Econometria

Valerio Potì

1

Introduzione e richiami di probabilità e statistica

- **Problema empirico:** Dimensione della classe e risultato dell'istruzione
  - Qual è l'effetto sui punteggi nei test (o su un'altra misura di risultato) se riduciamo la dimensione delle classi (poniamo) di uno studente per classe?
  - E se la riduciamo di 8 studenti per classe?
  - Dobbiamo utilizzare i dati per rispondere
- Richiami di probabilità e statistica

2

2

I dati dei punteggi nei test della California

Tutti i distretti scolastici K-8 della California (n = 420)

Variabili:

- TESTSCR (Punteggi nei test del quinto anno) := combinazione del risultato di prove di matematica e lettura, media del distretto
- STR (Rapporto studenti/insegnanti) := numero di studenti nel distretto diviso per numero di insegnanti a tempo pieno equivalente

3

3

Primo sguardo ai dati:
(dovreste già sapere come interpretare questa tabella)

Tabella 4.1 Sintesi della distribuzione del rapporto studenti/insegnanti e del punteggio nei test relativa al quinto grado d'istruzione (quinta elementare) per 420 distretti K-8 in California nel 1998.

			Percentile						
	Media	**Deviazione standard**	10%	25%	40%	50% (mediana)	60%	75%	90%
Rapporto studenti/insegnanti	19,6	1,9	17,3	18,6	19,3	19,7	20,1	20,9	21,9
Punteggio nei test	654,2	19,1	630,4	640,0	649,1	654,5	659,4	666,7	679,1

Questa tabella non ci dice nulla sulla relazione tra punteggi nei test e rapporto studenti/insegnanti.

4

4

Riprodurre la tabella in Gretl

1) Caricare i dati in Gretl usando 'file -> Apri dati -> File utente -> caschool.xls', avendo cura di cercare caschool.xls nel folder nel quale lo si è salvato.
2) Nella finestra di dialogo che si apre, seleziona prima TESTSCR e poi STR e si esegua poi per ciascuna 'Variabile -> Statistiche descrittive'.
3) Si applichi poi la funzione **quantile*** sia a TESTSCR che a STR per trovare i quantili seguenti:

0.1, .25, .4, .5, .6, .75, .9

*Si veda esempio nella slide che segue…

5

5

Riprodurre la tabella in Gretl

Esempio relativo al quantile 0.1 di STR:

- Il comando **quantile** è applicato a STR, con 0.1 come ulteriore argomento:

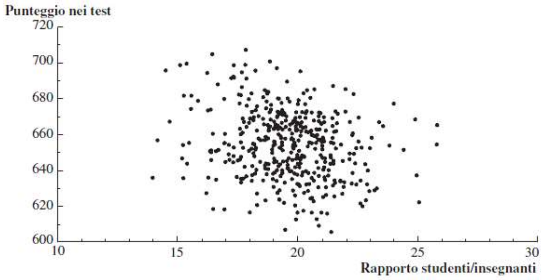
  quantile(STR,0.1)
- N.B.:
  - Gretl distingue tra minuscolo e maiuscolo per il nome delle variabili.
  - Le variabili nel file Excel di dati sono in carattere minuscolo e vengono caricate come tali in Gretl.
  - Quindi, occorre occorre scriverle in minuscolo nel comando di cui sopra.

6

6

I distretti con classi più piccole ottengono punteggi più elevati nei test?

Diagramma a nuvola di punteggio nei test e STS



Che cosa mostra questa figura?

7

7

Riprodurre la figura in Gretl

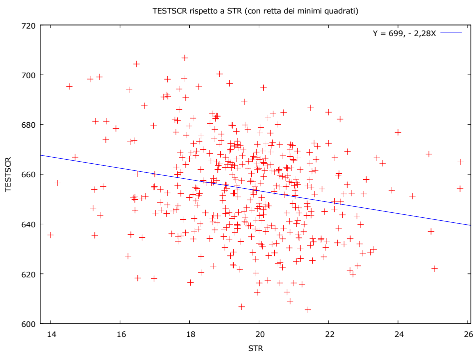
1) Avendo già caricato i dati in Gretl come da slide precedente, si selezionino (cliccandole) le variabili STR e TESTSCR
2) Facendo ‘right-click’, si fa saltare fuori una finestra di dialogo con diverse opzioni. Si selezioni quella denominata “Grafico X-Y a dispersione”
3) Nella finestra di dialogo successiva, che chiede di selezionare la variabile in ascissa (“Variabile asse X”), si scrollino le opzioni disponibili scegliendo “STR”.
4) Dovrebbe venir fuori una figura come quella nella slide successiva (che aggiunge per default la linea dei minimi quadrati).

In alternativa, si selezioni «Visuallizza->Grafico->X-Y a dispersione…» e si seguano le indicazioni sulla schermata che ne risulta.

8

8

Riprodurre la figura in Gretl



9

9

Dobbiamo capire <u>se</u>, <u>quanto</u> e <u>perché</u> i distretti con basso STR hanno punteggi nei test più alti – ma come?

1. Quanto?

- Confrontare i punteggi nei test nei distretti con basso STR a quelli con alto STR ("stima")
- Regressione

2. Ma è vero o è un'illusione?

- test di ipotesi
- intervalli di confidenza

3. Se è vero, come si spiega?

- Causalità

10

10

E ora…

- Per rispondere alla prima domanda, rivediamo ed approfondiamo alcuni elementi di **statistica descrittiva**
  - Media condizionata
  - Correlazione
  - Regressione
- Per rispondere alla seconda domanda, ci servono elementi di **statistica inferenziale**
  - Elementi della teoria alla base della stima dei modelli statistici e di verifica di ipotesi
- Per rispondere alla terza domanda, ci serve il **ragionamento** (**identificazione econometrica** del modello che genera i dati)

11

11

E ora…

- Ci occupiamo **<u>prima del primo e terzo</u>** di questi temi, e poi del secondo (più complesso sul piano matematico e statistico e forse anche più noiosetto perché meno saliente)
- Per poter ignorare (per ora) il secondo problema, assumiamo di avere a che fare con **osservazioni sull'intera popolazione, ovvero di un <u>campione che include tutti gli individui della popolazione</u>**
- Ciò ci consente di ignorare per il momento la incertezza che deriva dall'errore di campionamento
- Quando la introdurremo, rivisiteremo anche gli altri due temi

12

12

Capitoli 1, 2.1-2.3, 3.1, 3.5, 3.7

QUADRO DI RIFERIMENTO PROBABILISTICO E STATISTICHE DESCRITTIVE

13

13

Quadro di riferimento probabilistico e statistiche descrittive

a) Popolazione, variabile casuale e distribuzione
b) Momenti di una distribuzione (media, varianza, deviazione standard, covarianza, correlazione)
c) Distribuzione condizionata e media condizionata

Questi sono argomenti approfonditi negli **esercizi di richiamo** delle basi di probabilità e statistica descrittiva.

14

14

(a) Popolazione, variabile casuale e distribuzione

Popolazione
- Il gruppo o l'insieme di tutte le possibili unità che ci interessano (la totalità dei distretti scolastici)

Variabile casuale Y
- Rappresentazione numerica del risultato di un esperimento casuale (punteggio medio nei test del distretto, STR del distretto)

Distribuzione di Y nella popolazione
- Le probabilità dei diversi valori di Y che si verificano nella popolazione
  - Per esempio, quando Y è discreta, $Pr[Y = 650]$
- Oppure le probabilità di insiemi di questi valori
  - Per esempio, quando Y è continua o discreta, $Pr[640 \le Y \le 660]$

15

15

(b) Momenti di una distribuzione (in popolazione)

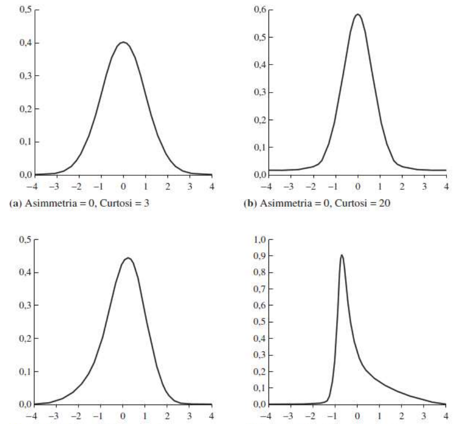
Media := valore atteso (aspettativa) di *Y*
$:= E(Y) = \sum_{i=1} Pr(Y = y_i)y_i$
$:= \mu_Y$

Varianza $:= E(Y - \mu_Y)^2 = \sum_{i=1} Pr(Y = y_i)(y_i - \mu_Y)^2$
$:= \sigma_Y^2$
:= misura della dispersione quadratica della distribuzione

Deviazione standard = $\sqrt{\text{varianza}} = \sigma_Y$
:= misura della dispersione della distribuzione

16

16



(a) Asimmetria = 0, Curtosi = 3
(b) Asimmetria = 0, Curtosi = 20
(c) Asimmetria = −0,1, Curtosi = 5
(d) Asimmetria = 0,6, Curtosi = 5

17

17

Momenti (continua)

- ***asimmetria*** (o skewness) = $\frac{E[(Y-\mu_Y)^3]}{\sigma_Y^3} = \frac{\sum_{i=1} Pr(Y=y_i)(y_i-\mu_Y)^3}{\sigma_Y^3}$

  = misura di asimmetria di una distribuzione

  - *asimmetria* = 0: la distribuzione è simmetrica
  - *asimmetria* > (<) 0: la distribuzione ha una coda lunga destra (sinistra)

- ***curtosi*** = $\frac{E[(Y-\mu_Y)^4]}{\sigma_Y^4} = \frac{\sum_{i=1} Pr(Y=y_i)(y_i-\mu_Y)^4}{\sigma_Y^4}$

  = misura di massa nelle code

  = misura di probabilità di valori grandi

  - *curtosi* > 3: code pesanti (“***leptocurtica***”)
  - distribuzione normale ⟹ {*asimmetria* = 0 & *curtosi* = 3}

- **Etc., etc.**

18

18

Momenti (distribuzione congiunta)

- Le variabili casuali X e Z hanno una ***distribuzione congiunta***
- La ***covarianza*** tra X e Z è

$$\text{cov}(X,Z) := E[(X - \mu_X)(Z - \mu_Z)] := \sigma_{XZ}$$

- La covarianza è una misura dell'associazione lineare tra X e Z; le sue unità sono unità di $X \times$ unità di Z
  - $\text{cov}(X,Z) > 0$ significa una relazione positiva tra X e Z
  - Se X e Z sono indipendentemente distribuite, allora $\text{cov}(X,Z) = 0$ (ma non vale il vice-versa!!)
- La covarianza di una variabile casuale con se stessa è la sua varianza:

$$\text{cov}(X,X) = E[(X - \mu_X)(X - \mu_X)] = E[(X - \mu_X)^2] = \sigma_X^2$$

19

19

Momenti (distribuzione congiunta)

- Risultato utile:

$$\begin{aligned}\text{cov}(X,Z) &:= E[(X - \mu_X)(Z - \mu_Z)] \\ &= E[XZ - X\mu_Z - Z\mu_X + \mu_X\mu_Z] \\ &= E(XZ) - E(X\mu_Z) - E(Z\mu_X) + E(\mu_X\mu_Z) \\ &= E(XZ) - E(X)\mu_Z - E(Z)\mu_X + \mu_X\mu_Z \\ &= E(XZ) - \mu_X\mu_Z - \mu_Z\mu_X + \mu_X\mu_Z \\ &= E(XZ) - \mu_X\mu_Z\end{aligned}$$

- Dunque:
  $\text{var}(X) = \text{E}(X^2) - \mu_X^2$

20

20

La covarianza tra *Punteggio nei test* e *Rapporto studenti/insegnanti* è negativa:

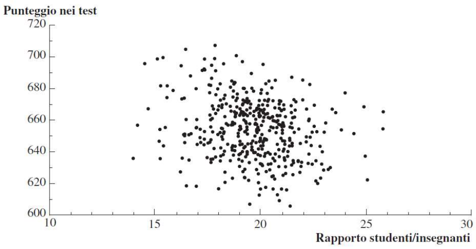


Figura 4.2
Diagramma a nuvola del punteggio nei test e del rapporto studenti/insegnanti (dati relativi ai distretti scolastici della California).
Dati per i 420 distretti scolastici della California. C'è una debole relazione negativa tra il rapporto studenti/insegnanti e il punteggio nei test: la correlazione campionaria è pari a −0,23.

E così la ***correlazione***…

21

21

Il ***coefficiente di correlazione*** è definito in termini di covarianza:

$$\text{corr}(X,Z) := \frac{\text{cov}(X,Z)}{\sqrt{\text{var}(X)\text{var}(Z)}} = \frac{\sigma_{X,Z}}{\sigma_X \sigma_Z} := r_{XZ}$$

$$-1 \leq \text{corr}(X,Z) \leq 1$$

corr(X,Z) = 1 significa associazione lineare positiva perfetta

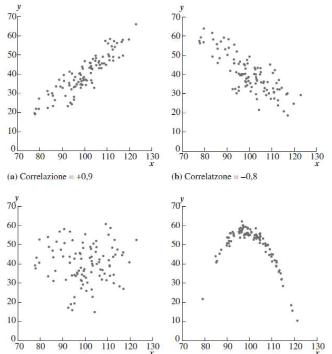
corr(X,Z) = −1 significa associazione lineare negativa perfetta

corr(X,Z) = 0 significa che non c'è associazione lineare

22

22

Il coefficiente di correlazione misura l'associazione lineare



(a) Correlazione = +0,9 (b) Correlazione = −0,8 (c) Correlazione = 0,0 (d) Correlazione = 0,0 (quadratica)

23

23

(c) Distribuzione e media condizionate

Distribuzione condizionata
- La distribuzione di Y dato il valore (o i valori) di un'altra variabile casuale X
- Es: la distribuzione dei punteggi nei test dato STR < 20

Media condizionata
- *Media condizionata* := media della distribuzione condizionata = $E(Y|X = x)$ (***concetto e notazione importanti***)

Altri momenti condizionati
- *Varianza condizionata* := varianza della distribuzione condizionata

24

24

Media condizionata (continua)

Esempi:
- Salari di tutte le lavoratrici femmine (Y = salari, X = genere)
- Tasso di mortalità di pazienti che ricevono una cura sperimentale (Y = vivo/morto; X = trattato/non trattato)
- Se $E(X|Z)$ = costante, allora corr(X,Z) = 0 (tuttavia non vale necessariamente il vice versa)

La media condizionata è un termine (forse nuovo) utilizzato per il concetto familiare di media di gruppo

25

25

Momenti (calcolo)

- Se osserviamo <u>tutta</u> la popolazione, la probabilità che le osservazioni sui caratteri degli individui ricadano in una data classe di misura coincide con la frequenza con cui ciò avviene
- Dunque la probabilità di ciascuna realizzazione della variabile aleatoria con cui descriviamo la distribuzione del carattere in popolazione coincide con la frequenza relativa della realizzazione stessa
- Quindi…

26

26

Momenti (calcolo)

- Quindi, poiché in una popolazione di n individui i valori di un dato carattere rappresentano tutti i valori possibili y_i della variabile aleatoria Y che ne si descrive la distribuzione, si ha che

$$\Pr(Y = y_i) = \frac{1}{n}\sum_{i=1}^{n} 1(Y = y_i)$$

- Per conseguenza, i momenti si possono calcolare come medie aritmetiche
- Ad esempio:
  - **Media** := $E(Y) := \sum_{i=1}^{n} p_i y_i = \frac{1}{n}\sum_{i=1}^{n} y_i$
  - **Varianza** := $E(Y - \mu_Y)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_Y)^2$
  - **Co-varianza** := $E[(Y - \mu_Y)(X - \mu_X)] = \frac{1}{n}\sum_{i=1}^{n}[(x_i - \mu_X)(y_i - \mu_Y)]$
  - **Etc.**

27

27

Esempio sulle scuole californiane

- Supponiamo le scuole di cui abbiamo i dati siano tutta la «popolazione» delle possibili scuole californiane
- Come calcoliamo l'impatto della dimensione delle classi sul rendimento scolastico?
- Un modo di definire quest'impatto è come differenza media del rendimento in classi piccole e grandi, ad esempio con meno e più di 20 studenti per insegnante (piccole e grandi)

28

28

Esempio sulle scuole californiane

- Ma la differenza media non è altro che la differenza tra le medie di due distribuzioni condizionate:

$$\Delta = E(\textit{Punteggio test}|\textit{STR} < 20) - E(\textit{Punteggio test}|\textit{STR} \geq 20)$$

- Ove
  - $E(\textit{Punteggio test}|\textit{STR} < 20)$:= media dei punteggi nei test tra i distretti con dimensioni delle classi piccole
  - $E(\textit{Punteggio test}|\textit{STR} \geq 20)$:= media dei punteggi nei test tra i distretti con dimensioni delle classi grandi

29

29

Analisi dei dati: confrontare i distretti con dimensioni delle classi "piccole" (STR < 20) e "grandi" (STR ≥ 20) :

Dimensione classe	*n*	Punteggio medio ($\bar{Y}$)
Piccola	238	657,4
Grande	182	650,0

***Ci interessa* Δ := differenza tra medie dei gruppi**

Per lavorare con un sotto-insieme delle osservazioni in Gretl, si fa così:
1. "Campione -> Imposta in base a condizione"
2. Specificare la condizione (per esempio, STR<20) nella finestra di dialogo risultante
3. Poi, per ciascun sottocampione, si selezione la variabile che interessa (per es., TESTSCR) e da menu si sceglie "Variabile -> Statistiche descrittive"

30

30

Esempio sulle scuole californiane

$\Delta = E(Punteggio\ test|STR < 20) - E(Punteggio\ test|STR \geq 20)$

$$= \bar{Y}_{piccola} - \bar{Y}_{grande} := \left(\frac{1}{n_{piccola}}\sum_{i=1}^{n_{piccola}} Y_i\right) - \left(\frac{1}{n_{grande}}\sum_{i=1}^{n_{grande}} Y_i\right)$$

$$= \quad 657{,}4 \quad - \quad 650{,}0$$

$$= 7{,}4$$

31

31

Media condizionata e regressione

- Un modo per calcolare una media condizionata, ad esempio E(Y|X), è quello di calcolare una regressione di Y su X
- Ci occuperemo soprattutto di regressioni lineari con uno o più regressori (le X)

32

32

Capitolo 4.1-4.3

REGRESSIONE LINEARE CON UN REGRESSORE

33

33

Sommario

1. Il modello di regressione lineare
2. Misure di adattamento (bontà) della regressione

34

34

La regressione lineare consente di calcolare la pendenza della retta di regressione.

- La pendenza della retta di regressione è l'effetto atteso su *Y* di una variazione unitaria in *X*.
- Il nostro scopo è quello di calcolare l'effetto su *Y* di una variazione unitaria in *X* e quindi questa pendenza
- Ci interessa anche sapere quanto si adatti bene questa retta ai dati sulle due variabili *Y* e *X* in una popolazione, cioè se sia una buona descrizione della loro relazione

35

35

Il modello di regressione lineare semplice

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \qquad i = 1,\ldots, n$$

- Ci sono n individui/unità, (X_i, Y_i) ed $i = 1,.., n,$ nella popolazione, ed abbiamo osservazioni su ciascuno/ciascuna
- *X* è la ***variabile indipendente*** ovvero il ***regressore***
- *Y* è la ***variabile dipendente***
- β_0 = ***intercetta***
- β_1 = ***pendenza***
- u_i = ***errore*** *di regressione*
  - L'errore di regressione è costituito da fattori omessi.
  - In generale questi fattori omessi sono altri fattori, diversi dalla variabile *X*, che influenzano *Y*.
  - L'errore di regressione include anche l'eventuale errore di misura di *Y*.

36

36

Condizione di identificazione (esogeneità debole)

$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1,\ldots, n$

Importantissimo:

l'errore u_i dev'essere tale che

$E(u_i | X_i) = 0$!!!!!!!

Solo così abbiamo che

$E(Y_i|X_i) = \beta_0 + \beta_1 X_i$

37

37

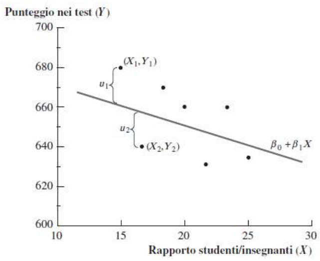
Il modello di regressione in un'immagine

- Supponiamo ci siano solo sette scuole (n = 7) nella "popolazione" di scuole californiane
- Mostriamo la retta di regressione e l'errore di regressione (il "termine d'errore"):

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

$E(u_i | X_1) = 0$

$E(Y_i | X_1) = \beta_0 + \beta_1 X_i$



38

38

Il metodo OLS per calcolare la regressione

Come possiamo calcolare β_0 e β_1 dai dati?

Possiamo usare il **metodo dei minimi quadrati (OLS, "*ordinary least squares*")** per trovare i parametri ignoti β_0 e β_1:

$$\min_{b_0,b_1} \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_i)]^2$$

Un caso speciale è la media aritmetica, $\bar{Y}$, che è il valore atteso di Y, μ_Y, nella popolazione secondo OLS

$$\min_{b_0} \sum_{i=1}^{n} (Y_i - b_0)^2$$

39

39

OLS: $$\min_{b_0,b_1} \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_i)]^2$$

- OLS minimizza la differenza quadratica media tra i valori di Y_i e valore atteso secondo la retta di regressione.
- Questo problema di minimizzazione si può risolvere con il calcolo differenziale
- **Il risultato sono $\boldsymbol{\beta_0}$ e $\boldsymbol{\beta_1}$.**
- **Si veda il secondo capitolo del "Compendio su OLS", dal titolo "Regressione (elementi essenziali"**

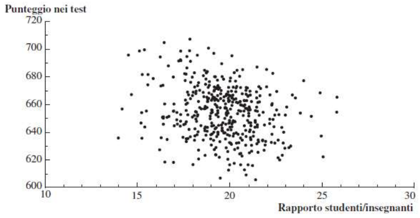
40

40

Applicazione ai dati sui punteggi nei test della California

La retta di regressione: $E(TestScore|STR) = \beta_0 + \beta_1\, STR$

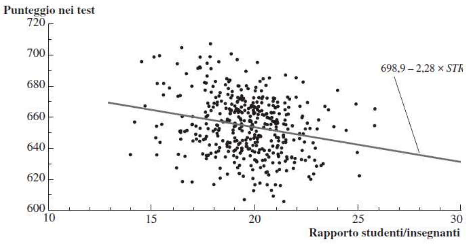
$$\beta_1 = \frac{\Delta E(TestScr|STR)}{\Delta STR} = ?? \quad \left(versione\ sofisticata\ di\ \ \frac{\Delta TestScore}{\Delta STR}\right)$$



41

41

Applicazione ai dati sui punteggi nei test della California



- Pendenza calcolata = $\boldsymbol{\beta_1}$ = – 2,28
- Intercetta calcolata = $\boldsymbol{\beta_0}$ = 698,9
- Retta di regressione: $E(TestScr|STR)$= 698,9 – 2,28 × STR

42

42

Interpretazione di pendenza e intercetta nell'esempio

$$E(TestScore|STR) = 698{,}9 - 2{,}28 \times STR$$

- I distretti con uno studente in più per insegnante in media ottengono punteggi nei test inferiori di 2,28 punti.
- Cioè:

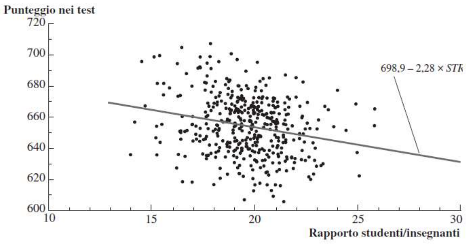
$$\frac{\Delta E(\text{TestScore}|\text{STR})}{\Delta STR} = -2.28$$

  - L'intercetta (letteralmente) significa che, secondo questa retta, i distretti con zero studenti per insegnante otterrebbero un punteggio nei test stimato in 698,9.
  - Ma questa interpretazione dell'intercetta non ha senso – estrapola la linea al di fuori dell'intervallo dei dati – in questo caso, l'intercetta non ha significato dal punto di vista economico.

43

43

Valori predetti ed errori/residui



Uno dei distretti nella banca dati è Antelope, CA, con *STR* = 19,33 e *TestScore* = 657,8

Valore atteso: $E(TestScore|STR) = 698{,}9 - 2{,}28 \times 19{,}33$

Errore: $u_{Antelope} = 657{,}8 - 654{,}8 = 3{,}0$

44

44

Regressione OLS: output di Gretl

```
Modello 1: OLS, usando le osservazioni 1-420
Variabile dipendente: TESTSCR

             coefficiente   errore std.   rapporto t    p-value
  ---------------------------------------------------------------
  const        698,933        9,46749        73,82      6,57e-242 ***
  STR           -2,27981      0,479826       -4,751     2,78e-06  ***

Media var. dipendente   654,1565   SQM var. dipendente      19,05335
Somma quadr. residui    144315,5   E.S. della regressione  18,58097
R-quadro                0,051240   R-quadro corretto        0,048970
F(1, 418)               22,57511   P-value(F)               2,78e-06
Log-verosimiglianza    -1822,250   Criterio di Akaike       3648,499
Criterio di Schwarz     3656,580   Hannan-Quinn             3651,693
Note: SQM = scarto quadratico medio; E.S. = errore standard
```

La regressione calcolata è

698,9 – 2.28 × *STR*

- Discuteremo più avanti la parte rimanente di questo output, che per il momento non ha senso perchè stiamo assumendo di avere dati su tutta la popolazione, e dunque di non avere errore di campionamento

45

45

Esercizi

- Svolgere esercizi da A1 ad A5 di «Esercizi II.pdf»

46

46

Capitolo 6.1-4

REGRESSIONE LINEARE MULTIPLA
(CON UNO O PIU' REGRESSORI)

47

47

Il modello di regressione multipla

Si consideri il caso di due variabili esplicative:

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

- Y è la *variabile dipendente*
- X_1, X_2 sono le due *variabili indipendenti* (*regressori*)
- (Y_i, X_{1i}, X_{2i}) denotano l'i-esima osservazione su Y, X_1 e X_2.
- β_0 = intercetta della popolazione ignota
- β_1 = effetto su E(Y| X_{1i}, X_{2i}) di una variazione in X_1, tenendo X_2 costante
- β_2 = effetto su E(Y| X_{1i}, X_{2i}) di una variazione in X_2, tenendo X_1 costante
- u_i = errore di regressione (fattori omessi)

NB: Come prima, assumiamo che le n unità sulle quali abbiamo osservazioni siano la popolazione intera

48

48

Condizione di identificazione (esogeneità debole)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + u_i, \ \ i = 1, \dots, n$$

Importantissimo:

l'errore u_i dev'essere tale che

$$E(u_i | X_1, \dots, X_k) = 0 \text{ !!!!!!!}$$

Così che

$$E(Y_i | X_1, \dots, X_k) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_k$$

49

49

Interpretazione dei coefficienti nella regressione multipla

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

Si consideri di variare X_1 di ΔX_1 tenendo X_2 costante:

- Retta attesa di regressione della popolazione ***prima*** della variazione:
  $$E(Y | X_{1i}, X_{2i}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$
- Retta attesa di regressione della popolazione ***dopo*** la variazione:
  $$E(Y | X_{1i}, X_{2i}) + \Delta E(Y | X_{1i}, X_{2i}) =$$
  $$= \beta_0 + \beta_1 (X_1 + \Delta X_1) + \beta_2 X_2$$

50

50

Prima: $E(Y | X_{1i}, X_{2i}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Dopo: $E(Y | X_{1i}, X_{2i}) + \Delta E(Y | X_{1i}, X_{2i}) = \beta_0 + \beta_1 (X_1 + \Delta X_1) + \beta_2 X_2$

Differenza: $\Delta E(Y | X_{1i}, X_{2i}) = \beta_1 \Delta X_1$

Quindi:

$\beta_1 = \frac{\Delta E(Y | X_{1i}, X_{2i})}{\Delta X_1}$, **tenendo X_2 costante**

$\beta_2 = \frac{\Delta E(Y | X_{1i}, X_{2i})}{\Delta X_2}$, **tenendo X_1 costante**

β_0 = valore predetto di Y quando $X_1 = X_2 = 0$

51

51

OLS per la regressione multipla

- Con due regressori, OLS risolve:

$$\min_{b_0,b_1,b_2} \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_{1i} + b_2 X_{2i})]^2$$

- OLS minimizza come prima la differenza quadratica media tra i valori attuali di Y_i e il valore predetto in base alla retta di regressione.
- Questo problema di minimizzazione si risolve usando l'analisi matematica
- **Così si ottengono β_0, β_1 e β_2.**

52

52

Esempio: i dati dei punteggi nei test della California

Regressione di *TestScore* su *STR*:

$$E(TestScore \mid STR) = 698{,}9 - 2{,}28 \times STR$$

- Ora includiamo la percentuale di studenti non di madrelingua nel distretto (*PctEL*):

$$E(TestScore \mid STR) = 686{,}0 - 1{,}10 \times STR - 0{,}65 \times PctEL$$

- Che cosa accade al coefficiente di *STR*?

NB: corr(*STR*, *PctEL*) = 0,19

53

53

Regressione multipla in Gretl

```
Modello 2: OLS, usando le osservazioni 1-420
Variabile dipendente: TESTSCR

             coefficiente   errore std.   rapporto t    p-value
  ---------------------------------------------------------------
  const       686,032        7,41131        92,57      3,87e-280 ***
  STR          -1,10130      0,380278       -2,896     0,0040    ***
  EL_PCT       -0,649777     0,0393425     -16,52      1,66e-047 ***

Media var. dipendente   654,1565   SQM var. dipendente     19,05335
Somma quadr. residui    87245,29   E.S. della regressione  14,46448
R-quadro                0,426431   R-quadro corretto       0,423680
F(2, 417)               155,0136   P-value(F)              4,62e-51
Log-verosimiglianza    -1716,561   Criterio di Akaike      3439,123
Criterio di Schwarz     3451,243   Hannan-Quinn            3443,913
Note: SQM = scarto quadratico medio; E.S. = errore standard
```

La regressione calcolata è

$$E(TestScore|STR) = 686{,}0 - 1{,}10 \times STR - 0{,}65\ PctEL$$

- *Più avanti torneremo su questo stampato per occuparci delle informazioni aggiuntive che fornisce (che per ora non servono per la ragione già esposta nel caso univariato).*

54

54

Bontà dell'adattamento

- Reale = predetto + errore: $Y_i = E(Y_i|X_1,..,X_k) + u_i$
- L'**R^2 della regressione** misura la frazione della varianza di Y spiegata dalle X
- E' priva di unità e può variare tra 0 (nessun adattamento) e 1 (adattamento perfetto)

55

55

L'R^2 della regressione è la frazione della varianza campionaria di Y_i "spiegata" dalla regressione.

$Y_i = E(Y_i|X_1,..,X_k) + u_i$

$= (\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki}) + u_i =$ previsione OLS + errore OLS

→ var. (Y) = var. ($E(Y_i|X_1,..,X_k)$) + var. (u_i) (*perché?*)

→ somma dei quadrati = SS "spiegata" + SS "residua"

Definizione di R^2:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}(\hat{Y}_i-\bar{\hat{Y}})^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i-\bar{\hat{Y}})^2}{\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2} := \frac{Var(\hat{Y})}{Var(Y)}$$

$$\hat{Y}_i := E(Y_i|\beta = \hat{\beta}, X_1,..,X_k) = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki}$$

Dunque

- $R^2 = 0$ significa $ESS = 0$
- $R^2 = 1$ significa $ESS = TSS$
- $0 \le R^2 \le 1$ a condizione che ci sia una constante
- Per la regressione con una singola X, R^2 = il quadrato del coefficiente di correlazione tra X e Y

56

56

Esempio di R^2

- Si confronti l'R^2 prima e dopo aver incluso PctEL, oltre a STR, tra i regressori:

$E(TestScr_i|STR) = 698{,}9 - 2{,}28 \times$ STR, **R^2 = 0,05**

$E(TestScr_i|STR, PctEL) =$

$= 686{,}0 - 1{,}10 \times$ STR − 0,65 PctEL **R^2 = ?**

- Che conclusione traiamo?
- Come potremmo rappresentare graficamente questo risultato?
- Cosa comporta in termini di policy?

57

57

L'R^2 della regressione aggiustato

Definizione:

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

- Contiene una penalizzazione per la mancanza di parsimonia, poiché il termine k appare al denominatore
- Quindi, ceteris paribus, $\bar{R}^2 < R^2$
- Se $k = 1$

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1} = 1 - (1 - R^2)\frac{n-1}{n-2} = 1 - (1 - R^2)\frac{n-1}{n-2}$$

$$= \frac{n-2-(1-R^2)\,(n-1)}{n-2} = \frac{n-2-(1-R^2)n+1-R^2}{n-2}$$

$$= \frac{n-2-n+nR^2+1-R^2}{n-2} = \frac{-1+nR^2-R^2}{n-2} = \frac{R^2(n-1)-1}{n-2}$$

58

58

Per capirci

- Supponiamo che il modello di regressione sia il seguente

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + e_i \qquad \beta_2 \neq 0$$

- Ma noi consideriamo il seguente

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

- Allora succede che la variabile omessa finisce nell'errore e dunque

$$u_i = \beta_2 Z_i + e_i$$

- Ne consegue che

$$\begin{aligned}\rho_{X,u} &= corr(X_i, u_i)\\ &= corr(X_i, \beta_2 Z_i + e_i)\\ &= \beta_2 corr(X_i, Z_i)\end{aligned}$$

- Dunque, a meno che $corr(X_i, Z_i) = 0$, si è in presenza di violazione dell'assunzione di esogeneità debole
- Vediamo qual è la conseguenza se $corr(X_i, Z_i) \neq 0$…

59

59

Per capirci

$$\boxed{\begin{gathered}Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + e_i \qquad \beta_2 \neq 0\\ Y_i = \beta_0 + \beta_1 X_i + u_i\end{gathered}}$$

- Abbiamo che

$$\begin{aligned}E(Y_i|X_i) &= \beta_0 + \beta_1 X_i + E(u_i|X_i)\\ &= \beta_0 + \beta_1 X_i + E(\beta_2 Z_i + e_i|X_i)\\ &= \beta_0 + \beta_1 X_i + E(\beta_2 Z_i|X_i)\\ &= \beta_0 + \beta_1 X_i + \beta_2 E(Z_i|X_i)\\ &= \beta_0 + \beta_1 X_i + \beta_2 \beta_{Z,X} X_i\\ &= \beta_0 + (\beta_1 + \beta_2 \beta_{Z,X})X_i\end{aligned}$$

$$\boxed{\begin{gathered}Z_i = \beta_{Z0} + \beta_{ZX} X_i + e_{Xi}\\ E(Z_i|X_i) = ? = \beta_{Z0} + \beta_{ZX} X_i\\ = \beta_{ZX} X_i \quad \beta_{Z0} = 0\end{gathered}}$$

$$\begin{aligned}Z_i &= \beta_{Z0} + \beta_{ZX}\\ -\beta_{ZX} X_i &= -Z_i +\\ X_i &= -\frac{1}{-\beta_{ZX}} Z_i + \frac{\beta_{Z}}{-\beta_{ZX}}\\ &= -\frac{\beta_{Z0}}{\beta_{ZX}} + \frac{1}{\beta_{ZX}} Z_i\end{aligned}$$

- Quindi, vi è una distorsione pari a

$$\beta_2 \beta_{Z,X} = \beta_2 \frac{Cov(X_i, Z_i)}{Var(X_i)} = \beta_2 \frac{corr(X_i, Z_i)\sigma_X \sigma_Z}{\sigma_X^2} = \beta_2 corr(X_i, Z_i)\frac{\sigma_Z}{\sigma_X}$$

- **Il <u>segno</u> della distorsione è dato dal <u>segno</u> di**

$$\boldsymbol{\beta_2 corr(X_i, Z_i)}$$

60

60

Tabella 6.1 Differenza tra i punteggi nei test dei distretti scolastici della California con bassi e alti rapporti studenti/insegnanti (STR), per percentuali diverse di studenti non di madrelingua inglese nel distretto.

	Rapporto studenti-insegnanti < 20		Rapporto studenti-insegnanti ≥ 20		Differenza tra punteggi, basso v/s alto STR	
	Media punteggi	n	Media punteggi	n	Differenza	Statistica t
Tutti i distretti	**657,4**	**238**	**650,0**	**182**	**7,4**	**4,04**
Percentuale di studenti non di madrelingua inglese	**Il problema è che i distretti con le classi più grandi hanno anche più studenti non di madrelingua**					
< 1,9%	664,5	76	665,4	27	–0,9	–0,30
1,9 – 8,8%	665,2	64	661,8	44	3,3	1,13
8,8 – 23,0%	654,9	54	649,7	50	5,2	1,72
> 23,0%	636,7	44	634,8	61	1,9	0,68

- I distretti con meno studenti non di madrelingua ottengono migliori punteggi nei testi.
- I distretti con una minore percentuale di studenti non di madrelingua hanno classi più piccole.
- Tra i distretti con percentuali di studenti non di madrelingua comparabili, l'effetto della dimensione delle classi è piccolo (si ricordi che complessivamente la "differenza di punteggio nei test" = 7.4).

61

61

Per capirci

- Dunque, poiché $\rho_{X,u} = \beta_2 corr(X_i, Z_i)$, la distorsione dipende proprio dalla correlazione tra errori e variabile inclusa:

$$\beta_2 \beta_{Z,X} = \rho_{X,u} \frac{\sigma_Z}{\sigma_X}$$

- Quindi, quando la assunzione relativa alla esogeneità sia violata, l'analisi di regressiona offre indicazioni fuorvianti
  - Nell'esempio, attribuiamo a STR parte del lavoro che viene fatto da ELPct
- Per un decisore questo può essere un problema

62

62

Distorsione da variabile omessa: Condizioni

- Dunque, PctEL verosimilmente soddisfa i due criteri per la distorsione da variabili omesse
- Ovvero $Z = PctEL$ è

1. Rilevante, in quanto un determinante di Y… ***ed anche…***
2. correlata con il regressore X.

63

63