

Principi di Econometria

lezione 8

AA 2016-2017

Paolo Brunori

a cosa mi
servono i gradi
di libertà?

regressione multipla con $n = k$

Immaginate di voler studiare i determinanti del voto
all'esame di econometria

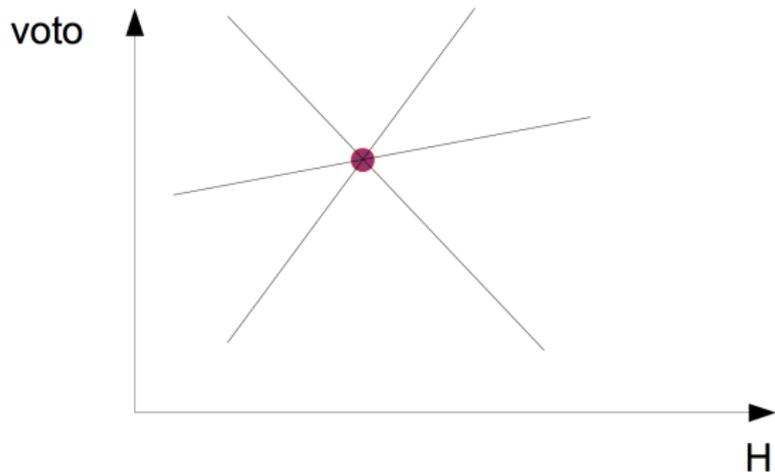
$$Y = \beta_1 X_1 + u$$

$$Y_i = \beta_1 H_i + u_i$$

H=ore studiate alla settimana

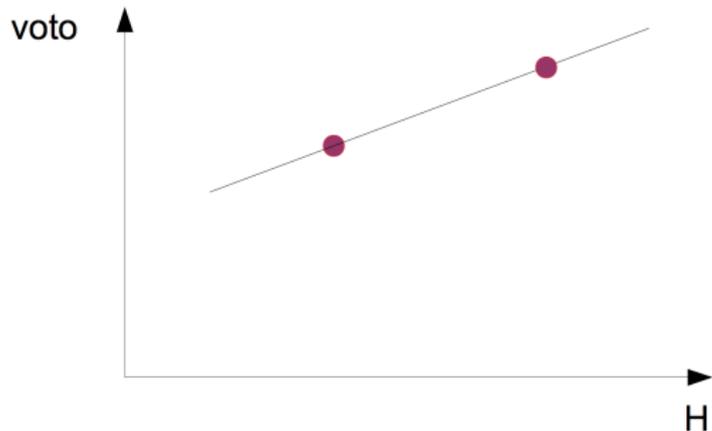
cosa accade se ho solo un'osservazione? ($n < k$)

a cosa mi
servono i gradi
di libertà?



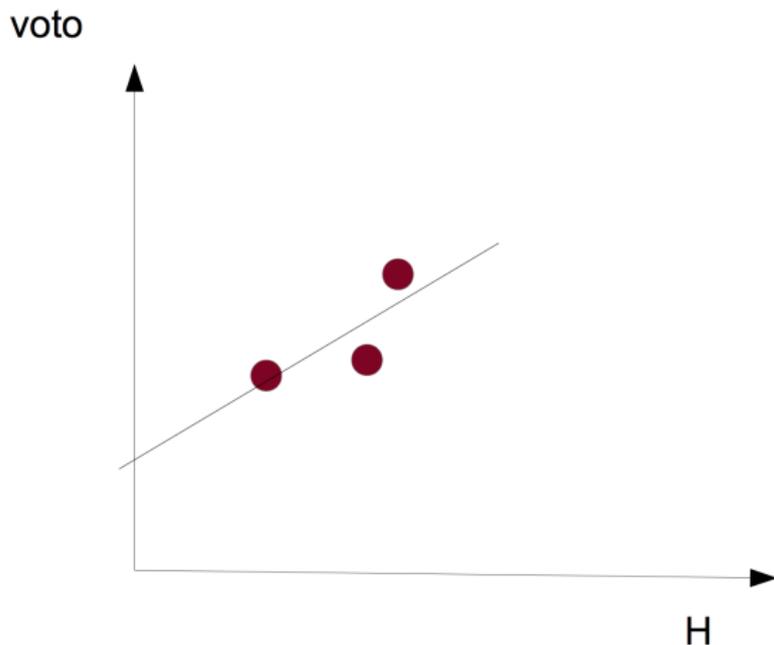
cosa accade se ho due osservazioni? ($n=k$)

a cosa mi
servono i gradi
di libertà?



a cosa mi
servono i gradi
di libertà?

cosa accade se ho tre osservazioni? ($n > k$)



L'informazione libera ci serve a cominciare a capire quanto è affidabile la nostra stima e (il nostro primo grado di libertà)

cosa accade se aumentano i regressori?

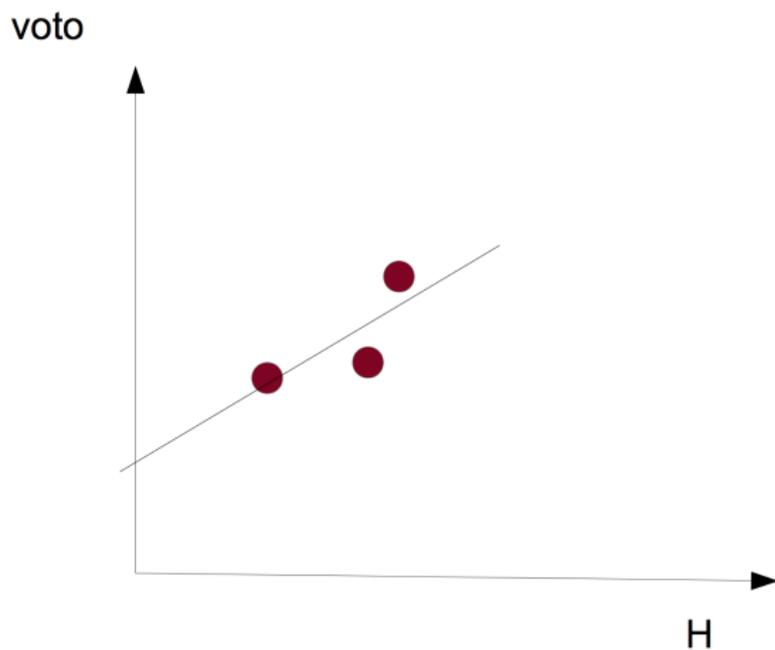
- ▶ aumentare i regressori vuol dire aumentare i parametri da misurare
- ▶ ogni parametri 'usa' un'informazione
- ▶ se ad esempio vogliamo stimare un modello di regressione lineare

$$Y_i = \beta_1 H_i + \beta_2 FREQU_i$$

- ▶ usiamo un grado di libertà e torniamo nella situazione di perfetta interpolazione e impossibilità di fare inferenza

cosa accade se ho tre osservazioni? ($n > k$)

a cosa mi
servono i gradi
di libertà?



Zero gradi di libertà = impossibilità di verificare l'entità dell'errore = zero informazione

a cosa mi
servono i gradi
di libertà?

regressione multipla con $n = k$

Immaginate di voler studiare i determinanti del voto all'esame di econometria

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$$

$$Y_i = \beta_1 H + \beta_2 \text{FREQ}_i + \beta_3 \text{COPIA}_i + \beta_4 \text{LICEO}_i + u_i$$

H=ore studiate alla settimana

FREQ=1 se frequentante

COPIA=1 se riesce a copiare

LICEO= 1 se ha licenza scientifica o classica

a cosa mi
servono i gradi
di libertà?

regressione multipla con $n = k$

immaginate di avere a disposizione un campione di 4
studenti

voto	h	freq	copia	liceo
18	2	1	1	0
22	4	0	0	1
29	5	1	0	0
30	3	0	1	1

questi dati possono essere visti come un sistema di
equazioni

regressione multipla con $n = k$

questi dati possono essere visti come un sistema di n equazioni di primo grado in $k = n$ incognite

voto	h	freq	copia	liceo
18	2	1	1	0
22	4	0	0	1
29	5	1	0	0
30	3	0	1	1

$$18 = 2 \beta_1 + 1\beta_2 + 1 \beta_3 + 0 \beta_4$$

$$22 = 4 \beta_1 + 0\beta_2 + 0 \beta_3 + 1 \beta_4$$

$$29 = 5 \beta_1 + 1\beta_2 + 0 \beta_3 + 0 \beta_4$$

$$30 = 3 \beta_1 + 0\beta_2 + 1 \beta_3 + 1 \beta_4$$

a cosa mi servono i gradi di libertà?

regressione multipla con $n = k$

il sistema ha una sola soluzione, cioè esiste una
combinazione di β_1, \dots, β_k che soddisfa tutte le equazioni

$$18 = 2 \beta_1 + 1\beta_2 + 1 \beta_3 + 0 \beta_4$$

$$22 = 4 \beta_1 + 0\beta_2 + 0 \beta_3 + 1 \beta_4$$

$$29 = 5 \beta_1 + 1\beta_2 + 0 \beta_3 + 0 \beta_4$$

$$30 = 3 \beta_1 + 0\beta_2 + 1 \beta_3 + 1 \beta_4$$

in questo caso: $\beta_1 = 9.5$, $\beta_2 = -18.5$, $\beta_3 = 17.5$, $\beta_4 = -16$

$$2 \times 9.5 + 1 \times -18.5 + 1 \times 17.5 + 0 \times -16 = 18$$

$$4 \times 9.5 + 0 \times -18.5 + 0 \times 17.5 + 1 \times -16 = 22$$

$$5 \times 9.5 + 1 \times -18.5 + 0 \times 17.5 + 0 \times -16 = 29$$

$$3 \times 9.5 + 0 \times -18.5 + 1 \times 17.5 + 1 \times -16 = 30$$

a cosa mi
servono i gradi
di libertà?

regressione multipla con $n = k$

l' R^2 di questa regressione è 1

voto	h	freq	copia	liceo
18	2	1	1	0
22	4	0	0	1
29	5	1	0	0
30	3	0	1	1

perchè non ci convince?

regressione multipla con $n = k$

Immaginate di estrarre un altro campione, intervistare 4 studenti diversi e ottenere:

voto	h	freq	copia	liceo
28	3.5	0	1	1
21	2	0	0	1
24	2	1	1	0
27	3.5	1	1	1

Se fate i calcoli otterrete un altro risultato:

$$\beta_1 = 36, \beta_2 = -1, \beta_3 = -47, \beta_4 = -51$$

regressione multipla con $n = k$

a cosa mi
servono i gradi
di libertà?

- quando $n = k$ otteniamo un'interpolazione perfetta dei dati perchè dobbiamo stimare tanti parametri quante osservazioni
- allo stesso tempo però non abbiamo nessuna idea di come i parametri stimati varierebbero se avessimo osservato un campione diverso
- stimando β_1 vogliamo stimare in modo affidabile la relazione fra H e voto
- per fare questo occorre osservare la relazione di interesse per tanti studenti
- questo permette di capire come il legame fra H e voto si distribuisce nel campione e sulla base di questo fare inferenza sulla popolazione

a cosa mi
servono i gradi
di libertà?

- le assunzioni sono le stesse già viste + 1

1. $E(u_i|X_i) = 0 \forall i = 1, \dots, k$
2. (X_1, \dots, X_n) sono i.i.d
3. gli outlier sono improbabili

\Rightarrow assenza di collinearità perfetta

a cosa mi
servono i gradi
di libertà?

- i regressori mostrano collinearità perfetta se uno dei regressori è funzione lineare degli altri
- se ad esempio $X_2 = a - bX_1$ o $X_2 = \frac{X_1}{100}$
- in presenza di collinearità perfetta non è possibile stimare la regressione OLS
- ma si tratta generalmente di un errore (materiale o logico)

a cosa mi
servono i gradi
di libertà?

- anche in questo caso gli stimatori sono delle variabili casuali con una distribuzione campionaria
- anche in questo caso sono stimatori corretti: convergono in probabilità al vero valore β_i
- inoltre per n abbastanza grande si può assumere una distribuzione approssimativamente normale del valore campionario

casi frequenti di collinearità perfetta

a cosa mi
servono i gradi
di libertà?

- X_1 = tasso di diplomati, X_2 tasso di drop-out o fuori-corso
- X_1 = residenti al nord, X_2 residenti al centro, X_3 residenti al sud
- i software generalmente danno un messaggio di errore in questi casi o rimuovono una delle variabili collineari

- se il legame fra una o più variabili è vicino ad un legame lineare la regressione OLS può essere stimata
- immaginate ad esempio di inserire una variabile ‘stipendio medio’ insieme alla variabile ‘PIL pro capite’
- in questo caso la stima dei coefficienti è corretta
- ma per almeno uno di questi è imprecisa

assunzioni necessarie per OLS

a cosa mi
servono i gradi
di libertà?

- le assunzioni sono le stesse già viste + 1

1. $E(u_i|X_i) = 0 \forall i = 1, \dots, k$
2. (X_1, \dots, X_n) sono i.i.d
3. gli outlier sono improbabili

\Rightarrow assenza di collinearità perfetta

a cosa mi
servono i gradi
di libertà?

- i regressori mostrano collinearità perfetta se uno dei regressori è funzione lineare degli altri
- se ad esempio $X_2 = a - bX_1$ o $X_2 = \frac{X_1}{100}$
- in presenza di collinearità perfetta non è possibile stimare la regressione OLS
- ma si tratta generalmente di un errore (materiale o logico)

- anche in questo caso gli stimatori sono delle variabili casuali con una distribuzione campionaria
- anche in questo caso sono stimatori corretti: convergono in probabilità al vero valore β_i
- inoltre per n abbastanza grande si può assumere una distribuzione approssimativamente normale del valore campionario

casi frequenti di collinearità perfetta

a cosa mi
servono i gradi
di libertà?

- X_1 = età, X_2 anno di nascita
- X_1 = residenti al nord, X_2 residenti al centro, X_3 residenti al sud
- i software generalmente danno un messaggio di errore in questi casi o rimuovono una delle variabili collineari

- se il legame fra una o più variabili è vicino ad un legame lineare la regressione OLS può essere stimata
- immaginate ad esempio di inserire una variabile ‘stipendio medio’ insieme alla variabile ‘PIL pro capite’
- in questo caso la stima dei coefficienti è corretta
- ma per almeno uno di questi è imprecisa