Il modello lineare inferenziale con due variabili esplicative

Si consideri il seguente modello:

$$X_{1i} = \alpha + \beta x_{2i} + \gamma x_{3i} + \epsilon_i$$

$$\epsilon_i \sim NID(0, \sigma^2)$$

Stimatori per i parametri hanno la seguente forma:

$$\hat{\alpha} = \frac{\sigma_{33}\sigma_{12} - \sigma_{13}\sigma_{23}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2}$$

$$\hat{\beta} = \frac{\sigma_{22}\sigma_{13} - \sigma_{12}\sigma_{23}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2}$$

$$\hat{\gamma} = \overline{X_1} - \hat{\beta}\overline{x_2} - \hat{\sigma}\overline{x_3}$$

Essi sono combinazioni lineari di variabili casuali normali, pertanto hanno distribuzione normale.

Pertanto valgono le seguenti:

$$\frac{\hat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{n}\cdot\frac{\sigma_{33}}{\sigma_{22}\sigma_{33}-\sigma_{23}^2}}}\sim N(0,1)$$

$$\frac{\hat{\gamma}-\gamma}{\sqrt{\frac{\sigma^2}{n}\cdot\frac{\sigma_{22}}{\sigma_{22}\sigma_{33}-\sigma_{23}^2}}}\sim N(0,1)$$

Inoltre, vale il seguente risultato:

$$\frac{\sum_1^n\left(X_{1i}-\hat{\alpha}-\hat{\beta}x_{2i}-\hat{\gamma}x_{3i}\right)^2}{\sigma^2}\sim chi^2{}_{n-3}$$

Pertanto, stimatore corretto per σ^2 avrà la forma:

$$\hat{\sigma}^2=\frac{\sum_1^n\left(X_{1i}-\hat{\alpha}-\hat{\beta}x_{2i}-\hat{\gamma}x_{3i}\right)^2}{n-3}$$

$$\frac{\sum_1^n\left(X_{1i}-\hat{\alpha}-\hat{\beta}x_{2i}-\hat{\gamma}x_{3i}\right)^2}{\sigma^2}=\frac{(n-3)\hat{\sigma}^2}{\sigma^2}\sim chi^2{}_{n-3}$$

Per i teoremi sulla distribuzione t di Student, vale:

$$T_2 = \frac{\hat{\beta} - \beta}{\sqrt{\frac{\hat{\sigma}^2}{n} \cdot \frac{\sigma_{33}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2}}} \sim t_{n-3}$$

$$T_3 = \frac{\hat{\gamma} - \gamma}{\sqrt{\frac{\hat{\sigma}^2}{n} \cdot \frac{\sigma_{22}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2}}} \sim t_{n-3}$$

Ciò permette di svolgere verifiche d'ipotesi sui coefficienti di regressione.

$$H_0: \beta = \beta_0$$

Allora:

$$H_1: \beta > \beta_0$$

Si rifiuta H_0 se

$$T_2 > t_{1-\alpha,n-3}$$

$$H_1: \beta < \beta_0$$

Si rifiuta H_0 se

$$T_2 < -t_{1-\alpha,n-3}$$

$$H_1: \beta \neq \beta_0$$

Si rifiuta H_0 se

$$|T_2| > t_{1-\alpha/2,n-3}$$

In modo analogo per verificare l'ipotesi $H_0: \gamma = \gamma_0$.

Ponendo $\beta_0 = 0$, qualora si accetti l'ipotesi nulla, significa che la variabile x_2 va eliminata dal modello in quanto non significativa.

Ponendo $\gamma_0 = 0$, qualora si accetti l'ipotesi nulla, significa che la variabile x_3 va eliminata dal modello in quanto non significativa.

Un'altra verifica d'ipotesi importante per tale modello è la seguente:

$$\begin{cases} H_0: \beta = \gamma = 0 \\ H_1: \beta \neq 0 \ \cup \gamma \neq 0 \end{cases}$$

E' noto il seguente risultato sulla devianza spiegata DS:

$\frac{DS}{\sigma^2} \sim chi^2_{3-1}$, sotto H_0

Inoltre, vale il seguente risultato sulla devianza residua:

$$\frac{DR}{\sigma^2} \sim chi^2_{n-3}$$

Inoltre DS e DR sono fra loro indipendenti.

Pertanto, sotto H_0:

$$V = \frac{DS/[\sigma^2(3-1]}{DR/[\sigma^2(n-3)]} = \frac{DS/[DT(3-1]}{DR/[DT(n-3)]} =$$
$$= \frac{I^2_{1.23}/2}{(1-I^2_{1.23})/(n-3} \sim F(2, n-3)$$

Pertanto si rifiuterà H_0 se $V > F_{1-\alpha}(2, n-3)$

Gli steps dell'analisi del modello saranno i seguenti:

a) Stima dei parametri
b) Verifica dell'ipotesi: $H_0 \colon \beta = \gamma = 0$
   Se si accetta l'ipotesi nulla, le variabili esplicative sono congiuntamente non significative ai fini della previsione di X_1. Se si rifiuta l'ipotesi nulla, l'analisi continua
c) Verifica della significatività dei coefficienti di regressione mediante le ipotesi
   $H_0 \colon \beta = 0$ e $H_0 \colon \gamma = 0$

Se si accetta l'ipotesi nulla, la corrispondente variabile esplicativa andrà omessa.

Il modello lineare generale: simbologia

$$y = X\beta + \epsilon$$

Ove,

X è una matrice n x k

y è un vettore n x 1

β è un vettore k x 1

ϵ è un vettore n x 1, a componenti normali indipendenti con varianza σ^2 (omoschedasticità)

Ponendo k=2 e $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{bmatrix}$, si ottiene il

modello lineare con una sola variabile esplicativa.

Ponendo k=3 e $X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \dots & \dots & - \\ 1 & x_{1n} & x_{2n} \end{bmatrix}$, si ottiene il modello lineare con due variabili esplicative.

Il modello lineare generale, o modello di regressione lineare multipla, contiene i modelli precedenti come casi particolari. Nel seguito, gli studenti scarichino o visualizzino il file da e-learning sulla regressione lineare multipla.