MODELLO DI REGRESSIONE LINEARE MULTIPLA

1. Modello e assunzioni
2. Stimatori OLS e proprietà
3. R^2, variabilità totale, spiegata e residua
4. Previsione
5. Test per la verifica di ipotesi
6. Variabili dummy
7. Multicollinearità
8. Eteroschedasticità

REGRESSIONE LINEARE MULTIPLA: IL PROBLEMA

- Ricerca di un modello matematico in grado di esprimere la relazione esistente tra una variabile di risposta y (quantitativa) e (ad esempio) k variabili esplicative
- Si tratta di una relazione asimmetrica del tipo

$$y = f(x_1 \ldots x_k)$$

Nel caso del modello di regr.lineare multipla

abbiamo che:

$$f(x_1 \ldots x_k) = \beta_1 x_1 + \beta_2 x_2 \ldots \beta_k x_k$$

che geometricamente corrisponde ad un iperpiano a k dimensioni

- Perché si studia tale modello

i) facilità con cui può essere interpretato un iperpiano a k dimensioni

ii) ii) Facilità di stima dei parametri incogniti β_j (j = 1…k)

Nella realtà studiamo un modello del tipo

$$y = f(x_1 \ldots x_k) + u$$

Componente sistematica — componente casuale

IL MODELLO

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots \beta_k x_{ik} + u_i$$

In forma matriciale

$$y = X\beta + u$$

dove

y : vettore (n x 1) di osservazioni sulla variabile dipendente

X : matrice (n x k) di osservazioni su k regressori

β : vettore (k x 1) di parametri incogniti

u : vettore (n x 1) di disturbi stocastici

Le matrici e i vettori sono così definiti

$$\underset{(n\times 1)}{y} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{pmatrix} \qquad \underset{(n\times k)}{X} = \begin{pmatrix} x_{11} & x_{12} & . & . & . & x_{1k} \\ x_{21} & x_{22} & . & . & . & x_{2k} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ x_{n1} & x_{n2} & . & . & . & x_{nk} \end{pmatrix}$$

$$\underset{(k\times 1)}{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{pmatrix} \qquad \underset{(n\times 1)}{u} = \begin{pmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_n \end{pmatrix}$$

N.B.

La matrice X ha la <u>**prima colonna unitaria**</u> nel caso in cui si consideri un modello con intercetta β_1 nel sistema di riferimento multidimensionale

ASSUNZIONI DEL MODELLO

1) Esiste legame lineare tra variabile dipendente e regressori
2) Le variabili sono tutte osservabili
3) I coefficienti β_i non sono v.c.
4) I regressori X sono non stocastici
5) Il termine u non è osservabile
6) $E(u_i) = 0$
7) $Cov(u_i, u_j) = \begin{cases} 0 & per \;\; i \neq j \\ \sigma^2 & per \;\; i = j \end{cases}$

$\Rightarrow$ le u_i sono omoschedastiche ed incorrelate

$$E[uu'] = \begin{pmatrix} \sigma^2 & 0 & 0 & . & . & 0 \\ 0 & \sigma^2 & 0 & . & . & 0 \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \sigma^2 \end{pmatrix}$$

8) X ha rango pieno rank (X) = k

condizione necessaria $\boxed{n \geq k}$

9) $u \quad N(0, \sigma^2 I)$ hp aggiuntiva da utilizzare nell'analisi inferenziale

STIMATORE OLS

$$Y = X\beta + u$$

Si cercherà quel vettore $\hat{\beta}$ che minimizza gli scarti al quadrato:

$$\min \sum_{i:1}^{n} \left(y_i - X_i\beta\right)^2$$

dove X_i è la riga i-esima di X

In forma matriciale

$$e_* = \hat{u} = \left(y - X\beta_*\right)$$

$$\min \;\; e_*'e_* \quad o \quad \min \;\; \left(y - X\beta_*\right)'\left(y - X\beta_*\right)$$

$$Q = e_*'e_* = \left(y - X\beta_*\right)'\left(y - X\beta_*\right)$$

$$= \left(y' - \beta_*'X'\right)\left(y - X\beta_*\right)$$

$$= y'y - \beta_*'\underbrace{X'y - y'X}_{=}\beta_* + \beta'X_*'X_*\beta_*$$

perché scalare

$$\frac{\partial Q}{\partial \beta_*} = -2X'y + 2X'X\beta_* = 0 \qquad (1)$$

perché

$$\beta' X' y = \overset{(1\times k)}{(\beta_1 \ldots \beta_k)} \overset{(k\times n)}{\begin{pmatrix} 1 & 1 & 1 & . & 1 \\ x_{21} & x_{22} & . & . & x_{2n} \\ x_{31} & x_{32} & . & . & x_{3n} \\ . & . & . & . & . \\ x_{k1} & . & . & . & x_{kn} \end{pmatrix}} \overset{(n\times 1)}{\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}}$$

è uno scalare

$$(\beta' X' y) = (\beta' X' y)' = (y' X \beta)$$

dalla (1) si ottiene

$$2X'X\beta_* = 2X'y$$

$$(X'X)\beta_* = X'y$$

pre-moltiplicando ambo i membri

$$(X'X)^{-1}(X'X)\beta_* = (X'X)^{-1}X'y$$

perché **rank *(X'X)* = rank *(X)* = k**

X'X è a rango pieno ovvero invertibile

$$\boxed{\hat{\beta} = (X'X)^{-1}X'y}$$ stimatore OLS di β

CARATTERISTICHE STIMATORE OLS

Teorema di Gauss-Markov

$\hat{\beta}$ è uno stimatore di tipo BLUE

Best Linear Unbiased Estimator

ovvero ha varianza minima nella classe degli stimatori Lineari e Corretti

1. $\hat{\beta} = (X'X)^{-1} X'y$

La matrice $(X'X)^{-1} X'$ è formata da elementi costanti per cui $\hat{\beta}$ è una trasformazione lineare di y.

2. $$\hat{\beta} = (X'X)^{-1} X'\underbrace{y} = (X'X)^{-1} X'\underbrace{(X\beta + u)}$$

$$= (X'X)^{-1} X'X\beta + (X'X)^{-1} X'u$$

$$= \beta + (X'X)^{-1} X'u$$

$$E(\hat{\beta}) = \beta + (X'X)^{-1} X' E(u) = \beta$$

È uno stimatore corretto

Inoltre: $(\hat{\beta} - \beta) = (X'X)^{-1} X'u$

3. $$Var\left(\hat{\beta}\right)=E\left[\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)'\right]$$

$$=E\left[\left(X'X\right)^{-1}X'uu'X\left(X'X\right)^{-1}\right]$$

$$=\left(X'X\right)^{-1}X'E\left[uu'\right]X\left(X'X\right)^{-1}$$

$$=\left(X'X\right)^{-1}X'\sigma^2 I\,X\left(X'X\right)^{-1}$$

$$=\sigma^2\left(X'X\right)^{-1}X'X\left(X'X\right)^{-1}=\sigma^2\left(X'X\right)^{-1}$$

Si consideri più in dettaglio $E\left[\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)'\right]$:

$$\begin{pmatrix} E\left(\hat{\beta}_1-\beta_1\right)^2 & E\left(\hat{\beta}_1-\beta_1\right)\left(\hat{\beta}_2-\beta_2\right) & . \quad . & E\left(\hat{\beta}_1-\beta_1\right)\left(\hat{\beta}_k-\beta_k\right) \\ E\left(\hat{\beta}_1-\beta_1\right)\left(\hat{\beta}_2-\beta_2\right) & E\left(\hat{\beta}_2-\beta_2\right)^2 & . \quad . & . \\ . & . & . \quad . & . \\ E\left(\hat{\beta}_k-\beta_k\right)\left(\hat{\beta}_1-\beta_1\right) & . & . \quad . & E\left(\hat{\beta}_k-\beta_k\right)^2 \end{pmatrix}$$

Pertanto la varianza $E\left(\hat{\beta}_j-\beta_j\right)^2$ di ogni parametro $\hat{\beta}_j$ si desume prendendo il corrispondente valore sulla diagonale principale della $\left(X'X\right)^{-1}$, moltiplicato per σ^2 :

$$Var\left(\hat{\beta}_j\right)=\left[\left(X'X\right)^{-1}\right]_{jj}\sigma^2$$

STIMA $\hat{\sigma}^2$ DI σ^2

Si può dimostrare che lo stimatore corretto della

varianza risulta essere:

$$\hat{\sigma}^2 = \frac{e'e}{(n-k)}$$

$$E\left[\hat{\sigma}^2\right] = \frac{1}{(n-k)}(n-k)\sigma^2 = \sigma^2$$

ESEMPIO (Greene p.200)

i : 1960 … 1986 , n = 27

G_i = consumo di benzina in $

P_{gi} = indice dei prezzi benzina

Y_i = reddito pro-capite in $

P_{qi} = indice dei prezzi auto nuove

Vettore y	x_1	x_2	x_3	x_4
121.01034	1	0.9250000	6036.0000	1.0450000
130.20306	1	0.9140000	6113.0000	1.0450000
136.62968	1	0.9190000	6271.0000	1.0410000
134.39852	1	0.9180000	6378.0000	1.0350000
150.34150	1	0.9140000	6727.0000	1.0320000
171.88391	1	0.9490000	7027.0000	1.0090000
175.44395	1	0.9700000	7280.0000	0.9910000
172.03874	1	1.0000000	7513.0000	1.0000000
198.65222	1	1.0470000	7891.0000	1.0440000
208.37573	1	1.0560000	8134.0000	1.0760000
214.38531	1	1.0630000	8322.0000	1.1200000
228.52113	1	1.0760000	8562.0000	1.1100000
237.37202	1	1.1810000	9042.0000	1.1110000
234.34193	1	1.5990000	8867.0000	1.1750000
222.32567	1	1.7080000	8944.0000	1.2760000
228.16247	1	1.7790000	9175.0000	1.3570000
242.33362	1	1.8820000	9381.0000	1.4290000
248.32557	1	1.9630000	9735.0000	1.5380000
240.93266	1	2.6560000	9829.0000	1.6600000
229.58893	1	3.6910000	9722.0000	1.7930000
227.13648	1	4.1090000	9769.0000	1.9020000
210.44373	1	3.8940000	9725.0000	1.9760000
236.85998	1	3.7640000	9930.0000	2.0260000
255.36365	1	3.7070000	10421.000	2.0850000
243.75057	1	3.7380000	10563.000	2.1520000
277.31965	1	2.9210000	10780.000	2.2400000

Matrice X'X;

27.000000	51.357000	229865.00	37.296000
51.357000	133.15081	473127.10	83.319118
229865.00	473127.10	2.0120502e+09	331319.22
37.296000	83.319118	331319.22	56.280428

Matrice inv (X'X);

2.6605735	0.51586178	-0.00029970528	-0.76246362
0.51586178	0.30384762	-6.4047001e-07	-0.78790617
-0.00029970528	-6.4047001e-07	6.6199636e-08	-0.00019015563
-0.76246362	-0.78790617	-0.00019015563	2.8089108

Stime b=inv(X'X) * X'y;

-89.761482
-12.588147
0.039938109
-14.443884

Y 121.01034 130.20306 136.62968 134.39852 150.34150 171.88391 175.44395 172.03874 198.65222 208.37573	n=10		
X1 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000	X2 0.92500000 0.91400000 0.91900000 0.91800000 0.91400000 0.94900000 0.97000000 1.00000000 1.04700000 1.05600000	X3 6036.0000 6113.0000 6271.0000 6378.0000 6727.0000 7027.0000 7280.0000 7513.0000 7891.0000 8134.0000	X4 1.0450000 1.0450000 1.0410000 1.0350000 1.0320000 1.0090000 0.9910000 1.0000000 1.0440000 1.0760000
(X'X) 10.000000 9.6120000 69370.000 10.318000	 9.6120000 9.2665480 67031.717 9.9199470	 69370.000 67031.717 4.8631105e+08 71575.421	 10.318000 9.9199470 71575.421 10.651854
Inv (X'X) 197.12839 -30.407072 0.00072941000 -167.53347	 -30.407072 489.93203 -0.034015993 -198.24254	 0.00072941000 -0.034015993 2.558142e-06 0.013782628	 -167.53347 -198.24254 0.013782628 254.38467
Beta = inv(X'X)*X'y -131.78025 -90.513381 0.045503884 61.076792			

ANOVA

Analisi della varianza

Se vogliamo testare simultaneamente ipotesi su tutti i parametri o coefficienti dei regressori andiamo a considerare la statistica F di **Fisher-Snedecor.**

Considerando il modello in forma di scarti

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \cdot \\ \cdot \\ \hat{\beta}_k \end{pmatrix} = \left(X'X\right)^{-1} X'y$$

$$R^2 = \frac{\hat{\beta}' X' y}{y' y}$$

$$y_i \qquad N\left(0, \sigma^2\right)$$

$$\hat{\beta}_i \qquad N\left(\beta_i , \sigma^2 \left[\left(X'X\right)^{-1}\right]_{ii}\right)$$

Si può dimostrare che

$$\frac{\left(\hat{\beta}-\beta\right)' X'y}{\sigma^2} \qquad \chi^2_{(k-1)}$$

e ricordando che

$$\frac{\chi^2_{p/p}}{\chi^2_{q/q}} \qquad F_{p,q}$$

$$\frac{\dfrac{\left(\hat{\beta}-\beta\right)' X'y}{\sigma^2} \Big/ (k-1)}{\dfrac{e'e}{\sigma^2} \Big/ (n-k)} \qquad F_{(k-1,n-k)}$$

Sotto $H_0 : \beta = 0$

$$\frac{\hat{\beta}' X'y / (k-1)}{e'e / (n-k)} = \frac{R^2 / (k-1)}{\left(1-R^2\right) / (n-k)} \qquad F_{k-1,n-k}$$

TABELLA ANOVA

Causa var.	Devianza	G.L.	Stime var.
Modello $x_2 \ldots\ldots x_k$	$\hat{\beta}' X' y = y' y R^2$	k-1	$\hat{\beta}' X' y / (k-1)$
Residuo	$e'e = y'y(1 - R^2)$	n-k	$e'e / (n-k)$
Totale	$y'y = \sum y_i^2$	n-1	

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

- Si costruisce la statistica F
- Si individua il 95% o il 99%

quantile della distribuzione $F_{(k-1),(n-k)}$

- Se $F > F_{(1-\alpha);(k-1)(n-k)}$ si rifiuta H_0

COEFFICIENTE DI CORRELAZIONE MULTIPLA

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{\sum e_i^2}{\sum Y_i^2} \le 1$$

$$0 \le R^2 \le 1$$

Il coefficiente di correlazione è un indicatore del legame lineare tra *Y* e i regressori.

Ha però un difetto:

Esso può aumentare anche se viene aggiunto un regressore che non "spiega" *y*.

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum e_i^2}{\sum Y_i^2}$$

Se dividiamo le devianze per i gradi di libertà andiamo a pesare il contributo a R^2 di ogni regressore

$$\hat{R}^2 = 1 - \frac{\sum e_i^2 / (n-k)}{\sum Y_i^2 / (n-1)}$$

$$\boxed{\hat{R}^2 = 1 - \frac{(n-1)}{(n-k)}\left(1 - R^2\right)}$$

APPLICAZIONE

$$Y = X\beta + u$$

n = 12

k = 3

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

Facendo riferimento ai valori

$$\bar{Y} = 9 \qquad \bar{X}_2 = 2 \qquad \bar{X}_3 = 1$$

$$\sum x_2^2 = 10 \qquad \sum x_3^2 = 15 \qquad \sum y^2 = 200$$

$$\sum x_2 y = 12 \qquad \sum x_3 y = 9 \qquad \sum x_2 x_3 = -11$$

Determinare il vettore di stime OLS

Se consideriamo il modello in forma di scarti dalle medie

$$\begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = (X'X)^{-1} X' y$$

Dove

$$X = \begin{bmatrix} x_{21} & x_{31} \\ x_{22} & x_{32} \\ . & . \\ . & . \\ x_{2n} & x_{3n} \end{bmatrix} \qquad \begin{aligned} x_{2i} &= X_{2i} - \bar{X}_2 \\ x_{3i} &= X_{3i} - \bar{X}_3 \end{aligned}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

$$(X'X) = \begin{bmatrix} \sum X_{2i}^2 & \sum X_{2i} X_{3i} \\ \sum X_{2i} X_{3i} & \sum X_{3i}^2 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{|X'X|} \begin{pmatrix} (-1)^2 \sum X_{3i}^2 & (-1)^3 \sum X_{2i} X_{3i} \\ (-1)^3 \sum X_{2i} X_{3i} & (-1)^4 \sum X_{2i}^2 \end{pmatrix}$$

$$= \frac{1}{\sum X_{2i}^2 \sum X_{3i}^2 - \left(\sum X_{2i} X_{3i}\right)^2} \begin{pmatrix} \sum X_{3i}^2 & -\sum X_{2i} X_{3i} \\ -\sum X_{2i} X_{3i} & \sum X_{2i}^2 \end{pmatrix}$$

$$X'y = \begin{pmatrix} \sum X_{2i} Y_i \\ \sum X_{3i} Y_i \end{pmatrix}$$

da cui

$$\begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \frac{1}{\sum X_2^2 \sum X_3^2 - \left(\sum X_2 X_3\right)^2} \times$$

$$\times \begin{pmatrix} \sum X_3^2 \sum X_2 Y - \sum X_2 X_3 \sum X_3 Y \\ \sum X_2^2 \sum X_3 Y - \sum X_2 X_3 \sum X_2 Y \end{pmatrix}$$

$$\hat{\beta}_2 = \frac{15 \times 12 - (-11) \times 9}{10 \times 15 - 121} = \frac{180 + 99}{29} \cong 9.62$$

$$\hat{\beta}_3 = \frac{10 \times 9 - (-11) \times 12}{10 \times 15 - 121} = \frac{90 + 132}{29} \cong 7.65$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 = 9 - 2 \times 9.62 - 7.65 = -17.89$$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} -17.89 \\ 9.62 \\ 7.65 \end{pmatrix}$$

RICAPITOLANDO

$$\hat{\beta} = (X'X)^{-1} X' y$$

$$E[\hat{\beta}] = \beta$$

$$V[\hat{\beta}] = E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)' \right) = (X'X)^{-1} \sigma^2$$

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k}$$

$$E[\hat{\sigma}^2] = \sigma^2$$

Fino ad ora nessuna ipotesi è stata posta per la distribuzione degli errori nel problema della stima.

Aggiungiamo :

$$u_i \qquad N(0, \sigma^2)$$

$$u \qquad N(0, \sigma^2 I)$$

TEST PER LA VERIFICA DI IPOTESI

Dal teorema di GAUSS-MARKOV :

$$\hat{\beta} \qquad N\left(\beta, \sigma^2 \left(X'X\right)^{-1}\right)$$

Vogliamo testare $\quad H_0 : \beta_i = 0$

Ovvero vogliamo verificare se il regressore X_i spiega effettivamente la variabile dipendente Y nel caso (improbabile) che sia nota σ^2

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 \left[\left(X'X\right)^{-1}\right]_{ii}}} \qquad N(0,1)$$

Sotto $\quad H_0 : \beta_i = 0 \quad$ andiamo a considerare la statistica

$$\frac{\hat{\beta}_i}{\sqrt{\sigma^2 \left[\left(X'X\right)^{-1}\right]_{ii}}}$$

Se il valore cade all'esterno dell'intervallo di confidenza al 95% della $N(0,1)$ (± 1.96) rifiutiamo H_0 ed il parametro β_i sarà "significativamente" diverso da zero.

In generale rifiuto H_0 al livello $100\varepsilon\%$ di significatività quando

$$\left| \frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 \left[(X'X)^{-1} \right]_{ii}}} \right| > q\,n_{\varepsilon/2}$$

QUANDO σ^2 NON E' NOTA

Utilizziamo la sua stima $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{e'e}{(n-k)}$$

Si può dimostrare che.

$$u \qquad N\left(0, \sigma^2 I\right)$$

$$v = P'u$$

$$v \qquad N\left(0, \sigma^2 I\right)$$

$$\frac{e'e}{\sigma^2} = \frac{v_1^2}{\sigma^2} + \frac{v_2^2}{\sigma^2} + \ldots + \frac{v_{n-k}^2}{\sigma^2} =$$

$$= \sum_{i=1}^{n-k} \left(\frac{v_i}{\sigma}\right)^2 = \sum_{i=1}^{n-k} N(0,1)^2 = \chi^2_{n-k}$$

e e $\hat{\beta}$ sono Normali e incorrelate quindi indipendenti ; lo saranno anche $\hat{\beta}$ e $\hat{\sigma}^2 = \frac{e'e}{n-k}$

N.B.

$$\boxed{\frac{N(0,1)}{\sqrt{\chi^2_{n-k}/(n-k)}} = t_{(n-k)}}$$

Quindi

$$\frac{\dfrac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2\left[(X'X)^{-1}\right]_{ii}}}}{\sqrt{\dfrac{e'e}{\sigma^2}\Big/(n-k)}} \qquad t_{n-k}$$

$$\frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}\sqrt{a_{ii}}} \qquad t_{n-k}$$

Le ipotesi su β_i possono essere verificate sostituendo i valori in tale formula. e controllando poi che la statistica superi o meno i valori della regione critica della distribuzione t_{n-k} .

RIPRENDIAMO L'ESERCIZIO
(Applicazione lucidi precedenti 18-20)

$$H_0: \ \beta_2 = \beta_3 = 0 \qquad (F_{0.01\,,\,2\,,\,9} = 8.02)$$

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \boxed{\frac{ESS/(k-1)}{RSS/(n-k)}}$$

Ricordiamo:
n = 12
k = 3 con intercetta $\Rightarrow$ 2 var. esplicative

$$R^2 = \frac{\hat{\beta}'X'y}{y'y}$$

in forma di scarti

$$\begin{pmatrix}\hat{\beta}_2 & \hat{\beta}_3\end{pmatrix} \times \left(\begin{array}{cccc} x_{21} & . & . & x_{2n} \\ x_{31} & . & . & x_{3n} \end{array}\right) \times \begin{pmatrix} y_1 \\ . \\ . \\ y_n \end{pmatrix} =$$

$$\begin{pmatrix}\hat{\beta}_2 & \hat{\beta}_3\end{pmatrix} \times \begin{pmatrix} \sum X_2 y \\ \sum X_3 y \end{pmatrix} = \hat{\beta}_2 \sum X_2 y + \hat{\beta}_3 \sum X_3 y$$

$$R^2 = \frac{9.62 \times 12 + 7.65 \times 9}{200} = \frac{184.29}{200} \cong 0.92$$

$$F = \frac{0.92/2}{(1-0.92)/9} = 11.5 \times \frac{9}{2} \cong 51.75$$

valore empirico di F

Si rifiuta H_0 con un livello di significatività del 99%

F empirico = 51.75 > $F_{0.01,2,9}$ = 8.02

Se avessimo voluto testare

$H_0: \ \beta_2 = 0$

Ovvero la significatività di X_2

$$t = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\hat{\sigma}^2 a_{22}}} \qquad t_{n-k} \qquad \left(o \quad F(1, n-k)\right)$$

($t_{99.9} = 2.82$)

$$a_{22} = \frac{\sum X_3^2}{\sum X_2^2 \sum X_3^2 - \left(\sum X_2 X_3\right)^2} = \frac{15}{150 - 121} = \frac{15}{29} \cong 0.51$$

$$\hat{\sigma}^2 = \frac{e'e}{n-k} = \frac{TSS - ESS}{9} = \frac{200 - 184.29}{9} \cong 1.74$$

$$t = \frac{\hat{\beta}_2}{\sqrt{\hat{\sigma}^2 a_{22}}} = \frac{9.62}{\sqrt{1.74 \times 0.51}} = \frac{9.62}{0.94} \cong 10.2$$

valore empirico di t

Anche adesso rifiutiamo H_0 $\Rightarrow$ il regressore X_2 è significativo