Principi di Econometria

lezione 4

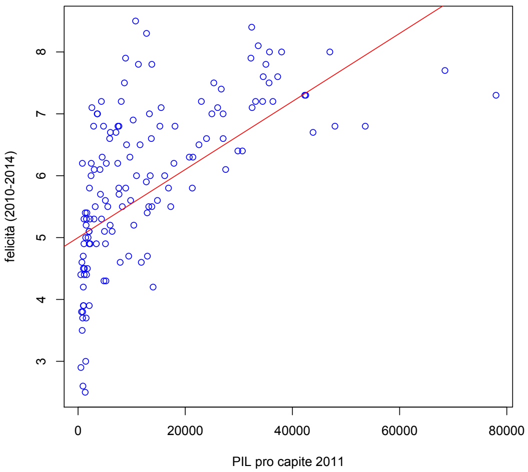
AA 2016-2017

Paolo Brunori

dove eravamo arrivati

- abbiamo individuato la regressione lineare semplice (OLS) come modo immediato per sintetizzare una relazione fra una variabile dipendente (Y) e una indipendente (X)
- i parametri β_0 e β_1 che consideriamo migliori sono quelli tali per cui la somma degli errori di interpilazione al quadrato è minima: $\hat{\beta}_0$ e $\hat{\beta}_1$
- ma le stime che otteniamo sono variabili casuali che dipendono dal campione osservato
- sotto alcune assunzioni però possiamo stare tranquilli, le stime che otteniamo hanno valore atteso pari ai veri parametri nella popolazione (in media ci azzecchiamo)

retta dei minimi quadrati



fonte: elaborazione su dati World Bank

assunzioni che rendono valido quanto detto fino ad ora

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

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

Inferenza

- 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:

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

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

inferenza

- il teorema del limite centrale ci consente di sfruttare le informazioni contenute nel campione per quantificare l'incertezza delle stime ottenute
- quanto è probabile che la relazione che osservo fra Y e X sia veramente positiva?
- è possibile, ammettendo un certo grado di incertezza, individuare una stima minima e massima per il valore del coefficiente?

inferenza

- il grado di incertezza delle nostre stime è quantificabile guardando alla varianza dei due stimatori:

$$var(\hat{\beta}_0) = \frac{1}{n}\frac{var(H_i u_i)}{[E(H_i^2)]^2}$$

dove $H_i = 1 - \left[\frac{\mu_x}{E(X_i^2)}\right] X_i$

$$var(\hat{\beta}_1) = \frac{1}{n}\frac{var[(X_i - \mu_x)u_i]}{[var(X_i)]^2}$$

dove σ^2 è la varianza del termine errore u

inferenza

- il grado di incertezza delle nostre stime è quantificabile guardando alla varianza dei due stimatori:

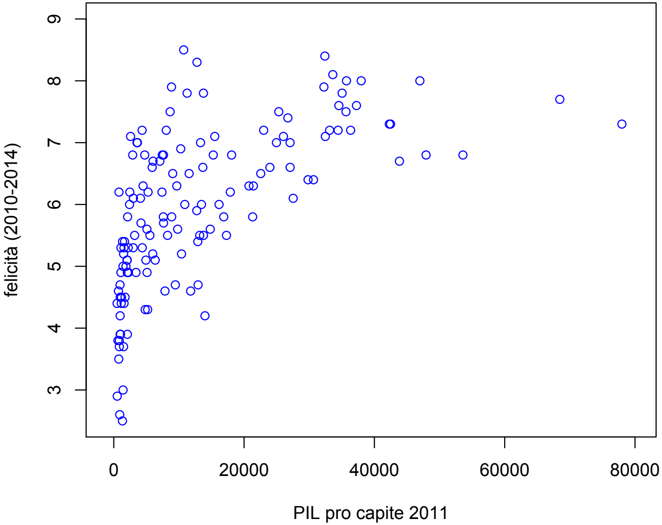
$$var(\hat{\beta}_0) = \sigma^2 \frac{\sum_i^n X_i^2}{\sum_i^n (X_i - \bar{X})^2}$$

$$var(\hat{\beta}_1) = \frac{\sigma^2}{n \sum_i^n (X_i - \bar{X})^2}$$

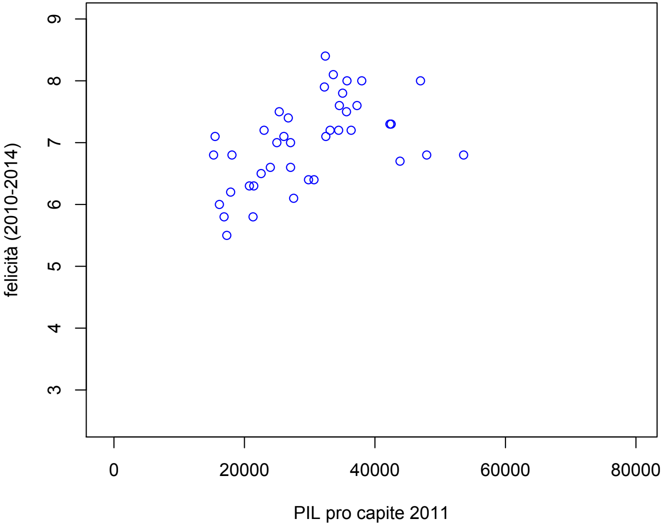
dove σ^2 è la varianza del termine errore u

- tanto maggiore σ^2 tanta più alta è l'incertezza riguardo ai coefficienti
- tanto minore la variabilità di X tanto più alta l'incertezza riguardo ai coefficienti

inferenza: ruolo della variabilità della X



inferenza: ruolo della variabilità della X



test delle ipotesi

- se è valido il teorema del limite centrale sarà anche possibile effettuare test delle ipotesi
- “con quale grado di confidenza possiamo affermare che un aumento del PIL si traduce in un aumento della felicità?”
- ciò equivale a testare l’ipotesi:

$$H_0 : \ \beta_1 = 0$$

contro l’ipotesi:

$$H_1 : \ \beta_1 \neq 0$$

test delle ipotesi

- per effettuare il test si deve costruire la statistica $\hat{t}$:

$$\hat{t} = \frac{\hat{\beta}_1 - \beta_1^{H_0}}{\sqrt{var(\hat{\beta}_1)}}$$

- rifiutiamo l'ipotesi nulla (H_0) se $\hat{t}$ è significativamente diversa da zero
- l'ipotesi H_0 può essere rigettata se:

$$|\hat{t}| > t^{\star}$$

- $t^{\star}$ è il valore critico che dipende dal numero di osservazioni (gradi di libertà) e dal livello di confidenza
- nel nostro caso 139 gradi di libertà e 95% → 1.96

intervalli di confidenza

- allo stesso modo possiamo individuare l'intervallo nel quale al 95% si trova il vero parametro β_1
- se l'intervallo vero al 95% include la nostra ipotesi nulla (in questo caso $\beta_1 = 0$) non possiamo rifiutarla
- l'intervallo si ottiene con la formula:

$$\left[\hat{\beta}_1 - \sqrt{var(\hat{\beta}_1)} \times |t^\star|, \quad \hat{\beta}_1 + \sqrt{var(\hat{\beta}_1)} \times |t^\star|\right]$$

$p - value$

- ci possiamo anche chiedere quale sia la probabilità che l'ipotesi nulla sia vera dato il valore $\hat{t}$ che osserviamo
- il *valore* $- p$ si ottiene confrontando la statistica $\hat{t}$ con la probabilità di osservare un valore così estremo nella distribuzione della t
- con 139 gradi di libertà un $p - value$ pari a 0.05 si associa a una statistica $t^\star = 1.96$
- se osserviamo un $\hat{t} = 9.314$ la probabilità che β_1 sia pari a 0.00000000000002% !
- se n è maggiore di qualche decina possiamo abbandonare la t e usare i valori critici della normale standardizzata z

inferenza: sommario

- 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 dell'errore
  3. dalla variabilità di X
- la stima della statistica $\hat{t}$ ci consente di:
  - testare ipotesi sul parametri
  - stimare intervalli di confidenza per i parametri
  - calcolare la probabilità che il vero parametro sia zero dato il campione che osserviamo

Felicità e reddito

	coefficiente	se	t	$p - value$
β_0	5.197	0.1184	43.893	0.0000
PIL (β_1)	0.00005	0.0000	9.314	0.0000

- gradi di libertà: 139 (141 osservazioni - uno per ogni parametro stimato)
- errore standard dei residui: 1.031
- $R^2 = 0.3843$