

PRINCIPI DI ECONOMETRIA

lezione 8

AA 2015-2016

Paolo Brunori

Ripasso

- oggi ripassiamo le cose essenziali viste fino ad ora
- per passare l'esame è necessaria padronanza di questi argomenti
- se avete dubbi tornate alle slide delle lezioni 1-7
- consultate risorse on-line o uno dei libri di testo

il mantra: correlazione non è causalità

- i nessi di causalità sono i meccanismi di funzionamento
- una domanda dell'esame potrebbe riguardare il riconoscimento di nessi di causalità
- di due fenomeni A e B, qual'è causa e quale effetto?
- potrebbero essere entrambi causa ed entrambi effetto?

il modello lineare

- in generale: $Y_i = \beta_0 + \beta_1 X_i + u_i$ è il modello di regressione lineare con un solo regressore
- Y : variabile dipendente
- X : variabile indipendente o regressore
- coefficienti o paramenti:
 - β_0 : intercetta ($Y|X = 0$)
 - β_1 : pendenza ($\Delta Y = \beta_1$ volte ΔX)
- u : errore di disturbo
- il pedice i si riferisce alla specifica osservazione di un individuo i

la migliore approssimazione lineare:

retta dei minimi quadrati

- migliore la capacità di spiegare i dati della retta minori gli errori commessi
- $u_i = Y_i - \beta_0 - \beta_1 X_i$ è l'errore di interpolazione per l'osservazione i-esima
- gli stimatori dei minimi quadrati $\hat{\beta}_0, \hat{\beta}_1$ minimizzano:

$$\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 = \sum_{i=1}^n u_i^2$$

stimatori OLS β_0, β_1

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

dove \bar{X}, \bar{Y} sono le medie delle due variabili nel campione

assunzioni che rendono valido quanto detto fino ad ora

- ▶ ovviamente $\hat{\beta}_0, \hat{\beta}_1$ se non osserviamo tutta la popolazione dipendono dal campione estratto... sono anch'essi variabili casuali
- ▶ questo metodo per stimare i parametri della retta non è sempre valido
- ▶ perché i parametri siano una buona stima di quelli veri, β_0, β_1 , devono essere soddisfatte alcune condizioni

assunzioni che rendono valido quanto detto fino ad ora

- 0) il DGP deve effettivamente essere lineare
- 1) la distribuzione di u_i condizionata a X_i ha media nulla
- 2) X, Y sono indipendentemente e identicamente distribuite
- 3) valori estremi (outlier) devono essere improbabili

DGP non lineare

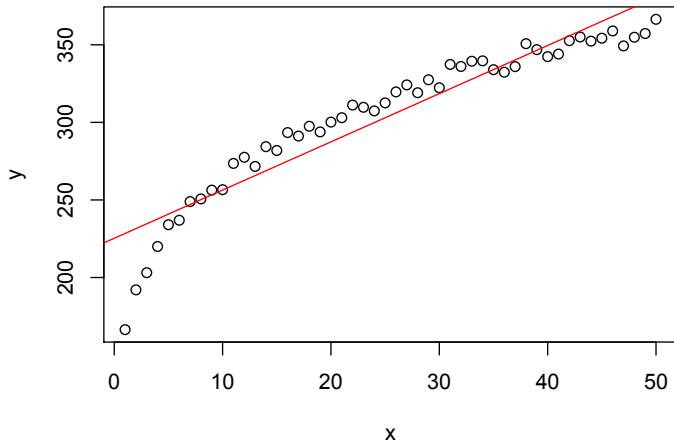
PRINCIPI DI
ECONOMETRIA

LEZIONE 8

I DATI

REGRESSIONE
LINEARE
UNIVARIATA

ASSUNZIONI OLS



Distribuzione di u_i condizionata a X_i non ha media nulla

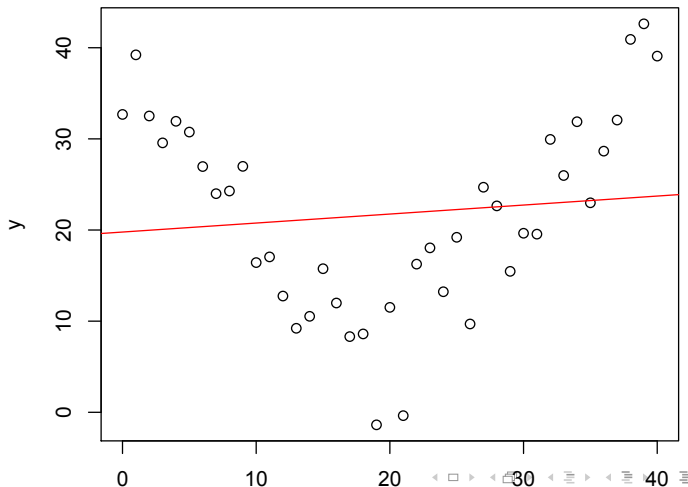
PRINCIPI DI
ECONOMETRIA

LEZIONE 8

I DATI

REGRESSIONE
LINEARE
UNIVARIATA

ASSUNZIONI OLS



Outlier non improbabili

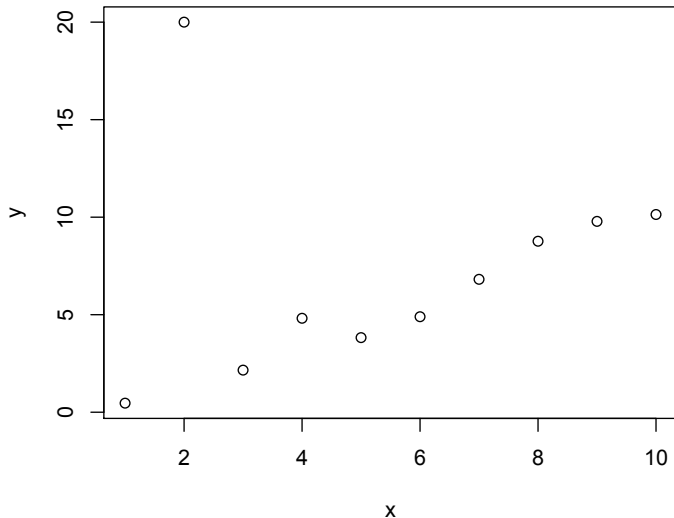
PRINCIPI DI
ECONOMETRIA

LEZIONE 8

I DATI

REGRESSIONE
LINEARE
UNIVARIATA

ASSUNZIONI OLS



- ▶ se le assunzioni 1-3 si verificano sono non distorti:
 $E(\hat{\beta}_0) = \beta_0$ e $E(\hat{\beta}_1) = \beta_1$
- ▶ inoltre se n è abbastanza grande la loro distribuzione è ben approssimata da una normale bivariata
- ▶ si può applicare il teorema del limite centrale:

$$\hat{\beta}_0 \sim (\beta_0, \sigma_{\beta_0})$$

$$\hat{\beta}_1 \sim (\beta_1, \sigma_{\beta_1})$$

- ▶ una volta stimato un modello di regressione OLS possiamo valutare il livello di certezza che si associa alle nostre stime che dipende:
 - 1 dalla numerosità campionaria
 - 2 dalla variabilità di X
- ▶ la stima della statistica $\hat{t} = \frac{\beta}{\sqrt{\text{var}(\hat{\beta})}}$ ci consente di:
 - testare ipotesi sui parametri
 - stimare intervalli di confidenza per i parametri
 - calcolare la probabilità che il vero parametro sia zero dato il campione che osserviamo

- è difficile ipotizzare che una sola variabile indipendente spieghi il comportamento della variabile dipendente
- in generale infatti u cattura tutte quelle variabili che influenzano Y ma non sono considerate/osservabili
- l'omissione di una variabile Z distorce lo stimatore OLS se si verificano due condizioni :
 1. Z è una delle variabili che determina Y
 2. $\text{corr}(X, Z) \neq 0$
- ogni volta che si stima una regressione ci si deve chiedere: è possibile che la stima dei coefficienti sia distorta da una o più variabili omesse?
- riesco a immaginare il segno della distorsione?

regressione multipla

- seppur difficile da vedere graficamente Y può essere una funzione di molte variabili
- per un numero generico k di regressori il modello di regressione multipla prende la forma:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_k X_{k,i} + u_i$$

stimatore OLS della regressione multipla

- gli stimatori $\hat{\beta}_i$ che vengono normalmente utilizzati sono quelli che minimizzano:

$$\sum_{i=1}^n (Y_i - b_0 - b_1 X_{1,i} - \dots - b_k X_{k,i})^2$$

- la terminologia è la stessa usata per il modello con un regressore:

stimatori OLS: $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$

valore predetto: \hat{Y}_i

residuo \hat{u}_i

assunzioni necessarie per OLS

- le assunzioni sono le stesse già viste + 1
- i regressori mostrano collinearità perfetta se uno dei regressori è funzione lineare degli altri
- se ad esempio $X_2 = a - bX_1$ o $X_2 = \frac{X_1}{100}$
- $\Delta X_1 = \Delta X_2 + \Delta X_3$, caso tipico quando queste sono variabili dicotomiche indicanti per esempio le tre aree di residenza (Nort, Centro, Sud)

Regressione multipla del consumo di tabacco

	coefficiente	errore standard	t	$p - value$
β_0	1.6572	0.1237	13.394	0.0000
β_Y	0.0003	0.0000	6.518	0.0000
β_P	-0.4231	0.096	-4.3662	0.0001

- per passare l'esame bisogna saper interpretare tutti i numeri di una tabella come questa!
- in caso di eteroschedasticità del termine errore i coefficienti rimangono corretti ma i loro errori standard sono sottostimati

errori omoschedastici

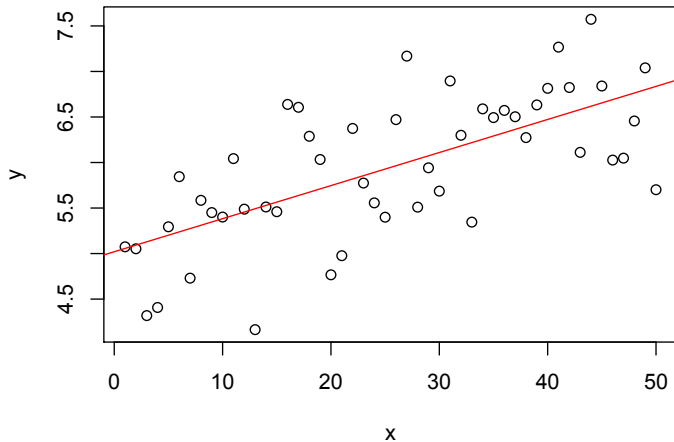
PRINCIPI DI
ECONOMETRIA

LEZIONE 8

I DATI

REGRESSIONE
LINEARE
UNIVARIATA

ASSUNZIONI OLS



errori eteroschedastici

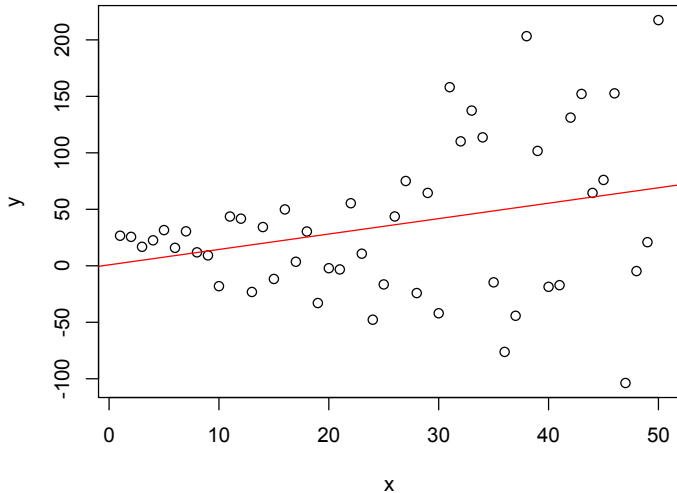
PRINCIPI DI
ECONOMETRIA

LEZIONE 8

I DATI

REGRESSIONE
LINEARE
UNIVARIATA

ASSUNZIONI OLS



Potere esplicativo del modello

- se tutte le assunzioni del modello sono vere possiamo interpretare i coefficienti
- può inoltre essere utile valutare quanto precisamente riusciamo a spiegare la variabilità del fenomeno
- $R^2 = 1 - \frac{\sum_i^n \hat{u}_i^2}{\sum_i^n (y_i - \bar{y})^2}$
- $adj - R^2$ o R^2 -corretto = $1 - (1 - R^2) \frac{n-1}{n-k-1}$
- $MSE = \frac{1}{n} \sum_i^n (y_i - \hat{y}_i)^2$ (scarto quadratico medio o errore quadratico medio)

Regressione multipla robusta per eteroschedasticità

	coefficiente	errore standard	t	$\text{valore} - p$
β_0	1.6572	0.1746	9.4859	0.0000
β_Y	0.0003	0.0000	3.599	0.0014
β_P	-0.4231	-0.24	-1.7494	0.09202

$$R^2 = 0.6406$$

$$\bar{R}^2 = 0.6129$$

$$\sqrt{MSE} = 0.1513$$

interazioni tra regressori

$$Y_i = \beta_0 + \beta_1 D_{1,i} + \beta_2 D_{2,i}$$

- dove $D_{1,i}$ = laureata/o e $D_{2,i}$ = sesso
- in questo modello la laurea ha un effetto identico per i due gruppi
- è possibile considerare un possibile effetto di interazione fra le due variabili:

$$Y_i = \beta_0 + \beta_1 D_{1,i} + \beta_2 D_{2,i} + \beta_3 (D_{1,i} \times D_{2,i})$$

l'effetto di avere una laurea

$$Y_i = \beta_0 + \beta_1 D_{1,i} + \beta_2 D_{2,i} + \beta_3 (D_{1,i} \times D_{2,i})$$

- $E(Y|D_{1,i} = 0, D_{2,i} = d_2) = \beta_0 + \beta_1 \times 0 + \beta_2 d_2 + \beta_3 \times 0 \times d_2 = \beta_0 + \beta_2 d_2$
- $E(Y|D_{1,i} = 1, D_{2,i} = d_2) = \beta_0 + \beta_1 + \beta_2 d_2 + \beta_3 \times d_2 = \beta_0 + \beta_1 + (\beta_2 + \beta_3) d_2$
- l'effetto di avere o meno una laurea D_1 dipende dal sesso (D_2): $\beta_1 + \beta_3 d_2$

interazione fra variabili binarie e continue

- D = laurea o meno, X =esperienza lavorativa.
- modello base:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i$$

- in pratica stiamo stimando due rette parallele di regressione
- ma nella realtà le prospettive di carriera dipendono dall'istruzione
- anche qui si può aggiungere un termine di interazione

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 D_i X_i + u_i$$

- se si tratta di un laureato/a la funzione di regressione è:

$$Y_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X_i + u_i$$

- se non è laureato/a:

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

- per cui ho intercette diverse e pendenze diverse