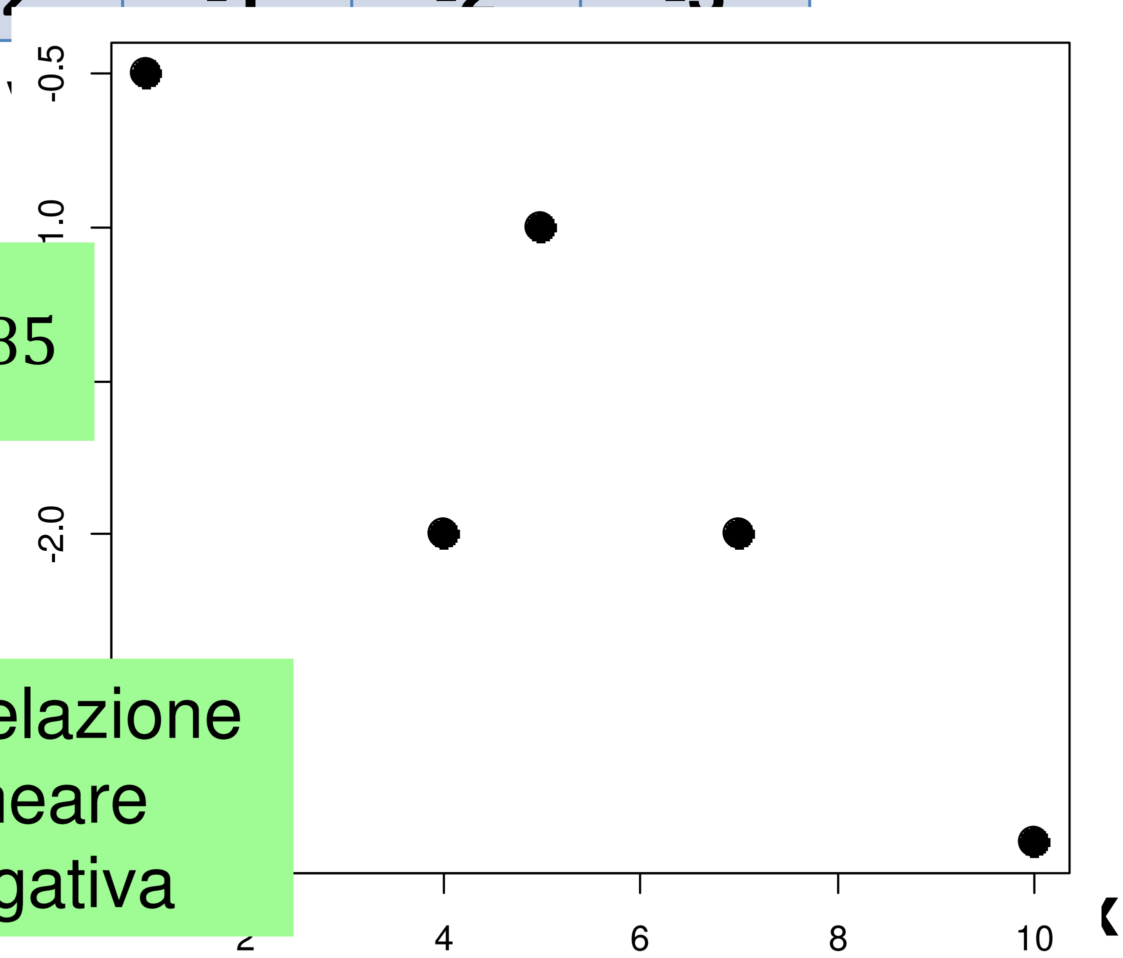
| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |

$$\rho_{xy} = \frac{-2.32}{\sqrt{9.04 \times 0.76}} = -0.885$$

$${\sigma_x}^2 = 9.04$$

$${\sigma_y}^2 = 0.76$$

$$\boxed{\sigma_{xy} = -2.32}$$




Correlazione lineare negativa

# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |

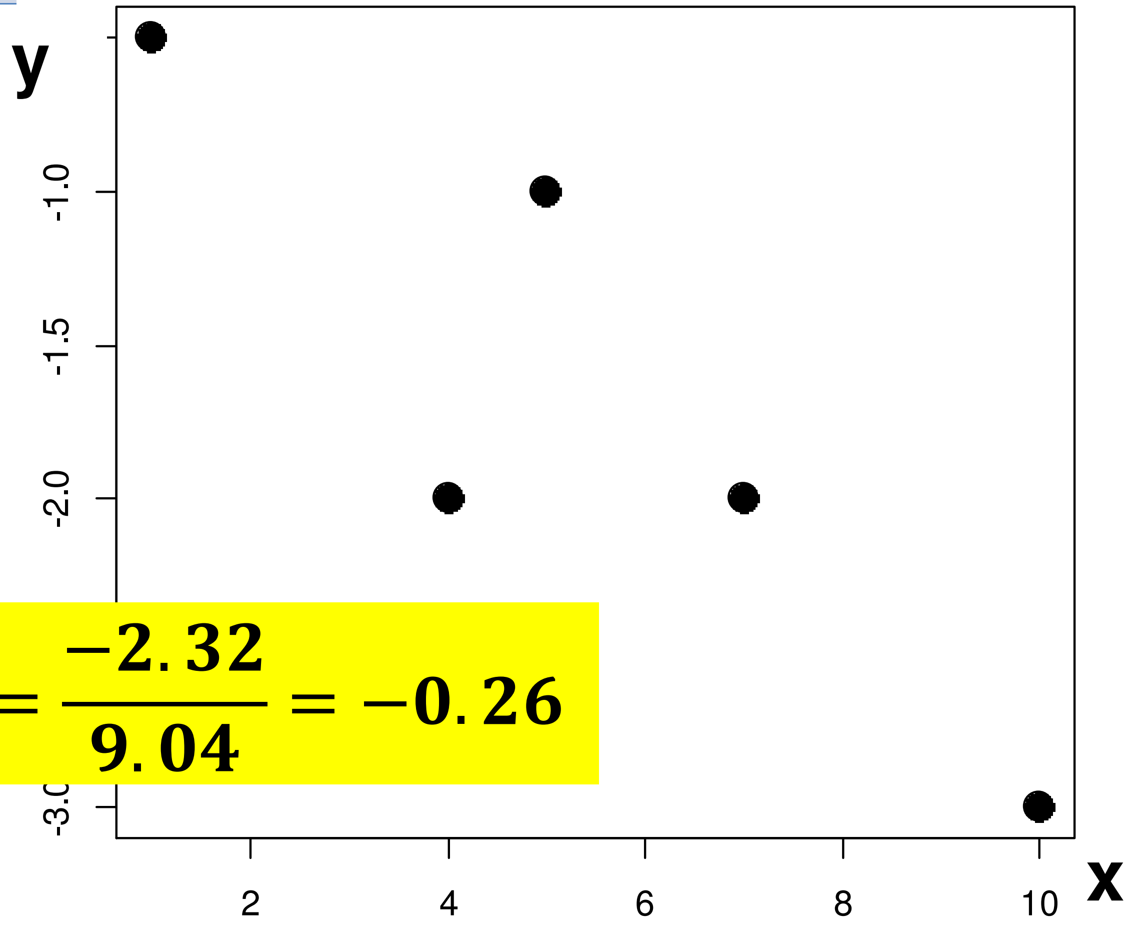
$\bar{x} = 5.4$

$\bar{y} = -1.7$

$\boldsymbol{\sigma_x}^2 = \mathbf{9.04}$

$\boldsymbol{\sigma_{xy}} = \mathbf{-2.32}$

$$\hat{b} = \frac{-2.32}{9.04} = -0.26$$



# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |

$$\bar{x} = 5.4$$

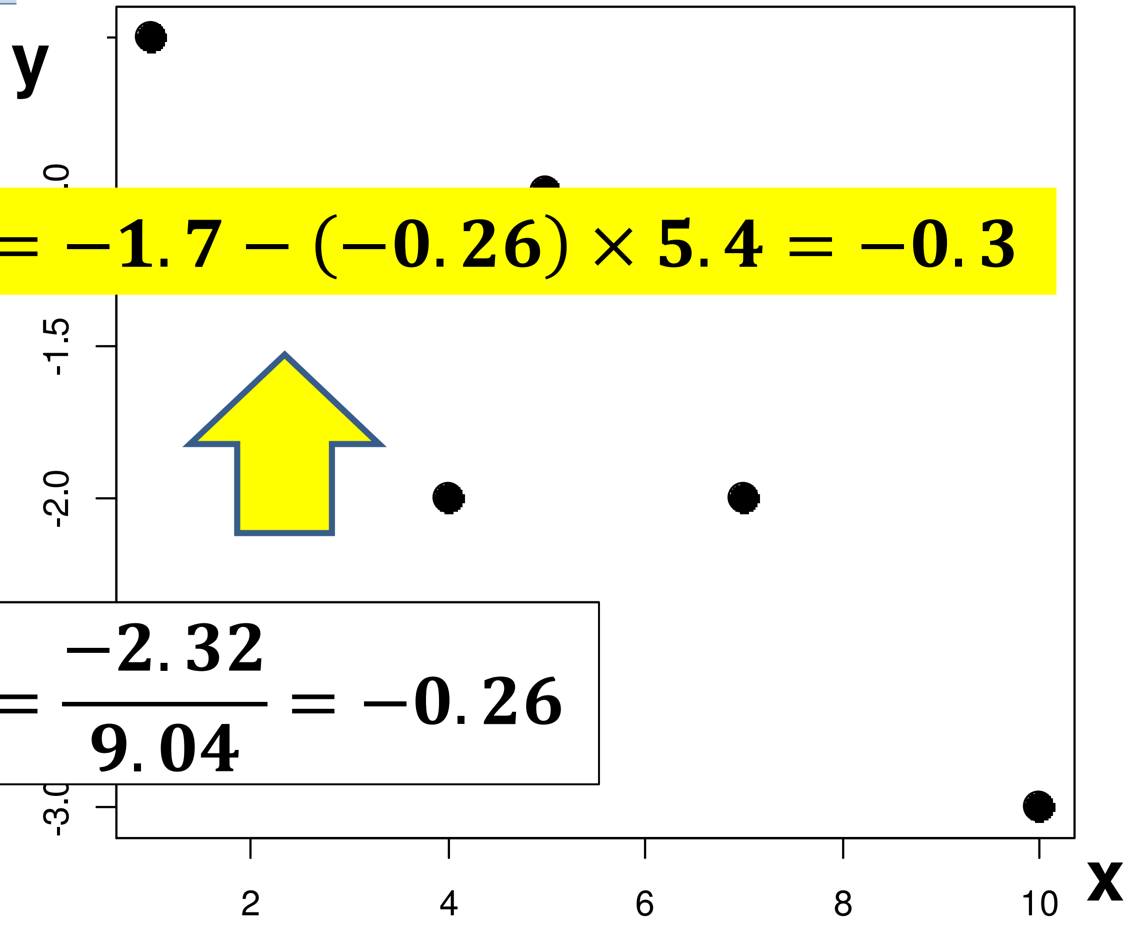
$$\bar{y} = -1.7$$

$$\sigma_x{}^2 = 9.04$$

$$\sigma_{xy} = -2.32$$

$$\hat{a} = -1.7 - (-0.26) \times 5.4 = -0.3$$

$$\hat{b} = \frac{-2.32}{9.04} = -0.26$$



# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |

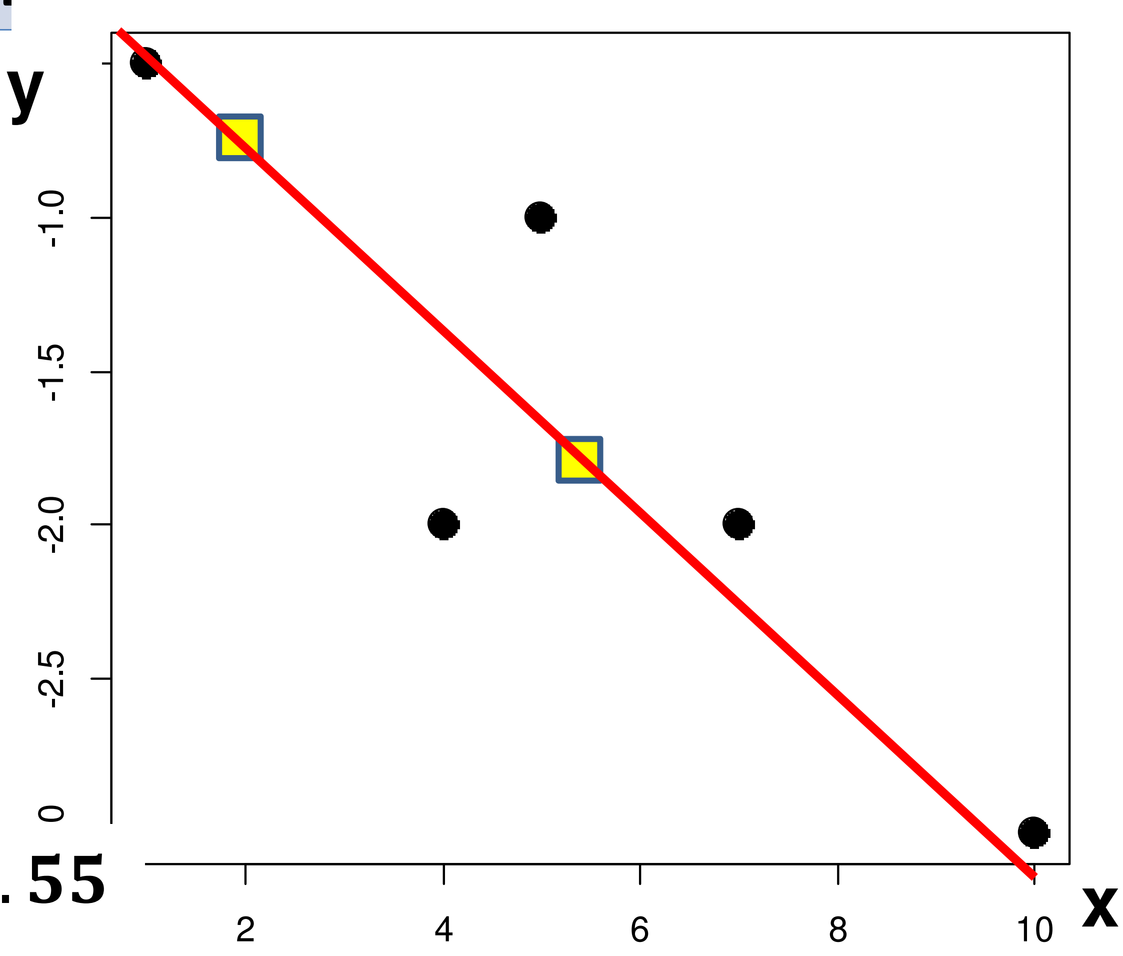
$$\bar{x} = 5.4$$

$$\bar{y} = -1.7$$

$$\boxed{\hat{a} = -0.3} \qquad \boxed{\hat{b} = -0.26}$$

$$x = 2$$

$$y = -0.3 - 0.26 \times 2 = -0.55$$



# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |

1. Rappresentare l'andamento congiunto di *Y* in funzione di *X* mediante un opportuno grafico: le due variabili sono correlate? Come e quanto?
2. Determinare le stime dei minimi quadrati dei parametri della retta interpolatrice dei dati
3. Determinare la significatività del modello di regressione lineare basato sulla retta al punto 2.
4. Valutare con un opportuno indice la bontà di adattamento del modello ai dati
5. Indicare entro quale intervallo sia possibile fare previsione sui valori di *Y*

# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |

1. Rappresentare l'andamento congiunto di *Y* in funzione di *X* mediante un opportuno grafico: le due variabili sono correlate? Come e quanto?
2. Determinare le stime dei minimi quadrati dei parametri della retta interpolatrice dei dati
3. Valutare con un opportuno indice la bontà di adattamento del modello ai dati

$$\boldsymbol{R^2 = \rho^2 = 0.885^2 = 0.78 > 0.70}$$

# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |

1. Rappresentare l'andamento congiunto di *Y* in funzione di *X* mediante un opportuno grafico: le due variabili sono correlate? Come e quanto?
2. Determinare le stime dei minimi quadrati dei parametri della retta interpolatrice dei dati
5. Indicare entro quale intervallo sia possibile fare previsione sui valori di *Y*

$$\boldsymbol{1 \leq x \leq 10}$$

# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |

1. Rappresentare l'andamento congiunto di *Y* in funzione di *X* mediante un opportuno grafico: le due variabili sono correlate? Come e quanto?
2. Determinare le stime dei minimi quadrati dei parametri della retta interpolatrice dei dati
3. Determinare la significatività del modello di regressione lineare basato sulla retta al punto 2.

$$\boldsymbol{H_0 : b = 0 \quad vs \quad H_1 \; : \; b \neq 0}$$

# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, $X$ ed $Y$:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |
| $\hat{y}_i$ | -0.56 | -1.34 | -1.6 | -2.12 | -2.9 |
| $e_i$ | 0.06 | -0.66 | 0.6 | 0.12 | -0.1 |

$$H_0 : b = 0 \quad vs \quad H_1 : \quad b \neq 0$$

$$\hat{a} = -0.3, \quad \hat{b} = -0.26 \Rightarrow \hat{y}_i = -0.3 - 0.26x_i$$

$$\sigma_x{}^2 = 9.04$$

# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |
| $\hat{y}_i$ | -0.56 | -1.34 | -1.6 | -2.12 | -2.9 |
| $e_i$ | 0.06 | -0.66 | 0.6 | 0.12 | -0.1 |

$$H_0 : b = 0 \quad vs \quad H_1 : b \neq 0$$

$$\hat{a} = -0.3, \quad \hat{b} = -0.26 \Rightarrow \hat{y}_i = -0.3 - 0.26x_i$$

$$s^2 = \frac{1}{3}\sum_{i=1}^{5} {e_i}^2 = 0.27$$

$${\sigma_x}^2 = 9.04$$

# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |
| $\hat{y}_i$ | -0.56 | -1.34 | -1.6 | -2.12 | -2.9 |
| $e_i$ | 0.06 | -0.66 | 0.6 | 0.12 | -0.1 |

$$\boldsymbol{H_0 : b = 0 \quad vs \quad H_1 \; : \; b \neq 0}$$

$$\hat{a} = -0.3, \qquad \hat{b} = -0.26 \Rightarrow \hat{y}_i = -0.3 - 0.26x_i$$

$$s^2 = \frac{1}{3}\sum_{i=1}^{5} e_i{}^2 = 0.27$$

$$\sigma_x{}^2 = 9.04$$

$$\boldsymbol{\frac{|\hat{b}|}{\sqrt{\frac{s^2}{n\sigma_x{}^2}}} = \frac{0.26}{\sqrt{\frac{0.27}{5 \times 9.04}}} = 3.36}$$

# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |
| $\hat{y}_i$ | -0.56 | -1.34 | -1.6 | -2.12 | -2.9 |
| $e_i$ | 0.06 | -0.66 | 0.6 | 0.12 | -0.1 |

$$\boldsymbol{H_0 : b = 0 \quad vs \quad H_1 \ : \ b \neq 0}$$

$$\hat{a} = -0.3, \quad \hat{b} = -0.26 \Rightarrow \hat{y}_i = -0.3 - 0.26x_i$$

$$s^2 = \frac{1}{3}\sum_{i=1}^{5} {e_i}^2 = 0.27$$

$$\boldsymbol{\frac{|\hat{b}|}{\sqrt{\frac{s^2}{n{\sigma_x}^2}}} = \frac{0.26}{\sqrt{\frac{0.27}{5 \times 9.04}}} = 3.36}$$

$$\boldsymbol{t(3)_{1-\frac{\alpha}{2}}}$$

$${\sigma_x}^2 = 9.04$$

# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| $y_i$ | -0.5 | -2 | -1 | -2 | -3 |
| $\hat{y}_i$ | -0.56 | -1.34 | -1.6 | -2.12 | -2.9 |
| $e_i$ | 0.06 | -0.66 | 0.6 | 0.12 | -0.1 |

$$H_0 : b = 0 \quad vs \quad H_1 : \quad b \neq 0$$

$$\hat{a} = -0.3, \quad \hat{b} = -0.26 \Rightarrow \hat{y}_i = -0.3 - 0.26x_i$$

$$s^2 = \frac{1}{3}\sum_{i=1}^{5} e_i{}^2 = 0.27$$

$$\sigma_x{}^2 = 9.04$$

$$\frac{|\hat{b}|}{\sqrt{\dfrac{s^2}{n\sigma_x{}^2}}} = \frac{0.26}{\sqrt{\dfrac{0.27}{5 \times 9.04}}} = 3.36$$

$$t(3)_{1-\frac{0.05}{2}} = 3.18$$

# Esercizio 3

Su un campione di $n = 5$ unità sono state osservate due variabili, *X* ed *Y*:

| $x_i$ | 1 | 4 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| | | | | -2 | -3 |
| | | | | 12 | -2.9 |
| | | | | 12 | -0.1 |

**LA REGRESSIONE E' SIGNIFICATIVA AL LIVELLO DEL 5%**

**IL p-valore E' < 0.025 x 2 = 0.05**

$\neq 0$

$\hat{a}$ … $6x_i$

$s^2 = \frac{1}{3}\sum_{i=1} e_i \;=\; 0.27$

$\sigma_x{}^2 = 9.04$

$$\sqrt{\frac{s^2}{n\sigma_x{}^2}} \qquad \sqrt{\frac{0.27}{5 \times 9.04}} = 3.36$$

$$t(3)_{1-\frac{0.05}{2}} = 3.18$$