

# PRINCIPI DI ECONOMETRIA

## lezione 15

AA 2015-2016

Paolo Brunori

# variabile dipendente ordinale

- quando abbiamo davanti una variabile ordinale l'unica cosa che sappiamo è l'ordine ma non conosciamo la distanza fra le risposte
- per questo è prudente essere cauti
- in questo caso il modello potrebbe o spiegare quali variabili sono buoni predittori di uno stato di felicità moderato o elevato
- ovvero costruiamo una variabile dicotomica felice = 1 se 'si' o 'decisamente sì' e pari a zero negli altri due casi

# aggregazione di variabili qualitative o ordinali

- spesso indicatori sintetici utilizzano una moltitudine di variabili:
- l'assegnazione della 'bandiera blu' ai comuni sulla costa
- indici di benessere nelle città
- indici di sviluppo economico alternativi al PIL
- in questi casi bisogna fare attenzione al procedimento di aggregazione delle dimensioni

# aggregazione di variabili qualitative o ordinali

- le variabili potrebbero essere cardinali ma espresse in unità di misura non comparabili
- le variabili potrebbero essere ordinali e sommarle o moltiplicarle equivale a dare un significato cardinale

# variabili espresse in scale non comparabili

- talvolta abbiamo variabili cardinali espresse in scale non comparabili
- se ad esempio vogliamo aggregare: *BAL* (% della costa balneabile) e *CIN* (numero di cinema per 1000 abitanti) in un unico indicatore di quanto una località turistica balneale è attraente
- *BAL* varia fra 0 e 100
- *CIN* potenzialmente è un qualsiasi numero reale positivo (potremmo avere 0.02 cinema per 1000 abitanti oppure 26 cinema per 1000 abitanti)
- nella realtà *BAL* è molto maggiore di *CI*
- esempio:  $BAL = 76$ ,  $CIN = 0.5$
- un'aggregazione basata sulla media
$$ATTRATTIVITA = \frac{1}{2}(BAL + CIN) = 38.25$$
 è fourviante perché significa far pesare molto più *BAL* che *CIN*

# standardizzazione di variabili espresse in scale non comparabili

- in questo caso una strada spesso usata è la standardizzazione
- un modo semplice di standardizzare consiste nel manipolare le due variabili in questo modo:

$$\tilde{B\ddot{A}R}_i = \frac{B\ddot{A}R_i - MIN_{B\ddot{A}R}}{MAX_{B\ddot{A}R} - MIN_{B\ddot{A}R}}$$

- in questo modo tutte le variabili variano fra 0 e 1 e per tutte:
  - 1 significa per tutte assumere valore massimo fra quelli osservati
  - 0 assumere valore minimo

# aggregazione di variabili qualitative o ordinali

- anche nel caso che le variabili siano tutte confrontabili e cardinali occorre ricordare che:
- sommare o moltiplicare significa dare a tutte le dimensioni lo stesso peso
- sommare significa assumere zero complementarità fra le dimensioni
- moltiplicare significa assumere un certo grado di complementarità fra le dimensioni
- se sono variabili dicotomiche moltiplicare significa assumere totale complementarità

# aggregazione di variabili qualitative o ordinali

- una volta creato l'indicatore compito fate attenzione a come lo spiegate
- spiegare l'indicatore di qualità della vita nelle città usando come il regressore il PIL pro-capite ad esempio
- è problematico se il PIL pro capite è uno degli indicatori che lo determinano
- significa spiegare un'identità e non scoprire un nesso causa-effetto



## INDICATORI COMPOSITI

REGRESSIONE CON  
DATI PANEL

Classifica finale				Tenore di vita	Servizi & Ambiente	Affari & Lavoro	Ordine pubblico	Popolazione	Tempo libero
Pos	Diff. pos.	Provincia	Punti	▼	▼	▼	▼	▼	▼
1	9 <span>▲</span>	Bolzano	603	4	22	4	32	3	10
2	6 <span>▲</span>	Milano	581	1	4	10	108	12	11
3	-1 <span>▼</span>	Trento	568	16	16	48	40	8	20
4	12 <span>▲</span>	Firenze	567	47	17	19	102	7	3
5	9 <span>▲</span>	Sondrio	563	28	11	37	11	27	40
6	14 <span>▲</span>	Olbia-Tempio	562	73	68	11	12	1	17
7	10 <span>▲</span>	Cuneo	561	22	36	3	25	13	44
8	-2 <span>▼</span>	Aosta	555	7	31	20	45	35	29
9	0 =	Siena	555	52	59	15	27	11	12
10	-9 <span>▼</span>	Ravenna	555	39	2	5	105	61	15
11	2 <span>▲</span>	Macerata	554	48	39	52	46	10	6
12	-5 <span>▼</span>	Bologna	553	9	3	16	110	20	21
13	6 <span>▲</span>	Parma	552	6	19	23	96	4	36

## Cosa spiega la posizione in classifica?

È possibile stimare una regressione che usa osservazioni a livello di comune per spiegare il valore della qualità di vita...

## INDICATORI COMPOSITI

REGRESSIONE CON  
DATI PANEL

[Home Page](#) > [Affari & Lavoro](#) > **Tasso di occupazione totale in % (15-64 anni) -2014**

### Tasso di occupazione totale in % (15-64 anni) -2014

**Impieghi su depositi totali in % -2014**

### Giovani imprenditori (18-29 anni) in rapporto ai giovani - 2013

Imprese registrate x 1

### Sofferenze sugli impieghi totali in % - 2014

**Tasso di occupazione totale in % (15-64 anni) -2014**

Pos	Provincia	Valore	Punti
1	Bolzano	71	1.000
2	Bologna	69	979
3	Firenze	69	971
4	Parma	68	967
5	Monza e Brianza	68	960
6	Belluno	68	954

... dopo aver controllato che i regressori non siano inclusi nella formula che calcola l'indice

# Dati longitudinali (panel)

- variabile dipendente:  $Y_{i,t}$   $i = 1, \dots, n$ ,  $t = 1, \dots, T$
- variabili indipendenti:  $X_{i,t}$
- variabili che non variano nel tempo ma fra osservazioni:  $Z_i$
- variabili che non variano fra osservazioni ma variano nel tempo:  $S_t$

# esempio del libro

- morti per il traffico e accise sull'alcool
- $y = \text{morti}/10.000$
- 48 stati Americani  $n = 48$
- 7 anni (1982,..., 1988)  $T = 7$
- panel bilanciato  $7 \times 48 = 336$  osservazioni

# relazione di interesse: effetto causale delle imposte sulla mortalità 1982

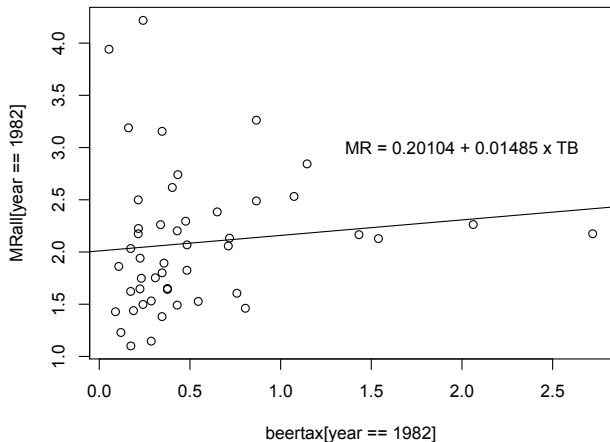
PRINCIPI DI  
ECONOMETRIA

LEZIONE 15

INDICATORI  
COMPOSITI

REGRESSIONE CON  
DATI PANEL

1982



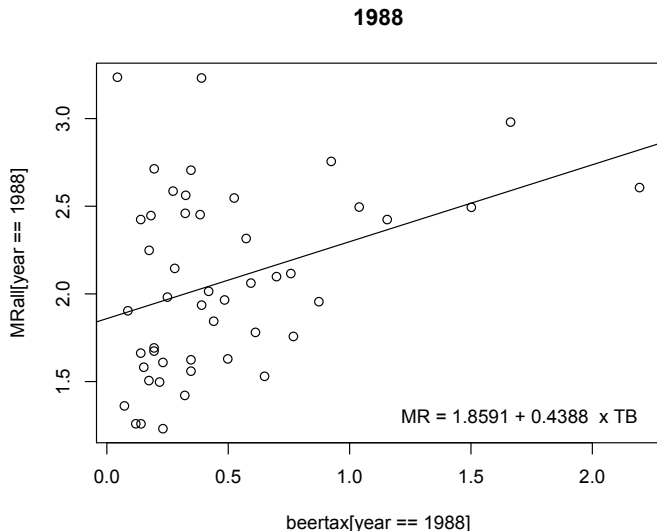
# relazione di interesse: effetto causale delle imposte sulla mortalità 1988

PRINCIPI DI  
ECONOMETRIA

LEZIONE 15

INDICATORI  
COMPOSITI

REGRESSIONE CON  
DATI PANEL



# altri dati disponibili

- età degli automobilisti
- qualità delle strade
- densità automobilisti
- età minima per consumo alcol
- tasso di disoccupazione
- ...

# distorsione da variabili omesse?

- è possibile che “intensità di traffico” sia una variabile omessa?
  - 1 determina  $Y$ ?
  - 2 potrebbe essere correlato con  $X$ ?
- sfruttando la struttura dei dati possiamo eliminare questa distorsione a patto che le variabili omesse non varino nel tempo



# ci sono altre possibili variabili omesse?

- cosa determina la probabilità di un incidente mortale?

# modello panel con 2 periodi

$$Y_{i,t} = \beta_0 + \beta_1 BT_{i,t} + \beta_2 Z_i + u_{i,t}$$

- $BT_{i,t}$  importo dell'accisa sulla birra
- $Z_i$  attitudine culturale (non cambia nel tempo)

# modello panel con 2 periodi

- ogni variazione dal 1988 al 1998 non può essere causata da  $Z_i$
- se stimiamo un modello per ogni  $t$

$$mort_{i,1982} = \beta_0 + \beta_1 BT_{i,1982} + \beta_2 Z_i + u_{i,1982}$$

$$mort_{i,1988} = \beta_0 + \beta_1 BT_{i,1988} + \beta_2 Z_i + u_{i,1988}$$

- assumiamo

$$E(u_{i,t} | BT_{i,t}, Z_i) = 0$$

# modello panel con 2 periodi

- quindi possiamo riscrivere i due modelli:

$$MRT_{i,1988} - MRT_{i,1982} =$$

$$= \beta_0 + \beta_1 BT_{i,1988} + \beta_2 Z_i + u_{i,1988} - \beta_0 + \beta_1 BT_{i,1982} + \beta_2 Z_i + u_{i,1982}$$

$$MRT_{i,1988} - MRT_{i,1982} = \beta_1 (BT_{i,1988} - BT_{i,1982}) + u_{i,1988} - u_{i,1982}$$

- $u_{i,t}$  non è correlato con i regressori  $\rightarrow$  l'equazione può essere stimata con OLS

# equazione in differenza

- $\Delta Y_i = \Delta X_i + u_i$
- anche se  $Z_i$  non è osservata non è una variabile omessa
- perchè?

# OLS in differenza

