COSTRUZIONE DI MODELLI
PER PREVEDERE UN OUTPUT
IN FUNZIONE DI PIÙ INPUTS

ETÀ
ISTRUZIONE
$\Rightarrow$
RETRIBUZIONE

$$y = f(X_1 X_2 \ldots X_p) + \varepsilon$$

SISTEMATICA

ERRORE
VARIABILITÀ
NON MISURABILE
NON PREVEDIBILE

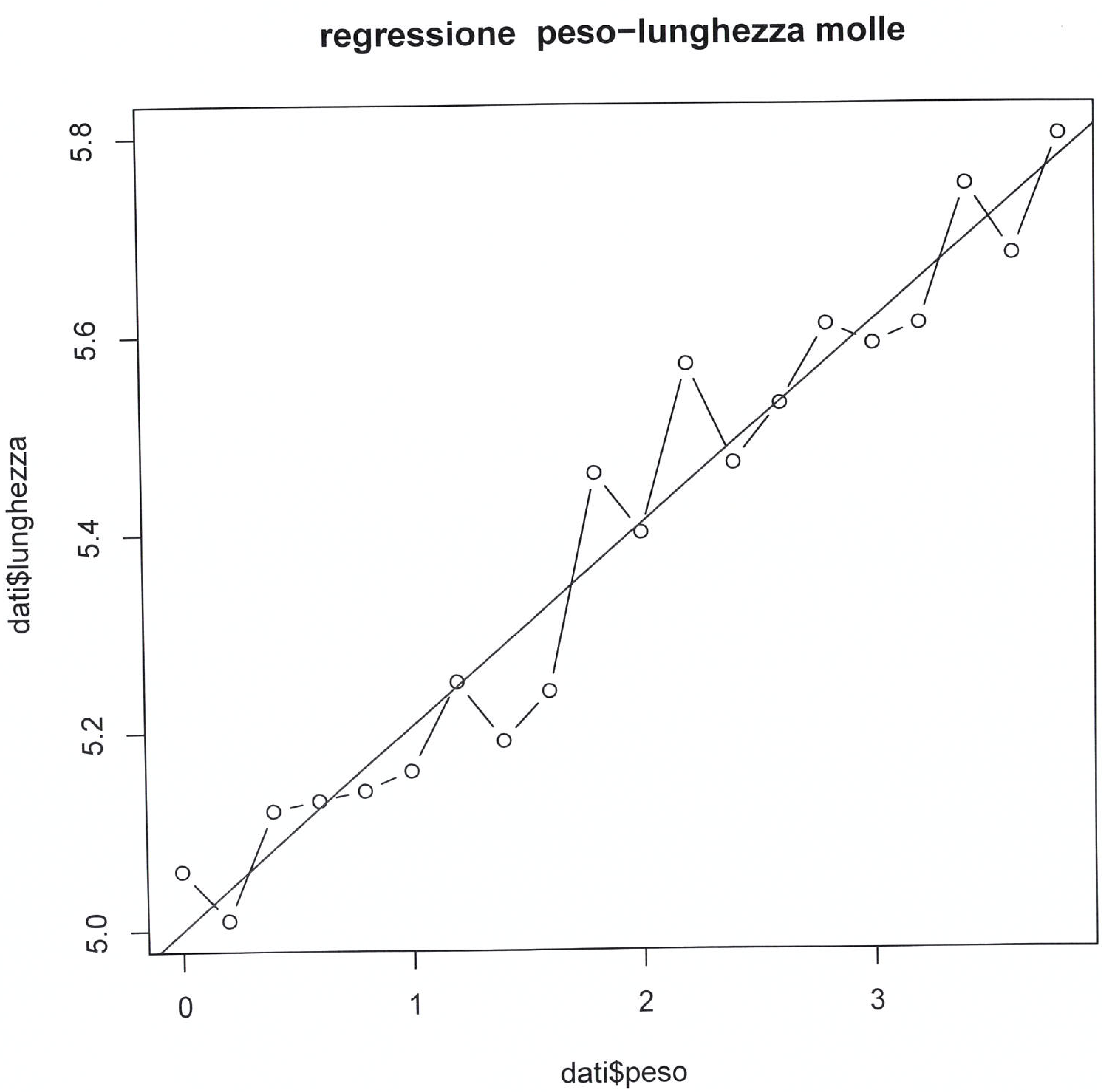
PREVISIONE o COMPRENSIONE

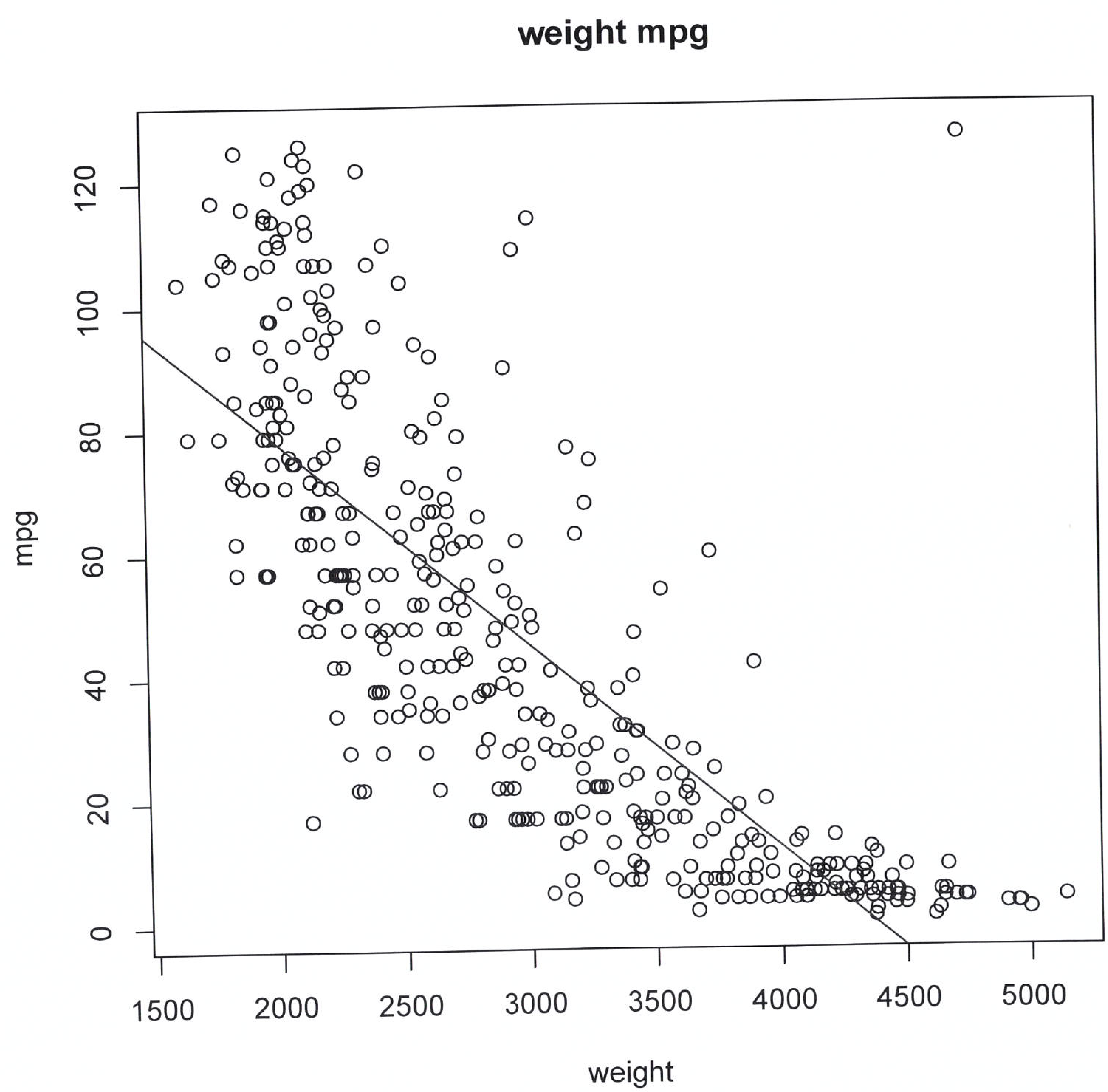
ACCURATEZZA
PREVISIONI

INTERPRETABILITÀ

SEMPLICITÀ

REGRESSIONE o CLASSIFICAZIONE


regressione peso–lunghezza molle
dati$lunghezza
5.0
5.2
5.4
5.6
5.8
0
1
2
3
dati$peso


weight mpg
mpg
0
20
40
60
80
100
120
1500
2000
2500
3000
3500
4000
4500
5000
weight

Modello di regressione

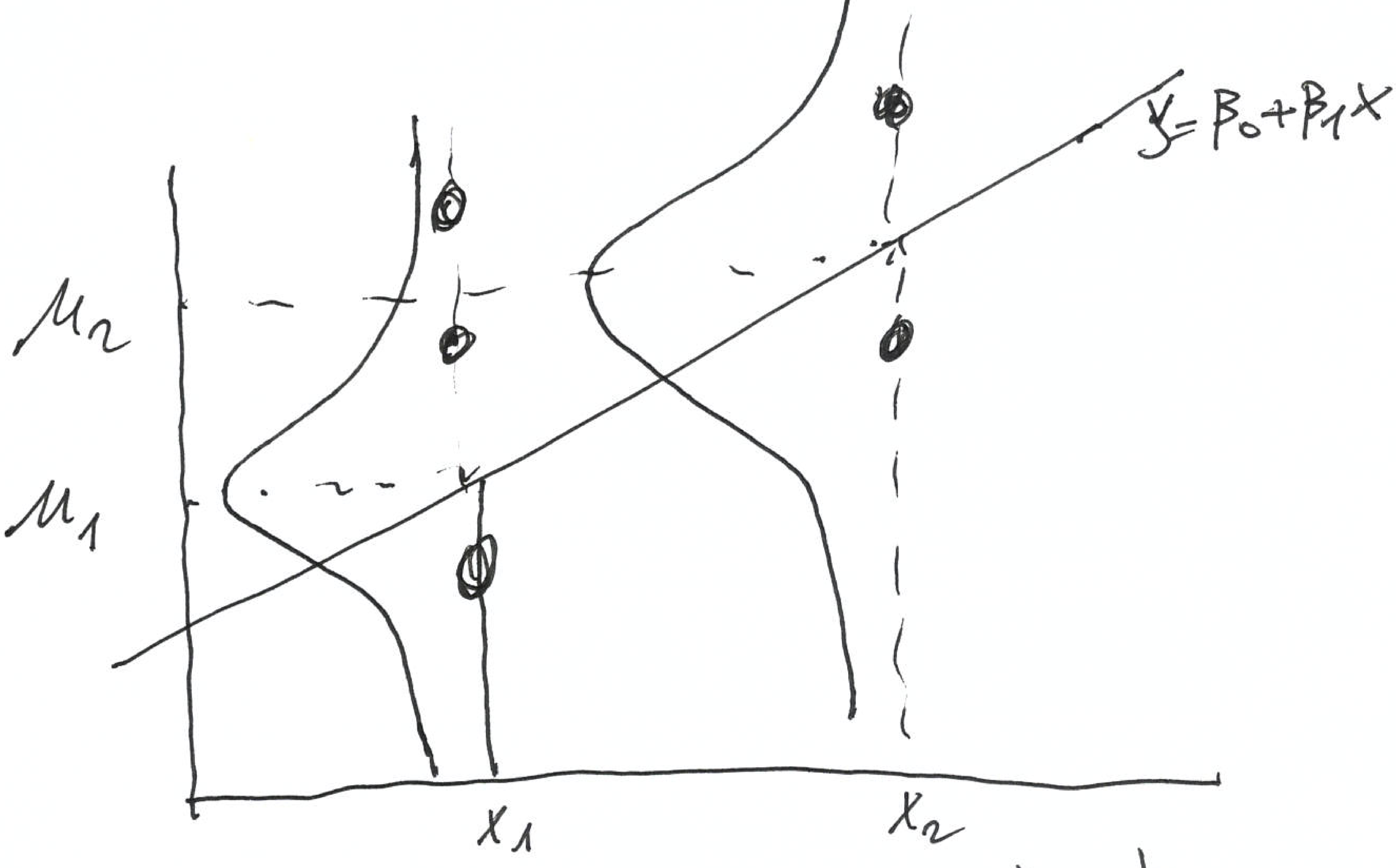
$$Y_i = \beta_0 + \beta_1 X_i + \sigma Z_i$$

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$Z_i \sim N(0,1)$ v.c. indipendenti

$\varepsilon_i \sim N(0,\sigma^2)$ identicamente distr.

$\mu_i = E(Y_i) = \beta_0 + \beta_1 X_i$ FUNZIONE DI REGRESSIONE



$x_1, x_2, \ldots\ldots x_n$ costanti

assunzioni

errori normali indipendenti identicamente distribuiti (stesso valore atteso stesse varianze)

funzione di regressione lineare

regressore (variabile indipendente) e valori che sono costanti note (il regressore non è una variabile casuale)

Densità di probabilità singole osservazione

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{y_i - \mu_i}{\sigma}\right)^2\right\}$$

y_i normali $E(y_i) = \mu$ $Var(y_i) = \sigma^2$

~~log~~ verosimiglianza

$$\prod_{1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{y_i-\mu_i}{\sigma}\right)^2\right\}$$

log verosimiglianza

$$-\frac{n}{2}\lg 2\pi - \frac{n}{2}\lg\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-\mu_i\right)^2$$

$$\mu_i = \beta_0 + \beta_1 X_i$$

stimatori di massima verosimiglianza

$$b_1 = \frac{Cov(X,Y)}{Var(X)} \quad ; \quad E(b_1) = \beta_1$$

$$b_0 = M_y - b_1 M_x \qquad E(b_0) = \beta_0$$

$$\tilde{s}_2 = \frac{\sum\left(y_i - b_0 - b_1 X_i\right)^2}{n-2} = \frac{\sum\left(y_i-\hat{y}_i\right)^2}{n-2}$$

Valori interpolati

$$\hat{y}_i = b_0 + b_1 X_i$$

$$E(\hat{y}_i) = E(b_0) + E(b_1) X_i =$$

$$= \beta_0 + \beta_1 X_i = \mu_i$$

$$\tilde{s}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}$$

Varianza

$$\sigma^2_{\beta_1} = \sigma^2 \cdot \frac{1}{n} \cdot \frac{1}{Var(X)}$$

$$\sigma^2_{\beta_0} = \sigma^2 \cdot \frac{\sum_i X_i^2}{n} \frac{1}{n\,Var(X)} = \sigma^2 \left[\frac{1}{n} + \frac{\mu_x^2}{n\,Var(X)}\right]$$

$$\sigma^2_{\hat{y}_i} = \sigma^2 \cdot \left[\frac{1}{n} + \frac{(\cancel{X_i} - \mu_x)^2}{n\,Var(X)}\right]$$

STIMATORI VARIANZE

$$\hat{\sigma}^2_{\beta_0} = \tilde{s}^2 \cdot \left[\frac{1}{n} + \frac{\mu_x^2}{n\,Var(X)}\right]$$

$$\hat{\sigma}^2_{\beta_1} = \tilde{s}^2 \cdot \frac{1}{n} \cdot \frac{1}{Var(X)}$$

$$\hat{\sigma}^2_{y_i} = \tilde{s}^2 \cdot \left[\frac{1}{n} + \frac{(X_i - \mu_x)^2}{n\,Var(X)}\right]$$

V.C PIVOTALI E STATISTICHE TEST

$$T_0 = \frac{b_0 - \beta_0}{\sqrt{\hat{\sigma}^2_{b_0}}}$$

$$T_1 = \frac{b_1 - \beta_1}{\sqrt{\hat{\sigma}^2_{b_1}}}$$

$$T_i = \frac{\hat{y}_i - \mu_i}{\sqrt{\hat{\sigma}^2_{y_i}}} \qquad i = 1, 2, \dots n$$

$\boxed{\text{T student con } n-2 \text{ g.d.l.}}$

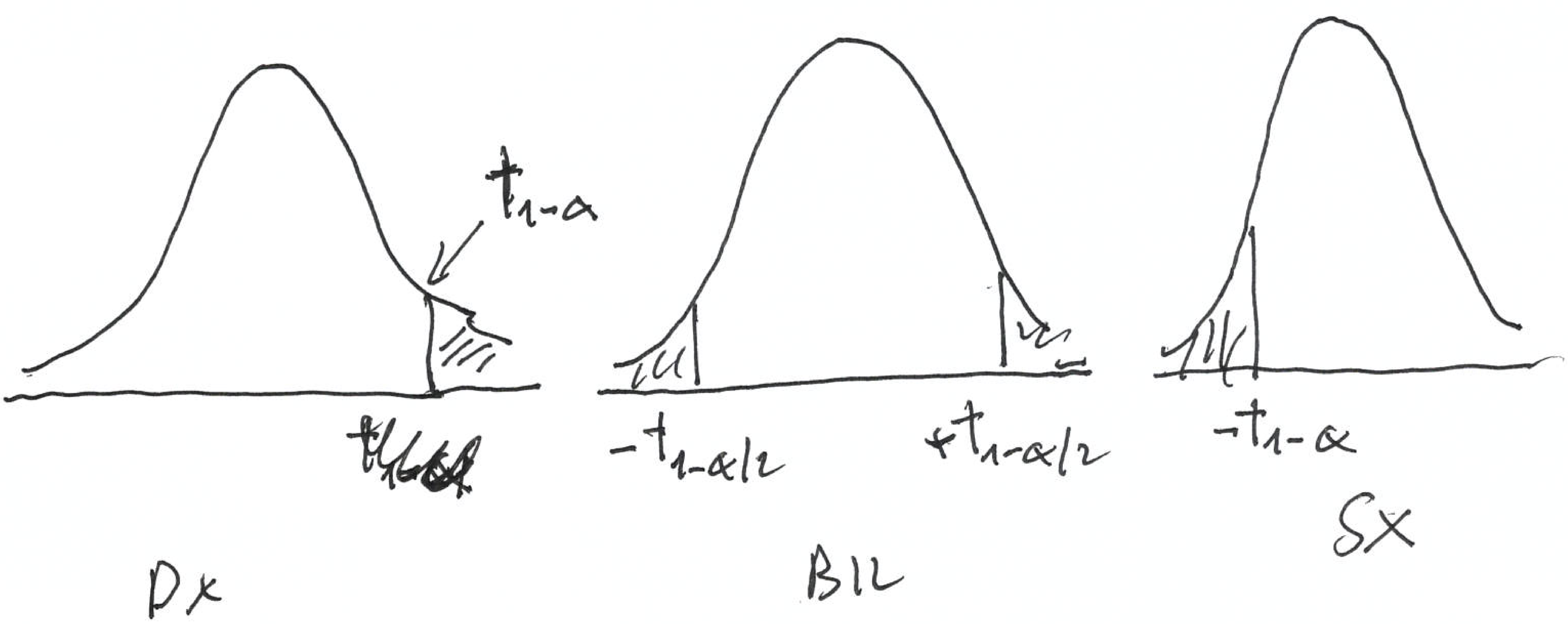
IC

$$b_0 \pm t_{1-\alpha/2;\,n-2}\sqrt{\hat{\sigma}^2_{b_0}}$$

$$b_1 \pm t_{1-\alpha/2;\,n-2}\sqrt{\hat{\sigma}^2_{b_1}}$$

$$\hat{y}_i \pm t_{1-\alpha/2;\,n-2}\sqrt{\hat{\sigma}^2_{y_i}}$$

TEST come al solito



SCOMPOSIZIONE VARIANZA

$$\frac{\sum (y_i - M_y)^2}{n} =$$

VAR TOTALE

$$= \frac{\sum (\hat{y}_i - M_y)^2}{n} + \frac{\sum (y_i - \hat{y})^2}{n}$$

VAR SPIEGATA

VAR RESIDUA

$$R^2 = \frac{\sum (\hat{y}_i - M_y)^2 / n}{\sum (y_i - M_y)^2 / n} =$$

$$= 1 - \frac{\sum (y_i - \hat{y}_i)^2 / n}{\sum (y_i - M_y)^2 / n}$$

FORMULE CALCOLO

$$r = \frac{Cov(x,y)}{\sqrt{Var(x)\,Var(y)}}$$

$$R^2 = r^2$$

$$\tilde{s}^2 = Var(y)(1-r^2) \times \frac{n}{n-2}$$

$$Var(y)(1-r^2) = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$