

Principi di Econometria

lezione 3

AA 2016-2017

Paolo Brunori

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

econometria: la ricerca dei processi che generano i dati

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

- domenica era brutto tempo
- se mi fermo alla pasticceria a fare colazione troverò:
più coda del solito? meno? la stessa?
- cosa determina quante persone decidono di fare
colazione al bar un certo giorno a una certa ora?

dove siamo arrivati?

- quello che vogliamo fare è capire se i dati ci possono insegnare qualcosa su come funzionano i fenomeni socioeconomici
- il primo passo - una volta identificata un possibile modello di funzionamento di un fenomeno - consiste nell'approssimare la relazione fra causa ed effetto con un'approssimazione lineare
- il fenomeno che ci interessa (felicità percepita nel nostro esempio) si avvicina ad un andamento costituito da una costante e una funzione crescente di una variabile indipendente (PIL pro capite)
- questa relazione è riscontrabile nei dati? di quanto aumenta la felicità per ogni dollaro in più di PIL?

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

la migliore approssimazione lineare: *retta dei minimi quadrati*

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

- OLS per Ordinary Least Squares in inglese
- 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:

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

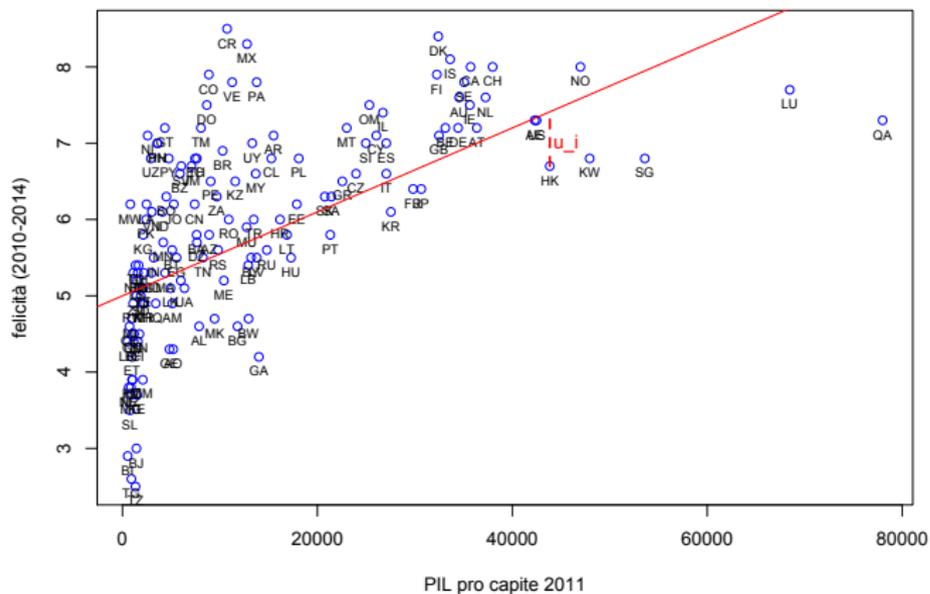
l'errore di previsione di un modello lineare

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS



fonte: World Bank (2011), World Happiness Report (2015)

interpretazione del modello lineare in generale

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

β_0 : intercetta, è il valore che ipotizziamo abbia la Y quando $X = 0$

β_1 : pendenza, mi dice di quanto aumenta Y quando X aumenta di un'unità

u_i : errore, mi indica di quanto sbaglio ad approssimare linearmente la relazione che lega X e Y

stimatori OLS β_0, β_1

- per ottenere gli stimatori $\hat{\beta}_0, \hat{\beta}_1$ dei due coefficienti si pongono pari a zero le due derivate parziali:

$$\frac{\partial}{\partial \beta_0} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 = 0$$

$$\frac{\partial}{\partial \beta_1} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 = 0$$

- la dimostrazione potete trovarla sul libro (e in fondo a queste slide)
- le soluzioni sono semplici e vanno ricordate

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

OLS

Bontà di
adattamento

assunzioni OLS

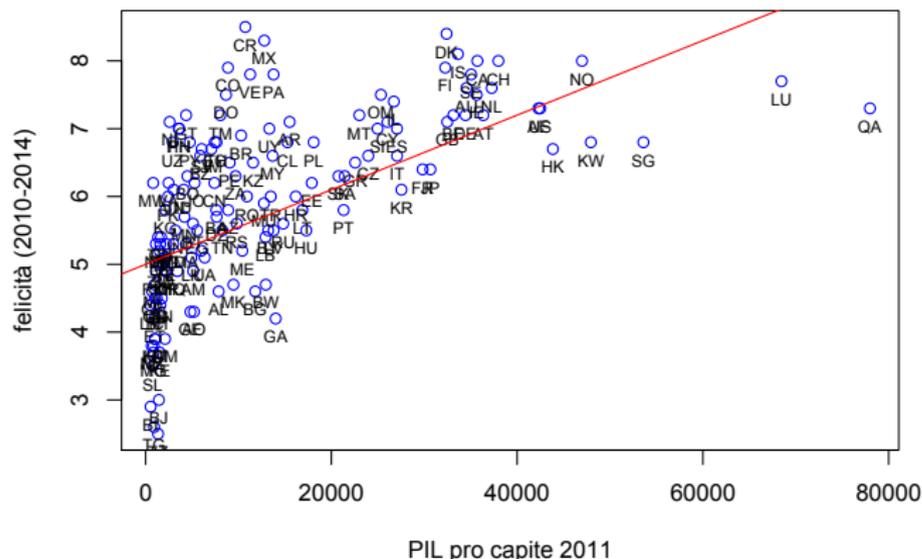
Appendice:
derivazione
stimatori OLSstimatori OLS β_0, β_1

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

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

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

retta dei minimi quadrati



fonte: elaborazione su dati World Bank e WHR

i parametri stimati sono: $\hat{\beta}_0 = 5.19$, $\hat{\beta}_1 = 0.00005$
come possono essere interpretati?

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

interpretazione β_0, β_1

- β_0 : se un paese ha PIL pro capite=0 la media della risposta della domanda di Cantril è 5.19
- β_1 : per ogni dollaro in più di PIL mi aspetto un aumento di 0.00005 della media dell'indicatore di Cantril (equivalente a dire che per ogni 10 mila dollari in più l'aumento è di 0.5 punti)

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

Come potete ottenere queste stime in pratica?

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

- otteniamo i dati
- sistemiamo i dati in modo leggibile da un software

i dati caricati su R

code	name	cantril	GDP_pc
AF	Afghanistan	4.1	NA
AL	Albania	4.6	7861.131
DZ	Algeria	5.7	7643.171
AD	Andorra	6.8	NA
AO	Angola	4.3	5201.309
AR	Argentina	7.1	15501.42
AM	Armenia	4.9	5112.398
AU	Australia	7.6	34548.32
AT	Austria	7.2	36353
AZ	Azerbaijan	5.8	8889.891
BD	Bangladesh	5.3	1568.438
BY	Belarus	5.5	13191.19
BE	Belgium	7.2	33126.52
BZ	Belize	6.6	5895.771
BJ	Benin	3	1427.694
BT	Bhutan	5.6	5095.597

dataset ottenuto unendo informazioni da: World Bank (2011), World Happiness Report (2015)

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

Come potete ottenere queste stime in pratica?

- il nostro modello è:

$$cantril = \beta_0 + \beta_1 GDP_pc$$

- per alcuni paesi la variabile indipendente GDP_pc (PIL pro capite) non è disponibile
- per questi dati non è possibile stimare la retta di regressione
- utilizziamo quelli per i quali osserviamo sia Y che X

Come potete ottenere queste stime in pratica?

- il modo più ovvio è ricorrere alle formule:

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

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

- calcoliamo \bar{X} e \bar{Y} e poi utilizziamo la formula
- i passaggi che vi faccio vedere ora sono disponibili sul file nella pagina del corso

- quanto bene la retta interpola i dati?
- R^2 della regressione è la frazione della varianza campionaria di Y spiegata da X ($var(\hat{Y}_i)/var(Y_i)$)
- somma dei quadrati spiegata (ESS: Explained Sum of Squares) = $\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$
- somma dei quadrati totali (TSS) = $\sum_{i=1}^n (Y_i - \bar{Y})^2$

$$R^2 = \frac{ESS}{TSS}$$

- somma dei quadrati dei residui (SSR) = $\sum_{i=1}^n \hat{u}_i^2$

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}$$

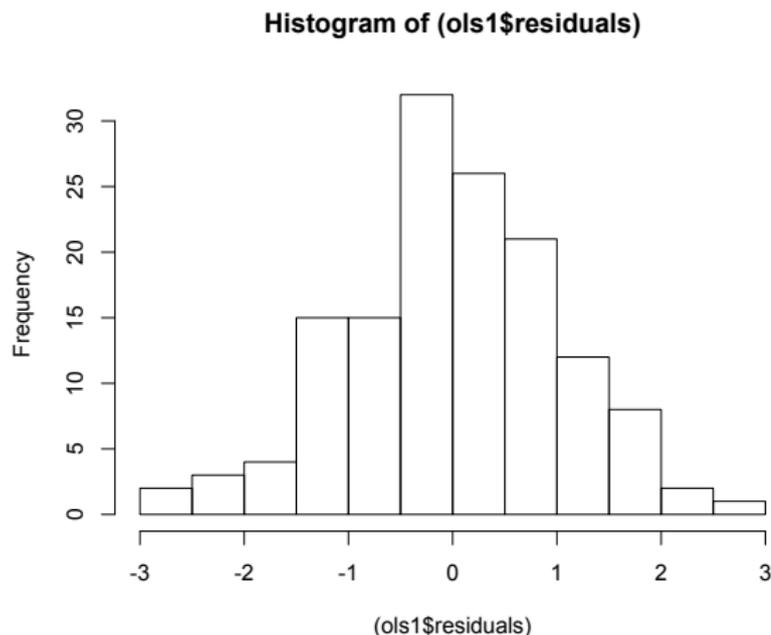
- se $\beta_1 = 0$ $R^2 = 0$
- se $\hat{Y}_i = Y_i \forall i = 1, \dots, n$ allora $R^2 = 1$

Quanto bene il reddito spiega la felicità?

- torniamo all'analisi empirica
- quanto sbagliamo ad approssimare la relazione fra X e Y?
- guardiamo ai residui u_i

Quanto bene il reddito spiega la felicità?

- il grafico riporta la distribuzione degli errori della stima di Y dato X
- come possiamo giudicare se si tratta di errori grandi o meno?



Quanto bene il reddito spiega la felicità?

- proviamo ad usare il buon senso, l'errore in valore assoluto in media è pari a 0.79
- in aggregato il 38.42% della variabilità è spiegata dal modello

assunzioni che rendono valido quanto detto fino ad ora

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

- questo metodo per stimare i parametri della retta $(\hat{\beta}_0, \hat{\beta}_1)$ 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

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

- 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

la distribuzione di u_i condizionata a X_i ha
media nulla

OLS

Bontà di
adattamento

assunzioni OLS

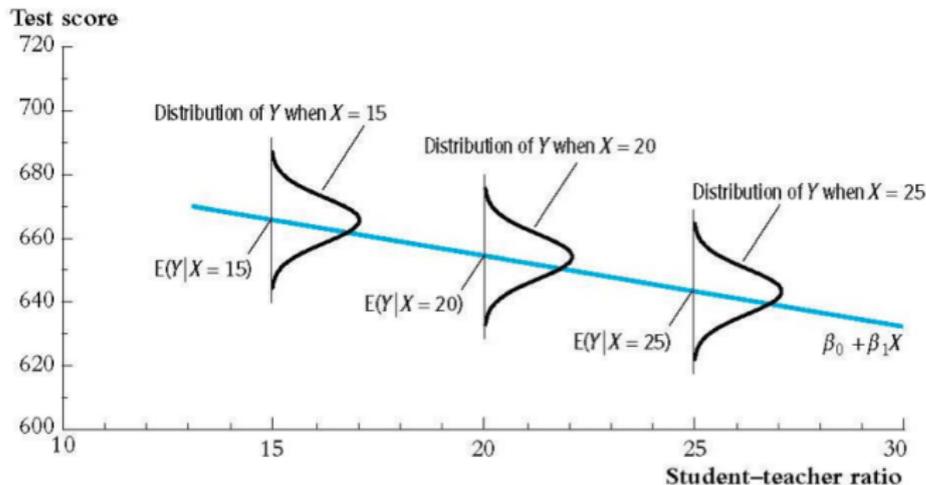
Appendice:
derivazione
stimatori OLS

$$1) E(u_i|X_i) = 0 \quad \forall i = 1, \dots, n$$

- gli ‘altri fattori’ che confluiscono in u e determinano Y non sono sistematicamente legati a X

la distribuzione di u_i condizionata a X_i ha
media nulla

$$1) E(u_i|X_i) = 0 \quad \forall i = 1, \dots, n$$



OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

X, Y i.i.d.

- campionamento semplice dalla stessa popolazione (età e altezza degli studenti di uniba) allora ogni osservazione si distribuisce alla stessa maniera
- se sono estratti in modo casuale sono anche indipendenti
- esistono casi di non-indipendenza: nel caso dei dati della Banca Mondiale ad esempio sono mancanti i valori di PIL per paesi molto arretrati
- in questi casi il campionamento non è casuale, il campione non rappresenta la popolazione

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

gli outlier sono improbabili

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

- outlier: misure fuori dall'intervallo 'normale'
- potrebbero essere dovuti a errori di imputazione dei dati

$\hat{\beta}_0, \hat{\beta}_1$ sono stimatori non-distorti

- $\hat{\beta}_0, \hat{\beta}_1$ sono variabili casuali: il loro valore dipende dal campione selezionato
- se valgono le condizioni il loro valore si distribuisce attorno al vero valore (β_0, β_1)
- come accade per la media di un campione: la media campionaria è uno stimatore non distorto della vera media della popolazione
- così accade per i parametri $\hat{\beta}_0, \hat{\beta}_1$ se si verificano le condizioni

Appendice: derivazione stimatori OLS

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

Le slide che seguono non fanno parte della parte essenziale del programma d'esame. Se avete la curiosità di capire come si ottengono gli stimatori dei minimi quadrati potete consultarle.

stimatori OLS β_0, β_1

si pongono pari a zero le due derivate parziali:

$$\frac{\partial}{\partial \beta_0} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 = 2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\frac{\partial}{\partial \beta_1} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) X_i = 0$$

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

dividendo per n :

$$\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \rightarrow \bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{X} = 0$$

$$\frac{1}{n} \sum_{i=1}^n (Y_i X_i) - \hat{\beta}_0 \bar{X} - \hat{\beta}_1 \frac{1}{n} \sum_{i=1}^n X_i^2 = 0$$

stimatori OLS $\hat{\beta}_0, \hat{\beta}_1$

sostituiamo

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

nella seconda equazione:

$$\frac{1}{n} \sum_{i=1}^n (Y_i X_i) - (\bar{Y} - \hat{\beta}_1 \bar{X}) \bar{X} - \hat{\beta}_1 \frac{1}{n} \sum_{i=1}^n X_i^2 = 0$$

$$\frac{1}{n} \sum_{i=1}^n (Y_i X_i) - \bar{Y} \bar{X} - \hat{\beta}_1 \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right) = 0$$

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

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

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

numeratore:

$$\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^n (X_i Y_i - X_i \bar{Y} - \bar{X} Y_i + \bar{X} \bar{Y}) =$$

$$\sum_{i=1}^n (X_i Y_i) - \bar{Y} \sum_{i=1}^n X_i - \bar{X} \sum_{i=1}^n Y_i + N \bar{X} \bar{Y} =$$

$$\sum_{i=1}^n (X_i Y_i) - \bar{Y} N \bar{X} - \bar{X} N \bar{Y} + N \bar{X} \bar{Y} = \sum_{i=1}^n (Y_i X_i) - N \bar{Y} \bar{X}$$

OLS

Bontà di
adattamento

assunzioni OLS

Appendice:
derivazione
stimatori OLS

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

denominatore

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i^2 - 2X_i \bar{X} + \bar{X}^2) = \\ &= \sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + N\bar{X}^2 = \sum_{i=1}^n X_i^2 - N\bar{X}^2 \end{aligned}$$