

Principi di Econometria

lezione 2

AA 2016-2017

Paolo Brunori

- la concomitanza di due eventi non ha significato statistico
- se una donna che indossa il velo è disoccupata questo non mi dice nulla sulla discriminazione dei datori del lavoro
- ma una sistematica covarianza di due fenomeni ci possono insegnare qualcosa sul funzionamento del mondo

- se le donne che indossano il velo hanno una probabilità minore di ottenere un lavoro a parità di altre condizioni allora c'è discriminazione
- uno studio pubblicato questo mese dall'economista austriaca Doris Weichselbaumer (2016) ha indagato se vi fosse una probabilità inferiore di essere contattati per un colloquio di lavoro
- ha mandato 1474 mail ad altrettanti posti di lavoro includendo CV (identico) e foto (diverse) di una candidata immaginaria
- verificando l'esistenza di una forte discriminazione verso donne con nome turco e che indossano il velo

correlazione che dimostra discriminazione



Sandra Bauer



Meryem Öztürk



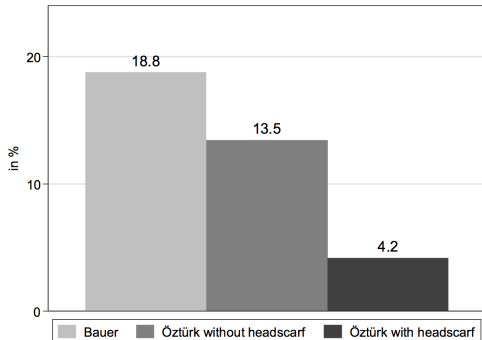
Meryem Öztürk

fonte: Weichselbaumer (2016)

correlazione che dimostra discriminazione

i dati

Regressione
lineare
univariata



fonte: Weichselbaumer (2016)

- il caso dello studio di Weichselbaumer è molto particolare
- siamo in una situazione simile a quella del laboratorio in cui si possono controllare tutte le variabili (il CV, la mail di accompagnamento, il tipo di lavoro)
- non siamo sempre così fortunati e spesso occorre accontentarsi di dati molto meno ad hoc
- quando non si possono considerare tutte le altre condizioni una correlazione potrebbe essere spuria
- una correlazione spuria è una correlazione che non deriva da un nesso causa-effetto

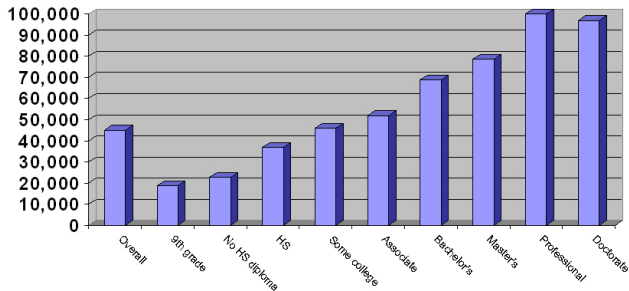
il mantra: correlazione non è causalità

- i nessi di causalità sono i meccanismi di funzionamento dei fenomeni
- un esercizio per cominciare a ragionare come un econometrico: 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: titolo di studio e reddito

i dati

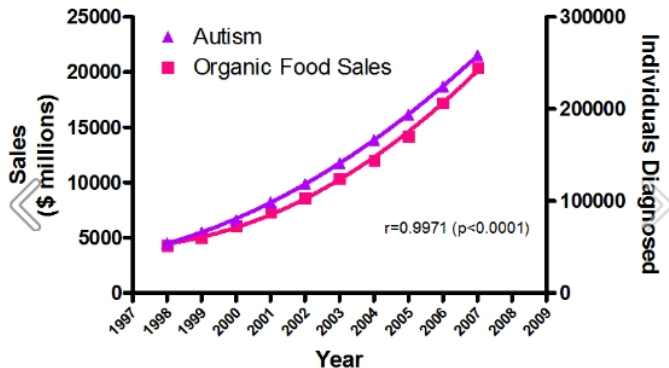
Regressione
lineare
univariata



causa ed effetto: cibo biologico e autismo

i dati

Regressione
lineare
univariata

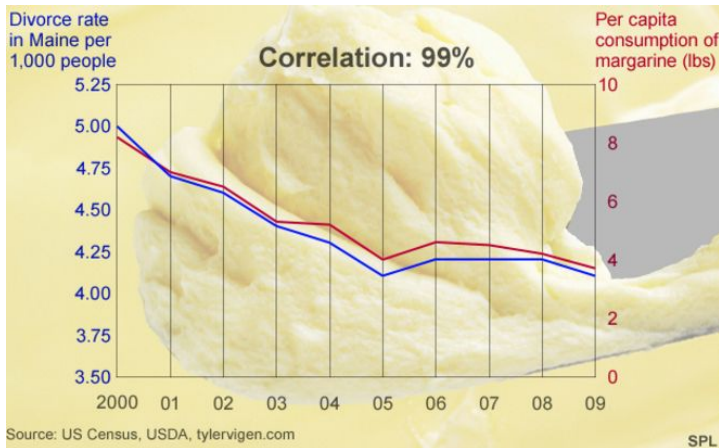


Sources: Organic Trade Association, 2011 Organic Industry Survey; U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS), OMB# 1820-0043. *Children with Disabilities Receiving Special Education Under Part B of the Individuals with Disabilities Education Act

causa ed effetto: margarina e divorzi

i dati

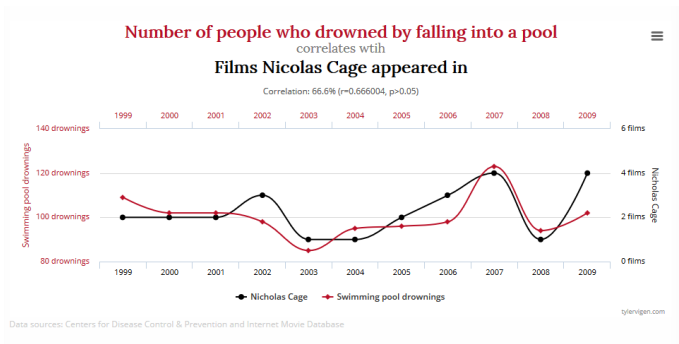
Regressione
lineare
univariata



causa ed effetto: annegati e Nicolas Cage

i dati

Regressione
lineare
univariata



il processo che genera i dati (DGP)

- i nostri modelli ci suggeriscono relazioni fra variabili:
 - ▶ $\uparrow \text{sport} \rightarrow \downarrow \text{peso}$
 - ▶ $\downarrow P \rightarrow \uparrow Q_D$
 - ▶ $\uparrow P_{VINO} \rightarrow \uparrow \text{Qualità}$
- 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 fertilità degli italiani se organizzo il Fertility Day?
 - ▶ cosa accade al PIL se diminuisco la spesa pubblica?
 - ▶ cosa accade alla disoccupazione se abolisco l'art. 18?
- in questi casi vogliamo capire se esiste una relazione e qual'è il segno della relazione

- prima approssimazione: la relazione ipotizziamo sia lineare
- equivalente a dire che $Y = \beta_0 + \beta_1 X$
- dove Y è l'effetto (variabile dipendente) e X la causa (variabile indipendente)

- tornando al problema di ieri
- possiamo verificare la relazione fra felicità (variabile dipendente) e PIL pro capite (variabile indipendente) guardando a due fonti di dati:
- la media della felicità percepita è pubblicato nel World Happiness Report
- il PIL pro capite dalla Banca Mondiale
- i due dati sono disponibili sulla pagina del corso

Assume that this ladder is a way of picturing your life. The top of the ladder represents the best possible life for you. The bottom rung of the ladder represents the worst possible life for you.

Indicate where on the ladder you feel you personally stand right now by marking the circle.



il modello lineare in questo caso

- il nostro DGP ipotizza una relazione:

$$\text{felicità media}_i = \beta_0 + \beta_1 \text{ PIL pro capite}_i$$

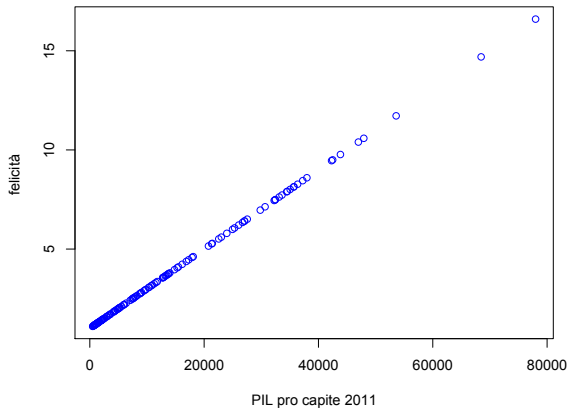
- *felicità*: variabile dipendente
- *PIL pro capite*: variabile indipendente (chiamata anche regressore o variabile esplicativa)

il modello lineare in questo caso

- immaginiamo ad esempio che in un paese con PIL pro capite zero tutti rispondano indicando il minimo possibile livello di felicità
- questo valore identifica il coefficiente β_0
($Y = \beta_0 + \beta_1 \times 0 = \beta_0$)
- immaginiamo poi che ogni 5.000 dollari di *PIL pro capite* in più la felicità salga di un gradino nella scala (un'unità)
- $\Delta X = 5000 \rightarrow \Delta Y = 1$
- questo identifica il coefficiente $\beta_1 = \frac{1}{5000} = 0.0002$

PIL pro capite e felicità: relazione perfettamente lineare

Se la realtà corrisponde esattamente alla nostra ipotesi
dovremmo osservare questo:



i dati

Regressione
lineare
univariata

il modello lineare in questo caso

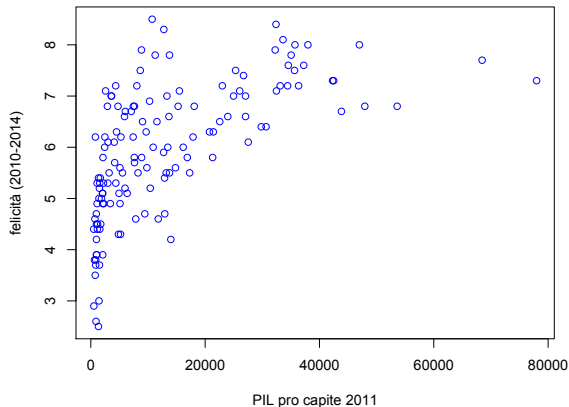
- in realtà sappiamo:
 - 1 che potrebbero esserci molte altre variabili che influenzano la felicità
 - 2 che la relazione fra le due potrebbe non essere lineare, stiamo facendo un'approssimazione
- quindi ci aspettiamo che i dati non siano perfettamente coerenti con il modello

$$\text{felicità media}_i = \beta_0 + \beta_1 \text{ PIL pro capite}_i + u_i$$

- dove u_i è la distanza fra quello che supponiamo essere il livello di felicità associato con un certo livello di PIL e la realtà

PIL pro capite e felicità

I dati reali, come sappiamo mostrano una relazione solo molto vagamente positiva e lineare:



i dati

Regressione
lineare
univariata

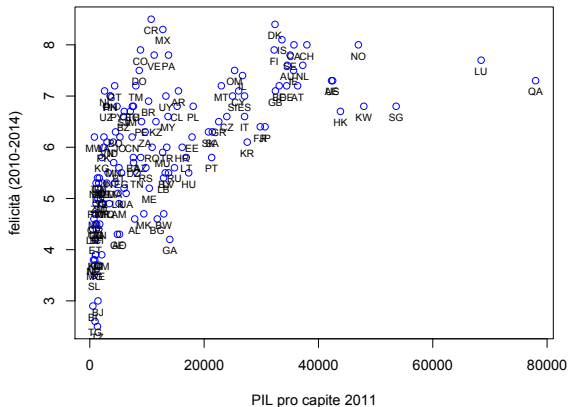
fonte: World Bank (2011), World Happiness Report (2015)

i dati

Regresione
lineare
univariata

PIL pro capite e felicità

I dati reali, come sappiamo mostrano una relazione solo molto vagamente positiva e lineare:

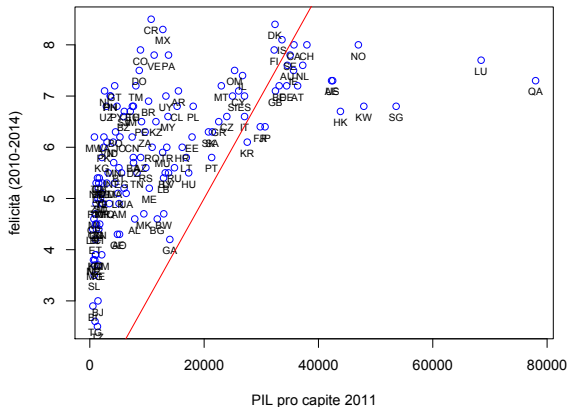


i dati

Regresione
lineare
univariata

PIL pro capite e felicità

la nostra ipotesi non fa un gran lavoro nell'approssimare questa relazione



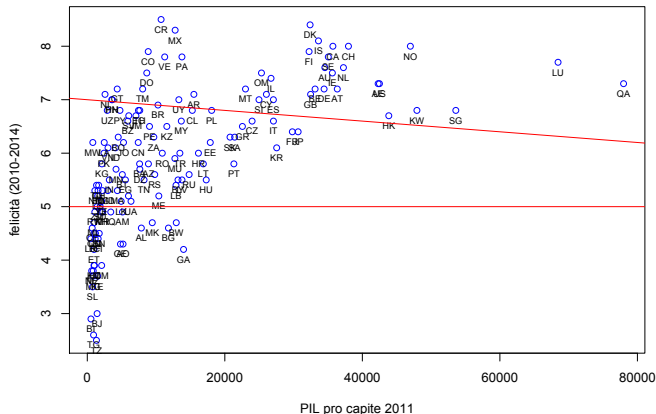
fonte: World Bank (2011), World Happiness Report (2015)

i dati

Regresione
lineare
univariata

la migliore approssimazione lineare

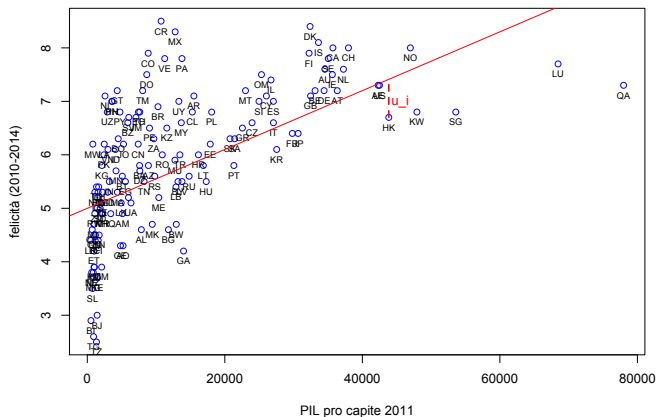
Per migliorare la situazione possiamo cercare la migliore possibile approssimazione lineare della relazione che vediamo nei dati



la migliore approssimazione lineare

i dati

Regresione
lineare
univariata



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

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:

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

- 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 (anche chiamata regressore o variabile esplicativa)
- coefficienti o parametri:
 - β_0 : intercetta
 - β_1 : pendenza
- u_i : errore di disturbo

interpretazione del modello lineare in generale

β_0 : intercetta, è il valore che ha 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