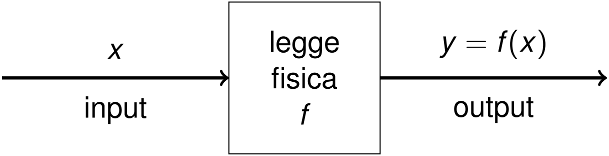
Regressione lineare

Alessandro Toigo

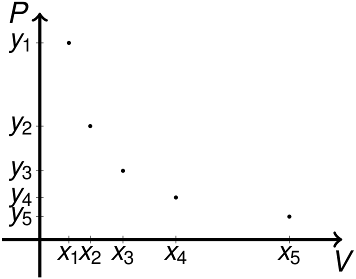
Programma

- Statistica descrittiva
  (riassumere molti dati attraverso poche caratteristiche essenziali)
- Probabilità
  (costruire un modello che preveda il risultato di un esperimento)
- Inferenza statistica
  (tarare i parametri del modello in base ai risultati dell'esperimento)
- Regressione lineare
  (riconoscere relazioni tra dati di tipo diverso)

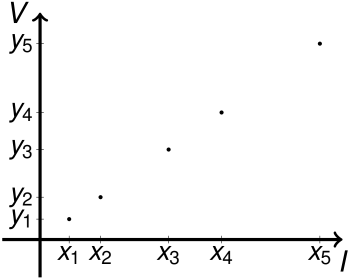
Modelli empirici



Idealmente:



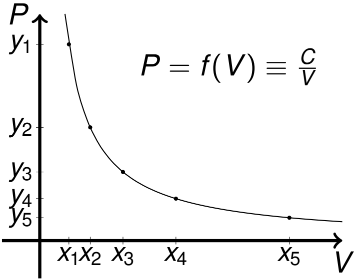
Gas perfetto



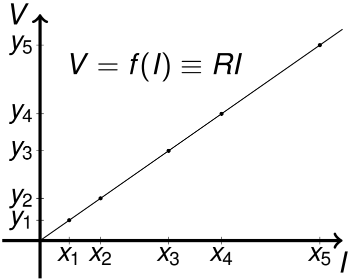
Corrente-tensione

Modelli empirici

Idealmente:

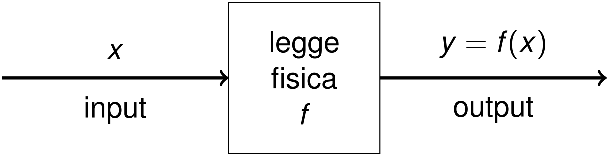


Gas perfetto

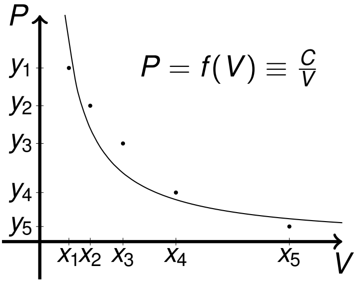


Corrente-tensione

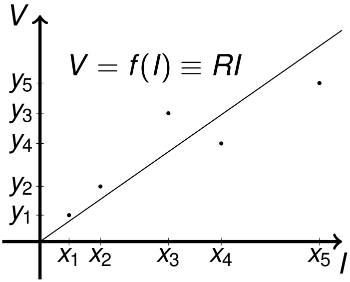
Modelli empirici



Realisticamente:



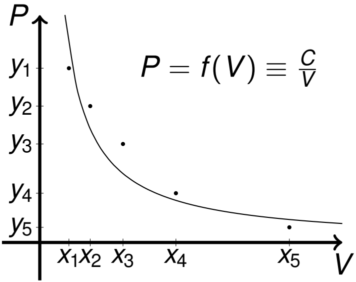
Gas perfetto



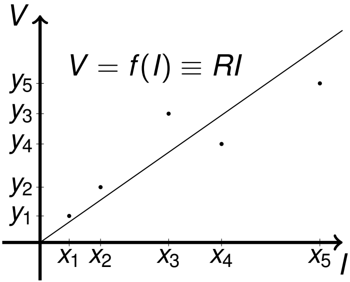
Corrente-tensione

Modelli empirici

Vogliamo trovare f!



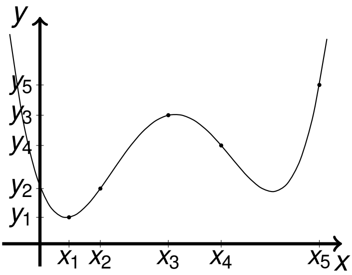
Gas perfetto



Corrente-tensione

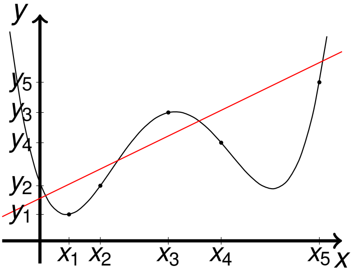
Modello lineare

Per n punti passa sempre un polinomio di grado $n - 1$:



Modello lineare

Per n punti passa sempre un polinomio di grado $n-1$:



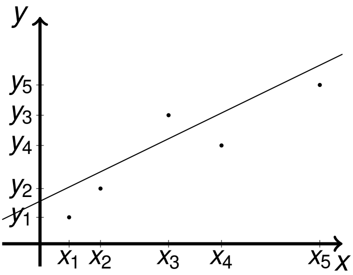
Ma noi vogliamo la funzione più semplice possibile: la retta

$$y = a + bx$$

parametrizzata solo da (a, b).

Modello lineare

Per n punti passa sempre un polinomio di grado $n - 1$:



Ma noi vogliamo la funzione più semplice possibile: la retta

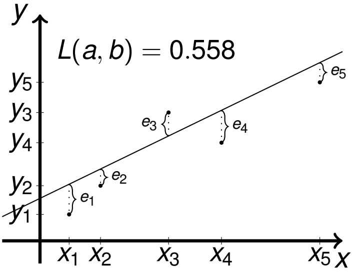
$$y = a + bx$$

parametrizzata solo da (a, b).

Come si capisce se la retta di parametri (a, b) approssima bene gli n punti?

Modello lineare

Per n punti passa sempre un polinomio di grado $n-1$:



Ma noi vogliamo la funzione più semplice possibile: la retta

$$y = a + bx$$

parametrizzata solo da (a, b).

Come si capisce se la retta di parametri (a, b) approssima bene gli n punti?

Dobbiamo introdurre gli errori (o *residui*)

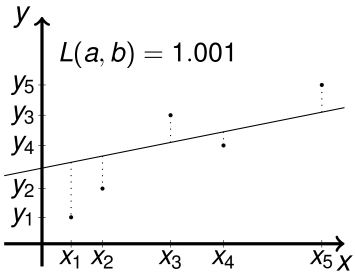
$$e_i := y_i - (a + bx_i)$$

e un funzionale di errore

$$L(a, b) := \sum_{i=1}^{n} (y_i - a - bx_i)^2 = \sum_{i=1}^{n} e_i^2$$

Modello lineare

Per n punti passa sempre un polinomio di grado $n-1$:



Ma noi vogliamo la funzione più semplice possibile: la retta

$$y = a + bx$$

parametrizzata solo da (a, b).

Come si capisce se la retta di parametri (a, b) approssima bene gli n punti?

Dobbiamo introdurre gli errori (o *residui*)

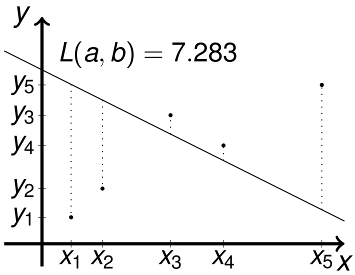
$$e_i := y_i - (a + bx_i)$$

e un funzionale di errore

$$L(a, b) := \sum_{i=1}^{n} (y_i - a - bx_i)^2 = \sum_{i=1}^{n} e_i^2$$

Modello lineare

Per n punti passa sempre un polinomio di grado $n-1$:



Ma noi vogliamo la funzione più semplice possibile: la retta

$$y = a + bx$$

parametrizzata solo da (a, b).

Come si capisce se la retta di parametri (a, b) approssima bene gli n punti?

Dobbiamo introdurre gli errori (o *residui*)

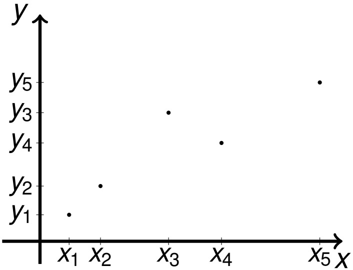
$$e_i := y_i - (a + bx_i)$$

e un funzionale di errore

$$L(a, b) := \sum_{i=1}^{n} (y_i - a - bx_i)^2 = \sum_{i=1}^{n} e_i^2$$

Modello lineare

Per n punti passa sempre un polinomio di grado $n-1$:



Ma noi vogliamo la funzione più semplice possibile: la retta

$$y = a + bx$$

parametrizzata solo da (a, b).

Come si capisce se la retta di parametri (a, b) approssima bene gli n punti?

Dobbiamo introdurre gli errori (o *residui*)

$$e_i := y_i - (a + bx_i)$$

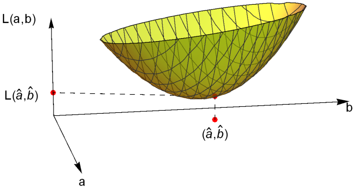
e un funzionale di errore

Perché si mette il quadrato?

$$L(a,b) := \sum_{i=1}^{n} (y_i - a - bx_i)^2 = \sum_{i=1}^{n} e_i^2$$

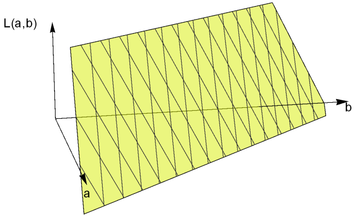
Modello lineare

Col quadrato:



- $L(a, b) \geq 0 \ \forall (a, b)$
- $L(\infty) = +\infty$
- $\exists!$ punto di minimo $(\hat{a}, \hat{b})$

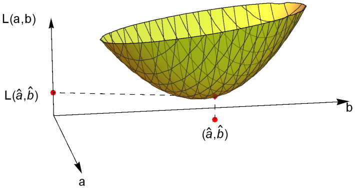
Senza quadrato:



- i residui si compensano
- $\nexists$ minimo

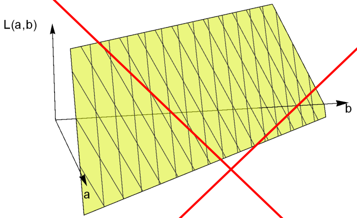
Modello lineare

Col quadrato:



- $L(a, b) \geq 0 \ \forall (a, b)$
- $L(\infty) = +\infty$
- $\exists!$ punto di minimo $(\hat{a}, \hat{b})$

Senza quadrato:



- i residui si compensano
- $\nexists$ minimo

$y = \hat{a} + \hat{b}x$ è la *retta dei minimi quadrati* (LSL = *least square line*)

Retta dei minimi quadrati

Cerchiamo $(\hat{a}, \hat{b})$ ponendo $\nabla L(\hat{a}, \hat{b}) \equiv 0$:

$$\frac{\partial L}{\partial a} = \frac{\partial}{\partial a} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

Retta dei minimi quadrati

Cerchiamo $(\hat{a}, \hat{b})$ ponendo $\nabla L(\hat{a}, \hat{b}) \equiv 0$:

$$\begin{aligned} \frac{\partial L}{\partial a} &= \frac{\partial}{\partial a} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \\ &= -2 \sum_{i=1}^{n} (y_i - a - bx_i) \end{aligned}$$

Retta dei minimi quadrati

Cerchiamo $(\hat{a}, \hat{b})$ ponendo $\nabla L(\hat{a}, \hat{b}) \equiv 0$:

$$\begin{aligned} \frac{\partial L}{\partial a} &= \frac{\partial}{\partial a} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \\ &= -2 \sum_{i=1}^{n} (y_i - a - bx_i) \\ &= -2(n\bar{y} - na - nb\bar{x}) \end{aligned}$$

dove

$$\bar{x} := \frac{1}{n} \sum_{i=1}^{n} x_i \qquad \bar{y} := \frac{1}{n} \sum_{i=1}^{n} y_i .$$

Retta dei minimi quadrati

Cerchiamo $(\hat{a}, \hat{b})$ ponendo $\nabla L(\hat{a}, \hat{b}) \equiv 0$:

$$
\begin{aligned}
\frac{\partial L}{\partial a} &= \frac{\partial}{\partial a} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \\
&= -2 \sum_{i=1}^{n} (y_i - a - bx_i) \\
&= -2(n\bar{y} - na - nb\bar{x}) \\
&\equiv 0
\end{aligned}
$$

dove

$$
\bar{x} := \frac{1}{n} \sum_{i=1}^{n} x_i \qquad \bar{y} := \frac{1}{n} \sum_{i=1}^{n} y_i .
$$

Retta dei minimi quadrati

Cerchiamo $(\hat{a}, \hat{b})$ ponendo $\nabla L(\hat{a}, \hat{b}) \equiv 0$:

$$\begin{aligned}
\frac{\partial L}{\partial a} &= \frac{\partial}{\partial a} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \\
&= -2 \sum_{i=1}^{n} (y_i - a - bx_i) \\
&= -2(n\bar{y} - na - nb\bar{x}) \\
&\equiv 0 \\
& \qquad \Rightarrow \quad a = \bar{y} - b\bar{x}
\end{aligned}$$

dove

$$\bar{x} := \frac{1}{n} \sum_{i=1}^{n} x_i \qquad \bar{y} := \frac{1}{n} \sum_{i=1}^{n} y_i.$$

Retta dei minimi quadrati

Cerchiamo $(\hat{a}, \hat{b})$ ponendo $\nabla L(\hat{a}, \hat{b}) \equiv 0$:

$$\frac{\partial L}{\partial b} = \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

Retta dei minimi quadrati

Cerchiamo $(\hat{a}, \hat{b})$ ponendo $\nabla L(\hat{a}, \hat{b}) \equiv 0$:

$$\begin{aligned}\frac{\partial L}{\partial b} &= \frac{\partial}{\partial b}\sum_{i=1}^{n}(y_i - a - bx_i)^2 \\ &= -2\sum_{i=1}^{n} x_i(y_i - a - bx_i)\end{aligned}$$

Retta dei minimi quadrati

Cerchiamo $(\hat{a}, \hat{b})$ ponendo $\nabla L(\hat{a}, \hat{b}) \equiv 0$:

$$
\begin{aligned}
\frac{\partial L}{\partial b} &= \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \\
&= -2 \sum_{i=1}^{n} x_i (y_i - a - bx_i) \\
&= -2 \left(\sum_{i=1}^{n} x_i y_i - na\bar{x} - b \sum_{i=1}^{n} x_i^2 \right)
\end{aligned}
$$

Retta dei minimi quadrati

Cerchiamo $(\hat{a}, \hat{b})$ ponendo $\nabla L(\hat{a}, \hat{b}) \equiv 0$:

$$\begin{aligned}
\frac{\partial L}{\partial b} &= \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \\
&= -2 \sum_{i=1}^{n} x_i (y_i - a - bx_i) \\
&= -2 \left(\sum_{i=1}^{n} x_i y_i - na\bar{x} - b \sum_{i=1}^{n} x_i^2 \right) \\
&= -2 \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}(\bar{y} - b\bar{x}) - b \sum_{i=1}^{n} x_i^2 \right) \qquad \text{con } a = \bar{y} - b\bar{x}
\end{aligned}$$

Retta dei minimi quadrati

Cerchiamo $(\hat{a}, \hat{b})$ ponendo $\nabla L(\hat{a}, \hat{b}) \equiv 0$:

$$
\begin{aligned}
\frac{\partial L}{\partial b} &= \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \\
&= -2 \sum_{i=1}^{n} x_i (y_i - a - bx_i) \\
&= -2 \left(\sum_{i=1}^{n} x_i y_i - na\bar{x} - b \sum_{i=1}^{n} x_i^2 \right) \\
&= -2 \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}(\bar{y} - b\bar{x}) - b \sum_{i=1}^{n} x_i^2 \right) \qquad \text{con } a = \bar{y} - b\bar{x} \\
&\equiv 0
\end{aligned}
$$

Retta dei minimi quadrati

Cerchiamo $(\hat{a}, \hat{b})$ ponendo $\nabla L(\hat{a}, \hat{b}) \equiv 0$:

$$\begin{aligned}
\frac{\partial L}{\partial b} &= \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \\
&= -2 \sum_{i=1}^{n} x_i (y_i - a - bx_i) \\
&= -2 \left(\sum_{i=1}^{n} x_i y_i - na\bar{x} - b \sum_{i=1}^{n} x_i^2 \right) \\
&= -2 \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}(\bar{y} - b\bar{x}) - b \sum_{i=1}^{n} x_i^2 \right) \qquad \text{con } a = \bar{y} - b\bar{x} \\
&\equiv 0 \\
&\qquad \Rightarrow \quad b = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_j^2 - n\bar{x}^2}
\end{aligned}$$

Retta dei minimi quadrati

Cerchiamo $(\hat{a}, \hat{b})$ ponendo $\nabla L(\hat{a}, \hat{b}) \equiv 0$:

$$
\begin{aligned}
\frac{\partial L}{\partial b} &= \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \\
&= -2 \sum_{i=1}^{n} x_i (y_i - a - bx_i) \\
&= -2 \left(\sum_{i=1}^{n} x_i y_i - na\bar{x} - b \sum_{i=1}^{n} x_i^2 \right) \\
&= -2 \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}(\bar{y} - b\bar{x}) - b \sum_{i=1}^{n} x_i^2 \right) \qquad \text{con } a = \bar{y} - b\bar{x} \\
&\equiv 0
\end{aligned}
$$

$$
\Rightarrow \quad b = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_j^2 - n\bar{x}^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_j - \bar{x})^2}
$$

Retta dei minimi quadrati

Dai punti $(x_1, y_1), \ldots, (x_n, y_n)$ distilliamo le 5 quantità

$$\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} := \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$s_{xx} := \sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad s_{xy} := \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \qquad s_{yy} := \sum_{i=1}^{n}(y_i - \bar{y})^2$$

Retta dei minimi quadrati

Dai punti $(x_1, y_1), \ldots, (x_n, y_n)$ distilliamo le 5 quantità

$$\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} := \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$s_{xx} := \sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad s_{xy} := \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \qquad s_{yy} := \sum_{i=1}^{n}(y_i - \bar{y})^2$$

Tutte le altre quantità della LSL sono funzioni di queste. P. es.:

$$\hat{a} = \bar{y} - \frac{s_{xy}}{s_{xx}}\,\bar{x} \qquad \hat{b} = \frac{s_{xy}}{s_{xx}}$$ parametri della LSL

Retta dei minimi quadrati

Dai punti $(x_1, y_1), \ldots, (x_n, y_n)$ distilliamo le 5 quantità

$$\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} := \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$s_{xx} := \sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad s_{xy} := \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \qquad s_{yy} := \sum_{i=1}^{n}(y_i - \bar{y})^2$$

Tutte le altre quantità della LSL sono funzioni di queste. P. es.:

$$\hat{a} = \bar{y} - \frac{s_{xy}}{s_{xx}}\,\bar{x} \qquad \hat{b} = \frac{s_{xy}}{s_{xx}} \qquad \text{parametri della LSL}$$

$$\hat{y}_i := \hat{a} + \hat{b}x_i = \bar{y} + \frac{s_{xy}}{s_{xx}}(x_i - \bar{x}) \qquad \text{output della LSL}$$

Retta dei minimi quadrati

Dai punti $(x_1, y_1), \ldots, (x_n, y_n)$ distilliamo le 5 quantità

$$\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} := \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$s_{xx} := \sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad s_{xy} := \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \qquad s_{yy} := \sum_{i=1}^{n}(y_i - \bar{y})^2$$

Tutte le altre quantità della LSL sono funzioni di queste. P. es.:

$$\hat{a} = \bar{y} - \frac{s_{xy}}{s_{xx}}\,\bar{x} \qquad \hat{b} = \frac{s_{xy}}{s_{xx}}$$ parametri della LSL

$$\hat{y}_i := \hat{a} + \hat{b}x_i = \bar{y} + \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})$$ output della LSL

$$\hat{e}_i := y_i - \hat{y}_i = (y_i - \bar{y}) - \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})$$ residui della LSL

Varianza totale

$$ss_t := s_{yy} = \sum_i (y_i - \bar{y})^2 \qquad \textit{total sum of squares}$$

Varianza totale

$$ss_t := s_{yy} = \sum_i (y_i - \bar{y})^2 \qquad \textit{total sum of squares}$$

- $ss_t \geq 0$

Varianza totale

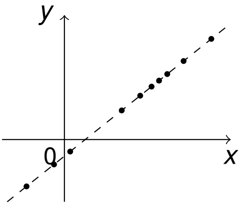
$$ss_t := s_{yy} = \sum_i (y_i - \bar{y})^2 \qquad \textit{total sum of squares}$$

- $ss_t \geq 0$
- ss_t non dipende dall'aver scelto il modello lineare

Varianza totale

$$ss_t := s_{yy} = \sum_i (y_i - \bar{y})^2 \qquad \textit{total sum of squares}$$

- $ss_t \geq 0$
- ss_t non dipende dall'aver scelto il modello lineare
- $ss_t \neq 0$ anche se tutti i punti stanno esattamente su una retta:

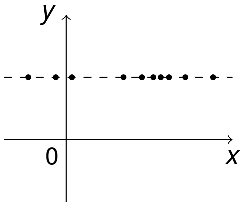


$ss_t = 11.98$

Varianza totale

$$ss_t := s_{yy} = \sum_i (y_i - \bar{y})^2 \qquad \textit{total sum of squares}$$

- $ss_t \geq 0$
- ss_t non dipende dall'aver scelto il modello lineare
- $ss_t \neq 0$ anche se tutti i punti stanno esattamente su una retta:



$ss_t = 0$

Varianza totale

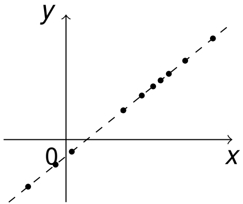
$$ss_t := s_{yy} = \sum_i (y_i - \bar{y})^2 \qquad \textit{total sum of squares}$$

- $ss_t \geq 0$
- ss_t non dipende dall'aver scelto il modello lineare
- $ss_t \neq 0$ anche se tutti i punti stanno esattamente su una retta
- $ss_t = ss_r + ss_e$, dove:
  - ss_r è intrinseca al modello lineare (varianza *spiegata*)
  - ss_e dipende dalla dispersione intorno alla LSL (varianza *residua*)

Varianza totale

$$ss_t := s_{yy} = \sum_i (y_i - \bar{y})^2 \qquad \textit{total sum of squares}$$

- $ss_t \geq 0$
- ss_t non dipende dall'aver scelto il modello lineare
- $ss_t \neq 0$ anche se tutti i punti stanno esattamente su una retta
- $ss_t = ss_r + ss_e$, dove:
  - ss_r è intrinseca al modello lineare (varianza *spiegata*)
  - ss_e dipende dalla dispersione intorno alla LSL (varianza *residua*)



$ss_r = 11.98$
$ss_e = 0$

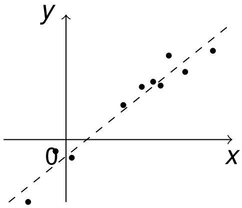
Varianza totale

$$ss_t := s_{yy} = \sum_i (y_i - \bar{y})^2 \qquad \textit{total sum of squares}$$

- $ss_t \geq 0$
- ss_t non dipende dall'aver scelto il modello lineare
- $ss_t \neq 0$ anche se tutti i punti stanno esattamente su una retta
- $ss_t = ss_r + ss_e$, dove:

  - ss_r è intrinseca al modello lineare (varianza *spiegata*)

  - ss_e dipende dalla dispersione intorno alla LSL (varianza *residua*)



$ss_r = 11.98$

$ss_e = 0.70$

Varianza residua

$$ss_e := \min_{(a,b)} L(a,b)$$

Varianza residua

$$ss_e \;:=\; \min_{(a,b)} L(a,b) \;=\; L(\hat{a},\hat{b})$$

Varianza residua

$$ss_e \;:=\; \min_{(a,b)} L(a,b) \;=\; L(\hat{a},\hat{b}) \;=\; \sum_i (y_i - \hat{a} - \hat{b}x_i)^2$$

Varianza residua

$$
\begin{aligned}
ss_e &:= \min_{(a,b)} L(a,b) = L(\hat{a},\hat{b}) = \sum_i (y_i - \hat{a} - \hat{b}x_i)^2 \\
&= \sum_i (y_i - \hat{y}_i)^2
\end{aligned}
$$

Varianza residua

$$
\begin{aligned}
ss_e \;&:=\; \min_{(a,b)} L(a,b) \;=\; L(\hat{a},\hat{b}) \;=\; \sum_i (y_i - \hat{a} - \hat{b}x_i)^2 \\
&=\; \sum_i (y_i - \hat{y}_i)^2 \;=\; \sum_i \hat{e}_i^2
\end{aligned}
$$

Varianza residua

$$
\begin{aligned}
ss_e &:= \min_{(a,b)} L(a,b) = L(\hat{a},\hat{b}) = \sum_i (y_i - \hat{a} - \hat{b}x_i)^2 \\
&= \sum_i (y_i - \hat{y}_i)^2 = \sum_i \hat{e}_i^2 \qquad \textit{error sum of squares}
\end{aligned}
$$

Varianza residua

$$
\begin{aligned}
ss_e &:= \min_{(a,b)} L(a,b) = L(\hat{a},\hat{b}) = \sum_i (y_i - \hat{a} - \hat{b}x_i)^2 \\
&= \sum_i (y_i - \hat{y}_i)^2 = \sum_i \hat{e}_i^2 \qquad \textit{error sum of squares}
\end{aligned}
$$

- $ss_e \geq 0$

Varianza residua

$$
\begin{aligned}
ss_e \;&:=\; \min_{(a,b)} L(a,b) \;=\; L(\hat{a},\hat{b}) \;=\; \sum_i (y_i - \hat{a} - \hat{b}x_i)^2 \\
&=\; \sum_i (y_i - \hat{y}_i)^2 \;=\; \sum_i \hat{e}_i^2 \qquad \textit{error sum of squares}
\end{aligned}
$$

- $ss_e \geq 0$
- $ss_e = 0$ se e solo se tutti i punti stanno esattamente su una retta

Varianza residua

$$
\begin{aligned}
ss_e \;&:=\; \min_{(a,b)} L(a,b) \;=\; L(\hat{a},\hat{b}) \;=\; \sum_i (y_i - \hat{a} - \hat{b}x_i)^2 \\
&=\; \sum_i (y_i - \hat{y}_i)^2 \;=\; \sum_i \hat{e}_i^2 \qquad \textit{error sum of squares}
\end{aligned}
$$

- $ss_e \geq 0$
- $ss_e = 0$ se e solo se tutti i punti stanno esattamente su una retta:

$$
\begin{aligned}
\Rightarrow: \quad ss_e = 0 \quad &\Rightarrow \quad (y_i - \hat{a} - \hat{b}x_i)^2 = 0 \text{ per ogni } i \\
&\Rightarrow \quad y_i = \hat{a} + \hat{b}x_i \text{ per ogni } i
\end{aligned}
$$

Varianza residua

$$ss_e := \min_{(a,b)} L(a,b) = L(\hat{a},\hat{b}) = \sum_i (y_i - \hat{a} - \hat{b}x_i)^2$$

$$= \sum_i (y_i - \hat{y}_i)^2 = \sum_i \hat{e}_i^2 \qquad \textit{error sum of squares}$$

- $ss_e \geq 0$
- $ss_e = 0$ se e solo se tutti i punti stanno esattamente su una retta:

$\Rightarrow$: $\quad ss_e = 0 \quad \Rightarrow \quad (y_i - \hat{a} - \hat{b}x_i)^2 = 0$ per ogni i

$\Rightarrow \quad y_i = \hat{a} + \hat{b}x_i$ per ogni i

$\Leftarrow$: $\quad y_i = a + bx_i$ per ogni $i \quad \Rightarrow \quad L(a,b) = 0$

$\Rightarrow \quad ss_e = \min_{(a,b)} L(a,b) = 0$

Varianza residua

$$
\begin{aligned}
ss_e &:= \min_{(a,b)} L(a,b) = L(\hat{a},\hat{b}) = \sum_i (y_i - \hat{a} - \hat{b}x_i)^2 \\
&= \sum_i (y_i - \hat{y}_i)^2 = \sum_i \hat{e}_i^2 \qquad \textit{error sum of squares}
\end{aligned}
$$

- $ss_e \geq 0$
- $ss_e = 0$ se e solo se tutti i punti stanno esattamente su una retta:

$$
\begin{aligned}
\Rightarrow: \quad ss_e = 0 \quad &\Rightarrow \quad (y_i - \hat{a} - \hat{b}x_i)^2 = 0 \text{ per ogni } i \\
&\Rightarrow \quad y_i = \hat{a} + \hat{b}x_i \text{ per ogni } i
\end{aligned}
$$

$$
\begin{aligned}
\Leftarrow: \quad y_i = a + bx_i \text{ per ogni } i \quad &\Rightarrow \quad L(a,b) = 0 \\
&\Rightarrow \quad ss_e = \min_{(a,b)} L(a,b) = 0
\end{aligned}
$$

- Se $ss_e = 0$, tutti i punti stanno sulla LSL.

Varianza spiegata

$$ss_r := \sum_i (\bar{y} - \hat{y}_i)^2$$

Varianza spiegata

$$ss_r := \sum_i (\bar{y} - \hat{y}_i)^2 = \sum_i \left[\bar{y} - \left(\bar{y} + \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})\right)\right]^2$$

Varianza spiegata

$$\begin{aligned} ss_r \;&:=\; \sum_i (\bar{y} - \hat{y}_i)^2 \;=\; \sum_i \left[\bar{y} - \left(\bar{y} + \frac{s_{xy}}{s_{xx}}\,(x_i - \bar{x})\right)\right]^2 \\ &=\; \left(\frac{s_{xy}}{s_{xx}}\right)^2 \sum_i (x_i - \bar{x})^2 \end{aligned}$$

Varianza spiegata

$$
\begin{aligned}
ss_r \;&:=\; \sum_i (\bar{y} - \hat{y}_i)^2 \;=\; \sum_i \left[\bar{y} - \left(\bar{y} + \frac{s_{xy}}{s_{xx}}\,(x_i - \bar{x})\right)\right]^2 \\
&=\; \left(\frac{s_{xy}}{s_{xx}}\right)^2 \sum_i (x_i - \bar{x})^2 \;=\; \frac{s_{xy}^2}{s_{xx}}
\end{aligned}
$$

Varianza spiegata

$$
\begin{aligned}
ss_r &:= \sum_i (\bar{y} - \hat{y}_i)^2 = \sum_i \left[\bar{y} - \left(\bar{y} + \frac{s_{xy}}{s_{xx}} (x_i - \bar{x}) \right) \right]^2 \\
&= \left(\frac{s_{xy}}{s_{xx}} \right)^2 \sum_i (x_i - \bar{x})^2 = \frac{s_{xy}^2}{s_{xx}} \qquad \textit{regression sum of squares}
\end{aligned}
$$

Varianza spiegata

$$
\begin{aligned}
ss_r \;&:=\; \sum_i (\bar{y} - \hat{y}_i)^2 \;=\; \sum_i \left[\bar{y} - \left(\bar{y} + \frac{s_{xy}}{s_{xx}} (x_i - \bar{x}) \right) \right]^2 \\
&=\; \left(\frac{s_{xy}}{s_{xx}} \right)^2 \sum_i (x_i - \bar{x})^2 \;=\; \frac{s_{xy}^2}{s_{xx}} \qquad \textit{regression sum of squares}
\end{aligned}
$$

- $ss_r \geq 0$

Varianza spiegata

$$
\begin{aligned}
ss_r \;:=\; \sum_i (\bar{y} - \hat{y}_i)^2 \;&=\; \sum_i \left[\bar{y} - \left(\bar{y} + \frac{s_{xy}}{s_{xx}} (x_i - \bar{x}) \right) \right]^2 \\
&= \left(\frac{s_{xy}}{s_{xx}} \right)^2 \sum_i (x_i - \bar{x})^2 \;=\; \frac{s_{xy}^2}{s_{xx}} \qquad \textit{regression sum of squares}
\end{aligned}
$$

- $ss_r \geq 0$
- $ss_t = ss_r + ss_e$

Varianza spiegata

$$ss_r \; := \; \sum_i (\bar{y} - \hat{y}_i)^2 \; = \; \sum_i \left[\bar{y} - \left(\bar{y} + \frac{s_{xy}}{s_{xx}} (x_i - \bar{x}) \right) \right]^2$$

$$= \left(\frac{s_{xy}}{s_{xx}} \right)^2 \sum_i (x_i - \bar{x})^2 \; = \; \frac{s_{xy}^2}{s_{xx}} \qquad \textit{regression sum of squares}$$

- $ss_r \geq 0$
- $ss_t = ss_r + ss_e$, perché

$$ss_e \; := \; \sum (y_i - \hat{y}_i)^2$$

Varianza spiegata

$$
\begin{aligned}
ss_r \;&:=\; \sum_i (\bar{y} - \hat{y}_i)^2 \;=\; \sum_i \left[\bar{y} - \left(\bar{y} + \frac{s_{xy}}{s_{xx}}\,(x_i - \bar{x})\right)\right]^2 \\
&= \left(\frac{s_{xy}}{s_{xx}}\right)^2 \sum_i (x_i - \bar{x})^2 \;=\; \frac{s_{xy}^2}{s_{xx}} \qquad \textit{regression sum of squares}
\end{aligned}
$$

- $ss_r \geq 0$
- $ss_t = ss_r + ss_e$, perché

$$
ss_e \;:=\; \sum (y_i - \hat{y}_i)^2 \;=\; \sum \left[y_i - \left(\bar{y} + \frac{s_{xy}}{s_{xx}}\,(x_i - \bar{x})\right)\right]^2
$$

Varianza spiegata

$$
\begin{aligned}
ss_r &:= \sum_i (\bar{y} - \hat{y}_i)^2 = \sum_i \left[\bar{y} - \left(\bar{y} + \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})\right)\right]^2 \\
&= \left(\frac{s_{xy}}{s_{xx}}\right)^2 \sum_i (x_i - \bar{x})^2 = \frac{s_{xy}^2}{s_{xx}} \qquad \textit{regression sum of squares}
\end{aligned}
$$

- $ss_r \geq 0$
- $ss_t = ss_r + ss_e$, perché

$$
\begin{aligned}
ss_e &:= \sum (y_i - \hat{y}_i)^2 = \sum \left[y_i - \left(\bar{y} + \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})\right)\right]^2 \\
&= \sum \left[(y_i - \bar{y}) - \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})\right]^2
\end{aligned}
$$

Varianza spiegata

$$ss_r \;:=\; \sum_i (\bar{y} - \hat{y}_i)^2 \;=\; \sum_i \left[\bar{y} - \left(\bar{y} + \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})\right)\right]^2$$
$$= \left(\frac{s_{xy}}{s_{xx}}\right)^2 \sum_i (x_i - \bar{x})^2 \;=\; \frac{s_{xy}^2}{s_{xx}} \qquad \textit{regression sum of squares}$$

- $ss_r \geq 0$
- $ss_t = ss_r + ss_e$, perché

$$ss_e \;:=\; \sum (y_i - \hat{y}_i)^2 \;=\; \sum \left[y_i - \left(\bar{y} + \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})\right)\right]^2$$
$$= \sum \left[(y_i - \bar{y}) - \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})\right]^2 \;=\; s_{yy} - 2\frac{s_{xy}}{s_{xx}} s_{xy} + \left(\frac{s_{xy}}{s_{xx}}\right)^2 s_{xx}$$

Varianza spiegata

$$ss_r := \sum_i (\bar{y} - \hat{y}_i)^2 = \sum_i \left[\bar{y} - \left(\bar{y} + \frac{s_{xy}}{s_{xx}} (x_i - \bar{x}) \right) \right]^2$$

$$= \left(\frac{s_{xy}}{s_{xx}} \right)^2 \sum_i (x_i - \bar{x})^2 = \frac{s_{xy}^2}{s_{xx}} \qquad \textit{regression sum of squares}$$

- $ss_r \geq 0$
- $ss_t = ss_r + ss_e$, perché

$$ss_e := \sum (y_i - \hat{y}_i)^2 = \sum \left[y_i - \left(\bar{y} + \frac{s_{xy}}{s_{xx}} (x_i - \bar{x}) \right) \right]^2$$

$$= \sum \left[(y_i - \bar{y}) - \frac{s_{xy}}{s_{xx}} (x_i - \bar{x}) \right]^2 = s_{yy} - 2 \frac{s_{xy}}{s_{xx}} s_{xy} + \left(\frac{s_{xy}}{s_{xx}} \right)^2 s_{xx}$$

$$= s_{yy} - \frac{s_{xy}^2}{s_{xx}}$$

Varianza spiegata

$$
\begin{aligned}
ss_r \;:=\; \sum_i (\bar{y} - \hat{y}_i)^2 \;&=\; \sum_i \left[\bar{y} - \left(\bar{y} + \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})\right)\right]^2 \\
&= \left(\frac{s_{xy}}{s_{xx}}\right)^2 \sum_i (x_i - \bar{x})^2 \;=\; \frac{s_{xy}^2}{s_{xx}} \qquad \textit{regression sum of squares}
\end{aligned}
$$

- $ss_r \geq 0$
- $ss_t = ss_r + ss_e$, perché

$$
\begin{aligned}
ss_e \;:=\; \sum (y_i - \hat{y}_i)^2 \;&=\; \sum \left[y_i - \left(\bar{y} + \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})\right)\right]^2 \\
&= \sum \left[(y_i - \bar{y}) - \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})\right]^2 \;=\; s_{yy} - 2\frac{s_{xy}}{s_{xx}} s_{xy} + \left(\frac{s_{xy}}{s_{xx}}\right)^2 s_{xx} \\
&= s_{yy} - \frac{s_{xy}^2}{s_{xx}} \;=\; ss_t - ss_r
\end{aligned}
$$

Varianza spiegata

$$ss_r := \sum_i (\bar{y} - \hat{y}_i)^2 = \sum_i \left[\bar{y} - \left(\bar{y} + \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})\right)\right]^2$$

$$= \left(\frac{s_{xy}}{s_{xx}}\right)^2 \sum_i (x_i - \bar{x})^2 = \frac{s_{xy}^2}{s_{xx}} \qquad \textit{regression sum of squares}$$

- $ss_r \geq 0$
- $ss_t = ss_r + ss_e$, perché

$$ss_e := \sum (y_i - \hat{y}_i)^2 = \sum \left[y_i - \left(\bar{y} + \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})\right)\right]^2$$

$$= \sum \left[(y_i - \bar{y}) - \frac{s_{xy}}{s_{xx}}(x_i - \bar{x})\right]^2 = s_{yy} - 2\frac{s_{xy}}{s_{xx}} s_{xy} + \left(\frac{s_{xy}}{s_{xx}}\right)^2 s_{xx}$$

$$= s_{yy} - \frac{s_{xy}^2}{s_{xx}} = ss_t - ss_r$$

Coefficiente di determinazione

$$r^2 := \frac{ss_r}{ss_t}$$

Coefficiente di determinazione

$$r^2 := \frac{SS_r}{SS_t} = \frac{SS_t - SS_e}{SS_t}$$

Coefficiente di determinazione

$$
\begin{aligned}
r^2 &:= \frac{SS_r}{SS_t} = \frac{SS_t - SS_e}{SS_t} \\
&= 1 - \frac{SS_e}{SS_t}
\end{aligned}
$$

Coefficiente di determinazione

$$
\begin{aligned}
r^2 &:= \frac{ss_r}{ss_t} = \frac{ss_t - ss_e}{ss_t} \\
&= 1 - \frac{ss_e}{ss_t} \qquad \textit{coefficiente di determinazione}
\end{aligned}
$$

Coefficiente di determinazione

$$
\begin{aligned}
r^2 &:= \frac{ss_r}{ss_t} = \frac{ss_t - ss_e}{ss_t} \\
&= 1 - \frac{ss_e}{ss_t} \qquad \textit{coefficiente di determinazione}
\end{aligned}
$$

- $0 \leq r^2 \leq 1$

Coefficiente di determinazione

$$\begin{aligned} r^2 &:= \frac{ss_r}{ss_t} = \frac{ss_t - ss_e}{ss_t} \\ &= 1 - \frac{ss_e}{ss_t} \qquad \textit{coefficiente di determinazione} \end{aligned}$$

- $0 \leq r^2 \leq 1$
- $r^2 \simeq 1 \quad \Rightarrow \quad$ i punti si dispongono bene su una retta

Coefficiente di determinazione

$$\begin{aligned} r^2 &:= \frac{ss_r}{ss_t} = \frac{ss_t - ss_e}{ss_t} \\ &= 1 - \frac{ss_e}{ss_t} \qquad \textit{coefficiente di determinazione} \end{aligned}$$

- $0 \leq r^2 \leq 1$
- $r^2 \simeq 1 \quad \Rightarrow \quad$ i punti si dispongono bene su una retta
- r^2 è la "percentuale di variabilità spiegata dal modello lineare"

Coefficiente di determinazione

$$r^2 := \frac{ss_r}{ss_t} = \frac{ss_t - ss_e}{ss_t}$$

$$= 1 - \frac{ss_e}{ss_t} \qquad \textit{coefficiente di determinazione}$$

- $0 \leq r^2 \leq 1$
- $r^2 \simeq 1 \quad \Rightarrow \quad$ i punti si dispongono bene su una retta
- r^2 è la "percentuale di variabilità spiegata dal modello lineare"
- $r^2 = \dfrac{s_{xy}^2 / s_{xx}}{s_{yy}}$

Coefficiente di determinazione

$$r^2 := \frac{ss_r}{ss_t} = \frac{ss_t - ss_e}{ss_t}$$

$$= 1 - \frac{ss_e}{ss_t} \qquad \textit{coefficiente di determinazione}$$

- $0 \leq r^2 \leq 1$
- $r^2 \simeq 1 \quad \Rightarrow \quad$ i punti si dispongono bene su una retta
- r^2 è la "percentuale di variabilità spiegata dal modello lineare"
- $r^2 = \dfrac{s_{xy}^2/s_{xx}}{s_{yy}} = \dfrac{s_{xy}^2}{s_{xx}s_{yy}}$

Coefficiente di correlazione

$$r := \frac{s_{xy}}{\sqrt{s_{xx} s_{yy}}}$$ *coefficiente di correlazione*

Coefficiente di correlazione

$$r := \frac{s_{xy}}{\sqrt{s_{xx} s_{yy}}} \qquad \textit{coefficiente di correlazione}$$

- $-1 \leq r \leq 1$

Coefficiente di correlazione

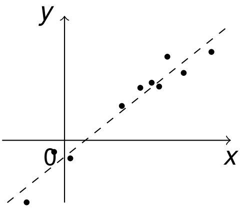
$$r := \frac{s_{xy}}{\sqrt{s_{xx} s_{yy}}} \qquad \textit{coefficiente di correlazione}$$

- $-1 \leq r \leq 1$
- r^2 è il coefficiente di determinazione

Coefficiente di correlazione

$$r := \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}} \qquad \textit{coefficiente di correlazione}$$

- $-1 \leq r \leq 1$
- r^2 è il coefficiente di determinazione
- $r > 0.9 \quad \Rightarrow$ i dati approssimano una retta di pendenza positiva

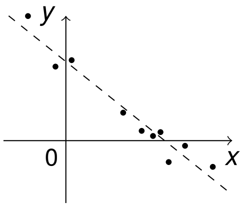


$$r = 0.972$$
$$r^2 = 0.945$$

Coefficiente di correlazione

$$r := \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}} \qquad \textit{coefficiente di correlazione}$$

- $-1 \leq r \leq 1$
- r^2 è il coefficiente di determinazione
- $r > 0.9 \quad \Rightarrow$ i dati approssimano una retta di pendenza positiva
- $r < -0.9 \quad \Rightarrow$ i dati approssimano una retta di pendenza negativa



$$r = -0.972$$
$$r^2 = 0.945$$

Il modello statistico

x

input deterministico
(numero reale)

Legge fisica
$y = a + bx$

$Y = a + bx + E$

output probabilistico
(v. a.)

Il modello statistico

x		$Y = a + bx + E$
$\longrightarrow$	Legge fisica $y = a + bx$	$\longrightarrow$
input deterministico (numero reale)		output probabilistico (v. a.)

Se facciamo n misure:

$$x_1, x_2, \ldots, x_n \quad \longrightarrow \quad Y_1, Y_2, \ldots, Y_n \quad \text{con} \quad Y_i = a + bx_i + E_i$$

Il modello statistico

$$x \xrightarrow{\text{input deterministico (numero reale)}} \boxed{\begin{array}{c}\text{Legge fisica}\\ y = a + bx\end{array}} \xrightarrow{\text{output probabilistico (v. a.)}} Y = a + bx + E$$

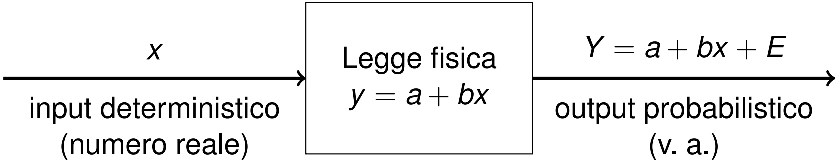
Se facciamo n misure:

$$x_1, x_2, \ldots, x_n \quad \longrightarrow \quad Y_1, Y_2, \ldots, Y_n \quad \text{con} \quad Y_i = a + bx_i + E_i$$

Ipotesi fondamentale del modello lineare:

- $E_1, E_2, \ldots, E_n$ i. i. d.
- $E_i \sim N(0, \sigma^2)$ con σ^2 incognita

Il modello statistico



Se facciamo n misure:

$$x_1, x_2, \ldots, x_n \quad \longrightarrow \quad Y_1, Y_2, \ldots, Y_n \quad \text{con} \quad Y_i = a + bx_i + E_i$$

Ipotesi fondamentale del modello lineare:

- $E_1, E_2, \ldots, E_n$ i. i. d.
- $E_i \sim N(0, \sigma^2)$ con σ^2 incognita

$$\Rightarrow \begin{cases} Y_i \sim N(a + bx_i\,,\, \sigma^2) \\ Y_1, \ldots, Y_n \text{ indipendenti} \end{cases}$$

Il modello statistico

$$\xrightarrow[\text{input deterministico (numero reale)}]{x} \boxed{\begin{array}{c}\text{Legge fisica}\\ y = a + bx\end{array}} \xrightarrow[\text{output probabilistico (v. a.)}]{Y = a + bx + E}$$

Se facciamo n misure:

$$x_1, x_2, \dots, x_n \quad \longrightarrow \quad Y_1, Y_2, \dots, Y_n \quad \text{con} \quad Y_i = a + bx_i + E_i$$

Ipotesi fondamentale del modello lineare:

- $E_1, E_2, \dots, E_n$ i. i. d.
- $E_i \sim N(0, \sigma^2)$ con σ^2 incognita

$$\Rightarrow \begin{cases} Y_i \sim N(a + bx_i\,,\, \sigma^2) \\ Y_1, \dots, Y_n \ \text{indipendenti} \end{cases}$$

$Y_1, \dots, Y_n$ non sono identicamente distribuite

Il modello statistico

Quali sono le varibili aleatorie nel modello lineare ?

Il modello statistico

Quali sono le varibili aleatorie nel modello lineare ?

Era:	Diventa:	Ora è una v.a.?
$\bar{x} = \frac{1}{n}\sum x_i$	$\bar{x} = \frac{1}{n}\sum x_i$	No

Il modello statistico

Quali sono le varibili aleatorie nel modello lineare ?

Era:	Diventa:	Ora è una v.a.?
$\bar{x} = \frac{1}{n} \sum x_i$	$\bar{x} = \frac{1}{n} \sum x_i$	No
$\bar{y} = \frac{1}{n} \sum y_i$	$\bar{Y} = \frac{1}{n} \sum Y_i$	Sì

Il modello statistico

Quali sono le varibili aleatorie nel modello lineare ?

Era:	Diventa:	Ora è una v.a.?
$\bar{x} = \frac{1}{n} \sum x_i$	$\bar{x} = \frac{1}{n} \sum x_i$	No
$\bar{y} = \frac{1}{n} \sum y_i$	$\bar{Y} = \frac{1}{n} \sum Y_i$	Sì
$s_{xx} = \sum (x_i - \bar{x})^2$	$s_{xx} = \sum (x_i - \bar{x})^2$	No

Il modello statistico

Quali sono le varibili aleatorie nel modello lineare ?

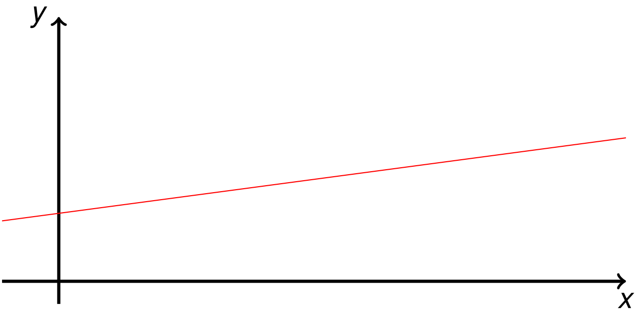
Era:	Diventa:	Ora è una v.a.?
$\bar{x} = \frac{1}{n}\sum x_i$	$\bar{x} = \frac{1}{n}\sum x_i$	No
$\bar{y} = \frac{1}{n}\sum y_i$	$\bar{Y} = \frac{1}{n}\sum Y_i$	Sì
$s_{xx} = \sum(x_i - \bar{x})^2$	$s_{xx} = \sum(x_i - \bar{x})^2$	No
$s_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y})$	$S_{xY} = \sum(x_i - \bar{x})(Y_i - \bar{Y})$	Sì

Il modello statistico

Quali sono le varibili aleatorie nel modello lineare ?

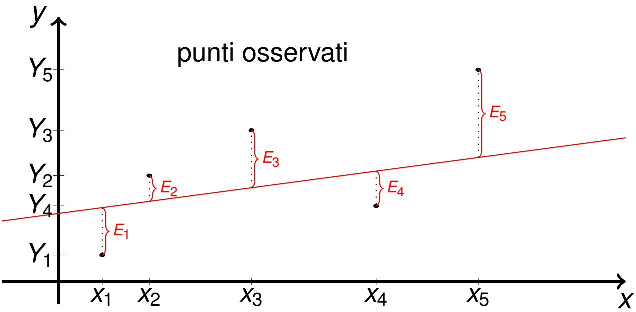
Era:	Diventa:	Ora è una v.a.?
$\bar{x} = \frac{1}{n}\sum x_i$	$\bar{x} = \frac{1}{n}\sum x_i$	No
$\bar{y} = \frac{1}{n}\sum y_i$	$\bar{Y} = \frac{1}{n}\sum Y_i$	Sì
$s_{xx} = \sum(x_i - \bar{x})^2$	$s_{xx} = \sum(x_i - \bar{x})^2$	No
$s_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y})$	$S_{xY} = \sum(x_i - \bar{x})(Y_i - \bar{Y})$	Sì
$s_{yy} = \sum(y_i - \bar{y})^2$	$S_{YY} = \sum(Y_i - \bar{Y})^2$	Sì
Ecc.		

Il modello statistico



$y = a + bx$
(legge vera)

Il modello statistico

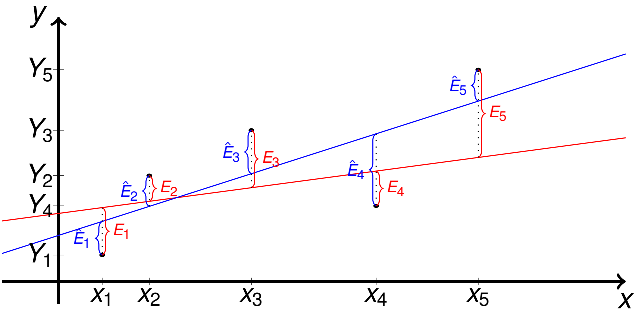


$y = a + bx$
(legge vera)

E_i = errori di misura (residui di $y = a + bx$) NON MISURABILI

perché a, b incogniti

Il modello statistico



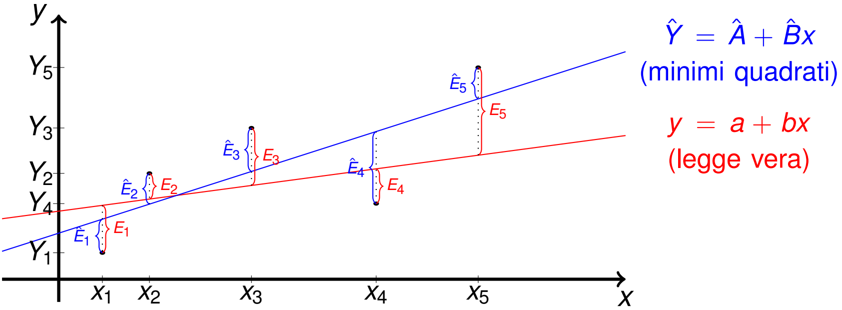
$\hat{Y} = \hat{A} + \hat{B}x$
(minimi quadrati)

$y = a + bx$
(legge vera)

E_i = errori di misura (residui di $y = a + bx$)

$\hat{Y} = \hat{A} + \hat{B}x$ interpola i punti $(x_1, Y_1), \ldots, (x_n, Y_n)$

Il modello statistico

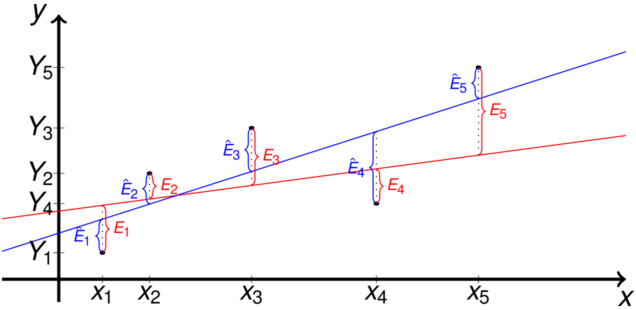


E_i = errori di misura (residui di $y = a + bx$)

$\hat{Y} = \hat{A} + \hat{B}x$ interpola i punti $(x_1, Y_1), \dots, (x_n, Y_n)$

Riguarderemo $(\hat{A}, \hat{B})$ come stimatori di (a, b)

Il modello statistico



$\hat{Y} = \hat{A} + \hat{B}x$
(minimi quadrati)

$y = a + bx$
(legge vera)

E_i = errori di misura (residui di $y = a + bx$)

$\hat{Y} = \hat{A} + \hat{B}x$ interpola i punti $(x_1, Y_1), \ldots, (x_n, Y_n)$

Riguarderemo $(\hat{A}, \hat{B})$ come stimatori di (a, b)

$\hat{E}_i$ = residui di $\hat{Y} = \hat{A} + \hat{B}x$ MISURABILI

Inferenza su b

$$\hat{B} := \frac{S_{xY}}{s_{xx}} = \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_j - \bar{x})^2} \sim N\left(b\,, \frac{\sigma^2}{s_{xx}}\right)$$

Inferenza su b

$$\hat{B} := \frac{S_{xY}}{s_{xx}} = \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_j - \bar{x})^2} \sim N\left(b\,,\, \frac{\sigma^2}{s_{xx}}\right)$$

DIMOSTRAZIONE:

$\hat{B}$ comb. lineare di gaussiane indipendenti $\quad \Rightarrow \quad \hat{B} \sim N$

Inferenza su b

$$\hat{B} := \frac{S_{xY}}{s_{xx}} = \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_j - \bar{x})^2} \sim N\left(b, \frac{\sigma^2}{s_{xx}}\right)$$

DIMOSTRAZIONE:

$\hat{B}$ comb. lineare di gaussiane indipendenti $\Rightarrow$ $\hat{B} \sim N$

$$\begin{aligned} \mathbb{E}\left[\bar{Y}\right] &= \mathbb{E}\left[\frac{1}{n}\sum Y_i\right] = \frac{1}{n}\sum \mathbb{E}\left[Y_i\right] = \frac{1}{n}\sum(a + bx_i) \\ &= a + b\bar{x} \end{aligned}$$

Inferenza su b

$$\hat{B} := \frac{S_{xY}}{s_{xx}} = \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_j - \bar{x})^2} \sim N\left(b, \frac{\sigma^2}{s_{xx}}\right)$$

DIMOSTRAZIONE:

$\hat{B}$ comb. lineare di gaussiane indipendenti $\quad \Rightarrow \quad \hat{B} \sim N$

$$\begin{aligned}
\mathbb{E}\left[\bar{Y}\right] &= \mathbb{E}\left[\frac{1}{n}\sum Y_i\right] = \frac{1}{n}\sum \mathbb{E}\left[Y_i\right] = \frac{1}{n}\sum(a + bx_i) \\
&= a + b\bar{x}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\hat{B}\right] &= \frac{\sum(x_i - \bar{x})(\mathbb{E}\left[Y_i\right] - \mathbb{E}\left[\bar{Y}\right])}{\sum(x_j - \bar{x})^2} = \frac{\sum(x_i - \bar{x})[a + bx_i - (a + b\bar{x})]}{\sum(x_j - \bar{x})^2} \\
&= b
\end{aligned}$$

Inferenza su b

$$\hat{B} := \frac{S_{xY}}{s_{xx}} = \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_j - \bar{x})^2} \sim N\left(b, \frac{\sigma^2}{s_{xx}}\right)$$

DIMOSTRAZIONE:

$\hat{B}$ comb. lineare di gaussiane indipendenti $\Rightarrow$ $\hat{B} \sim N$

$$\begin{aligned}\mathbb{E}\left[\bar{Y}\right] &= \mathbb{E}\left[\frac{1}{n}\sum Y_i\right] = \frac{1}{n}\sum \mathbb{E}\left[Y_i\right] = \frac{1}{n}\sum(a + bx_i) \\ &= a + b\bar{x}\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[\hat{B}\right] &= \frac{\sum(x_i - \bar{x})(\mathbb{E}\left[Y_i\right] - \mathbb{E}\left[\bar{Y}\right])}{\sum(x_j - \bar{x})^2} = \frac{\sum(x_i - \bar{x})[a + bx_i - (a + b\bar{x})]}{\sum(x_j - \bar{x})^2} \\ &= b\end{aligned}$$

Inferenza su b

$$\hat{B} := \frac{S_{xY}}{s_{xx}} = \frac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_j - \bar{x})^2} \sim N\left(b, \frac{\sigma^2}{s_{xx}}\right)$$

DIMOSTRAZIONE:

$\hat{B}$ comb. lineare di gaussiane indipendenti $\Rightarrow$ $\hat{B} \sim N$

$$\mathbb{E}\left[\bar{Y}\right] = \mathbb{E}\left[\frac{1}{n}\sum Y_i\right] = \frac{1}{n}\sum \mathbb{E}\left[Y_i\right] = \frac{1}{n}\sum (a + bx_i)$$

$$= a + b\bar{x}$$

$$\mathbb{E}\left[\hat{B}\right] = \frac{\sum (x_i - \bar{x})(\mathbb{E}\left[Y_i\right] - \mathbb{E}\left[\bar{Y}\right])}{\sum (x_j - \bar{x})^2} = \frac{\sum (x_i - \bar{x})[a + bx_i - (a + b\bar{x})]}{\sum (x_j - \bar{x})^2}$$

$$= b$$

$$\mathrm{Var}\left[\hat{B}\right] = \frac{\sigma^2}{s_{xx}}$$ (più complicato)

Inferenza su b

$$\hat{B} \;:=\; \frac{S_{xY}}{s_{xx}} \;=\; \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_j - \bar{x})^2} \;\sim\; N\left(b\,,\,\frac{\sigma^2}{s_{xx}}\right)$$

$\Rightarrow \quad \hat{B}$ stimatore non distorto di b

Se conoscessimo σ^2, potremmo usare la v. a.

$$\frac{\hat{B} - b}{\sqrt{\sigma^2 / s_{xx}}} \sim N(0, 1)$$

per costruire IC e fare test per il parametro b

Inferenza su b

$$\hat{B} \ := \ \frac{S_{xY}}{s_{xx}} \ = \ \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_j - \bar{x})^2} \ \sim \ N\left(b\,,\ \frac{\sigma^2}{s_{xx}}\right)$$

$$\Rightarrow \quad \hat{B} \ \text{stimatore non distorto di } b$$

Ma non conoscendo σ^2, la stimiamo con lo stimatore non distorto:

$$\hat{\Sigma}^2 := \frac{SS_E}{n-2} = \frac{1}{n-2}\left(S_{YY} - \frac{S_{xY}^2}{s_{xx}}\right)$$

Inferenza su b

$$\hat{B} \; := \; \frac{S_{xY}}{s_{xx}} \; = \; \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_j - \bar{x})^2} \; \sim \; N\left(b \, , \, \frac{\sigma^2}{s_{xx}}\right)$$

$$\Rightarrow \quad \hat{B} \text{ stimatore non distorto di } b$$

Ma non conoscendo σ^2, la stimiamo con lo stimatore non distorto:

$$\hat{\Sigma}^2 := \frac{SS_E}{n-2} = \frac{1}{n-2}\left(S_{YY} - \frac{S_{xY}^2}{s_{xx}}\right)$$

Si dimostra che

$$\frac{(n-2)\hat{\Sigma}^2}{\sigma^2} \sim \chi^2(n-2) \quad \text{indipendente da } \hat{B}$$

Inferenza su b

$$\hat{B} := \frac{S_{xY}}{s_{xx}} = \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_j - \bar{x})^2} \sim N\left(b, \frac{\sigma^2}{s_{xx}}\right)$$

$$\Rightarrow \quad \hat{B} \text{ stimatore non distorto di } b$$

Ma non conoscendo σ^2, la stimiamo con lo stimatore non distorto:

$$\hat{\Sigma}^2 := \frac{SS_E}{n-2} = \frac{1}{n-2}\left(S_{YY} - \frac{S_{xY}^2}{s_{xx}}\right)$$

Si dimostra che

$$\frac{(n-2)\hat{\Sigma}^2}{\sigma^2} \sim \chi^2(n-2) \quad \text{indipendente da } \hat{B}$$

e quindi

$$\frac{\hat{B} - b}{\sqrt{\hat{\Sigma}^2 / s_{xx}}} \sim t(n-2)$$

Inferenza su b

$$\hat{B} := \frac{S_{xY}}{s_{xx}} = \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_j - \bar{x})^2} \sim N\left(b, \frac{\sigma^2}{s_{xx}}\right)$$

$\Rightarrow \quad \hat{B}$ stimatore non distorto di b

Ma non conoscendo σ^2, la stimiamo con lo stimatore non distorto:

$$\hat{\Sigma}^2 := \frac{SS_E}{n-2} = \frac{1}{n-2}\left(S_{YY} - \frac{S_{xY}^2}{s_{xx}}\right)$$

Si dimostra che

$$\frac{(n-2)\hat{\Sigma}^2}{\sigma^2} \sim \chi^2(n-2) \quad \text{indipendente da } \hat{B}$$

e quindi

$$\frac{\frac{\hat{B}-b}{\sqrt{\sigma^2/s_{xx}}}}{\frac{1}{\sqrt{n-2}}\sqrt{\frac{(n-2)\hat{\Sigma}^2}{\sigma^2}}} = \frac{\hat{B}-b}{\sqrt{\hat{\Sigma}^2/s_{xx}}} \sim t(n-2)$$

$N(0,1)$

indip.

$\chi^2(n-2)$

Inferenza su b

$$\frac{\hat{B} - b}{\hat{\Sigma}/\sqrt{s_{xx}}} \sim \; t(n-2)$$

$$\Rightarrow \quad \left(\hat{b} \pm t_{\frac{1+\gamma}{2}}(n-2)\frac{\hat{\sigma}}{\sqrt{s_{xx}}} \right) \quad \text{è un} \;\; IC_b(\gamma)$$

Inferenza su b

$$\frac{\hat{B} - b}{\hat{\Sigma}/\sqrt{s_{xx}}} \sim t(n-2)$$

$$\Rightarrow \quad \left(\hat{b} \pm t_{\frac{1+\gamma}{2}}(n-2)\frac{\hat{\sigma}}{\sqrt{s_{xx}}}\right) \quad \text{è un} \quad IC_b(\gamma)$$

DIMOSTRAZIONE:

$$\mathbb{P}\left(b \in \left(\hat{B} \pm t_{\frac{1+\gamma}{2}}(n-2)\frac{\hat{\Sigma}}{\sqrt{s_{xx}}}\right)\right) =$$

$$= \mathbb{P}\left(-t_{\frac{1+\gamma}{2}}(n-2) \leq \frac{\hat{B} - b}{\hat{\Sigma}/\sqrt{s_{xx}}} \leq t_{\frac{1+\gamma}{2}}(n-2)\right)$$

$$= \gamma$$

Inferenza su b

$$\frac{\hat{B} - b}{\hat{\Sigma}/\sqrt{s_{xx}}} \sim \; t(n-2)$$

$$\Rightarrow \quad \text{“ Rifiuto } H_0 \text{ se } \left| \frac{\hat{B} - b_0}{\hat{\Sigma}/\sqrt{s_{xx}}} \right| \geq t_{1-\frac{\alpha}{2}}(n-2) \text{”}$$

è un test di livello α per le ipotesi

$$H_0 : b = b_0 \qquad \text{vs.} \qquad H_1 : b \neq b_0$$

Inferenza su b

$$\frac{\hat{B} - b}{\hat{\Sigma}/\sqrt{s_{xx}}} \sim \; t(n-2)$$

$$\Rightarrow \quad \text{“ Rifiuto } H_0 \text{ se } \left| \frac{\hat{B} - b_0}{\hat{\Sigma}/\sqrt{s_{xx}}} \right| \geq t_{1-\frac{\alpha}{2}}(n-2) \text{”}$$

è un test di livello α per le ipotesi

$$H_0 : b = b_0 \qquad \text{vs.} \qquad H_1 : b \neq b_0$$

DIMOSTRAZIONE:

$$\mathbb{P}_{H_0 \text{ vera}}(\text{“rifiuto } H_0\text{”}) = \mathbb{P}_{b=b_0}\left(\underbrace{\left| \frac{\hat{B} - b_0}{\hat{\Sigma}/\sqrt{s_{xx}}} \right|}_{t(n-2)} > t_{1-\frac{\alpha}{2}}(n-2) \right)$$

$$= \alpha$$

Inferenza su b

$$\frac{\hat{B} - b}{\hat{\Sigma}/\sqrt{s_{xx}}} \sim \; t(n-2)$$

$$\Rightarrow \quad \text{“ Rifiuto } H_0 \text{ se } \left| \frac{\hat{B} - b_0}{\hat{\Sigma}/\sqrt{s_{xx}}} \right| \geq t_{1-\frac{\alpha}{2}}(n-2)\text{”}$$

è un test di livello α per le ipotesi

$$H_0 : b = b_0 \qquad \text{vs.} \qquad H_1 : b \neq b_0$$

Se $b_0 = 0$, il test per le ipotesi

$$H_0 : b = 0 \qquad \text{vs.} \qquad H_1 : b \neq 0$$

verifica che le Y dipendano realmente da x (*significatività del modello*)

Inferenza su a

$$\hat{A} \;=\; \bar{Y} - \frac{S_{xY}}{s_{xx}}\,\bar{x} \;\sim\; N\left(a\,,\,\sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}\right)\right)$$

$\Rightarrow$ $\hat{A}$ stimatore non distorto di a

(dimostrazioni simili al caso per $\hat{B}$)

Inferenza su a

$$\hat{A} \;=\; \bar{Y} - \frac{S_{xY}}{s_{xx}}\,\bar{x} \;\sim\; N\left(a\,,\,\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{s_{xx}}\right)\right)$$

$\Rightarrow\quad \hat{A}$ stimatore non distorto di a

$$\frac{\hat{A}-a}{\hat{\Sigma}\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{s_{xx}}}} \;\sim\; t(n-2)$$

$\Rightarrow\quad$ IC e test per il parametro a

(dimostrazioni simili al caso per $\hat{B}$)

Terminologia

$$\hat{\Sigma}^2 \;:=\; \frac{1}{n-2}\left(S_{YY} - \frac{S_{xY}^2}{s_{xx}}\right) \quad \text{(stimatore non distorto di } \sigma^2\text{)}$$

Terminologia

$$\hat{\Sigma}^2 := \frac{1}{n-2}\left(S_{YY} - \frac{S_{xY}^2}{s_{xx}}\right) \quad \text{(stimatore non distorto di } \sigma^2\text{)}$$

$\hat{\Sigma} := \sqrt{\hat{\Sigma}^2}$ *residual standard error*

(stimatore *approx.* non distorto di σ)

Terminologia

$$\hat{\Sigma}^2 := \frac{1}{n-2}\left(S_{YY} - \frac{S_{xY}^2}{s_{xx}}\right)$$ (stimatore non distorto di σ^2)

$$\hat{\Sigma} := \sqrt{\hat{\Sigma}^2}$$ *residual standard error*

(stimatore *approx.* non distorto di σ)

$$\mathrm{se}(\hat{A}) := \hat{\Sigma}\sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}\right)}$$ *standard error di* $\hat{A}$

(stimatore *approx.* non distorto di $\sqrt{\mathrm{Var}[\hat{A}]}$)

Terminologia

$$\hat{\Sigma}^2 := \frac{1}{n-2}\left(S_{YY} - \frac{S_{xY}^2}{s_{xx}}\right) \quad \text{(stimatore non distorto di } \sigma^2\text{)}$$

$$\hat{\Sigma} := \sqrt{\hat{\Sigma}^2} \quad \textit{residual standard error}$$

(stimatore *approx.* non distorto di σ)

$$\mathrm{se}(\hat{A}) := \hat{\Sigma}\sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}\right)} \quad \textit{standard error di } \hat{A}$$

(stimatore *approx.* non distorto di $\sqrt{\mathrm{Var}[\hat{A}]}$)

$$\mathrm{se}(\hat{B}) := \frac{\hat{\Sigma}}{\sqrt{s_{xx}}} \quad \textit{standard error di } \hat{B}$$

(stimatore *approx.* non distorto di $\sqrt{\mathrm{Var}[\hat{B}]}$)

Terminologia

La terminologia corretta sarebbe (p.es. per $\hat{B}$)

$\mathrm{se}(\hat{B}) = \sqrt{\mathrm{Var}[\hat{B}]} = \dfrac{\sigma}{\sqrt{s_{xx}}}$	*errore* (o *deviazione*) *standard* di $\hat{B}$
$\widehat{\mathrm{SE}(\hat{B})} = \dfrac{\hat{\Sigma}}{\sqrt{s_{xx}}}$	stimatore approx. non distorto di $\mathrm{se}(\hat{B})$
$\widehat{\mathrm{se}(\hat{B})} = \dfrac{\hat{\sigma}}{\sqrt{s_{xx}}}$	stima (= realizzazione) di $\widehat{\mathrm{SE}(\hat{B})}$

Terminologia

La terminologia corretta sarebbe (p.es. per $\hat{B}$)

$$\mathrm{se}(\hat{B}) = \sqrt{\mathrm{Var}[\hat{B}]} = \frac{\sigma}{\sqrt{s_{xx}}}$$ *errore* (o *deviazione*) *standard* di $\hat{B}$

$$\widehat{\mathrm{SE}(\hat{B})} = \frac{\hat{\Sigma}}{\sqrt{s_{xx}}}$$ stimatore approx. non distorto di $\mathrm{se}(\hat{B})$

$$\widehat{\mathrm{se}(\hat{B})} = \frac{\hat{\sigma}}{\sqrt{s_{xx}}}$$ stima (= realizzazione) di $\widehat{\mathrm{SE}(\hat{B})}$

Ma con abuso di notazione, diremo semplicemente

$$\mathrm{se}(\hat{B}) = \frac{\hat{\Sigma}}{\sqrt{s_{xx}}}$$ *standard error di* $\hat{B}$

$$\mathrm{se}(\hat{b}) = \frac{\hat{\sigma}}{\sqrt{s_{xx}}}$$ (realizzazione dello) *standard error di* $\hat{B}$

Terminologia

In termini delle quantità precedenti:

$\left(\hat{a} \pm t_{\frac{1+\gamma}{2}}(n-2)\,\mathrm{se}(\hat{a})\right)$	$IC(\gamma)$ per a
$\dfrac{\hat{A}-a_0}{\mathrm{se}(\hat{A})} \sim t(n-2)$	statistica test per un test su a
$\left(\hat{b} \pm t_{\frac{1+\gamma}{2}}(n-2)\,\mathrm{se}(\hat{b})\right)$	$IC(\gamma)$ per b
$\dfrac{\hat{B}-b_0}{\mathrm{se}(\hat{B})} \sim t(n-2)$	statistica test per un test su b

Output della regressione in R

$se(\hat{a})$ $\qquad \dfrac{\hat{a}}{se(\hat{a})}$

```
Call:
lm(formula = Allungamento ~ I(Tensione^2), data = Data)

Residuals:
     Min       1Q   Median       3Q      Max
-14.7339  -3.5666   0.0062   3.6649  15.2774

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.6979121  0.8078622   0.864    0.389
I(Tensione^2) 0.0197671  0.0006648  29.736   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.519 on 148 degrees of freedom
Multiple R-squared:  0.8566,    Adjusted R-squared:  0.8557
F-statistic: 884.2 on 1 and 148 DF,  p-value: < 2.2e-16
```

$\hat{a}$ $\qquad$ p-value test su a

$\hat{b}$ $\qquad$ p-value test su b

r^2 $\qquad \hat{\sigma}$ $\qquad se(\hat{b})$ $\qquad n-2$ $\qquad \dfrac{\hat{b}}{se(\hat{b})}$

Test per la bontà del modello lineare

Le ipotesi su cui si basa il modello sono

(1) $Y_i = a + bx_i + E_i$

(2) $E_i \sim N(0, \sigma^2)$

(3) $E_1, \ldots, E_n$ i.i.d.

Test per la bontà del modello lineare

Le ipotesi su cui si basa il modello sono

(1) $Y_i = a + bx_i + E_i$

(2) $E_i \sim N(0, \sigma^2)$

(3) $E_1, \ldots, E_n$ i.i.d.

$\iff$

(A) $Y_i \sim N(a + bx_i\,,\, \sigma^2)$

(B) $Y_1, \ldots, Y_n$ indipendenti

Test per la bontà del modello lineare

Le ipotesi su cui si basa il modello sono

(1) $Y_i = a + bx_i + E_i$

(2) $E_i \sim N(0, \sigma^2)$

(3) $E_1, \ldots, E_n$ i.i.d.

$\Longleftrightarrow$

(A) $Y_i \sim N(a + bx_i\,,\, \sigma^2)$

(B) $Y_1, \ldots, Y_n$ indipendenti

Ma come facciamo a verificarle?

Test per la bontà del modello lineare

Le ipotesi su cui si basa il modello sono

(1) $Y_i = a + bx_i + E_i$

(2) $E_i \sim N(0, \sigma^2)$

(3) $E_1, \ldots, E_n$ i.i.d.

$\iff$

(A) $Y_i \sim N(a + bx_i\,, \sigma^2)$

(B) $Y_1, \ldots, Y_n$ indipendenti

Ma come facciamo a verificarle?

Per verificare almeno (1) - (2), servirebbe un test di normalità sugli E_i

residui di $a + bx$

Test per la bontà del modello lineare

Le ipotesi su cui si basa il modello sono

(1) $Y_i = a + bx_i + E_i$

(2) $E_i \sim N(0, \sigma^2)$

(3) $E_1, \ldots, E_n$ i.i.d.

$\Longleftrightarrow$

(A) $Y_i \sim N(a + bx_i\,, \sigma^2)$

(B) $Y_1, \ldots, Y_n$ indipendenti

Ma come facciamo a verificarle?

Per verificare almeno (1) - (2), servirebbe un test di normalità sugli E_i

MA

dai dati sappiamo ricavare solo gli $\hat{E}_i = Y_i - \hat{Y}_i$

residui della LSL

Test per la bontà del modello lineare

Le ipotesi su cui si basa il modello sono

(1) $Y_i = a + bx_i + E_i$

(2) $E_i \sim N(0, \sigma^2)$

(3) $E_1, \ldots, E_n$ i.i.d.

$\iff$

(A) $Y_i \sim N(a + bx_i\,,\ \sigma^2)$

(B) $Y_1, \ldots, Y_n$ indipendenti

Teorema (non dimostrato)

Se le (1) - (2) - (3) sono vere, allora i *residui standardizzati*

$$R_i := \frac{\hat{E}_i}{\hat{\Sigma}} = \frac{Y_i - (\hat{A} + \hat{B}x_i)}{\hat{\Sigma}}$$

- sono (approssimativamente) i.i.d.
- $R_i \approx N(0, 1)$

Test per la bontà del modello lineare

Le ipotesi su cui si basa il modello sono

(1) $Y_i = a + bx_i + E_i$

(2) $E_i \sim N(0, \sigma^2)$

(3) $E_1, \dots, E_n$ i.i.d.

$\iff$

(A) $Y_i \sim N(a + bx_i\,,\, \sigma^2)$

(B) $Y_1, \dots, Y_n$ indipendenti

Teorema (non dimostrato)

Se le (1) - (2) - (3) sono vere, allora i *residui standardizzati*

$$R_i := \frac{\hat{E}_i}{\hat{\Sigma}} = \frac{Y_i - (\hat{A} + \hat{B}x_i)}{\hat{\Sigma}}$$

- sono (approssimativamente) i.i.d.
- $R_i \approx N(0, 1)$

$\Rightarrow$ Possiamo fare un test di normalità per gli $R_1, \dots, R_n$!

Test per la bontà del modello lineare

Assumendo $Y_1, \ldots, Y_n$ indipendenti, un test per le ipotesi

$$H_0 : \exists\, a,\, b,\, \sigma \quad \text{tali che} \quad Y_i \sim N(a + bx_i\,,\, \sigma^2) \quad \forall i = 1, \ldots, n$$

vs.

$$H_1 : H_0 \text{ è falsa}$$

può quindi essere un qualunque test di normalità sugli $R_1, \ldots, R_n$

Test per la bontà del modello lineare

Assumendo $Y_1, \ldots, Y_n$ indipendenti, un test per le ipotesi

$$H_0 : \exists\, a, b, \sigma \quad \text{tali che} \quad Y_i \sim N(a + bx_i\,, \sigma^2) \quad \forall i = 1, \ldots, n$$

vs.

$$H_1 : H_0 \text{ è falsa}$$

può quindi essere un qualunque test di normalità sugli $R_1, \ldots, R_n$

Analizzando lo scatterplot degli $r_1, \ldots, r_n$:

- non deve esserci una struttura definita (*omoschedasticità*)

Test per la bontà del modello lineare

Assumendo $Y_1, \ldots, Y_n$ indipendenti, un test per le ipotesi

$$H_0 : \exists\, a, b, \sigma \quad \text{tali che} \quad Y_i \sim N(a + bx_i\,, \sigma^2) \quad \forall i = 1, \ldots, n$$

vs.

$$H_1 : H_0 \text{ è falsa}$$

può quindi essere un qualunque test di normalità sugli $R_1, \ldots, R_n$

Analizzando lo scatterplot degli $r_1, \ldots, r_n$:

- non deve esserci una struttura definita (*omoschedasticità*)
- i residui devono disporsi “a nuvola” intorno alla linea orizzontale

Test per la bontà del modello lineare

Assumendo $Y_1, \ldots, Y_n$ indipendenti, un test per le ipotesi

$$H_0 : \exists\, a, b, \sigma \quad \text{tali che} \quad Y_i \sim N(a + bx_i\,,\, \sigma^2) \quad \forall i = 1, \ldots, n$$

$$\text{vs.}$$

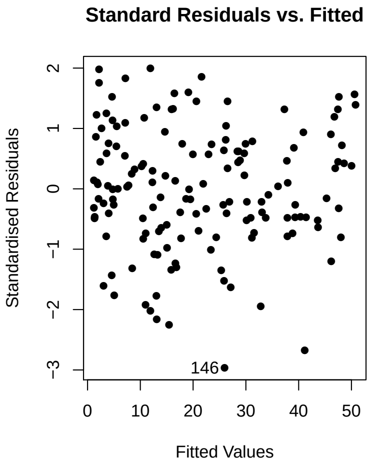
$$H_1 : H_0 \text{ è falsa}$$

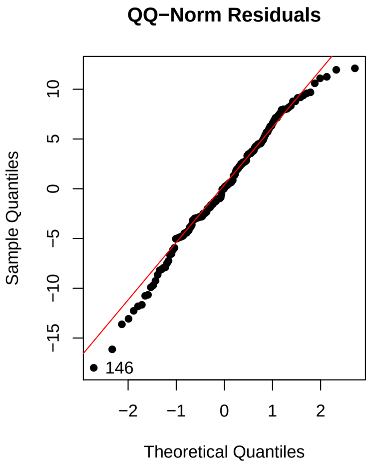
può quindi essere un qualunque test di normalità sugli $R_1, \ldots, R_n$

Analizzando lo scatterplot degli $r_1, \ldots, r_n$:

- non deve esserci una struttura definita (*omoschedasticità*)
- i residui devono disporsi "a nuvola" intorno alla linea orizzontale
- circa il 95% dei residui deve essere compreso tra -2 e $+2$

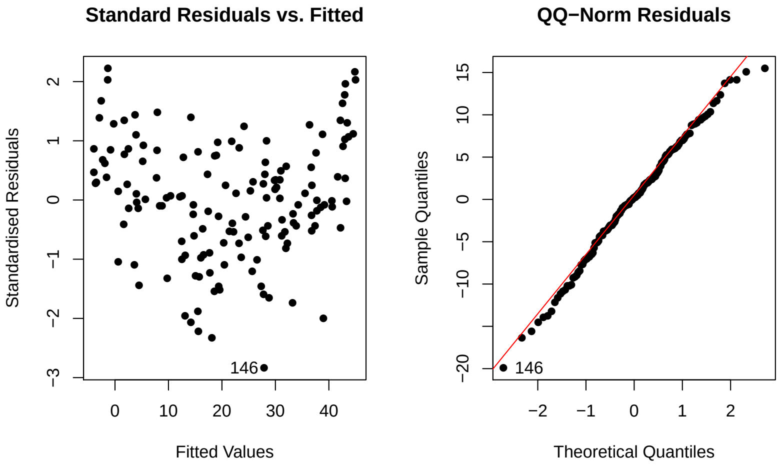
Due esempi





OK modello lineare!

Due esempi



NO modello lineare (forse quadratico)

IC per il valore atteso di una nuova osservazione

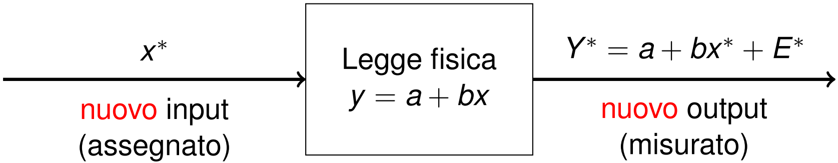
$$x_1, \dots, x_n \longrightarrow \boxed{\begin{array}{c}\text{Legge fisica}\\ y = a + bx\end{array}} \longrightarrow \begin{array}{c} Y_1, \dots, Y_n \\ Y_i = a + bx_i + E_i \end{array}$$

input (assegnati) — output (misurati)

Ipotesi:

- $E_1, \dots, E_n$ indipendenti $\Leftrightarrow$ $Y_1, \dots, Y_n$ indipendenti
- $E_i \sim N(0, \sigma^2)$ $\Leftrightarrow$ $Y_i \sim N(a + bx_i\,,\, \sigma^2)$

IC per il valore atteso di una nuova osservazione

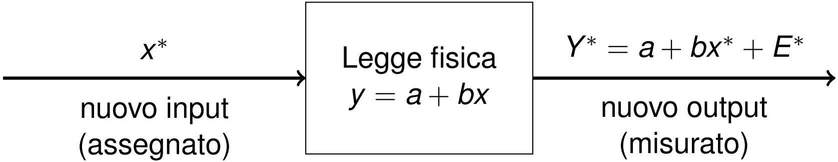


Ipotesi:

- $E_1, \ldots, E_n, E^*$ indipendenti $\quad\Leftrightarrow\quad$ $Y_1, \ldots, Y_n, Y^*$ indipendenti
- $E_i \sim N(0, \sigma^2)$ $\quad\Leftrightarrow\quad$ $Y_i \sim N(a + bx_i\,,\, \sigma^2)$

  $E^* \sim N(0, \sigma^2)$ $\quad\Leftrightarrow\quad$ $Y^* \sim N(a + bx^*\,,\, \sigma^2)$

IC per il valore atteso di una nuova osservazione



Ipotesi:

- $E_1, \ldots, E_n, E^*$ indipendenti $\quad\Leftrightarrow\quad$ $Y_1, \ldots, Y_n, Y^*$ indipendenti
- $E_i \sim N(0, \sigma^2)$ $\quad\Leftrightarrow\quad$ $Y_i \sim N(a + bx_i\,,\,\sigma^2)$

  $E^* \sim N(0, \sigma^2)$ $\quad\Leftrightarrow\quad$ $Y^* \sim N(a + bx^*\,,\,\sigma^2)$

Vogliamo trovare IC per il parametro

$$\mathbb{E}\left[Y^*\right] = a + bx^*$$

IC per il valore atteso di una nuova osservazione

IPOTESI:

- $Y_1, \ldots, Y_n, Y^*$ indipendenti
- $Y_i \sim N(a + bx_i\,,\, \sigma^2)$, $Y^* \sim N(a + bx^*\,,\, \sigma^2)$

IC per il valore atteso di una nuova osservazione

IPOTESI:

- $Y_1, \ldots, Y_n, Y^*$ indipendenti
- $Y_i \sim N(a + bx_i\,, \sigma^2)$, $Y^* \sim N(a + bx^*\,, \sigma^2)$

CONSEGUENZE:

$\hat{A} + \hat{B}x^* \sim N$

somma normali indipendenti

IC per il valore atteso di una nuova osservazione

IPOTESI:

- $Y_1, \dots, Y_n, Y^*$ indipendenti
- $Y_i \sim N(a + bx_i\,, \sigma^2)$, $Y^* \sim N(a + bx^*\,, \sigma^2)$

CONSEGUENZE:

$$\hat{A} + \hat{B}x^* \sim N\left(a + bx^* \qquad\qquad\qquad\qquad\right)$$

somma normali indipendenti

$$\mathbb{E}[\hat{A}] = a$$
$$\mathbb{E}[\hat{B}] = b$$

IC per il valore atteso di una nuova osservazione

IPOTESI:

- $Y_1, \ldots, Y_n, Y^*$ indipendenti
- $Y_i \sim N(a + bx_i\,,\, \sigma^2)$, $Y^* \sim N(a + bx^*\,,\, \sigma^2)$

CONSEGUENZE:

$$\hat{A} + \hat{B}x^* \sim N\left(a + bx^*\,,\, \sigma^2 \left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right)$$

somma normali indipendenti

$\mathbb{E}[\hat{A}] = a$
$\mathbb{E}[\hat{B}] = b$

più complicato

IC per il valore atteso di una nuova osservazione

IPOTESI:

- $Y_1, \ldots, Y_n, Y^*$ indipendenti
- $Y_i \sim N(a + bx_i\,,\, \sigma^2),\;\; Y^* \sim N(a + bx^*\,,\, \sigma^2)$

CONSEGUENZE:

$$\hat{A} + \hat{B}x^* \sim N\left(a + bx^*\,,\, \sigma^2\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right)$$

$\Rightarrow \quad \hat{Y}^* := \hat{A} + \hat{B}x^*$ stimatore non distorto di $a + bx^*$

IC per il valore atteso di una nuova osservazione

IPOTESI:

- $Y_1, \ldots, Y_n, Y^*$ indipendenti
- $Y_i \sim N(a + bx_i \,, \sigma^2)$, $Y^* \sim N(a + bx^* \,, \sigma^2)$

CONSEGUENZE:

$$\hat{A} + \hat{B}x^* \sim N\left(a + bx^* \,, \sigma^2 \left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right)$$

$$\Rightarrow \quad \hat{Y}^* := \hat{A} + \hat{B}x^* \text{ stimatore non distorto di } a + bx^*$$

$$\frac{\hat{A} + \hat{B}x^* - (a + bx^*)}{\sqrt{\sigma^2 \left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)}} \sim N(0, 1)$$

IC per il valore atteso di una nuova osservazione

IPOTESI:

- $Y_1, \dots, Y_n, Y^*$ indipendenti
- $Y_i \sim N(a + bx_i\,,\, \sigma^2)$, $\;Y^* \sim N(a + bx^*\,,\, \sigma^2)$

CONSEGUENZE:

$$\hat{A} + \hat{B}x^* \sim N\left(a + bx^*\,,\, \sigma^2\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right)$$

$\Rightarrow \quad \hat{Y}^* := \hat{A} + \hat{B}x^*$ stimatore non distorto di $a + bx^*$

$$\frac{\hat{A} + \hat{B}x^* - (a + bx^*)}{\sqrt{\hat{\Sigma}^2\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)}} \sim t(n-2)$$

IC per il valore atteso di una nuova osservazione

IPOTESI:

- $Y_1, \dots, Y_n, Y^*$ indipendenti
- $Y_i \sim N(a + bx_i\,,\, \sigma^2)$, $Y^* \sim N(a + bx^*\,,\, \sigma^2)$

CONSEGUENZE:

$$\hat{A} + \hat{B}x^* \sim N\left(a + bx^*\,,\, \sigma^2\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right)$$

$$\Rightarrow \quad \hat{Y}^* := \hat{A} + \hat{B}x^* \text{ stimatore non distorto di } a + bx^*$$

$$\frac{\hat{A} + \hat{B}x^* - (a + bx^*)}{\sqrt{\hat{\Sigma}^2\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)}} \sim t(n-2)$$

$$\Rightarrow \left(\hat{a} + \hat{b}x^* \pm t_{\frac{1+\gamma}{2}}(n-2)\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)}\right) \quad \text{è un } IC_{a+bx^*}(\gamma)$$

Intervalli di predizione per una nuova osservazione

Definizione

Siano $L = \ell(Y_1, \ldots, Y_n)$ e $U = u(Y_1, \ldots, Y_n)$ due statistiche tali che

$$\mathbb{P}\,(L < Y^* < U) = \gamma \qquad\qquad (\text{con } \gamma \in (0, 1) \text{ fissato})$$

Intervalli di predizione per una nuova osservazione

Definizione

Siano $L = \ell(Y_1, \dots, Y_n)$ e $U = u(Y_1, \dots, Y_n)$ due statistiche tali che

$$\mathbb{P}\,(L < Y^* < U) = \gamma \qquad\qquad \text{(con } \gamma \in (0,1) \text{ fissato)}$$

Allora, se $y_1, \dots, y_n$ sono le realizzazioni di $Y_1, \dots, Y_n$, si dice che

$$\big(\ell(y_1, \dots, y_n)\,,\, u(y_1, \dots, y_n)\big)$$

è un *intervallo di predizione* di livello γ per la v.a. Y^* ($IP_{Y^*}(\gamma)$)

Intervalli di predizione per una nuova osservazione

Definizione

Siano $L = \ell(Y_1, \ldots, Y_n)$ e $U = u(Y_1, \ldots, Y_n)$ due statistiche tali che

$$\mathbb{P}\,(L < Y^* < U) = \gamma \qquad \text{(con } \gamma \in (0,1) \text{ fissato)}$$

Allora, se $y_1, \ldots, y_n$ sono le realizzazioni di $Y_1, \ldots, Y_n$, si dice che

$$\big(\ell(y_1, \ldots, y_n)\,,\ u(y_1, \ldots, y_n)\big)$$

è un *intervallo di predizione* di livello γ per la v.a. Y^* ($IP_{Y^*}(\gamma)$)

$Y^* \sim N(a + bx^*\,,\, \sigma^2)$	
$a + bx^* \in \mathbb{R}$ parametro incognito $\mathbb{P}\,(L < a + bx^* < U) = \gamma$ $\Rightarrow (l, u)$ è un $IC_{a+bx^*}(\gamma)$	

Intervalli di predizione per una nuova osservazione

Definizione

Siano $L = \ell(Y_1, \dots, Y_n)$ e $U = u(Y_1, \dots, Y_n)$ due statistiche tali che

$$\mathbb{P}\,(L < Y^* < U) = \gamma \qquad\qquad \text{(con } \gamma \in (0,1) \text{ fissato)}$$

Allora, se $y_1, \dots, y_n$ sono le realizzazioni di $Y_1, \dots, Y_n$, si dice che

$$\big(\ell(y_1, \dots, y_n)\,,\, u(y_1, \dots, y_n)\big)$$

è un *intervallo di predizione* di livello γ per la v.a. Y^* ($IP_{Y^*}(\gamma)$)

<table>
<tr><td colspan="2" align="center">$Y^* \sim N(a + bx^*\,,\, \sigma^2)$</td></tr>
<tr><td>$a + bx^* \in \mathbb{R}$ parametro incognito
$\mathbb{P}\,(L < a + bx^* < U) = \gamma$
$\Rightarrow (l, u)$ è un $IC_{a+bx^*}(\gamma)$</td><td>Y^* variabile aleatoria
$\mathbb{P}\,(L < Y^* < U) = \gamma$
$\Rightarrow (l, u)$ è un $IP_{Y^*}(\gamma)$</td></tr>
</table>

Intervalli di predizione per una nuova osservazione

$$\left.\begin{array}{l} Y^* \sim N(a + bx^*\,,\, \sigma^2) \\ \hat{Y}^* = \hat{A} + \hat{B}x^* \sim N\left(a + bx^*\,,\, \sigma^2\left(\dfrac{1}{n} + \dfrac{(x^* - \bar{x})^2}{s_{xx}}\right)\right) \end{array}\right\} \quad \text{indipendenti}$$

funzioni di $Y_1, \ldots, Y_n$

Intervalli di predizione per una nuova osservazione

$$\left.\begin{aligned} &Y^* \sim N(a + bx^*, \sigma^2) \\ &\hat{Y}^* = \hat{A} + \hat{B}x^* \sim N\left(a + bx^*, \sigma^2\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right) \end{aligned}\right\} \quad \text{indipendenti}$$

$$\Rightarrow \quad Y^* - \hat{Y}^* \sim N\left(0, \sigma^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right)$$

Intervalli di predizione per una nuova osservazione

$$\left.\begin{array}{l} Y^* \sim N(a + bx^*\,,\, \sigma^2) \\ \hat{Y}^* = \hat{A} + \hat{B}x^* \sim N\left(a + bx^*\,,\, \sigma^2\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right) \end{array}\right\} \quad \text{indipendenti}$$

$$\Rightarrow \quad Y^* - \hat{Y}^* \sim N\left(0\,,\, \sigma^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right)$$

$$\frac{Y^* - \hat{Y}^*}{\sqrt{\hat{\Sigma}^2\left(1 + \frac{1}{n} + \frac{(x^*-\bar{x})^2}{s_{xx}}\right)}} \sim t(n-2)$$

Intervalli di predizione per una nuova osservazione

$$\left.\begin{aligned} & Y^* \sim N(a + bx^*\,,\,\sigma^2) \\ & \hat{Y}^* = \hat{A} + \hat{B}x^* \sim N\left(a + bx^*\,,\,\sigma^2\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right) \end{aligned}\right\} \quad \text{indipendenti}$$

$$\Rightarrow \quad Y^* - \hat{Y}^* \sim N\left(0\,,\,\sigma^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right)$$

$$\frac{Y^* - \hat{Y}^*}{\sqrt{\hat{\Sigma}^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)}} \sim t(n-2)$$

$$\Rightarrow \mathbb{P}\left(Y^* \in \left(\hat{Y}^* \pm t_{\frac{1+\gamma}{2}}(n-2)\sqrt{\hat{\Sigma}^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)}\right)\right) = \gamma$$

Intervalli di predizione per una nuova osservazione

$$\left.\begin{aligned} &Y^* \sim N(a + bx^*\,,\, \sigma^2) \\ &\hat{Y}^* = \hat{A} + \hat{B}x^* \sim N\left(a + bx^*\,,\, \sigma^2\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right) \end{aligned}\right\} \quad \text{indipendenti}$$

$$\Rightarrow \quad Y^* - \hat{Y}^* \sim N\left(0\,,\, \sigma^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right)$$

$$\frac{Y^* - \hat{Y}^*}{\sqrt{\hat{\Sigma}^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)}} \sim t(n-2)$$

$$\Rightarrow \mathbb{P}\left(Y^* \in \left(\hat{Y}^* \pm t_{\frac{1+\gamma}{2}}(n-2)\sqrt{\hat{\Sigma}^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)}\right)\right) = \gamma$$

$$\Rightarrow \left(\hat{y}^* \pm t_{\frac{1+\gamma}{2}}(n-2)\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)}\right) \quad \text{è un } IP_{Y^*}(\gamma)$$

Intervalli di predizione per una nuova osservazione

$$\left.\begin{aligned} & Y^* \sim N(a + bx^*\,,\,\sigma^2) \\ & \hat{Y}^* = \hat{A} + \hat{B}x^* \sim N\left(a + bx^*\,,\,\sigma^2\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right) \end{aligned}\right\} \quad \text{indipendenti}$$

$$\Rightarrow \quad Y^* - \hat{Y}^* \sim N\left(0\,,\,\sigma^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)\right)$$

$$\frac{Y^* - \hat{Y}^*}{\sqrt{\hat{\Sigma}^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)}} \sim t(n-2)$$

$$\Rightarrow \mathbb{P}\left(Y^* \in \left(\hat{Y}^* \pm t_{\frac{1+\gamma}{2}}(n-2)\sqrt{\hat{\Sigma}^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)}\right)\right) = \gamma$$

$$\Rightarrow \left(\hat{y}^* \pm t_{\frac{1+\gamma}{2}}(n-2)\underbrace{\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}\right)}}_{\text{più largo dell'}IC_{a+bx^*}(\gamma)}\right) \quad \text{è un } IP_{Y^*}(\gamma)$$

Regressione lineare multipla

$$\begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}$$

input
(vettore in $\mathbb{R}^k$)

Legge fisica

$$y = a + b_1 x_1 + \ldots + b_k x_k$$

Regressione lineare multipla

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}$$

input
(vettore in $\mathbb{R}^k$)

Legge fisica

$$y = a + b_1 x_1 + \ldots + b_k x_k$$
$$= a + \mathbf{x}^t \mathbf{b}$$
$$(a \in \mathbb{R},\ \mathbf{b} \in \mathbb{R}^k)$$

Regressione lineare multipla

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}$$

input
(vettore in $\mathbb{R}^k$)

Legge fisica

$$y = a + b_1 x_1 + \ldots + b_k x_k$$
$$= a + \mathbf{x}^t \mathbf{b}$$
$$(a \in \mathbb{R},\ \mathbf{b} \in \mathbb{R}^k)$$

$$Y = a + \mathbf{x}^t \mathbf{b} + E$$

output
(variabile aleatoria)

Regressione lineare multipla

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}$$

input
(vettore in $\mathbb{R}^k$)

Legge fisica

$$y = a + b_1 x_1 + \ldots + b_k x_k$$
$$= a + \mathbf{x}^t \mathbf{b}$$
$$(a \in \mathbb{R},\ \mathbf{b} \in \mathbb{R}^k)$$

$$Y = a + \mathbf{x}^t \mathbf{b} + E$$

output
(variabile aleatoria)

predittori
(o *regressori*)

Regressione lineare multipla

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} \longrightarrow \boxed{\begin{array}{c} \text{Legge fisica} \\ y = a + b_1 x_1 + \ldots + b_k x_k \\ \quad = a + \mathbf{x}^t \mathbf{b} \\ (a \in \mathbb{R},\ \mathbf{b} \in \mathbb{R}^k) \end{array}} \xrightarrow{Y = a + \mathbf{x}^t \mathbf{b} + E}$$

input
(vettore in $\mathbb{R}^k$)

output
(variabile aleatoria)

Se facciamo n misure:

$$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \quad \longrightarrow \quad Y_1, Y_2, \ldots, Y_n \quad \text{con} \quad Y_i = a + \mathbf{x}_i^t \mathbf{b} + E_i$$

Regressione lineare multipla

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}$$

input
(vettore in $\mathbb{R}^k$)

Legge fisica

$$y = a + b_1 x_1 + \ldots + b_k x_k$$
$$= a + \mathbf{x}^t \mathbf{b}$$
$$(a \in \mathbb{R},\ \mathbf{b} \in \mathbb{R}^k)$$

$$Y = a + \mathbf{x}^t \mathbf{b} + E$$

output
(variabile aleatoria)

Se facciamo n misure:

$$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \quad \longrightarrow \quad Y_1, Y_2, \ldots, Y_n \quad \text{con} \quad Y_i = a + \mathbf{x}_i^t \mathbf{b} + E_i$$

Ipotesi fondamentale del modello multilineare:

- $E_1, \ldots, E_n$ indipendenti $\quad \Leftrightarrow \quad$ $Y_1, \ldots, Y_n$ indipendenti
- $E_i \sim N(0, \sigma^2)\ \forall i \quad \Leftrightarrow \quad Y_i \sim N(a + \mathbf{x}_i^t \mathbf{b}\, ,\, \sigma^2)\ \forall i$

Regressione lineare multipla

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}$$

input
(vettore in $\mathbb{R}^k$)

Legge fisica

$$y = a + b_1 x_1 + \ldots + b_k x_k$$
$$= a + \mathbf{x}^t \mathbf{b}$$
$$(a \in \mathbb{R},\ \mathbf{b} \in \mathbb{R}^k)$$

$$Y = a + \mathbf{x}^t \mathbf{b} + E$$

output
(variabile aleatoria)

Se facciamo n misure:

$$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \quad \longrightarrow \quad Y_1, Y_2, \ldots, Y_n \quad \text{con} \quad Y_i = a + \mathbf{x}_i^t \mathbf{b} + E_i$$

Ipotesi fondamentale del modello multilineare:

- $E_1, \ldots, E_n$ indipendenti $\Leftrightarrow$ $Y_1, \ldots, Y_n$ indipendenti
- $E_i \sim N(0, \sigma^2)\ \ \forall i$ $\Leftrightarrow$ $Y_i \sim N(a + \mathbf{x}_i^t \mathbf{b}\, ,\, \sigma^2)\ \ \forall i$

Parametri incogniti: a, $\mathbf{b}$, σ^2

Iperpiano dei minimi quadrati

PROBLEMA: Valutare quanto bene l'iperpiano

$$y = a + b_1 x_1 + \ldots + b_k x_k = a + \mathbf{x}^t \mathbf{b}$$

interpola n punti $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ assegnati in $\mathbb{R}^{k+1}$

Iperpiano dei minimi quadrati

PROBLEMA: Valutare quanto bene l'iperpiano

$$y = a + b_1 x_1 + \ldots + b_k x_k = a + \mathbf{x}^t \mathbf{b}$$

interpola n punti $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ assegnati in $\mathbb{R}^{k+1}$

$e_i := y_i - (a + \mathbf{x}_i^t \mathbf{b})$ $\qquad$ *i*-esimo residuo

Iperpiano dei minimi quadrati

PROBLEMA: Valutare quanto bene l'iperpiano

$$y = a + b_1 x_1 + \ldots + b_k x_k = a + \mathbf{x}^t \mathbf{b}$$

interpola n punti $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ assegnati in $\mathbb{R}^{k+1}$

$$e_i := y_i - (a + \mathbf{x}_i^t \mathbf{b}) \qquad i\text{-esimo residuo}$$

$$L(a, \mathbf{b}) := \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left[y_i - (a + \mathbf{x}_i^t \mathbf{b}) \right]^2 \qquad \text{funzionale di errore}$$

Iperpiano dei minimi quadrati

PROBLEMA: Valutare quanto bene l'iperpiano

$$y = a + b_1 x_1 + \ldots + b_k x_k = a + \mathbf{x}^t \mathbf{b}$$

interpola n punti $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ assegnati in $\mathbb{R}^{k+1}$

$$e_i := y_i - (a + \mathbf{x}_i^t \mathbf{b}) \qquad i\text{-esimo residuo}$$

$$L(a, \mathbf{b}) := \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left[y_i - (a + \mathbf{x}_i^t \mathbf{b}) \right]^2 \qquad \text{funzionale di errore}$$

$$\tilde{\mathbf{x}} = \underbrace{\begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}}_{n \times (k+1)} \qquad \boldsymbol{\beta} = \underbrace{\begin{pmatrix} a \\ b_1 \\ \vdots \\ b_k \end{pmatrix}}_{(k+1) \times 1}$$

Iperpiano dei minimi quadrati

PROBLEMA: Valutare quanto bene l'iperpiano

$$y = a + b_1 x_1 + \ldots + b_k x_k = a + \mathbf{x}^t \mathbf{b}$$

interpola n punti $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ assegnati in $\mathbb{R}^{k+1}$

$$e_i := y_i - (a + \mathbf{x}_i^t \mathbf{b}) \qquad i\text{-esimo residuo}$$

$$L(a, \mathbf{b}) := \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left[y_i - (a + \mathbf{x}_i^t \mathbf{b}) \right]^2 \qquad \text{funzionale di errore}$$

$$\tilde{\mathbf{x}} = \underbrace{\begin{pmatrix} 1 & \mathbf{x}_1^t \\ 1 & \mathbf{x}_2^t \\ \cdots & \cdots \\ 1 & \mathbf{x}_n^t \end{pmatrix}}_{n \times (k+1)} \qquad \boldsymbol{\beta} = \underbrace{\begin{pmatrix} a \\ \\ \mathbf{b} \\ \\ \end{pmatrix}}_{(k+1) \times 1}$$

Iperpiano dei minimi quadrati

PROBLEMA: Valutare quanto bene l'iperpiano

$$y = a + b_1 x_1 + \ldots + b_k x_k = a + \mathbf{x}^t \mathbf{b}$$

interpola n punti $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ assegnati in $\mathbb{R}^{k+1}$

$$e_i \; := y_i - (a + \mathbf{x}_i^t \mathbf{b}) = y_i - (\tilde{\mathbf{x}}\boldsymbol{\beta})_i \qquad i\text{-esimo residuo}$$

$$L(a, \mathbf{b}) \; := \sum_{i=1}^{n} e_i^2 \; = \; \sum_{i=1}^{n} \left[y_i - (a + \mathbf{x}_i^t \mathbf{b}) \right]^2 \qquad \text{funzionale di errore}$$

$$= \sum_i (y_i - (\tilde{\mathbf{x}}\boldsymbol{\beta})_i)^2$$

$$\tilde{\mathbf{x}} = \underbrace{\begin{pmatrix} 1 & \mathbf{x}_1^t \\ 1 & \mathbf{x}_2^t \\ \cdots & \cdots \\ 1 & \mathbf{x}_n^t \end{pmatrix}}_{n \times (k+1)} \qquad \boldsymbol{\beta} = \underbrace{\begin{pmatrix} a \\ \\ \mathbf{b} \\ \\ \end{pmatrix}}_{(k+1) \times 1} =: \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

Iperpiano dei minimi quadrati

PROBLEMA: Valutare quanto bene l'iperpiano

$$y = a + b_1 x_1 + \ldots + b_k x_k = a + \mathbf{x}^t \mathbf{b}$$

interpola n punti $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ assegnati in $\mathbb{R}^{k+1}$

$$e_i := y_i - (a + \mathbf{x}_i^t \mathbf{b}) = y_i - (\tilde{\mathbf{x}}\boldsymbol{\beta})_i \qquad i\text{-esimo residuo}$$

$$\begin{aligned} L(a, \mathbf{b}) &:= \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left[y_i - (a + \mathbf{x}_i^t \mathbf{b})\right]^2 \qquad \text{funzionale di errore} \\ &= \sum_i (y_i - (\tilde{\mathbf{x}}\boldsymbol{\beta})_i)^2 = \|\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}\|^2 \end{aligned}$$

$$\tilde{\mathbf{x}} = \underbrace{\begin{pmatrix} 1 & \mathbf{x}_1^t \\ 1 & \mathbf{x}_2^t \\ \cdots & \cdots \\ 1 & \mathbf{x}_n^t \end{pmatrix}}_{n\times(k+1)} \qquad \boldsymbol{\beta} = \underbrace{\begin{pmatrix} a \\ \\ \mathbf{b} \\ \\ \end{pmatrix}}_{(k+1)\times 1} =: \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \qquad \mathbf{y} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{n\times 1}$$

Iperpiano dei minimi quadrati

Minimizziamo $L(\boldsymbol{\beta}) = \|\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}\|^2$ rispetto a $\boldsymbol{\beta}$:

$$\nabla_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) = ?????$$

Iperpiano dei minimi quadrati

Minimizziamo $L(\beta) = \|\mathbf{y} - \tilde{\mathbf{x}}\beta\|^2$ rispetto a β:

$$\nabla_\beta L(\beta) \; = \; -2\tilde{\mathbf{x}}^t(\mathbf{y} - \tilde{\mathbf{x}}\beta)$$

Iperpiano dei minimi quadrati

Minimizziamo $L(\boldsymbol{\beta}) = \|\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}\|^2$ rispetto a $\boldsymbol{\beta}$:

$$\nabla_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t(\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t\mathbf{y} + 2\tilde{\mathbf{x}}^t\tilde{\mathbf{x}}\boldsymbol{\beta}$$

Iperpiano dei minimi quadrati

Minimizziamo $L(\beta) = \|\mathbf{y} - \tilde{\mathbf{x}}\beta\|^2$ rispetto a β:

$$\nabla_{\beta} L(\beta) \; = \; -2\tilde{\mathbf{x}}^t(\mathbf{y} - \tilde{\mathbf{x}}\beta) \; = \; -2\tilde{\mathbf{x}}^t\mathbf{y} + 2\tilde{\mathbf{x}}^t\tilde{\mathbf{x}}\beta \; \equiv \; 0$$

Iperpiano dei minimi quadrati

Minimizziamo $L(\beta) = \|\mathbf{y} - \tilde{\mathbf{x}}\beta\|^2$ rispetto a β:

$$\nabla_\beta L(\beta) \;=\; -2\tilde{\mathbf{x}}^t(\mathbf{y} - \tilde{\mathbf{x}}\beta) \;=\; -2\tilde{\mathbf{x}}^t\mathbf{y} + 2\tilde{\mathbf{x}}^t\tilde{\mathbf{x}}\beta \;\equiv\; 0$$

$$\overset{\text{se } \tilde{\mathbf{x}}^t\tilde{\mathbf{x}} \text{ è invertibile}}{\Longrightarrow} \qquad \beta = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{y}$$

Iperpiano dei minimi quadrati

Minimizziamo $L(\beta) = \|\mathbf{y} - \tilde{\mathbf{x}}\beta\|^2$ rispetto a β:

$$\nabla_\beta L(\beta) \;=\; -2\tilde{\mathbf{x}}^t(\mathbf{y} - \tilde{\mathbf{x}}\beta) \;=\; -2\tilde{\mathbf{x}}^t\mathbf{y} + 2\tilde{\mathbf{x}}^t\tilde{\mathbf{x}}\beta \;\equiv\; 0$$

$$\xRightarrow{\text{se } \tilde{\mathbf{x}}^t\tilde{\mathbf{x}} \text{ è invertibile}} \qquad \beta = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{y}$$

P. es., con $k = 1$ predittore:

$$\tilde{\mathbf{x}}^t\tilde{\mathbf{x}} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & s_{xx} + n\bar{x}^2 \end{pmatrix} \quad \text{è invertibile} \quad \Leftrightarrow \quad s_{xx} \neq 0$$

Altrimenti: *collinearità*

Iperpiano dei minimi quadrati

Minimizziamo $L(\boldsymbol{\beta}) = \|\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}\|^2$ rispetto a $\boldsymbol{\beta}$:

$$\nabla_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t(\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t\mathbf{y} + 2\tilde{\mathbf{x}}^t\tilde{\mathbf{x}}\boldsymbol{\beta} \;\equiv\; 0$$

$$\overset{\text{se } \tilde{\mathbf{x}}^t\tilde{\mathbf{x}} \text{ è invertibile}}{\Longrightarrow} \qquad \boldsymbol{\beta} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{y} =: \hat{\boldsymbol{\beta}}$$

$$y \;=\; \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k \;=\; (1\ \mathbf{x}^t)\,\hat{\boldsymbol{\beta}}$$

iperpiano dei minimi quadrati (LSH)

Iperpiano dei minimi quadrati

Minimizziamo $L(\boldsymbol{\beta}) = \|\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}\|^2$ rispetto a $\boldsymbol{\beta}$:

$$\nabla_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t(\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t\mathbf{y} + 2\tilde{\mathbf{x}}^t\tilde{\mathbf{x}}\boldsymbol{\beta} \;\equiv\; 0$$

$$\xrightarrow{\text{se } \tilde{\mathbf{x}}^t\tilde{\mathbf{x}} \text{ è invertibile}} \qquad \boldsymbol{\beta} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{y} =: \hat{\boldsymbol{\beta}}$$

$$y \;=\; \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k \;=\; (1\ \mathbf{x}^t)\,\hat{\boldsymbol{\beta}}$$

iperpiano dei minimi quadrati (LSH)

Come nel caso semplice:

$$\hat{y}_i \;:=\; (1\ \mathbf{x}_i^t)\,\hat{\boldsymbol{\beta}}$$

output dell'LSH

Iperpiano dei minimi quadrati

Minimizziamo $L(\boldsymbol{\beta}) = \|\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}\|^2$ rispetto a $\boldsymbol{\beta}$:

$$\nabla_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t(\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t\mathbf{y} + 2\tilde{\mathbf{x}}^t\tilde{\mathbf{x}}\boldsymbol{\beta} \;\equiv\; 0$$

$$\xRightarrow{\text{se } \tilde{\mathbf{x}}^t\tilde{\mathbf{x}} \text{ è invertibile}} \qquad \boldsymbol{\beta} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{y} =: \hat{\boldsymbol{\beta}}$$

$$y \;=\; \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k \;=\; (1\ \mathbf{x}^t)\,\hat{\boldsymbol{\beta}}$$

iperpiano dei minimi quadrati (LSH)

Come nel caso semplice:

$$\hat{y}_i \;:=\; (1\ \mathbf{x}_i^t)\,\hat{\boldsymbol{\beta}} \qquad \text{output dell'LSH}$$

$$\hat{e}_i \;:=\; y_i - \hat{y}_i \qquad \text{residui dell'LSH}$$

Iperpiano dei minimi quadrati

Minimizziamo $L(\boldsymbol{\beta}) = \|\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}\|^2$ rispetto a $\boldsymbol{\beta}$:

$$\nabla_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t(\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t\mathbf{y} + 2\tilde{\mathbf{x}}^t\tilde{\mathbf{x}}\boldsymbol{\beta} \;\equiv\; 0$$

$$\stackrel{\text{se } \tilde{\mathbf{x}}^t\tilde{\mathbf{x}} \text{ è invertibile}}{\Longrightarrow} \qquad \boldsymbol{\beta} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{y} =: \hat{\boldsymbol{\beta}}$$

$$y \;=\; \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k \;=\; (1\ \mathbf{x}^t)\,\hat{\boldsymbol{\beta}}$$

iperpiano dei minimi quadrati (LSH)

Come nel caso semplice:

$\hat{y}_i \;:=\; (1\ \mathbf{x}_i^t)\,\hat{\boldsymbol{\beta}}$	output dell'LSH
$\hat{e}_i \;:=\; y_i - \hat{y}_i$	residui dell'LSH
$ss_r \;:=\; \sum(\bar{y} - \hat{y}_i)^2$	varianza spiegata

Iperpiano dei minimi quadrati

Minimizziamo $L(\boldsymbol{\beta}) = \|\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}\|^2$ rispetto a $\boldsymbol{\beta}$:

$$\nabla_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t(\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t\mathbf{y} + 2\tilde{\mathbf{x}}^t\tilde{\mathbf{x}}\boldsymbol{\beta} \;\equiv\; 0$$

$$\overset{\text{se } \tilde{\mathbf{x}}^t\tilde{\mathbf{x}} \text{ è invertibile}}{\Longrightarrow} \qquad \boldsymbol{\beta} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{y} =: \hat{\boldsymbol{\beta}}$$

$$y \;=\; \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k \;=\; (1\ \mathbf{x}^t)\,\hat{\boldsymbol{\beta}}$$

iperpiano dei minimi quadrati (LSH)

Come nel caso semplice:

$\hat{y}_i \;:=\; (1\ \mathbf{x}_i^t)\,\hat{\boldsymbol{\beta}}$	output dell'LSH
$\hat{e}_i \;:=\; y_i - \hat{y}_i$	residui dell'LSH
$ss_r \;:=\; \sum(\bar{y} - \hat{y}_i)^2$	varianza spiegata
$ss_e \;:=\; \sum(y_i - \hat{y}_i)^2$	varianza residua

Iperpiano dei minimi quadrati

Minimizziamo $L(\boldsymbol{\beta}) = \|\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}\|^2$ rispetto a $\boldsymbol{\beta}$:

$$\nabla_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t(\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t\mathbf{y} + 2\tilde{\mathbf{x}}^t\tilde{\mathbf{x}}\boldsymbol{\beta} \;\equiv\; 0$$

$$\overset{\text{se } \tilde{\mathbf{x}}^t\tilde{\mathbf{x}} \text{ è invertibile}}{\Longrightarrow} \qquad \boldsymbol{\beta} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{y} =: \hat{\boldsymbol{\beta}}$$

$$y \;=\; \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k \;=\; (1\ \mathbf{x}^t)\,\hat{\boldsymbol{\beta}}$$

iperpiano dei minimi quadrati (LSH)

Come nel caso semplice:

$\hat{y}_i \;:=\; (1\ \mathbf{x}_i^t)\,\hat{\boldsymbol{\beta}}$	output dell'LSH
$\hat{e}_i \;:=\; y_i - \hat{y}_i$	residui dell'LSH
$ss_r \;:=\; \sum(\bar{y} - \hat{y}_i)^2$	varianza spiegata
$ss_e \;:=\; \sum(y_i - \hat{y}_i)^2$	varianza residua
$ss_t \;:=\; \sum(y_i - \bar{y})^2$	varianza totale

Iperpiano dei minimi quadrati

Minimizziamo $L(\boldsymbol{\beta}) = \|\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}\|^2$ rispetto a $\boldsymbol{\beta}$:

$$\nabla_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t(\mathbf{y} - \tilde{\mathbf{x}}\boldsymbol{\beta}) \;=\; -2\tilde{\mathbf{x}}^t\mathbf{y} + 2\tilde{\mathbf{x}}^t\tilde{\mathbf{x}}\boldsymbol{\beta} \;\equiv\; 0$$

$$\xRightarrow{\text{se } \tilde{\mathbf{x}}^t\tilde{\mathbf{x}} \text{ è invertibile}} \qquad \boldsymbol{\beta} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{y} =: \hat{\boldsymbol{\beta}}$$

$$y \;=\; \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k \;=\; (1\ \mathbf{x}^t)\,\hat{\boldsymbol{\beta}}$$

iperpiano dei minimi quadrati (LSH)

Come nel caso semplice:

$\hat{y}_i \;:=\; (1\ \mathbf{x}_i^t)\,\hat{\boldsymbol{\beta}}$	output dell'LSH
$\hat{e}_i \;:=\; y_i - \hat{y}_i$	residui dell'LSH
$ss_r \;:=\; \sum(\bar{y} - \hat{y}_i)^2$	varianza spiegata
$ss_e \;:=\; \sum(y_i - \hat{y}_i)^2$	varianza residua
$ss_t \;:=\; \sum(y_i - \bar{y})^2 \;\equiv\; ss_r + ss_e$	varianza totale

Iperpiano dei minimi quadrati

OVERFITTING = ridurre ss_e aumentando i regressori

Iperpiano dei minimi quadrati

OVERFITTING = ridurre ss_e aumentando i regressori:

$$ss_e = L(\hat{\beta}) = \min_{\boldsymbol{\beta}} L(\beta)$$

Iperpiano dei minimi quadrati

OVERFITTING = ridurre ss_e aumentando i regressori:

$$ss_e \;=\; L(\hat{\beta}) \;=\; \min_{\boldsymbol{\beta}} L(\boldsymbol{\beta})$$

$$L_k(\boldsymbol{\beta}) \;=\; \sum_i (y_i - \beta_0 - \beta_1 x_{i,1} - \ldots - \beta_k x_{i,k})^2 \qquad \boldsymbol{\beta} \in \mathbb{R}^{k+1}$$

$$L_{k+1}(\boldsymbol{\beta}') \;=\; \sum_i (y_i - \beta'_0 - \beta'_1 x_{i,1} - \ldots - \beta'_{k+1} x_{i,k+1})^2 \qquad \boldsymbol{\beta}' \in \mathbb{R}^{k+2}$$

Iperpiano dei minimi quadrati

OVERFITTING = ridurre ss_e aumentando i regressori:

$$ss_e = L(\hat{\boldsymbol{\beta}}) = \min_{\boldsymbol{\beta}} L(\boldsymbol{\beta})$$

$$L_k(\boldsymbol{\beta}) = \sum_i (y_i - \beta_0 - \beta_1 x_{i,1} - \ldots - \beta_k x_{i,k})^2 \qquad \boldsymbol{\beta} \in \mathbb{R}^{k+1}$$

$$L_{k+1}(\boldsymbol{\beta}') = \sum_i (y_i - \beta'_0 - \beta'_1 x_{i,1} - \ldots - \beta'_{k+1} x_{i,k+1})^2 \qquad \boldsymbol{\beta}' \in \mathbb{R}^{k+2}$$

$$L_k(\boldsymbol{\beta}) \equiv L_{k+1}\begin{pmatrix} \boldsymbol{\beta} \\ 0 \end{pmatrix}$$

Iperpiano dei minimi quadrati

OVERFITTING = ridurre ss_e aumentando i regressori:

$$ss_e = L(\hat{\boldsymbol{\beta}}) = \min_{\boldsymbol{\beta}} L(\boldsymbol{\beta})$$

$$L_k(\boldsymbol{\beta}) = \sum_i (y_i - \beta_0 - \beta_1 x_{i,1} - \ldots - \beta_k x_{i,k})^2 \qquad \boldsymbol{\beta} \in \mathbb{R}^{k+1}$$

$$L_{k+1}(\boldsymbol{\beta}') = \sum_i (y_i - \beta'_0 - \beta'_1 x_{i,1} - \ldots - \beta'_{k+1} x_{i,k+1})^2 \qquad \boldsymbol{\beta}' \in \mathbb{R}^{k+2}$$

$$L_k(\boldsymbol{\beta}) \equiv L_{k+1}\begin{pmatrix} \boldsymbol{\beta} \\ 0 \end{pmatrix} \quad \Rightarrow \quad \min_{\boldsymbol{\beta}} L_k(\boldsymbol{\beta}) \geq \min_{\boldsymbol{\beta}'} L_{k+1}(\boldsymbol{\beta}')$$

Iperpiano dei minimi quadrati

OVERFITTING = ridurre ss_e aumentando i regressori:

$$ss_e = L(\hat{\boldsymbol{\beta}}) = \min_{\boldsymbol{\beta}} L(\boldsymbol{\beta})$$

$$L_k(\boldsymbol{\beta}) = \sum_i (y_i - \beta_0 - \beta_1 x_{i,1} - \ldots - \beta_k x_{i,k})^2 \qquad \boldsymbol{\beta} \in \mathbb{R}^{k+1}$$

$$L_{k+1}(\boldsymbol{\beta}') = \sum_i (y_i - \beta'_0 - \beta'_1 x_{i,1} - \ldots - \beta'_{k+1} x_{i,k+1})^2 \qquad \boldsymbol{\beta}' \in \mathbb{R}^{k+2}$$

$$L_k(\boldsymbol{\beta}) \equiv L_{k+1}\begin{pmatrix} \boldsymbol{\beta} \\ 0 \end{pmatrix} \quad \Rightarrow \quad \min_{\boldsymbol{\beta}} L_k(\boldsymbol{\beta}) \geq \min_{\boldsymbol{\beta}'} L_{k+1}(\boldsymbol{\beta}')$$

Perciò si preferisce usare l'*r^2-adjusted*

$$r_A^2 := 1 - \frac{ss_e}{ss_t} \frac{n-1}{n-1-k}$$

Iperpiano dei minimi quadrati

OVERFITTING = ridurre ss_e aumentando i regressori:

$$ss_e = L(\hat{\boldsymbol{\beta}}) = \min_{\boldsymbol{\beta}} L(\boldsymbol{\beta})$$

$$L_k(\boldsymbol{\beta}) = \sum_i (y_i - \beta_0 - \beta_1 x_{i,1} - \ldots - \beta_k x_{i,k})^2 \qquad \boldsymbol{\beta} \in \mathbb{R}^{k+1}$$

$$L_{k+1}(\boldsymbol{\beta}') = \sum_i (y_i - \beta_0' - \beta_1' x_{i,1} - \ldots - \beta_{k+1}' x_{i,k+1})^2 \qquad \boldsymbol{\beta}' \in \mathbb{R}^{k+2}$$

$$L_k(\boldsymbol{\beta}) \equiv L_{k+1}\begin{pmatrix} \boldsymbol{\beta} \\ 0 \end{pmatrix} \quad \Rightarrow \quad \min_{\boldsymbol{\beta}} L_k(\boldsymbol{\beta}) \geq \min_{\boldsymbol{\beta}'} L_{k+1}(\boldsymbol{\beta}')$$

Perciò si preferisce usare l'r^2-*adjusted*

$$r_A^2 := 1 - \frac{ss_e}{ss_t}\,\frac{n-1}{n-1-k} \begin{cases} \leq r^2 \\ \text{decrescente in } k \end{cases}$$

Stimatori, IC e test per i parametri

IPOTESI: $\begin{cases} Y_1, \ldots, Y_n \quad \text{indipendenti} \\ Y_i \sim N((1\ \mathbf{x}_i^t)\,\boldsymbol{\beta}\,,\,\sigma^2) \quad \forall i \end{cases}$ $\qquad \boldsymbol{\beta}$, σ^2 parametri incogniti

Stimatori, IC e test per i parametri

IPOTESI: $\begin{cases} Y_1, \ldots, Y_n \quad \text{indipendenti} \\ Y_i \sim N((1\ \mathbf{x}_i^t)\,\boldsymbol{\beta}\,,\,\sigma^2) \quad \forall i \end{cases}$ $\qquad \boldsymbol{\beta}$, σ^2 parametri incogniti

CONSEGUENZE:

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{y} \quad \underset{\text{vettore di v.a.}}{\overset{\text{diventa il}}{\longrightarrow}} \quad \hat{\boldsymbol{B}} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{Y} \quad \text{con} \quad \mathbf{Y} = (Y_1, \ldots, Y_n)^t$$

Stimatori, IC e test per i parametri

IPOTESI: $\begin{cases} Y_1, \ldots, Y_n \quad \text{indipendenti} \\ Y_i \sim N((1\ \mathbf{x}_i^t)\,\boldsymbol{\beta}\,,\ \sigma^2) \quad \forall i \end{cases}$ $\boldsymbol{\beta}$, σ^2 parametri incogniti

CONSEGUENZE:

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{y} \quad \xrightarrow[\text{vettore di v.a.}]{\text{diventa il}} \quad \hat{\boldsymbol{B}} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{Y} \quad \text{con} \quad \mathbf{Y} = (Y_1, \ldots, Y_n)^t$$

$$\hat{B}_r \sim N\left(\beta_r\,,\ \sigma^2\,[(\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}]_{rr}\right) \quad \Rightarrow \quad \hat{B}_r \ \text{stimatore corretto di } \beta_r$$

IPOTESI: $\begin{cases} Y_1, \ldots, Y_n & \text{indipendenti} \\ Y_i \sim N((1\ \mathbf{x}_i^t)\,\boldsymbol{\beta}\,,\,\sigma^2) & \forall i \end{cases}$ $\boldsymbol{\beta}$, σ^2 parametri incogniti

CONSEGUENZE:

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{y} \quad \underset{\text{vettore di v.a.}}{\overset{\text{diventa il}}{\longrightarrow}} \quad \hat{\boldsymbol{B}} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{Y} \quad \text{con} \quad \mathbf{Y} = (Y_1, \ldots, Y_n)^t$$

$$\hat{B}_r \sim N\left(\beta_r\,,\,\sigma^2\,[(\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}]_{rr}\right) \quad \Rightarrow \quad \hat{B}_r \text{ stimatore corretto di } \beta_r$$

$$\frac{SS_E}{\sigma^2} \sim \chi^2(n-k-1) \quad \text{indipendente da } \hat{B}_r$$

$$\Rightarrow \hat{\Sigma}^2 = \frac{SS_E}{n-k-1} \qquad \text{stimatore corretto di } \sigma^2$$

Stimatori, IC e test per i parametri

IPOTESI: $\begin{cases} Y_1, \ldots, Y_n \quad \text{indipendenti} \\ Y_i \sim N((1 \ \mathbf{x}_i^t)\,\boldsymbol{\beta}\,,\, \sigma^2) \quad \forall i \end{cases}$ $\qquad \boldsymbol{\beta}$, σ^2 parametri incogniti

CONSEGUENZE:

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{y} \quad \underset{\text{vettore di v.a.}}{\overset{\text{diventa il}}{\longrightarrow}} \quad \hat{\boldsymbol{B}} = (\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\mathbf{Y} \quad \text{con} \quad \mathbf{Y} = (Y_1, \ldots, Y_n)^t$$

$$\hat{B}_r \sim N\left(\beta_r\,,\, \sigma^2\,[(\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}]_{rr}\right) \quad \Rightarrow \quad \hat{B}_r \text{ stimatore corretto di } \beta_r$$

$$\frac{SS_E}{\sigma^2} \sim \chi^2(n-k-1) \quad \text{indipendente da } \hat{B}_r$$

$$\Rightarrow \ \hat{\Sigma}^2 = \frac{SS_E}{n-k-1} \qquad \text{stimatore corretto di } \sigma^2$$

$$\Rightarrow \ \mathrm{se}(\hat{B}_r) = \sqrt{\hat{\Sigma}^2[(\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}]_{rr}} \quad \text{stimatore approx. corretto di } \sqrt{\mathrm{Var}[\hat{B}_r]}$$

Stimatori, IC e test per i parametri

$$\frac{\hat{B}_r - \beta_r}{\mathrm{se}(\hat{B}_r)} \sim t(n-k-1)$$

Stimatori, IC e test per i parametri

$$\frac{\hat{B}_r - \beta_r}{\mathrm{se}(\hat{B}_r)} \sim t(n-k-1)$$

$$\swarrow$$

$$\left(\hat{\beta}_r \pm t_{\frac{1+\gamma}{2}}(n-k-1)\,\mathrm{se}(\hat{\beta}_r)\right)$$

è un $IC(\gamma)$ per β_r

Stimatori, IC e test per i parametri

$$\frac{\hat{B}_r - \beta_r}{\mathrm{se}(\hat{B}_r)} \sim t(n-k-1)$$

$$\swarrow \qquad\qquad\qquad \searrow$$

$$\left(\hat{\beta}_r \pm t_{\frac{1+\gamma}{2}}(n-k-1)\,\mathrm{se}(\hat{\beta}_r)\right)$$

è un $IC(\gamma)$ per β_r

"Rifiuto H_0 se

$$\left|\frac{\hat{B}_r - \beta_{r0}}{\mathrm{se}(\hat{B}_r)}\right| > t_{1-\frac{\alpha}{2}}(n-k-1)\text{"}$$

è un test di livello α per le ipotesi

$H_0 : \beta_r = \beta_{r0}$ vs. $H_1 : \beta_r \neq \beta_{r0}$

Eliminazione a ritroso dei predittori non significativi

$$\boxed{k = k_0, \quad \text{predittori } x_1, \ldots, x_{k_0}}$$

Eliminazione a ritroso dei predittori non significativi

$k = k_0, \quad$ predittori $x_1, \ldots, x_{k_0}$

$\downarrow$

Con R, ricavare l'output della regressione

Eliminazione a ritroso dei predittori non significativi

$k = k_0$, predittori $x_1, \ldots, x_{k_0}$

↓

Con R, ricavare l'output della regressione

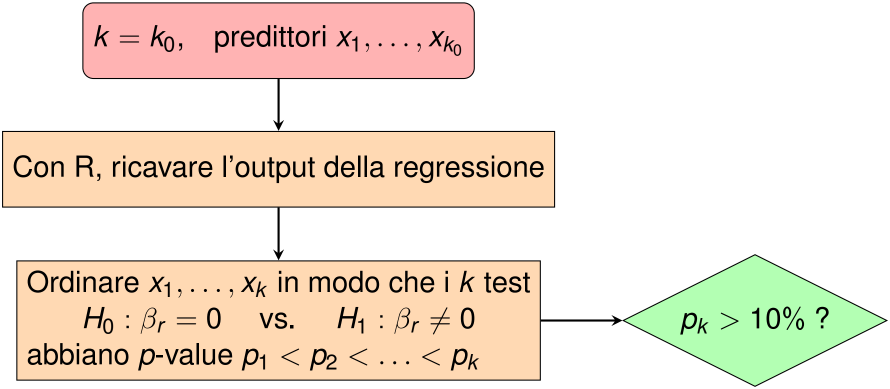
↓

Ordinare $x_1, \ldots, x_k$ in modo che i k test

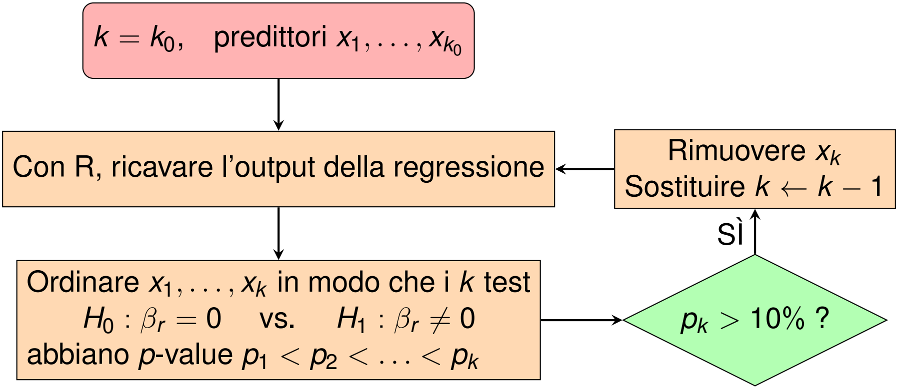
$$H_0 : \beta_r = 0 \quad \text{vs.} \quad H_1 : \beta_r \neq 0$$

abbiano p-value $p_1 < p_2 < \ldots < p_k$

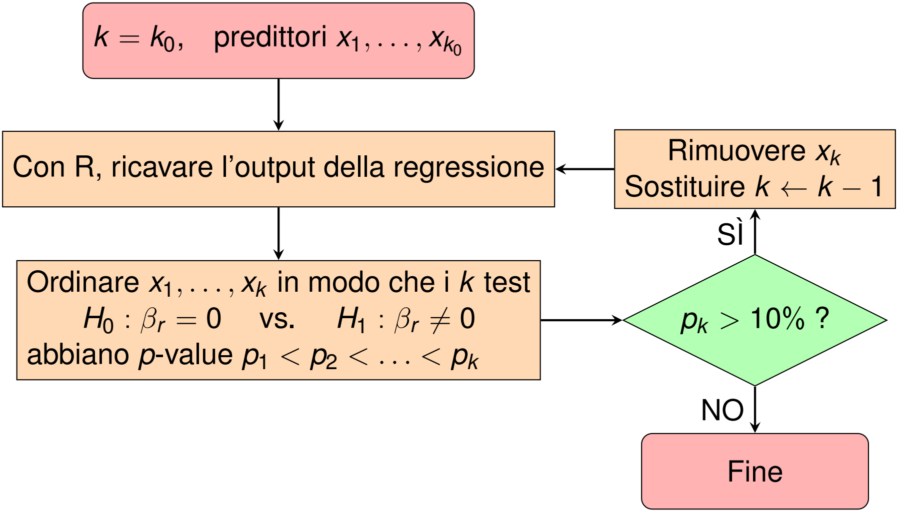
Eliminazione a ritroso dei predittori non significativi



Eliminazione a ritroso dei predittori non significativi



Eliminazione a ritroso dei predittori non significativi



Output della regressione in R

```
Call:
lm(formula = Distance ~ Temperature + Fuel + I(Temperature^2),
    data = D)

Residuals:
    Min      1Q  Median      3Q     Max
-32.078  -8.482  -0.377   9.223  34.466

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.743e+02  7.902e+01  -2.206   0.0308 *
Temperature       2.321e+01  5.066e+00   4.581 2.03e-05 ***
Fuel             -2.981e-03  7.689e-02  -0.039   0.9692
I(Temperature^2) -4.356e-01  8.139e-02  -5.352 1.10e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.27 on 68 degrees of freedom
Multiple R-squared:  0.631,     Adjusted R-squared:  0.6147
F-statistic: 38.76 on 3 and 68 DF,  p-value: 1.012e-14
```

p-value troppo alto!
Va eliminato

r_A^2 basso
(poca variabilità spiegata)

Test per la bontà del modello multilineare

Ipotesi del modello multilineare:

(1) $Y_1, \ldots, Y_n$ indipendenti

(2) $Y_i \sim N((1\ \mathbf{x}_i^t)\,\boldsymbol{\beta}\,,\,\sigma^2)$

Test per la bontà del modello multilineare

Ipotesi del modello multilineare:

(1) $Y_1, \dots, Y_n$ indipendenti

(2) $Y_i \sim N((1\ \mathbf{x}_i^t)\,\boldsymbol{\beta}\,,\,\sigma^2)$

Supponendo vera (1), testiamo (2):

$$H_0 :\ \exists\,\boldsymbol{\beta}, \sigma \quad \text{t. c.} \quad Y_i \sim N((1\ \mathbf{x}_i^t)\,\boldsymbol{\beta}\,,\,\sigma^2) \quad \forall i = 1, \dots, n$$

$$H_1 :\ H_0 \text{ è falsa}$$

Test per la bontà del modello multilineare

Ipotesi del modello multilineare:

(1) $Y_1, \ldots, Y_n$ indipendenti

(2) $Y_i \sim N((1\ \mathbf{x}_i^t)\,\boldsymbol{\beta}\,,\,\sigma^2)$

Supponendo vera (1), testiamo (2):

$$H_0 : \ \exists\, \boldsymbol{\beta}, \sigma \quad \text{t. c.} \quad Y_i \sim N((1\ \mathbf{x}_i^t)\,\boldsymbol{\beta}\,,\,\sigma^2) \quad \forall i = 1, \ldots, n$$

$$H_1 : \ H_0 \text{ è falsa}$$

Se (1) - (2) sono vere, allora i *residui studentizzati*

$$R_i' := \frac{Y_i - \hat{Y}_i}{\hat{\Sigma}\sqrt{1 - h_{ii}}} = \frac{R_i}{\sqrt{1 - h_{ii}}} \quad \text{con} \quad h_{ii} = \left[\tilde{\mathbf{x}}(\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\right]_{ii}$$

- sono (approssimativamente) i.i.d.
- $R_i' \approx N(0, 1)$

Test per la bontà del modello multilineare

Ipotesi del modello multilineare:

(1) $Y_1, \dots, Y_n$ indipendenti

(2) $Y_i \sim N((1\ \mathbf{x}_i^t)\,\boldsymbol{\beta}\,,\,\sigma^2)$

Supponendo vera (1), testiamo (2):

$$H_0 : \ \exists\, \boldsymbol{\beta}, \sigma \quad \text{t. c.} \quad Y_i \sim N((1\ \mathbf{x}_i^t)\,\boldsymbol{\beta}\,,\,\sigma^2) \quad \forall i = 1, \dots, n$$

$$H_1 : \ H_0 \text{ è falsa}$$

Se (1) - (2) sono vere, allora i *residui studentizzati*

$$R_i' := \frac{Y_i - \hat{Y}_i}{\hat{\Sigma}\sqrt{1 - h_{ii}}} = \frac{R_i}{\sqrt{1 - h_{ii}}} \quad \text{con} \quad h_{ii} = \left[\tilde{\mathbf{x}}(\tilde{\mathbf{x}}^t\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^t\right]_{ii}$$

- sono (approssimativamente) i.i.d.
- $R_i' \approx N(0, 1)$

$\Rightarrow$ test di normalità per $R_1', \dots, R_n'$

Stima e IP per una nuova osservazione

$$\mathbf{x}^* = \begin{pmatrix} x_1^* \\ \vdots \\ x_k^* \end{pmatrix}$$

nuovo input

Legge fisica

$$y = \beta_0 + \beta_1 x_1^* + \ldots + \beta_k x_k^*$$
$$= (1\ \mathbf{x}^{*t})\,\boldsymbol{\beta}$$

$$Y^* = (1\ \mathbf{x}^{*\,t})\,\boldsymbol{\beta} + E^*$$

nuovo output

Stima e IP per una nuova osservazione

$$\mathbf{x}^* = \begin{pmatrix} x_1^* \\ \vdots \\ x_k^* \end{pmatrix}$$

nuovo input

Legge fisica

$$y = \beta_0 + \beta_1 x_1^* + \ldots + \beta_k x_k^* = (1\ \mathbf{x}^{*t})\,\boldsymbol{\beta}$$

$$Y^* = (1\ \mathbf{x}^{*\,t})\,\boldsymbol{\beta} + E^*$$

nuovo output

IPOTESI: $E^* \sim N(0, \sigma^2)$ e indipendente da $E_1, \ldots, E_n$

$\Rightarrow$ $Y^* \sim N(\beta_0 + \beta_1 x_1^* + \ldots + \beta_k x_k^*\,,\,\sigma^2)$ e indipendente da $Y_1, \ldots, Y_n$

Stima e IP per una nuova osservazione

$$\mathbf{x}^* = \begin{pmatrix} x_1^* \\ \vdots \\ x_k^* \end{pmatrix}$$

nuovo input

Legge fisica

$$y = \beta_0 + \beta_1 x_1^* + \ldots + \beta_k x_k^* = (1\ \mathbf{x}^{*t})\,\boldsymbol{\beta}$$

$$Y^* = (1\ \mathbf{x}^{*\,t})\,\boldsymbol{\beta} + E^*$$

nuovo output

IPOTESI: $E^* \sim N(0, \sigma^2)$ e indipendente da $E_1, \ldots, E_n$

$\Rightarrow\ Y^* \sim N(\beta_0 + \beta_1 x_1^* + \ldots + \beta_k x_k^*\,,\ \sigma^2)$ e indipendente da $Y_1, \ldots, Y_n$

$\hat{Y}^* := \hat{B}_0 + \hat{B}_1 x_1^* + \ldots + \hat{B}_k x_k^*$ stimatore non distorto del parametro

$\mathbb{E}\,[Y^*] = \beta_0 + \beta_1 x_1^* + \ldots + \beta_k x_k^*$

Stima e IP per una nuova osservazione

$$\mathbf{x}^* = \begin{pmatrix} x_1^* \\ \vdots \\ x_k^* \end{pmatrix}$$

nuovo input

Legge fisica

$$y = \beta_0 + \beta_1 x_1^* + \ldots + \beta_k x_k^* = (1\ \mathbf{x}^{*t})\,\boldsymbol{\beta}$$

$$Y^* = (1\ \mathbf{x}^{*\,t})\,\boldsymbol{\beta} + E^*$$

nuovo output

IPOTESI: $E^* \sim N(0, \sigma^2)$ e indipendente da $E_1, \ldots, E_n$

$\Rightarrow\ Y^* \sim N(\beta_0 + \beta_1 x_1^* + \ldots + \beta_k x_k^*\,,\ \sigma^2)$ e indipendente da $Y_1, \ldots, Y_n$

$\hat{Y}^* := \hat{B}_0 + \hat{B}_1 x_1^* + \ldots + \hat{B}_k x_k^*$ stimatore non distorto del parametro

$$\mathbb{E}\,[Y^*] = \beta_0 + \beta_1 x_1^* + \ldots + \beta_k x_k^*$$

$IC_{\mathbb{E}[Y^*]}$ e IP_{Y^*} $\longleftarrow$ comando `predict()` di R

Test per la significatività di un gruppo di predittori

Per vedere se il gruppo dei primi r predittori è significativo, si testano

$$H_0 : \underbrace{\beta_1 = \beta_2 = \ldots = \beta_r = 0}_{\text{tutte contemporaneamente}} \qquad H_1 : \underbrace{\beta_s \neq 0 \text{ per qualche } s = 1, \ldots, r}_{\text{almeno una } \neq 0}$$

Test per la significatività di un gruppo di predittori

Per vedere se il gruppo dei primi r predittori è significativo, si testano

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_r = 0 \qquad H_1 : \beta_s \neq 0 \text{ per qualche } s = 1, \ldots, r$$

Un test di livello α è dato dalla regola

$$\text{“rifiuto } H_0 \text{ se } \quad \frac{[SS_E(\text{ridotto}) - SS_E(\text{completo})]/r}{SS_E(\text{completo})/(n-k-1)} > f_{1-\alpha}(r, n-k-1)\text{”}$$

dove

$SS_E(\text{completo}) :=$ varianza residua del modello con tutti i k predittori

$SS_E(\text{ridotto}) :=$ varianza residua del modello senza i primi r predittori

Test per la significatività di un gruppo di predittori

Per vedere se il gruppo dei primi r predittori è significativo, si testano

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_r = 0 \qquad H_1 : \beta_s \neq 0 \text{ per qualche } s = 1, \ldots, r$$

Un test di livello α è dato dalla regola

"rifiuto H_0 se $\dfrac{[SS_E(\mathrm{ridotto}) - SS_E(\mathrm{completo})]/r}{SS_E(\mathrm{completo})/(n-k-1)} > f_{1-\alpha}(r, n-k-1)$"

dove

$SS_E(\mathrm{completo}) :=$ varianza residua del modello con tutti i k predittori

$SS_E(\mathrm{ridotto}) :=$ varianza residua del modello senza i primi r predittori

Se $k = r$, otteniamo un test per la significatività dell'intero modello:

"rifiuto H_0 se $\dfrac{[SS_R(\mathrm{completo})]/k}{SS_E(\mathrm{completo})/(n-k-1)} > f_{1-\alpha}(k, n-k-1)$"

perchè $SS_E(\mathrm{ridotto}) = SS_T(\mathrm{completo})$ in questo caso

Output della regressione in R

```
Call:
lm(formula = Distance ~ Temperature + Fuel + I(Temperature^2),
    data = D)

Residuals:
    Min      1Q  Median      3Q     Max
-32.078  -8.482  -0.377   9.223  34.466

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.743e+02  7.902e+01  -2.206   0.0308 *
Temperature       2.321e+01  5.066e+00   4.581 2.03e-05 ***
Fuel             -2.981e-03  7.689e-02  -0.039   0.9692
I(Temperature^2) -4.356e-01  8.139e-02  -5.352 1.10e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.27 on 68 degrees of freedom
Multiple R-squared:  0.631,     Adjusted R-squared:  0.6147
F-statistic: 38.76 on 3 and 68 DF,  p-value: 1.012e-14
```

k

$n - k - 1$

$$\frac{[SS_R(\text{completo})]/k}{SS_E(\text{completo})/(n-k-1)}$$

p-value del test

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$

$H_1 : \beta_s \neq 0$ per qualche $s = 1, \ldots, k$

Generalizzazioni

Caso non lineare	Sostituzione	Linearizzazione
$y = a + bf(x)$	$x' = f(x)$	$y = a + bx'$

Generalizzazioni

Caso non lineare	Sostituzione	Linearizzazione
$y = a + bf(x)$	$x' = f(x)$	$y = a + bx'$
$y = a + b_1 f(x_1) + b_2 g(x_2)$	$x_1' = f(x_1)$ $x_2' = g(x_2)$	$y = a + b_1 x_1' + b_2 x_2'$

Generalizzazioni

Caso non lineare	Sostituzione	Linearizzazione
$y = a + bf(x)$	$x' = f(x)$	$y = a + bx'$
$y = a + b_1 f(x_1) + b_2 g(x_2)$	$x'_1 = f(x_1)$ $x'_2 = g(x_2)$	$y = a + b_1 x'_1 + b_2 x'_2$
$y = a + b_1 x_1^2 + b_2 x_2$	$x'_1 = x_1^2$	$y = a + b_1 x'_1 + b_2 x_2$

Generalizzazioni

Caso non lineare	Sostituzione	Linearizzazione
$y = a + bf(x)$	$x' = f(x)$	$y = a + bx'$
$y = a + b_1 f(x_1) + b_2 g(x_2)$	$x'_1 = f(x_1)$ $x'_2 = g(x_2)$	$y = a + b_1 x'_1 + b_2 x'_2$
$y = a + b_1 x_1^2 + b_2 x_2$	$x'_1 = x_1^2$	$y = a + b_1 x'_1 + b_2 x_2$
$y = a + b_1 x_1 + b_2 x_1 x_2$	$x'_2 = x_1 x_2$	$y = a + b_1 x_1 + b_2 x'_2$

$x'_2 = x_1 x_2$ *termine di interazione*

Generalizzazioni

Caso non lineare	Sostituzione	Linearizzazione
$y = a + bf(x)$	$x' = f(x)$	$y = a + bx'$
$y = a + b_1 f(x_1) + b_2 g(x_2)$	$x'_1 = f(x_1)$ $x'_2 = g(x_2)$	$y = a + b_1 x'_1 + b_2 x'_2$
$y = a + b_1 x_1^2 + b_2 x_2$	$x'_1 = x_1^2$	$y = a + b_1 x'_1 + b_2 x_2$
$y = a + b_1 x_1 + b_2 x_1 x_2$	$x'_2 = x_1 x_2$	$y = a + b_1 x_1 + b_2 x'_2$

A noi interessa solo fare inferenza su a, b, b_1, b_2

$\Rightarrow$ Ci basta che $\tilde{\mathbf{x}}^t \tilde{\mathbf{x}}$ sia invertibile!

Generalizzazioni

$x_k \in \{c_1, c_2, \ldots, c_l\}$ predittore categorico

(p.es., $x_k \in \{\text{rosso}, \text{verde}, \text{blu}\}$)

Generalizzazioni

$$x_k \in \{c_1, c_2, \ldots, c_l\} \quad \text{predittore categorico}$$

Definiamo $l-1$ nuovi predittori $z_k,\, z_{k+1}, \ldots,\, z_{k+l-2}$ con $z_r \in \{0,1\}$:

z_k	z_{k+1}	z_{k+2}	$\ldots$	x_k
0	0	0	$\ldots$	c_1
1	0	0	$\ldots$	c_2
0	1	0	$\ldots$	c_3
0	0	1	$\ldots$	c_4
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

e facciamo la regressione per i $k+l-2$ predittori numerici

$$x_1, \ldots,\, x_{k-1},\, z_k, \ldots,\, z_{k+l-2}$$

Generalizzazioni

$$x_k \in \{c_1, c_2, \ldots, c_l\} \quad \text{predittore categorico}$$

Definiamo $l-1$ nuovi predittori $z_k,\ z_{k+1}, \ldots,\ z_{k+l-2}$ con $z_r \in \{0,1\}$:

z_k	z_{k+1}	z_{k+2}	$\ldots$	x_k
0	0	0	$\ldots$	c_1
1	0	0	$\ldots$	c_2
0	1	0	$\ldots$	c_3
0	0	1	$\ldots$	c_4
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

e facciamo la regressione per i $k+l-2$ predittori numerici

$$x_1, \ldots,\ x_{k-1},\ z_k, \ldots,\ z_{k+l-2}$$

Facciamo così perché vogliamo che $\tilde{\mathbf{x}}^t\tilde{\mathbf{x}}$ sia invertibile.

Generalizzazioni

ESEMPIO: Due modelli diversi:

$$y = \begin{cases} a + bx & \text{se la caratteristica } z = 0 \\ c + dx & \text{se la caratteristica } z = 1 \end{cases}$$

Generalizzazioni

ESEMPIO: Due modelli diversi:

$$y = \begin{cases} a + bx & \text{se la caratteristica } z = 0 \\ c + dx & \text{se la caratteristica } z = 1 \end{cases}$$

$$\Rightarrow \quad y = \; a \; + \; b \quad x \; + (c - a) \;\; z \;\; + (d - b) \;\; xz$$

Generalizzazioni

ESEMPIO: Due modelli diversi:

$$y = \begin{cases} a + bx & \text{se la caratteristica } z = 0 \\ c + dx & \text{se la caratteristica } z = 1 \end{cases}$$

$$\Rightarrow \quad y = \underbrace{a}_{\beta_0} + \underbrace{b}_{\beta_1}\underbrace{x}_{x_1} + \underbrace{(c-a)}_{\beta_2}\underbrace{z}_{x_2} + \underbrace{(d-b)}_{\beta_3}\underbrace{xz}_{x_3}$$

$$\Rightarrow \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Un esempio

Il famoso esploratore e statistico Jack Bettazzi, recatosi in Groenlandia per condurre la sua ricerca sulle specie locali di foche, dopo mesi di rilievi e osservazioni riesce a misurare i seguenti caratteri di $n = 30$ esemplari: il numero di `anni` di vita dell'esemplare; il suo `peso` in Kg; e infine la sua `specie`, che può essere bianca, marrone o nera. Decide di impostare questo modello linare per spiegare la variabile `peso` tramite l'età e la specie:

$$\begin{aligned}\texttt{peso}_i &= \beta_0 + \beta_1 \texttt{anni}_i + \beta_2 \texttt{speciemarrone}_i + \beta_3 \texttt{specienera}_i \\ &\quad + \beta_4 \texttt{anni}_i \cdot \texttt{speciemarrone}_i + \beta_5 \texttt{anni}_i \cdot \texttt{specienera}_i + \varepsilon_i\end{aligned}$$

con $\varepsilon_1, \ldots, \varepsilon_{30}$ i.i.d. e $\varepsilon_i \sim N(0, \sigma^2)$. Nel modello, il regressore categorico `specie` è codificato come segue:

$$(\texttt{speciemarrone}, \texttt{specienera}) = \begin{cases} (0,\, 0) & \text{se la specie è bianca} \\ (1,\, 0) & \text{se la specie è marrone} \\ (0,\, 1) & \text{se la specie è nera} \end{cases}$$

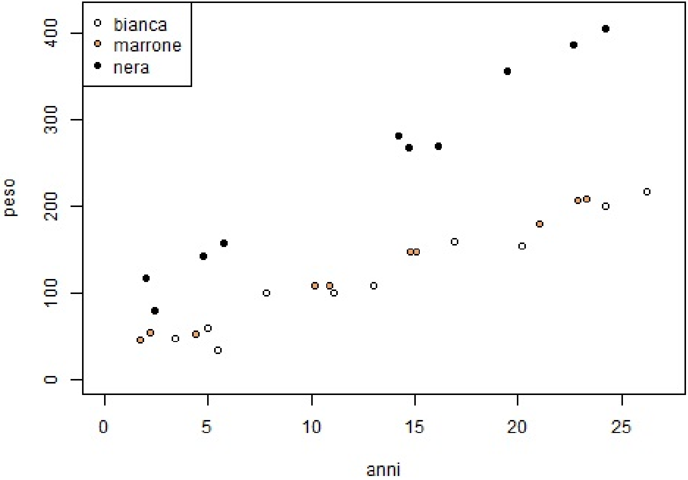
Un esempio

Questi sono i dati che ha trovato:

```
> anni
 [1]  7.8 24.2  2.4 16.1  2.0 10.2 19.5 14.7  2.2  5.5  3.4 10.9  1.7 11.1
[15] 14.2  5.0 22.7  5.8 15.1 20.2 13.0 24.2 22.9 14.8 16.9 23.3  4.4  4.8
[29] 21.0 26.2
> specie
 [1] "bianca"  "nera"    "nera"    "nera"    "nera"    "marrone" "nera"
 [8] "nera"    "marrone" "bianca"  "bianca"  "marrone" "marrone" "bianca"
[15] "nera"    "bianca"  "nera"    "nera"    "marrone" "bianca"  "bianca"
[22] "bianca"  "marrone" "marrone" "bianca"  "marrone" "marrone" "nera"
[29] "marrone" "bianca"
> peso
 [1] 100.1 406.2  80.5 270.0 117.6 108.0 356.4 268.1  55.1  34.4  47.1
[12] 108.8  45.7 100.6 282.5  60.3 386.7 158.4 148.3 155.1 109.0 200.4
[23] 206.3 148.2 160.3 209.6  52.7 143.0 180.5 216.7
```

Un esempio

E questo è il loro scatterplot:



Un esempio

```
Call:
lm(formula = peso ~ anni + specie + anni:specie)

Residuals:
    Min      1Q  Median      3Q     Max 
-25.447  -6.275   2.071   5.379  23.053 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)         18.71817    8.20343   2.282   0.0317 *  
anni                 7.47801    0.53113  14.080 4.28e-13 ***
speciemarrone       12.05817   11.39650   1.058   0.3006    
specienera          53.24377   11.32394   4.702 8.86e-05 ***
anni:speciemarrone   0.07485    0.75225   0.100   0.9216    
anni:specienera      6.36519    0.74566   8.536 9.82e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.1 on 24 degrees of freedom
Multiple R-squared:  0.9861,    Adjusted R-squared:  0.9832 
F-statistic: 340.6 on 5 and 24 DF,  p-value: < 2.2e-16
```

R usa i due punti per il termine d'interazione

$$\texttt{peso}_i = \beta_0 + \beta_1 \texttt{anni}_i + \beta_2 \texttt{speciemarrone}_i + \beta_3 \texttt{specienera}_i$$

$$+ \beta_4 \texttt{anni}_i \cdot \texttt{speciemarrone}_i + \beta_5 \texttt{anni}_i \cdot \texttt{specienera}_i + \varepsilon_i$$

Un esempio

```
Call:
lm(formula = peso ~ anni + specie + anni:specie)

Residuals:
    Min      1Q  Median      3Q     Max 
-25.447  -6.275   2.071   5.379  23.053 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)          18.71817    8.20343   2.282   0.0317 *  
anni                  7.47801    0.53113  14.080 4.28e-13 ***
speciemarrone        12.05817   11.39650   1.058   0.3006    
specienera           53.24377   11.32394   4.702 8.86e-05 ***
anni:speciemarrone    0.07485    0.75225   0.100   0.9216    
anni:specienera       6.36519    0.74566   8.536 9.82e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.1 on 24 degrees of freedom
Multiple R-squared:  0.9861,    Adjusted R-squared:  0.9832 
F-statistic: 340.6 on 5 and 24 DF,  p-value: < 2.2e-16
```

E decide lui come codificare le categorie bianca, marrone e nera (qui ha scelto la codifica di prima)

$$\texttt{peso}_i = \beta_0 + \beta_1 \texttt{anni}_i + \beta_2 \texttt{specie}\textcolor{red}{\texttt{marrone}}_i + \beta_3 \texttt{specie}\textcolor{red}{\texttt{nera}}_i$$

$$+ \beta_4 \texttt{anni}_i \cdot \texttt{specie}\textcolor{red}{\texttt{marrone}}_i + \beta_5 \texttt{anni}_i \cdot \texttt{specie}\textcolor{red}{\texttt{nera}}_i + \varepsilon_i$$

Un esempio

```
Call:
lm(formula = peso ~ anni + specie + anni:specie)

Residuals:
    Min      1Q  Median      3Q     Max
-25.447  -6.275   2.071   5.379  23.053

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         18.71817    8.20343   2.282   0.0317 *
anni                 7.47801    0.53113  14.080 4.28e-13 ***
speciemarrone       12.05817   11.39650   1.058   0.3006
specienera          53.24377   11.32394   4.702 8.86e-05 ***
anni:speciemarrone   0.07485    0.75225   0.100   0.9216
anni:specienera      6.36519    0.74566   8.536 9.82e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.1 on 24 degrees of freedom
Multiple R-squared:  0.9861,    Adjusted R-squared:  0.9832
F-statistic: 340.6 on 5 and 24 DF,  p-value: < 2.2e-16
```

p-value bassissimo!
Forte evidenza che la specie nera ha un'intercetta diversa dalle altre

$$\texttt{peso}_i = \beta_0 + \beta_1 \texttt{anni}_i + \beta_2 \texttt{speciemarrone}_i + {\color{red}\beta_3} \texttt{specienera}_i$$

$$+ \beta_4 \texttt{anni}_i \cdot \texttt{speciemarrone}_i + \beta_5 \texttt{anni}_i \cdot \texttt{specienera}_i + \varepsilon_i$$

Un esempio

```
Call:
lm(formula = peso ~ anni + specie + anni:specie)

Residuals:
    Min      1Q  Median      3Q     Max 
-25.447  -6.275   2.071   5.379  23.053 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)         18.71817    8.20343   2.282   0.0317 *  
anni                 7.47801    0.53113  14.080 4.28e-13 ***
speciemarrone       12.05817   11.39650   1.058   0.3006    
specienera          53.24377   11.32394   4.702 8.86e-05 ***
anni:speciemarrone   0.07485    0.75225   0.100   0.9216    
anni:specienera      6.36519    0.74566   8.536 9.82e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.1 on 24 degrees of freedom
Multiple R-squared:  0.9861,	Adjusted R-squared:  0.9832 
F-statistic: 340.6 on 5 and 24 DF,  p-value: < 2.2e-16
```

Ed evidenza ancora più forte che la specie nera ha pure il coefficiente angolare diverso dalle altre

$$\texttt{peso}_i = \beta_0 + \beta_1 \texttt{anni}_i + \beta_2 \texttt{speciemarrone}_i + \beta_3 \texttt{specienera}_i$$

$$+ \beta_4 \texttt{anni}_i \cdot \texttt{speciemarrone}_i + {\color{red}\beta_5} \texttt{anni}_i \cdot \texttt{specienera}_i + \varepsilon_i$$

Un esempio

```
Call:
lm(formula = peso ~ anni + specie + anni:specie)

Residuals:
    Min      1Q  Median      3Q     Max
-25.447  -6.275   2.071   5.379  23.053

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          18.71817    8.20343   2.282   0.0317 *
anni                  7.47801    0.53113  14.080 4.28e-13 ***
speciemarrone        12.05817   11.39650   1.058   0.3006
specienera           53.24377   11.32394   4.702 8.86e-05 ***
anni:speciemarrone    0.07485    0.75225   0.100   0.9216
anni:specienera       6.36519    0.74566   8.536 9.82e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.1 on 24 degrees of freedom
Multiple R-squared:  0.9861,	Adjusted R-squared:  0.9832
F-statistic: 340.6 on 5 and 24 DF,  p-value: < 2.2e-16
```

Invece, non sembra esserci nessuna evidenza che l'intercetta e il coefficiente angolare della specie bianca e di quella marrone siano diversi.
Però bisognerebbe eliminarli uno alla volta!

$$\mathtt{peso}_i = \beta_0 + \beta_1 \mathtt{anni}_i + \cancel{\beta_2 \mathtt{speciemarrone}_i} + \beta_3 \mathtt{specienera}_i$$
$$+ \cancel{\beta_4 \mathtt{anni}_i \cdot \mathtt{speciemarrone}_i} + \beta_5 \mathtt{anni}_i \cdot \mathtt{specienera}_i + \varepsilon_i$$