Principi di Econometria

lezione 5

AA 2016-2017

Paolo Brunori

dove eravamo arrivati

- le stime OLS ci consentono di approssimare linearmente la relazione fra una variabile dipendente (Y) e una indipendente (X)
- i parametri stimati su un campione $\hat{\beta}_0$ e $\hat{\beta}_1$ sono variabili casuali che dipendono dal campione osservato
- sotto alcune assunzioni hanno valore atteso pari ai veri parametri nella popolazione e varianza tanto più piccola tanto maggiore il campione osservato

PIL e diseguaglianza nel mondo

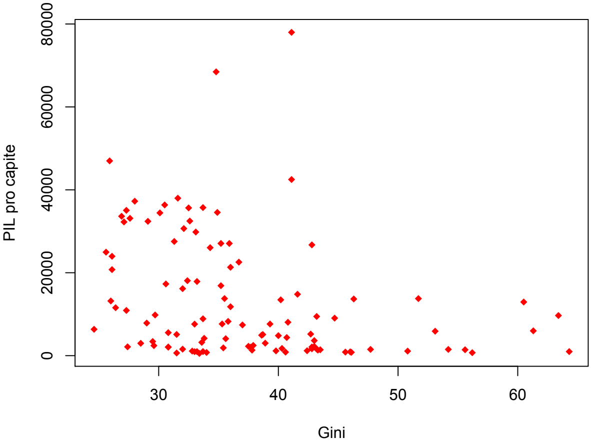
Aggiungiamo una colonna ai nostri dati

row.names	code	name	happiness	GDP_pc	gini
2	AL	Albania	4.6	7861.131	29
3	DZ	Algeria	5.7	7643.171	35.3
5	AO	Angola	4.3	5201.309	42.7
7	AM	Armenia	4.9	5112.398	31.5
8	AU	Australia	7.6	34548.32	34.9
9	AT	Austria	7.2	36353	30.5
10	AZ	Azerbaijan	5.8	8889.891	33.7
11	BD	Bangladesh	5.3	1568.438	32
12	BY	Belarus	5.5	13191.19	26
13	BE	Belgium	7.2	33126.52	27.6
14	BZ	Belize	6.6	5895.771	53.1
15	BJ	Benin	3	1427.694	43.5
16	BT	Bhutan	5.6	5095.597	38.7
18	BA	Bosnia Herzegovina	5.8	7607.442	33
19	BW	Botswana	4.7	12938.95	60.5
21	BG	Bulgaria	4.6	11799.46	36
22	BF	Burkina Faso	4.4	1148.518	39.8
23	BI	Burundi	2.9	533.2931	33.4
24	KH	Cambodia	4.9	2079.985	30.8
25	CM	Cameroon	3.9	2090.283	42.8

Gini

- cos' è l'indice di Gini?
- un'ottima referenza per chi non ricorda è la pagina di Wikipedia
- la domanda che si sono posti molti economisti è se la diseguaglianza favorisca o meno lo sviluppo
- quale potrebbe essere il processo di generazione dei dati?
- $Gini \rightarrow GDP\ p.c.$
- $GDP\ p.c. \rightarrow Gini$

PIL e diseguaglianza nel mondo



Correlazioni

quella fra Gini e PIL pro capite è una correlazione forte?

il termine correlazione è generico ma spesso ci si riferisce all'indice di correlazione di Pearson

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$-1 \geq r \geq 1$

Correlazioni

- relazioni molto diverse hanno stessa correlazione lineare

fonte: Wikipedia

Correlazioni

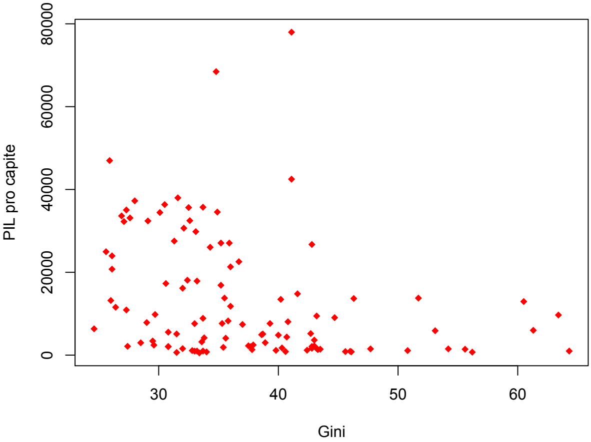
- relazioni molto fort hanno zero correlazione lineare

fonte: Wikipedia

Correlazioni: esercitarsi

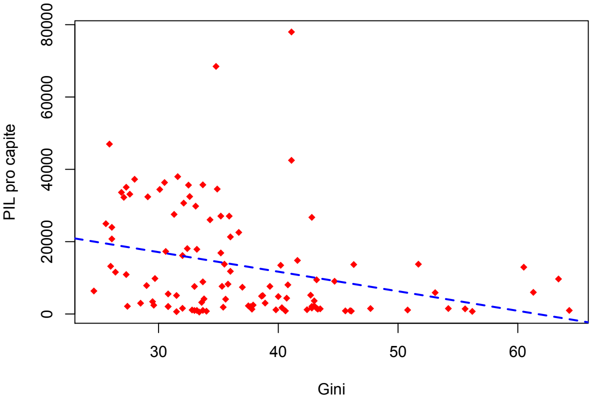
- una delle domande multiple dell'esame riguarda l'identificazione della forza della correlazione fra due variabili

PIL e diseguaglianza nel mondo



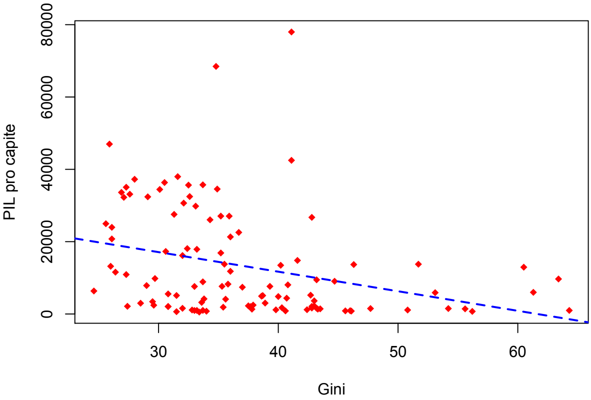
la correlazione è $r = -0.3108$

OLS



che valori per $\hat{\beta}_0$ e $\hat{\beta}_1$?

OLS



$\hat{\beta}_0 = 33335$ e $\hat{\beta}_1 = -541$

effetto ΔX su Y

- come varia Y al variare di X?

$$Y_t = \beta_0 + \beta_1 X_{1,t} \rightarrow Y_{t+1} = \beta_0 + \beta_1 X_{1,t+1}$$

$$\Delta X_1 = X_{1,t+1} - X_{1,t}$$

$$Y_{t+1} = \beta_0 + \beta_1 (X_{1,t} + \Delta X_1)$$

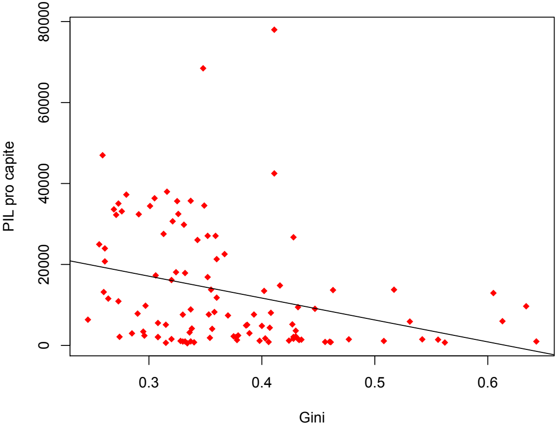
$$Y_{t+1} = \beta_0 + \beta_1 X_{1,t} + \beta_1 \Delta X_1$$

$$Y_{t+1} - Y_t = \beta_1 \Delta X_1$$

$$\Delta Y = \beta_1 \Delta X_1$$

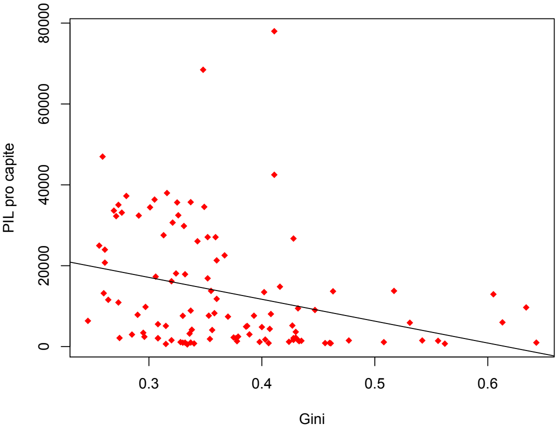
$$\beta_1 = \frac{\Delta Y}{\Delta X_1}$$

OLS: unità di misura



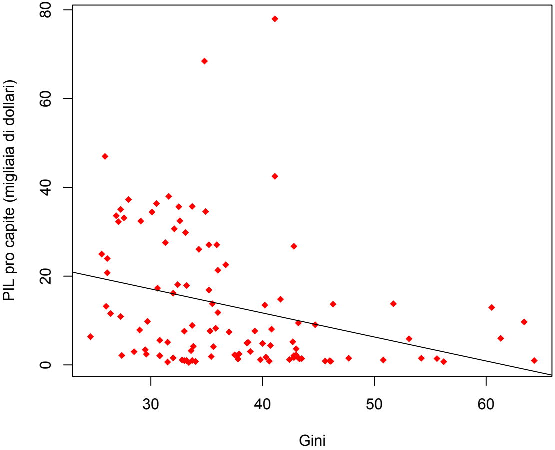
cosa accade ai coefficienti se Gini è espresso nell'intervallo [0,1]?

OLS: unità di misura



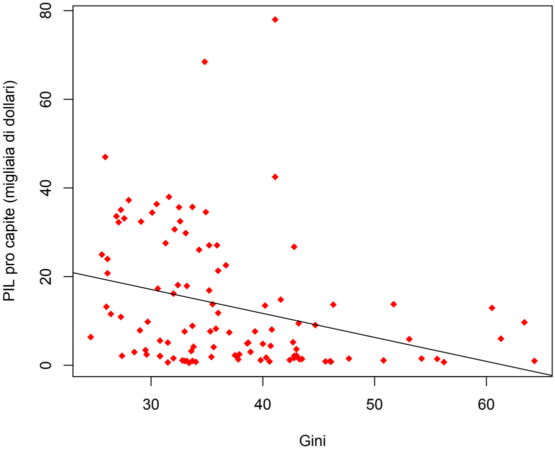
$\hat{\beta}_0 = 33335$ e $\hat{\beta}_1 = -54097$

OLS: unità di misura



cosa accade ai coefficienti se il PIL è espresso in migliaia di dollari?

OLS: unità di misura



$\hat{\beta}_0 = 33.335$ e $\hat{\beta}_1 = -0.541$

Regola generale

- se X è espresso in un'unità di misura superiore (il valore di X è diviso per un multiplo di 10) β_1 si moltiplica per lo stesso multiplo di 10

  - se Y è espresso in un'unità di misura superiore (il valore di Y è diviso per un multiplo di 10) sia β_0 che β_1 si riducono proporzionalmente
- effetto inverso quando le variabili sono espresse in unità di misura inferiore (moltiplicate)

PIL e diseguaglianza nel mondo: inferenza

- domanda 1: quanto bene il nostro modello spiega la relazione?
- domanda 2: i due parametri stimati si riferiscono a un campione con che livello di incertezza possono essere estesi alla popolazione
- domanda 3: siamo sicuri che la direzione della relazione $Gini \rightarrow PIL$ sia quella giusta?

Felicità e reddito

	coefficiente	se	t	$p - value$
β_0	33335.3	6098.5	5.466	0.0000
Gini (%) (β_1)	-541.0	159.9	-3.382	0.0013

- gradi di libertà: 107 (109 osservazioni - uno per ogni parametro stimato)
- $\sqrt{MSE} = 14220$
- $R^2 = 0.09659$

intervallo di confidenza per β_1

usiamo la formula:

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

	coefficiente	se	t	$p - value$
β_0	33335.3	6098.5	5.466	0.0000
Gini (%) (β_1)	-541.0	159.9	-3.382	0.0013

intervallo di confidenza per β_1

usiamo la formula:

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

$$[-541 - 159.9 \times 1.96, \quad -541 + 159.9 \times 1.96]$$

$$\beta_1 \in [-854.40, \quad -227.59] \text{ con probabilità} \geq 95\%$$