Principi di Econometria

lezione 10

AA 2015-2016

Paolo Brunori

Regressione multipla redditi svedesi: quale polinomio usare?

	coefficiente	errore standard	t	$p-value$
β_0	-14693.729	4149.33	-3.541	0.0004
β_{age}	1119.13	171.52	6.522	0.0000
β_{age^2}	-10.407	1.785	-5.831	0.0000
β_{sex}	1054.07	122.67	8.592	0.0000
β_{edu}	1776.77	856.584	2.074	0.0383

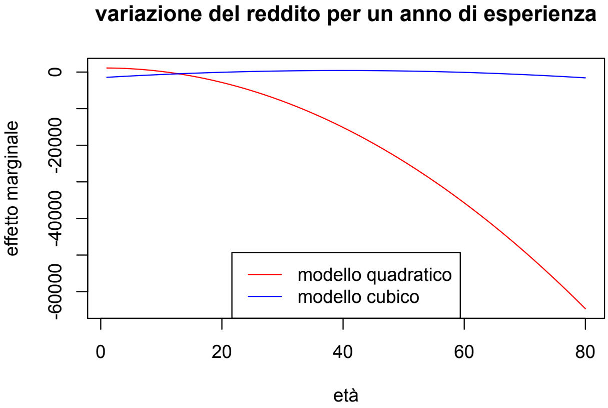
- $R^2 = 0.138$, R^2corretto $= 0.1339$
- errore standard di regressione $= \sqrt{\frac{1}{n-k}\sum_i^n u_i^2}$=12370

Regressione multipla redditi svedesi: quale polinomio usare?

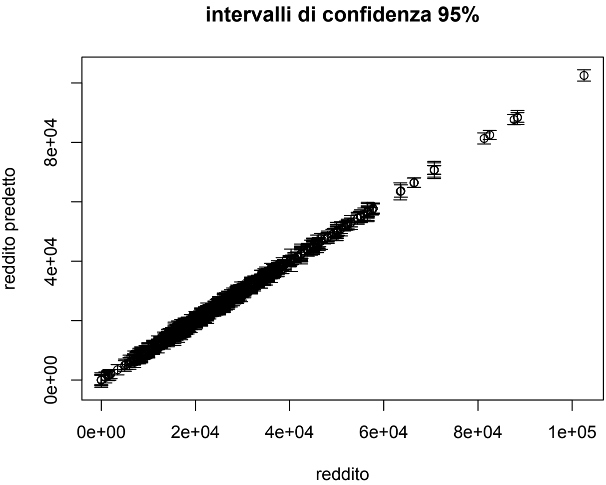
	coefficiente	errore standard	t	$p - value$
β_0	19625.42	10123.76	1.939	0.05289
β_{age}	-1486.15	722.42	-2.057	0.0399
β_{age^2}	48.51	15.97	3.037	0.00247
β_{age^3}	-0.412	0.1111	-3.711	0.00022
β_{edu}	1094.39	122.23	8.953	0.0000
β_{sex}	1851.75	850.32	2.178	0.02971

- $R^2 = 0.1521$, R^2corretto $= 0.147$
- errore standard di regressione $= \sqrt{\frac{1}{n-k}\sum_i^n u_i^2}$=12270

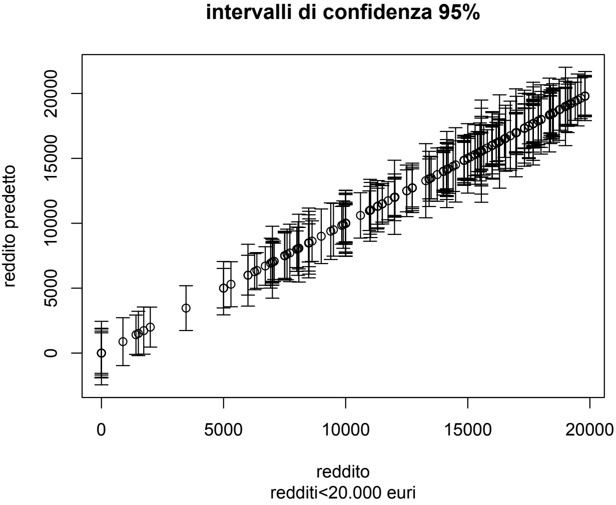
L'effetto di un anno di esperienza per i due modelli



I valore predetti (modello quadratico)

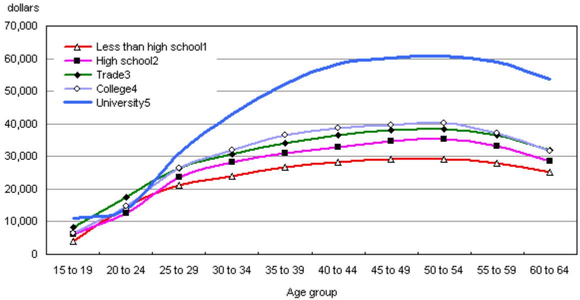


I valore predetti (modello quadratico)



Interazione di variabili

- potrebbe essere che l'effetto dell'esperienza dipenda dal livello di istruzione



stime su dati Canadesi: www.statcan.gc.ca

Interazione di variabili

- in questo caso vorremmo poter tener conto del fatto che la distanza fra un istruito e un non istruito aumenta all'aumentare dell'esperienza
- ci serve una variabile che a parità di età sia più elevata se il livello di istruzione è alto
- analogamente: sia più alta quando l'età è elevata a parità di istruzione
- in un certo senso è una situazione simile a quella vista per l'interazione fra variabili dicotomiche e non dicotomiche (o fra due dicotomiche)
- anche in questo caso il termine $età \times edu$ può essere usato per risolvere il problema

interazione fra variabili continue: formula generale

- in generale per capire quale relazione coglie una trasformazione non lineare di un regressore si deve svolgere l'equazione $\Delta Y = f(X_1, ..., X_j + \Delta X_j, ..., X_k) - f(X_1, ..., X_j, ..., X_k)$
- e calcolare la variazione di Y dovuta alla variazione di X_j
- se $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2)$ abbiamo:

$$\begin{aligned}\Delta Y = \beta_0 + \beta_1 X_1 + \beta_2 (X_2 + \Delta X_2) + \beta_3 [X_1 \times (X_2 + \Delta X_2)] + \\ -[\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2)]\end{aligned}$$

$$\Delta Y = \beta_2 \Delta X_2 + \beta_3 (X_1 \times \Delta X_2)$$

interazione fra variabili continue: formula generale

$$\Delta Y = \beta_2 \Delta X_2 + \beta_3 (X_1 \times \Delta X_2)$$

$$\frac{\Delta Y}{\Delta X_2} = \beta_2 + \beta_3 X_1$$

Per cui la variazione di Y dovuta ad una variazione di X_2 dipende dal valore di X_1

interazione fra variabili continue

- il nuovo modello sarà:

$$Y_I = \beta_0 + \beta_1 age + \beta_2 edu + \beta_3 sex + \beta_4 eta^2 + \beta_5 (eta \times edu)$$

	coefficiente	errore standard	t	$p - value$
β_0	-1694.34	7644.765	-0.222	0.8246
β_{age}	847.69	217.519	3.897	0.0001
β_{age^2}	-9.81	1.805	-5.43	0.0000
β_{edu}	158.68	459.22	0.346	0.7297
$\beta_{age \times edu}$	16.798	8.30	2.023	0.0433
β_{sex}	1673.6	856.5	1.954	0.0510

- $R^2 = 0.1423$, $R^2 -$ corretto $= 0.1371$
- errore standard di regressione $= 12340$

funzioni di logaritmi

- spesso si esprimono relazioni economiche in termini percentuali: l'elasticità è l'esempio più comune:

$$\eta_{Y,X} = \frac{\Delta Y / Y}{\Delta X / X} = \frac{\Delta \% Y}{\Delta \% X}$$

- l'uso della funzione logaritmica di un regressore ci permette di cogliere una relazione non lineare e di interpretare il coefficiente in termini di variazioni %
- la funzione logaritmica è la funzione inversa della funzione esponenziale

$$f(x) = e^{x} \;\rightarrow\; f^{-1}(x) = log_{e}(x) = ln(x)$$

- ovvero

$$x = ln(e^{x})$$

proprietà di ln

- $ln(1/x) = -ln(x)$
- $ln(ax) = ln(a) + ln(x)$
- $ln(x/a) = ln(x) - ln(a)$
- $ln(x^a) = a ln(x)$

ln e variazioni %

- la proprietà che ci interessa di più (valida se ΔX è piccolo)

$$ln(x + \Delta x) - ln(x) \cong \frac{\Delta x}{x}$$

- dove il simbolo $\cong$ indica 'approssimativamente uguale a'

regressione di funzioni logaritmiche

- 3 modelli possibile
  1. X è espressa in logaritmi Y no
  2. Y è espressa in logaritmi X no
  3. X e Y sono entrambe espresse in logaritmi

modello lineare-logaritmico

- se X è logaritmica e Y no

$$Y_i = \beta_0 + \beta_1 ln(X_i) + u_i$$

- ad una variazione dell'1% di X si associa una variazione di $0.01\beta_1$ di Y
- infatti

$$\Delta X \rightarrow \Delta Y = [\beta_0 + \beta_1 ln(X + \Delta X)] - [\beta_0 + \beta_1 ln(X)] =$$

$$\Delta Y = \beta_1 ln(X + \Delta X) - \beta_1 ln(X) \cong \beta_1 \frac{\Delta X}{X}$$

modello log-lineare

- se è la Y ad essere espressa in forma logaritmica ma la X no
- $ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$
- a seguito di ΔX il valore di Y è $ln(Y + \Delta Y)$:

$$ln(Y + \Delta Y) - ln(Y) = [\beta_0 + \beta_1(X + \Delta X)] - [\beta_0 + \beta_1 X]$$

$$\frac{\Delta Y}{Y} \cong \beta_1 \Delta X$$

modello log-log

- nel caso in cui entrambe le variabili siano espresse in termini logaritmici

$$ln(Y_i) = \beta_0 + \beta_1 ln(X_i) + u_i$$

- β è la variazione percentuale di Y dovuta ad una variazione percentuale di X: $(\eta_{Y,X}!)$

Modello log-log

$$ln(Y_i) = \beta_0 + \beta_1 ln(X_i) + u_i$$

$$ln(Y + \Delta Y) - ln(Y) = \beta_0 + \beta_1 ln(X + \Delta X) - [\beta_0 + \beta_1 ln(X)]$$

$$\frac{\Delta Y}{Y} \cong \beta_1 \frac{\Delta X}{X}$$

$$\beta_1 = \eta_{X,Y}$$

Elasticità del reddito all'età (Svezia)

	coefficiente	errore standard	t	$p-value$
β_0	9.4536	0.1983	47.655	0.000
$\beta_{ln(age)}$	0.16609	0.05163	3.217	0.0013

$R^2 = 0.01232$, $R^2 -$ corretto $= 0.01113$

errore standard di regressione $= 0.5324$