

6089 ECONOMETRIA

A.A. 2008 – 2009

Lezione Info1

Obiettivo di questa lezione applicata è studiare come le spese per la campagna elettorale influiscono sui risultati delle elezioni. Per farlo abbiamo a disposizione dati relativi alle elezioni del Congresso degli Stati Uniti del 1988 per 173 distretti elettorali in cui la campagna elettorale ha visto coinvolti solo due partiti (e quindi solo due candidati). I dati sono contenuti nel file *vote.dta*.

Alcuni utili riferimenti possono essere reperiti nelle seguenti pagine web:

http://en.wikipedia.org/wiki/United_States_House_of_Representatives_elections,_1988

http://commons.wikimedia.org/wiki/Image:Census_Regions_and_Divisions.PNG

1. Operazioni preliminari

- a) Create la directory “C:\econometria\” e copiatevi il file *vote.dta*.
- b) Nella finestra di comando di Stata digitate “**cd c:\econometria**”: da questo momento questa diventa la directory di riferimento per le operazioni che svolgeremo con Stata.
- c) Create un file in cui saranno contenuti tutti i passaggi svolti durante la lezione digitando “**log using info1.log, replace**”.

2. Un primo sguardo ai dati

- a) Aprite il file *vote.dta* utilizzando File → Open, oppure scrivendo nella riga di comando: “**use vote,clear**”.
- b) Il comando “**describe**” (“**des**”) fornisce una descrizione del contenuto del dataset (numero di osservazioni, numero di variabili, tipo e descrizione (*label*) delle variabili), mentre il comando “**summarize**” (“**sum**”) fornisce alcune statistiche descrittive delle variabili.
- c) I dati possono essere direttamente visionati selezionando Data → Data Editor.
- d) I dati sono riferiti ad alcuni distretti (“*district*”) negli stati degli USA, specificati nella variabile “*state*”, che possono essere raggruppati in quattro regioni come specificato in “*region*”. Per verificare la distribuzione nelle regioni, digitate “**tabulate region**” (“**tab region**”). Il comando “**tab region,s(expendA)**” fornisce poi la media della spesa elettorale del candidato A nelle quattro regioni geografiche. E’ quindi possibile anche vedere come si distribuisce la spesa elettorale complessiva con il comando “**tab region,s(expend)**”, dove la variabile “*expend*” può essere costruita attraverso il comando “**gen expend=expendA+expendB**”. Ad ogni variabile può essere associata una descrizione (*label*) con il seguente comando: **label variable**

expend “total expenditures in \$1000”. Volendo si possono recuperare ulteriori informazioni (oltre alla media) sulla distribuzione di “expend” nelle varie regioni: **tabstat expend,by(region) s(mean sd min max)**.

- e) Visto che esiste una dummy che ci dice se il candidato A è democratico (o repubblicano), possiamo creare una variabile con il voto percentuale per il candidato democratico, in due passi: **gen voteDem=voteA*democA; replace voteDem=(100-voteA) if democA==0**. Dopodiché, possiamo guardare alla distribuzione dei voti democratici in diverse regioni con i comandi introdotti sopra.

3. Un modello di regressione lineare (caso bivariato)

- a) Create una variabile uguale alla spesa di campagna elettorale sostenuta dal candidato A come percentuale delle spese complessive, digitando **“generate shareA=100*(expendA/(expendA+expendB))”**. Fate la stessa cosa per il candidato B, creando l’analogha variabile “shareB”. Anche in questo caso, alle variabili create possiamo assegnare una label. Ad esempio, per la variabile shareA: **label variable shareA “share of camp. expends. by A over total expends.”**.
- b) Stimiamo con OLS la relazione $voteA_i = \beta_1 + \beta_2 shareA_i + \beta_3 shareB_i + \varepsilon_i$, utilizzando il comando **“reg voteA shareA shareB”**. Cosa accade? Perché?
- c) Stimiamo dunque: $voteA_i = \beta_1 + \beta_2 shareA_i + \varepsilon_i$.
- Visualizziamo la relazione tra “voteA” e “shareA” attraverso il comando **“twoway (scatter voteA shareA)”** (in alternativa, si può selezionare Graphics → Twoway graph (scatter, line, etc.) → create, selezionare la Y – voteA – e la X – share A – e ottenere il grafico selezionando “accept” e “ok”).
 - Stimiamo il modello con OLS: **reg voteA shareA**. Commentate segno e significatività del coefficiente.
 - Otteniamo i valori fittati (\hat{y}): **predict voteA_hat**.
 - Mettiamo i valori effettivi della variabile dipendente ed i valori fittati sullo stesso grafico, così da evidenziare la retta di regressione: **twoway (scatter voteA shareA) (line voteA_hat shareA)**.

4. Un modello di regressione lineare multipla

- a) Stimiamo ora: $voteA_i = \beta_1 + \beta_2 expendA_i + \beta_3 expendB_i + \beta_4 prtystrA_i + \varepsilon_i$.
- Verifichiamo innanzitutto la correlazione tra le variabili coinvolte nella regressione: **correlate voteA expendA expendB prtystrA**.
 - Effettuiamo la regressione OLS: **reg voteA expendA expendB prtystrA**.
 - Commentate segno e significatività dei coefficienti, significatività della regressione, coefficiente di determinazione. Che interpretazione date ai coefficienti di “expendA” ed “expendB”?
- b) Stimate: $voteA_i = \alpha_1 + \alpha_2 expendAd100_i + \alpha_3 expendBd100_i + \beta_4 prtystrA_i + \varepsilon_i$, dove **expendAd100=expendA/100** ed **expendBd100=expendB/100**.

- Utilizzate il comando “generate” per costruire le variabili “expendAd100” ed “expendBd100”.
 - Stimare il modello utilizzando il comando “regress”.
 - Che relazione c’è tra α_2 (α_3) e β_2 (β_3)? Come interpretate i coefficienti di *expendAd100* e *expendBd100*?
- c) Stimare ora: $voteA_i = \gamma_1 + \gamma_2 \text{lexpend}A_i + \gamma_3 \text{lexpend}B_i + \gamma_4 \text{prtystr}A_i + \varepsilon_i$, dove “*lexpendA=log(expendA)*” e “*lexpendB=log(expendB)*”.
- Utilizzate il comando “generate” per costruire le variabili “*lexpendA*” e “*lexpendB*”.
 - Stimare il modello utilizzando il comando “regress”.
 - Come interpretate i coefficienti di “*lexpendA*” e “*lexpendB*”? Come potete ricavare una stima per β_2 e β_3 a partire dai valori stimati di γ_2 e γ_3 ?
- d) Vogliamo testare l’ipotesi $H_0: \gamma_2 = -\gamma_3$ nei seguenti modi:
- con il test t, a partire dalla regressione appena svolta, utilizzando la matrice stimata di varianza-covarianza dello stimatore OLS, che potete ottenere con il comando “**matrix list e(V)**”;
 - con il test F, stimando il modello ristretto;
 - usando il comando di Stata “**test lexpendA=-lexpendB**”.
- e) Inserite ora la dummy *democA*, uguale a 1 se il candidato A è democratico: $voteA_i = \gamma_1 + \gamma_2 \text{lexpend}A_i + \gamma_3 \text{lexpend}B_i + \gamma_4 \text{prtystr}A_i + \gamma_5 \text{democ}A + \varepsilon_i$.
- Stimare il nuovo modello utilizzando il comando “regress”.
 - Come interpretate il coefficiente di “*democA*”? Commentate anche la significatività dei coefficienti in relazione alla regressione effettuata sopra.

5. Operazioni di chiusura

Salvate il file dei dati con le nuove variabili costruite e chiudete il log file: **save vote_new.dta,replace; log close.**