Principi di Econometria

lezione 2

AA 2015-2016

Paolo Brunori

causa ed effetto

- quando piove molte persone usano l'ombrello

causa ed effetto

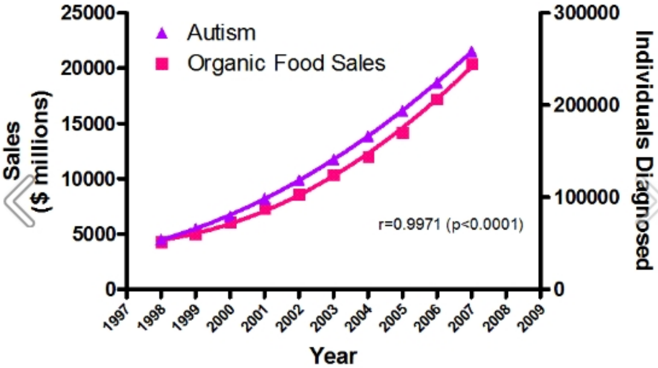
- abbiamo chiaro qualè la causa e quale l'effetto
- ma la correlazione non equivale a causalità
- se osservo gente che usa l'ombrello non è detto che piova
- magari è vero l'opposto

causa ed effetto

il mantra: correlazione non è causalità

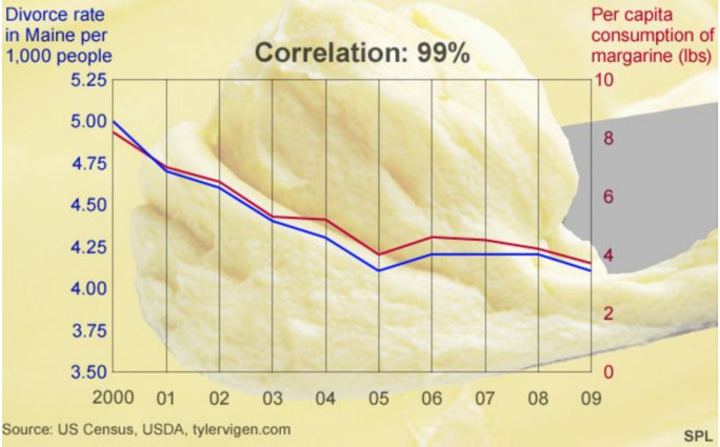
- se vogliamo capire come funziona il mondo (e provare a renderlo migliore) occorre capire come funziona
- i nessi di causalità sono i meccanismi di funzionamento
- quando vedete due cose correlate chiedetevi sempre “è correlazione o causalità?”
- possiamo pensare una teoria che spieghi come uno fenomeno implichi l’altro?
- la nostra teoria è credibile?

causa ed effetto

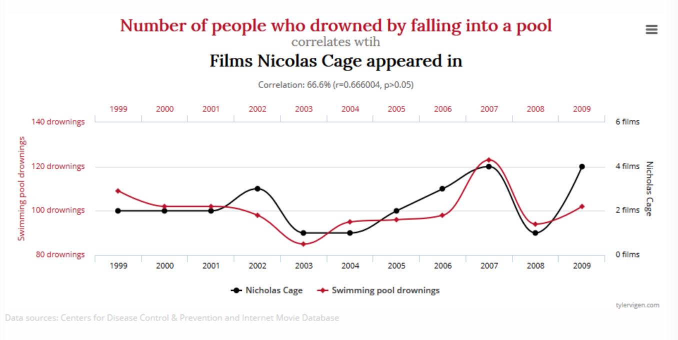


Sources: Organic Trade Association, 2011 Organic Industry Survey; U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS), OMB# 1820-0043: "Children with Disabilities Receiving Special Education Under Part B of the Individuals
with Disabilities Education Act

causa ed effetto



causa ed effetto



il processo che genera i dati (DGP)

- i nostri modelli ci suggeriscono relazioni fra variabili:
  - $\uparrow$ sport $\rightarrow \downarrow$ peso
  - $\downarrow P \rightarrow \uparrow Q_D$
  - $\uparrow P_{VINO} \rightarrow \uparrow$ Qualità
- in questi casi potremmo essere interessati a conoscere l'entità del legame fra le due variabili

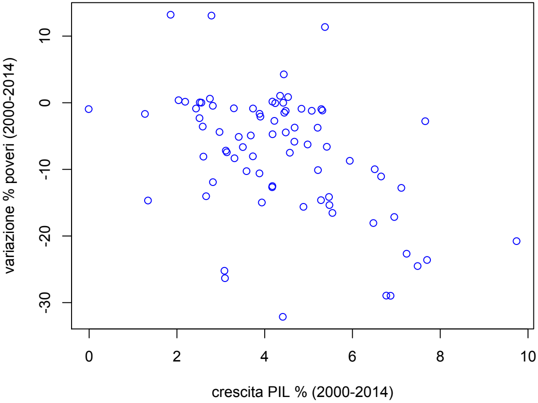
il processo che genera i dati (DGP)

- in alcuni casi è difficile sapere in anticipo qual'è la direzione della relazione:
  - cosa accade alla popolarità del governo se approva le adozioni alle coppie omossessuali?
  - cosa accade al PIL se diminuisco la spesa pubblica?
  - qual'è il momento migliore per comprare un biglietto aereo?
- in questi casi vogliamo capire se esiste una relazione e qual'è il segno della relazione

regressione lineare

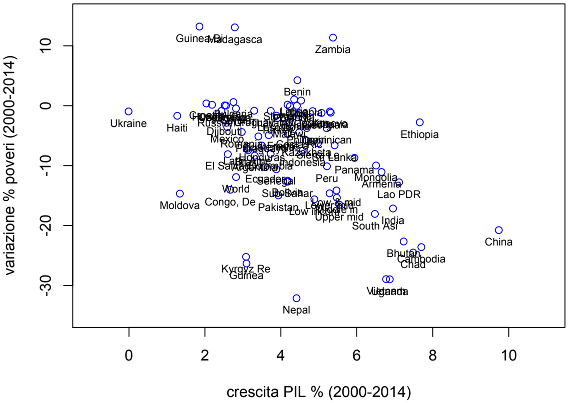
- prima approssimazione: la relazione ipotiziamo sia lineare
- equivalente a dire che $Y = \beta_0 + \beta_1 X$
- dove β_1 è l'effetto su Y di una variazione unitaria di X
- β_0, β_1 sono caratteristiche ignote della distribuzione congiunta X, Y
- ma è possibile stimare β_0, β_1 sulla base di un campione

Diminuzione della povertà e crescita economica



fonte: World Bank

Diminuzione della povertà e crescita economica



fonte: World Bank

da dove provengono questi dati?

- tasso di crescita annuo del PIL
- % di poveri (meno di 1.90 $ al giorno)
- http://data.worldbank.org/

il modello lineare in questo caso

- il nostro DGP ipotizza una relazione:

$$\Delta \text{ Poveri}_i = \beta_0 + \beta_1 \Delta \text{ PIL}_i + u_i$$

- Δ Poveri: variabile dipendente
- Δ PIL: variabile indipendente o regressore
  - β_0: intercetta $(Y|X = 0)$
  - β_1: pendenza (che sospettiamo essere negativa)
- u_i: errore di disturbo

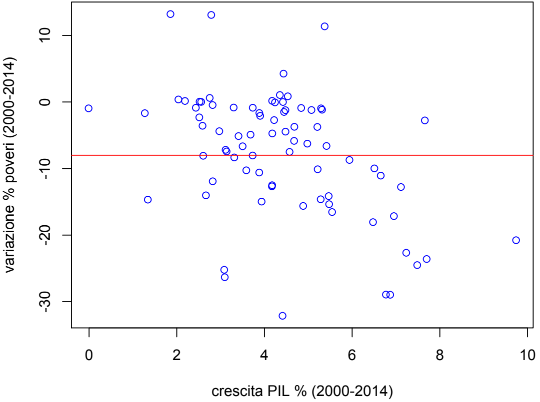
il modello lineare in generale

- 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 paramentri:

  β_0: intercetta $(Y|X = 0)$

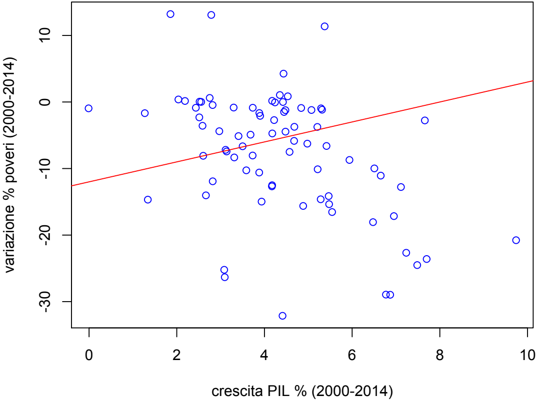
  β_1: pendenza
- u_i: errore di disturbo

la migliore approssimazione lineare



fonte: World Bank

la migliore approssimazione lineare



fonte: World Bank

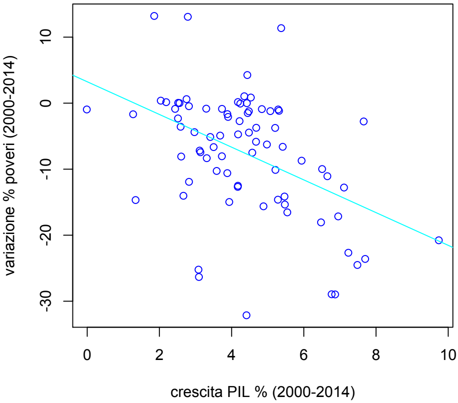
la migliore approssimazione lineare:
retta dei minimi quadrati

- 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:

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

la migliore approssimazione lineare:

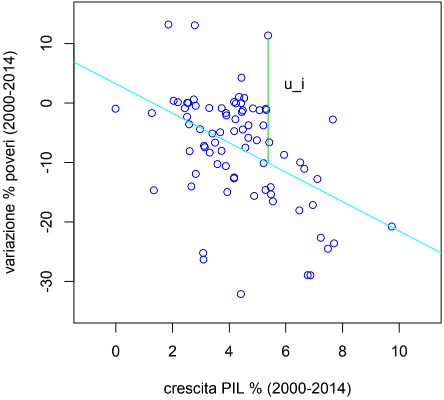
retta dei minimi quadrati



fonte: World Bank

la migliore approssimazione lineare:

retta dei minimi quadrati



fonte: World Bank

stimatori OLS β_0, β_1

- si pongono pari a zero le due derivate parziali:

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

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

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

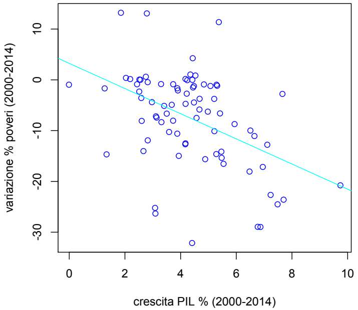
stimatori OLS β_0, β_1

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

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

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

retta dei minimi quadrati



fonte: elaborazione su dati World Bank

i parametri stimati sono:

$\hat{\beta}_0 = 3.246$, $\hat{\beta}_1 = -2.478$

come possono essere interpretati?

interpretazione β_0, β_1

- β_0: se un paese non è cresciuto dal 2000 al 2014 possiamo attenderci un aumento della % di poveri di 3.246 nello stesso periodo

- β_1: per ogni 1% aggiuntivo di crescita mi possiamo aspettarci una riduzione di 2.478% di poveri

Misure di bontà del'adattamento

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

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

misure di bontà del'adattamento

- 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 $\hat{Y}_i = Y_i$ $\forall i = 1, ..., n$ allora $R^2 = 1$

stimatori OLS β_0, β_1

si pongono pari a zero le due derivate parziali:

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

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

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

dividendo per *n*:

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

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

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

sostituiamo

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

nella seconda equazione:

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

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

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

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

numeratore:

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

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

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

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

deniminatore

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X} + \bar{X}^2) =$$

$$\sum_{i=1}^{n}X_i^2 - 2\bar{X}\sum_{i=1}^{n}X_i + N\bar{X}^2 = \sum_{i=1}^{n}X_i^2 - N\bar{X}^2$$