Principi di Econometria
prof. Brunori
01/02/2017

55' di tempo
Nome e cognome ____________________
Matricola ____________________

Potete consegnare solo le risposte multiple o sia le risposte multiple che quelle aperte. Nel secondo caso per passare l'esame è necessario superare 18 in entrambe le parti.

Risposta multipla

1. per spiegare la probabilità di contrarre l'influenza si stima un modello basato su due soli regressori $Y = \beta_0 + \beta_1 X_{eta<50} + \beta_2 X_{eta\geq 50} + u$, dove $X_{eta<50}$ e $X_{eta\geq 50}$ sono variabili che assumono valore 1 o 0 in base all'età del soggetto.

   A β_1 misura l'elasticità della probabilità di contarre all'età ☐
   B il modello stimato non è un modello lineare di probabilità ☐
   C il modello soffre del problema della perfetta multicollinearità ☐
   D nessuna delle precedenti è vera ☐

2. un modello lineare di probabilità

   A è un modello che spiega una variabile come funzione di una serie di probabilià ☐
   B stima una serie di effetti fissi di osservazione ☐
   C predice sempre probabilità comprese fra 0 e 1 ☐
   D potrebbe predire probabilità maggiori di 1 o minori di zero ☐

3. se una variabile esplicativa X è una variabile dicotomica (dummy):

   A il suo coefficiente sarà stimato con errore ☐
   B il suo coefficiente si interpreta come la variazione di Y dovuta alla variazione unitaria di X ($0 \rightarrow 1$) ☐
   C il coefficiente è sempre negativo ☐
   D il suo coefficiente è interpretabile come variazione dell'inclinazione della retta di regressione ☐

4. la relazione che lega reddito e soddisfazione nella vita mostra un andamento non lineare, al di sopra di un certo livello la soddisfazione non aumenta con incrementi ulteriori del reddito. Quale delle seguenti specifiche scegliereste per stimare la relazione fra queste due variabili?

   A $Y_{reddito} = \beta_0 + \beta_1 X_{soddisfazione} + \beta_2 X^2_{soddisfazione} + u$ ☐
   B $Y_{reddito} + Y^2_{reddito} = \beta_0 + \beta_1 X_{soddisfazione} + u$ ☐
   C $Y_{soddisfazione} = \beta_0 + \beta_1 X_{reddito} + \beta_2 X^2_{reddito} + u$ ☐
   D nessuna delle precedenti è vera ☐

5. avendo stimato l'elasticità della domanda al reddito con l'equazione $lnQ = \beta_0 + \beta_1 ln(R) + u$

   A β_0 è interpretabile come l'elasticità ☐
   B ad un aumento unitario di R corrisponde un aumento di β_1 di Q ☐
   C l'elasticità di Q a R è pari a β_0% ☐
   D l'elasticità di Q a R è pari a β_1 ☐

6. due possibili indicatori compositi della qualità di un cappuccino sono ottenuti con le formula: $Y_1 = X_{Latte} \times X_{caffe}$ e $Y_2 = (X_{Latte} + X_{caffe}) \times 5$. Le due dimensioni - X_{latte}, X_{caffe} - sono due scale di qualità dei due ingredienti che vanno da 0 a 10:

   A Y_1 assegna un peso minore alla qualità del caffè ☐
   B la qualità del caffè e del latte sono perfetti sostituti nei due indicatori ☐
   C Y_1 penalizza di più la presenza di un ingrediente di pessima qualità ☐
   D nessuna delle precedenti è vera ☐

7. la validità esterna di uno studio di regressione si basa su una serie di accorgimenti:

   A lo studio deve essere valido internamente ☐
   B lo studio deve basarsi su un campione rappresentativo della popolazione di interesse ☐
   C il meccanismo di selezione del campione deve essere casuale ☐
   D tutte le precedenti sono vere ☐

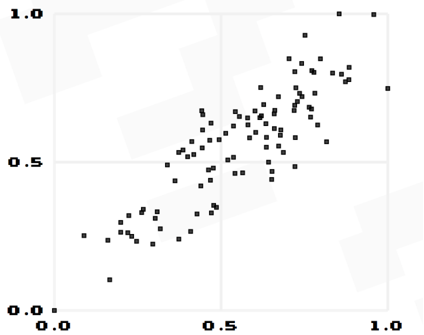
8. in un modello di regressione: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 - \beta_3 X_2^2$ la variazione di Y dovuta ad una variazione unitaria di X_2 è pari a:

   A è sempre minore di 0 ☐
   B $\beta_3 + \beta_2$ ☐
   C dipende dal valore iniziale di X_2 ☐
   D $\beta_2 + \beta_3 \times X_1$ ☐

9. vogliamo stimare un modello che predice il voto all'esame di econometria e abbiamo a disposizione i voti di tutti gli studenti in tutti gli esami del corso di economia e commercio a bari, quale modello è il migliore fra i seguenti:

   A un modello panel con effetti fissi di studente ☐
   B un modello panel con effetti a livello d'esame ☐
   C un modello con effetti fissi di studente e effetti fissi di esame ☐
   D è improbabile che il modello migliore sia un modello panel ☐

10. individua in quale intervallo si trova l'indice di correlazione lineare del grafico sotto



   A 0.2-0.4 ☐
   B 0.4-0.6 ☐
   C 0.6-0.8 ☐
   D 0.8 - 1 ☐

Risposte aperte

1. Come si interpreta il coefficiente β_3 nel modello di regressione: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2 \times X_3 + u$. Discutete specificando i casi un cui X_3 è una variabile continua e quando invece non lo è.

come Versione A

2. Gli errori di misurazione delle variabili indipendenti e di quella dipendente hanno effetti sui risultati di uno studio di regressione. Spiegate quali.

Gli errori di misurazione nella variabile indipendente hanno sempre un effetto distorsivo sulle stime. Nel caso di errore classico, ovvero non correlato con il valore di X, il segno della distorsione è sempre di ridurne il valore assoluto (attenuation bias). Nel caso di errore non classico il segno della distorsione dipende dalla correlazione dell'errore con Y.

Se l'errore di misura riguarda la variabile dipendente e l'errore è classico non vi sarà distorsione del coefficiente ma soltanto minore precisione nella stima. Se l'errore non è classico invece il segno della sua distorsione dipende dalla correlazione dell'errore con X.

	coefficiente	SE	t	$p-value$
β_0	0.3399	0.08973	5.242	0.0000
β_{CASA}	0.1259	0.05228	2.717	0.0082
β_{NS}	-0.2821	0.31332	-0.816	0.4325
$\beta_{NS \times CASA}$	0.0042	0.0284	1.523	0.0481
$\beta_{INFORTUNATI}$	-0.0232	0.0184	-2.323	0.0086

3. commentate il modello che spiega quali sono le probabilità di vittoria in una partita di calcio: $CASA$ assume valore 1 se la squadra gioca in casa e 0 altrimenti, NS sono gli spettatori paganti presenti allo stadio, $INFORTUNATI$ è il numero di giocatori infortunati che non possono essere schierati in campo e devono essere sostituiti da riserve.

Il modello indica che la squadra che gioca in casa ha una maggior probabilità di vincere la partita (12,59% in più). Il numero di spettatori non ha effetto sulle probabilità di vittoria (coefficiente negativo ma non significativo). Se però la partita viene giocata in casa la situazione cambia, in quel caso il coefficiente è positivo e significativo (+0,4% per ogni aumento unitario di NS, probabilmente espresso in migliaia di spettatori visto che sarebbe altrimenti irrealistico pensare che si tratti dell'effetto di un singolo spettatore in più). Infine ogni giocatore titolare infortunato riduce la probabilità di vittoria dello 0,02%, si tratta di un coefficiente significativo solo al 90%, non al 95%.

Principi di Econometria
prof. Brunori
01/02/2017

55' di tempo
Nome e cognome ____________________
Matricola ____________________

Potete consegnare solo le risposte multiple o sia le risposte multiple che quelle aperte. Nel secondo caso per passare l'esame è necessario superare 18 in entrambe le parti.

Risposta multipla

1. il termine errore u in uno studio di regressione:

   A misura l'errore standard nella stima dell'intercetta ☐

   B è sempre compreso fra 0 e 1 ☐

   C rappresenta la differenza fra il valore predetto di Y e il valore osservato nel campione ☒

   D è sempre nullo quando $R^2 = 0$ ☐

2. un modello lineare di probabilità

   A è un modello che spiega una variabile come funzione di una serie di probabilià ☐

   B stima una serie di effetti fissi di osservazione ☐

   C potrebbe predire probabilità maggiori di 1 o minori di zero ☒

   D predice sempre probabilità comprese fra 0 e 1 ☐

3. per spiegare la probabilità di contrarre l'influenza si stima un modello basato su due soli regressori $Y = \beta_0 + \beta_1 X_{eta<50} + \beta_2 X_{eta\geq 50} + u$, dove $X_{eta<50}$ e $X_{eta\geq 50}$ sono variabili che assumono valore 1 o 0 in base all'età del soggetto.

   A il modello soffre del problema della perfetta multicollinearità ☒

   B il modello stimato non è un modello lineare di probabilità ☐

   C β_1 misura l'elasticità della probabilità X a Y ☐

   D nessuna delle precedenti è vera ☐

4. la relazione che lega reddito e soddisfazione nella vita mostra un andamento non lineare, al di sopra di un certo livello la soddisfazione non aumenta con aumenti successivi del reddito. Quale delle seguenti specifiche sceglireste per stimare la relazione fra queste due variabili?

   A $Y_{soddisfazione} = \beta_0 + \beta_1 X_{reddito} + \beta_2 X^2_{reddito} + u$ ☒

   B $Y_{reddito} + Y^2_{reddito} = \beta_0 + \beta_1 X_{soddisfazione} + u$ ☐

   C $Y_{reddito} = \beta_0 + \beta_1 X_{soddisfazione} + \beta_2 X^2_{soddisfazione} + u$ ☐

   D nessuna delle precedenti è vera ☐

5. avendo stimato l'elasticità del reddito (Y) agli anni di istruzione (S) con l'equazione: $lnY = \beta_0 + \beta_1 ln(S) + u$

   A β_1 è interpretabile come l'aumento percentuale del reddito dovuto ad un anno di istruzione ☐

   B β_1 è interpretabile come l'aumento percentuale di Y dovuto ad un aumento percentuale di S ☒

   C l'elasticità di Y a S è pari a β_0 ☐

   D l'elasticità di S a Y è pari a $\beta_1\%$ ☐

6. Quando la variabile dipendente Y del modello $Y = \beta_0 + \beta_1 X + u$ è stimata con errore e l'errore è di tipo classico:

   A il coefficiente β_1 sarà sottostimato

   B il coefficiente β_1 sarà stimato correttamente ma in modo impreciso

   C se anche X è stimato con errore i due effetti si annullano

   D implica stime distorte di β_1

7. la validità interna di uno studio di regressione si basa su una serie di accorgimenti:

   A l'errore u non deve essere correlato con i regressori X

   B lo studio deve basarsi su un campione rappresentativo della popolazione di interesse

   C il meccanismo di selezione del campione non deve essere casuale

   D tutte le precedenti sono vere

8. in un modello di regressione: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 - \beta_3 X_2^2$ la variazione di Y dovuta ad una variazione unitaria di X_2 è pari a:

   A è sempre minore di 0

   B $\beta_3 + \beta_2$

   C dipende dal valore iniziale di X_2

   D $\beta_2 + \beta_3 \times X_1$

9. due possibili indicatori compositi della qualità di un cappuccino sono ottenuti con le formula: $Y_1 = X_{Latte} \times X_{caffe}$ e $Y_2 = 0.3X_{Latte} + 0.7X_{caffe}$. Le due dimensioni - X_{latte}, X_{caffe} - sono due scale di qualità dei due ingredienti che vanno da 0 a 10:

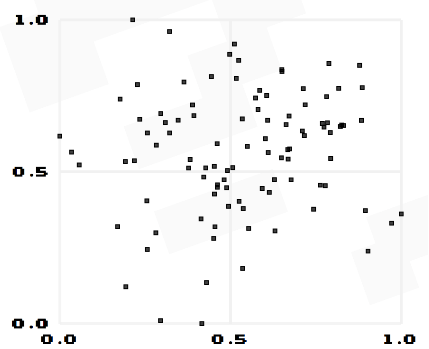
   A si tratta di due indicatori normalizzati compresu fra zero e uno

   B Y_1 penalizza di più la presenza di un ingrediente di pessima qualità

   C la qualità del caffè è più importante di quella del latte per entrambi gli indicatori

   D nessuna delle precedenti è vera

10. individua in quale intervallo si trova l'indice di correlazione lineare del grafico sotto



   A 0.2-0.4

   B 0.4-0.6

   C 0.6-0.8

   D 0.8 - 1

Risposte aperte

1. Come si interpreta il coefficiente β_3 nel modello di regressione: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 \times X_2 + u$. Discutete specificando i casi un cui X_1 sia X_2 che X_3 sono variabili dicotomiche e il caso in cui X_1 è dicotmica ma X_2 è una variabile continua.

beta_3 è il coefficiente di una variabile ottenuta interagendo due regressori. In caso di variabili interagite il coefficiente non coglie l'effetto di una variazione unitaria di uno dei due regressori indipendentemente dall'altro, coglie invece l'effetto della variazione di uno in funzione del valore dell'altro. Nel caso che si tratti di due variabili dicotomiche il coefficiente si interpreta come l'ulteriore traslazione dell'intercetta dovuta alla compresenza di entrambe le caratteristiche (X_1=1 e X_2=1). Per cui, se Y=reddito, X_1=donna e X_2=laureato, beta_3 è interpretabile come l'aumento di reddito dovuto alla laurea per una donna rispetto all'effetto della laurea per un uomo. Se una delle due variabili è continua, ad esempio X_2=anni di studio, allora beta_3 è si interpreta come la diversa inclinazione della retta di regressione che stima il rendimento di un anno di istruzione per le donne rispetto agli uomini.

2. commentate il modello cheo spiega quali sono le probabilità di vittoria in una partita di calcio: $CASA$ assume valore 1 se la squadra gioca in casa e 0 altrimenti, NS sono gli spettatori paganti presenti allo stadio, $INFORTUNATI$ è il numero di giocatori infortunati che non possono essere schierati in campo e devono essere sostituiti da riserve, $INOFRTUNATI^2$ è lo stesso numero al quadrato.

	coefficiente	SE	t	$p-value$
β_0	0.4399	0.08973	6.0432	0.0000
β_{CASA}	0.0954	0.03228	2.917	0.0062
β_{NS}	0.0821	0.09332	0.716	0.5325
$\beta_{INFORTUNATI}$	-0.0132	0.0084	-1.723	0.0362
$\beta_{INFORTUNATI^2}$	-0.0009	0.0004	-2.223	0.0076

L'output di regressione mostra alcune variabili significativamente legate alla probabilià di vittoria in una partita di calcio. Giocare in casa ha un effetto positivo (+9,5%) sulle chances di vittoria. Il numero di spettatori non ha invece un effetto significativo. Come immaginabile avere giocatori infortunati diminuisce la probabilità di vittoria. Questo andamento non è lineare, come si vede dalla presenza di coefficienti significativi sia per il numero di infortunati sia per lo stesso numero al quadrato. L'effetto di un infortunio a un titolare dipende dal numero di infortunati. Se vi è un solo infortunato questa variazione è bari a (- 1.41% somma dei due coefficienti) se gli infortunati sono già molti l'effetto è più forte perché l'effetto negativo degli infortunati al quadrato è maggiore. Se ad esempio gli infortunati passano da 8 a 9 la riduzione della probabilità di vittoria è pari a:

-0.00132-0.0009 x(81-64)= -2.85%

3. Cosa si intende per regressione delle minime devianze assolute? In quali casi potrebbe essere utile utilizzarla?

La regressione delle minime devianze assolute (LAD) stima una retta che interpola i dati minimizzando la somma del valore assoluto delle distanze fra retta e valore predetto.
La differenza fra LAD e OLS è spesso minima, un caso particolare si ha in presenza campioni nei quali siano presenti valori anomali (outlier). In questo caso la regressione OLS rischia di essere fortemente influenzata da un solo valore difforme. Un singolo errore molto grande infatti ha un peso molto elevato quando si stia minimizzando la somma dell'errore di predizione al quadrato. Se si minimizza la somma delle deviazioni in valore assoluto questo effetto è meno pronunciato. In assenza di outlier è preferibile invece usare OLS. Anche se entrambe le stime sono non distorte quella OLS è infatti più efficiente.