Economia e Politica delle Reti d'Impresa
Basi statistiche

Prof. Alessandro Sapio [1]

[1]Università degli Studi di Napoli Parthenope, DiSAE

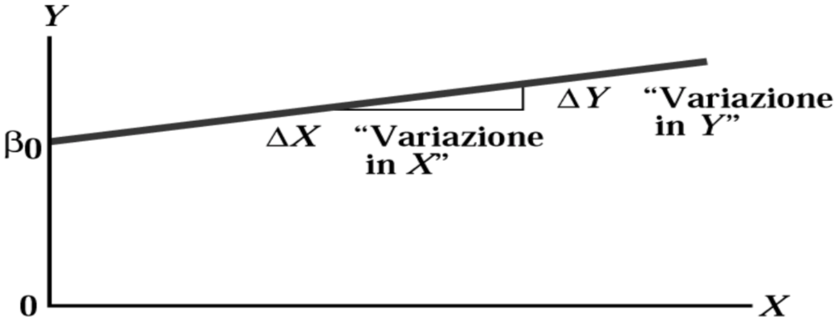
A.A. 2022-2023

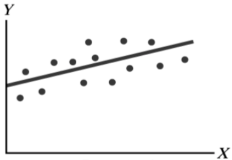
Piano della lezione

Ripasso dei seguenti argomenti:

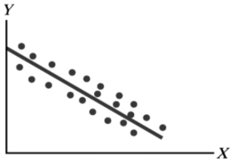
1. Modello di regressione lineare semplice
2. Metodo dei minimi quadrati ordinari
3. Misure di variabilità nella regressione
4. Le ipotesi del modello di regressione
5. Test di ipotesi sui coefficienti di regressione

Riferimenti: qualunque buon libro sull'analisi di regressione statistica

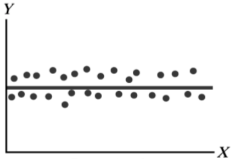

Y
β0
ΔX
"Variazione in X"
ΔY
"Variazione in Y"
0
X



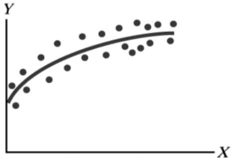
Riquadro A
Esempio di relazione lineare diretta



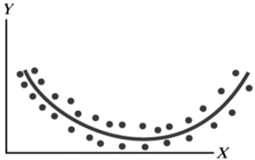
Riquadro B
Esempio di relazione lineare inversa



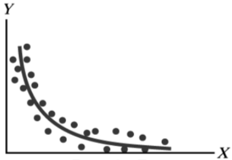
Riquadro C
Nessuna relazione tra X e Y



Riquadro D
Esempio di relazione polinomiale diretta



Riquadro E
Esempio di relazione curvilinea a U



Riquadro F
Esempio di relazione polinomiale inversa

Per stimare i coefficienti della retta di regressione...

si tratterà di *minimizzare* la somma dei loro quadrati:

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

dove

Y_i = il vero valore di Y per l'osservazione di i
$\hat{Y}_i$ = il valore previsto di Y per l'osservazione di i

Dal momento che in base al modello proposto $\hat{Y}_i = b_0 + b_1 X_i$, si tratta di minimizzare la seguente espressione:

$$\sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_i)]^2$$

rispetto alle due incognite b_0 e b_1.

L'equazione campionaria del modello di regressione lineare

La previsione di Y in base al modello di regressione lineare è data dalla somma tra l'intercetta campionaria e il prodotto tra il valore di X e l'inclinazione campionaria

$$\hat{Y}_i = b_0 + b_1 X_i \tag{9.2}$$

dove

$$\hat{Y}_i = \text{previsione di } Y \text{ per l'osservazione } i$$

$$X_i = \text{valore di } X \text{ per l'osservazione } i$$

Le misure di variabilità nella regressione

$$\text{Somma totale dei quadrati} = \text{somma dei quadrati della regressione} + \text{somma dei quadrati degli errori}$$

$$SQT = SQR + SQE \tag{9.3}$$

La somma totale dei quadrati (*SQT*)

La somma totale dei quadrati (*SQT*) è data dalla somma dei quadrati delle differenze tra i valori osservati di Y e la loro media.

$$SQT = \text{variabilità totale} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \tag{9.4}$$

La somma dei quadrati della regressione (*SQR*)

La somma dei quadrati della regressione (*SQR*) è data dalla somma dei quadrati delle differenze tra i valori previsti di Y e la media di Y.

$$SQR = \text{variabilità spiegata} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = SQT - SQE \tag{9.5}$$

La somma dei quadrati degli errori (*SQE*)

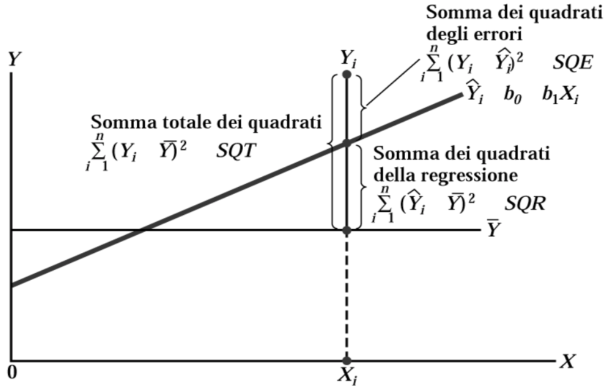
La somma dei quadrati degli errori (*SQE*) è data dalla somma dei quadrati delle differenze tra i valori osservati e i valori previsti di Y

$$SQE = \text{variabilità non spiegata} = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{9.6}$$

Il coefficiente di determinazione

Il coefficiente di determinazione è dato dal rapporto tra la somma dei quadrati della regressione e la somma totale dei quadrati.

$$r^2 = \frac{SQR}{SQT} \tag{9.7}$$

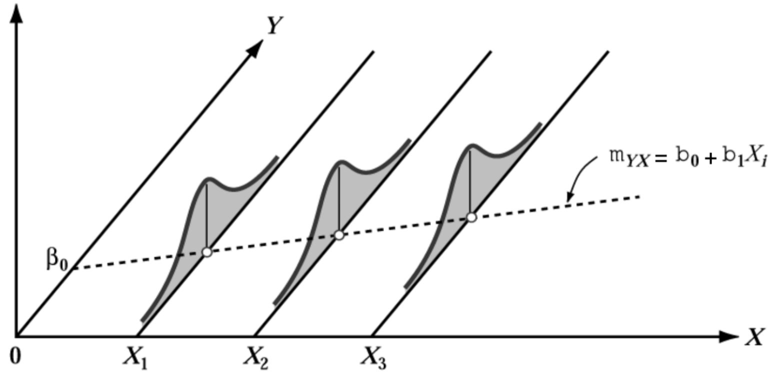

Somma dei quadrati
degli errori
SQE
Somma totale dei quadrati
SQT
Somma dei quadrati
della regressione
SQR
Y
X
0

L'errore standard della stima

$$S_{YX} = \sqrt{\frac{SQE}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}}$$

dove

$$\begin{aligned} Y_i &= \text{il valore di } Y \text{ in corrispondenza } X_i \\ \hat{Y}_i &= \text{il valore previsto di } Y \text{ in corrispondenza di } X_i \\ SQE &= \text{somma dei quadrati degli errori} \end{aligned}$$


Y
β_0
$m_{YX} = b_0 + b_1 X_i$
0
X_1
X_2
X_3
X

Il residuo

Il residuo è uguale alla differenza tra valore osservato e il valore previsto di Y:

$$e_i = Y_i - \hat{Y}_i \tag{9.9}$$

Riquadro 9.1 Le ipotesi del modello di regressione

✓ **1.** Distribuzione normale degli errori.

✓ **2.** Omoschedasticità.

✓ **3.** Indipendenza degli errori.

$$H_0\colon \beta_1 = 0 \text{ (non vi è una relazione lineare)}$$
$$H_1\colon \beta_1 \neq 0 \text{ (vi è una relazione lineare)}$$

Il test t per la verifica di ipotesi sull'inclinazione β_1

La statistica t è data dalla differenza tra l'inclinazione campionaria e l'inclinazione ipotizzata della popolazione, il tutto diviso per l'errore standard dell'inclinazione.

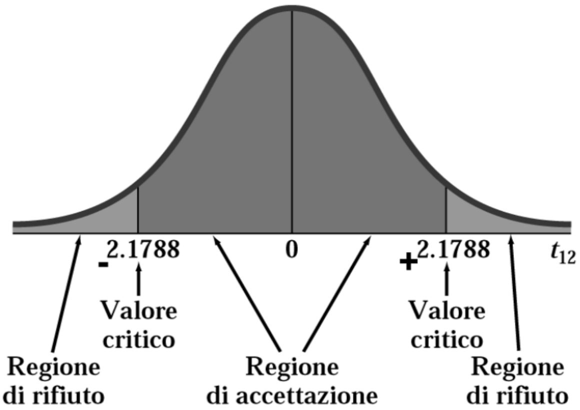
$$t = \frac{b_1 - \beta_1}{S_{b_1}}$$

dove

$$S_{b_1} = \frac{S_{YX}}{\sqrt{SQX}} \tag{9.11}$$

$$SQX = \sum_{i=1}^{n} (X_i - \overline{X})^2$$

La statistica t ha una distribuzione t di Student con $n - 2$ gradi di libertà.


-2.1788
0
+2.1788
t_{12}
Valore critico
Valore critico
Regione di rifiuto
Regione di accettazione
Regione di rifiuto

L'intervallo di confidenza per l'inclinazione

L'intervallo di confidenza per β_1 si ottiene addizionando e sottraendo all'inclinazione campionaria b_1 il prodotto tra il valore critico della statistica t e l'errore standard dell'inclinazione.

$$b_1 \pm t_{n-2} S_{b_1} \tag{9.13}$$

Grazie mille per l'attenzione