Principi di Econometria

lezione 5

AA 2015-2016

Paolo Brunori

dove siamo arrivati?

- se siamo interessati a studiare l'andamento congiunto di due fenomeni economici
- possiamo provare a misurare i due fenomeni e poi usare la regressione lineare semplice per valutare quanto il legame fra due fenomeni sia ben approssimato da una relazione lineare
- questo metodo ci fornisce: una stima dei due parametri che definiscono la forza della relazione fra i due fenomeni, il livello di incertezza delle stime, una quantificazione della bontà di adattamento dei dati al modello (quanto bene il modello 'spiega' i dati)

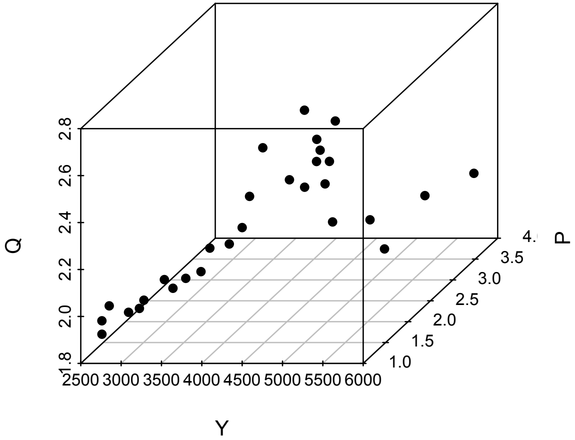
variabili omesse

- come visto nel caso dei dati di consumo di tabacco in Turchia, è difficile ipotizzare che una sola variabile indipendente spieghi il comportamento della dipendete
- in generale infatti u cattura tutte quelle variabili che influenzano Y ma non sono considerate/osservabili
- l'omissione di una variabile Z distorce lo stimatore OLS se si verificano due condizioni :
  1. Z è una delle variabili che determina Y
  2. $corr(X, Z) \neq 0$

esempio: $Y=$ consumo tabacco, $X=$ prezzo, $Z=$ reddito pro capite

- condizione 1: Z ha un impatto su Y?
- condizione 2: Z potrebbe essere correlato con X?
- qual'è la direzione attesa della distorsione?

Consumo di tabacco, prezzo, reddito in Turchia



distorsione per variabili omesse

- $lim_{n \to +\infty}(\hat{\beta}_1 - \beta_1) = 0$ sotto l'assunzione:

$$cov(X_i, u_i) = 0$$

- in caso contrario:

$$lim_{n \to +\infty}(\hat{\beta}_1 - \beta_1) = \frac{\sigma_u}{\sigma_X}\rho_{u,X}$$

cosa vogliamo misurare con β_1?

- β_1 può essere semplicemente la pendenza della retta che interpola un grafico a dispersione
- β_1 può invece servire a predirre $Y|X$
- β_1 come effetto **causale** di X su Y

che cosa si intende per nesso di causalità?

regressione con due regressori

- Y, X_1, X_2
- siamo alla ricerca della relazione:

$$E(Y_i | X_{1,i} = x_i, X_{2,i} = x_2)$$

- se lineare ha la forma generica:

$$E(Y_i | X_{1,i} = x_i, X_{2,i} = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

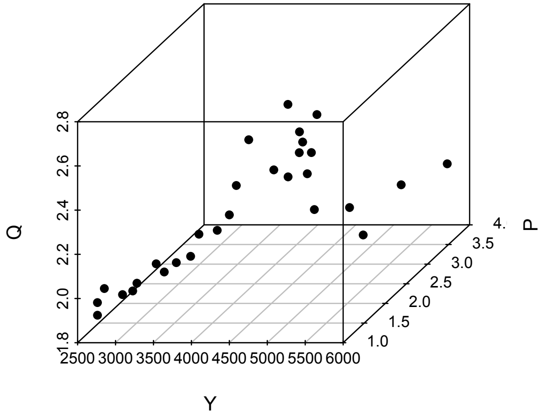
definizioni

- $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ è la retta di regressione della popolazione
- β_0 è l'intercetta
- β_1, β_2 sono i coefficienti associati alle variabili X_1, X_2
- analogamente a quanto visto per un solo regressore:

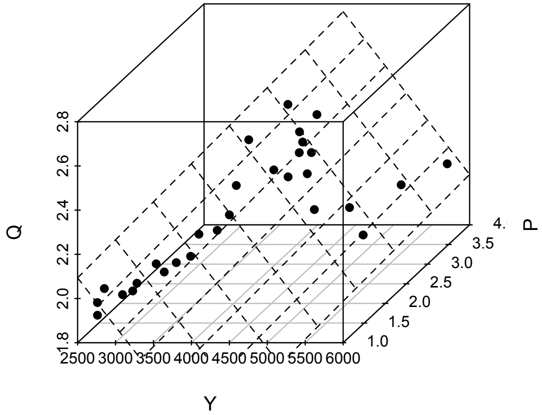
$$Y + \Delta Y = \beta_0 + \beta_1(X_1 + \Delta X_1) + \beta_2 X_2$$

- quindi $\beta_1 = \frac{\Delta Y}{\Delta X_1}$ è l'effetto parziale di X_1 (tenendo costante X_2)

il piano di regressione



il piano di regressione



regressione multipla

- seppur difficile da vedere graficamente Y può essere una funzione di molte variabili
- per un numero generico k di regressori il modello di regressione multipla prende la forma:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i} + u_i$$

stimatore OLS della regressione multipla

- gli stimatori $\hat{\beta}_i$ che vengono normalmente utilizzati sono quelli che minimizzano:

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_{1,i} - ... - b_k X_{k,1})^2$$

- le formule per ottenerli sono in formula matriciale e potete consultarli qui
- la terminologia è la stessa usata per il modello con un regressore:

  stimatori OLS: $\hat{\beta}_0, \hat{\beta}_1, ...$
  valore predetto: $\hat{Y}_i$
  residuo $\hat{u}_i$

Regressione multipla del consumo di tabacco

	coefficiente	errore standard	t	$valore - p$
β_0	1.6572	0.1237	13.394	0.0000
β_Y	0.0003	0.0000	6.518	0.0000
β_P	-0 .0423	0.0096	-4.3662	0.0001

- come posso interpretare questi coefficienti?

R^2

- è la frazione della varianza campionaria spiegata dai regressori
$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}$$
- $R^2 \in [0, 1]$
- nella regressione multipla R^2 cresce all'aumentare dei regressori
- $k = n - 1$ regressori mi garantiscono sempre un'interpolazione perfetta (se i regressori sono perchè c'è anche l'intercetta)
- nel modello del fumo di tabacco $R^2 = 0.6406$

R^2 corretto

- è possibile tener conto del numero di regressori

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}$$

- $R^2 > \bar{R}^2$ sempre
- due effetti dell'aggiunta di un regressorie: $\uparrow\downarrow$
- $\bar{R}^2$ può essere minore di zero
- nel modello del fumo di tabacco $\bar{R}^2 = 0.6129$