

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

- nel modello  $\log Y = \beta_0 + \beta_1 \log X + u$ 
  - l'effetto di X su Y è non lineare
  - $\beta_1$  misura l'elasticità di Y a X
  - $\beta_1$  misura la variazione percentuale di Y dovuta ad una variazione percentuale di X
  - tutte le precedenti sono vere
- $R^2$  corretto
  - è pari al numero di osservazioni  $n$  meno il numero di regressori  $k$
  - misura la % della variabilità della Y catturata dalla variabilità dei regressori
  - è sempre minore o al massimo uguale all' $R^2$
  - è pari a  $R^2 \times (n - k)$
- se una variabile esplicativa  $X$  è una variabile dicotomica (dummy):
  - il suo coefficiente si interpreta come traslazione della retta di regressione rispetto all'intercetta
  - il suo coefficiente non può essere stimato per via della multicollinearità perfetta
  - il suo coefficiente risulta distorto sempre verso il basso
  - il suo coefficiente è interpretabile come variazione dell'inclinazione della retta di regressione
- In una regressione lineare l'eteroschedasticità dell'errore
  - implica che le stime dei coefficienti sono distorte
  - viola una delle assunzioni necessarie alla correttezza dello stimatore OLS
  - rende più complessa la quantificazione dell'errore standard dei coefficienti
  - nessuna delle precedenti è vera
- Per determinare il miglior grado del polinomio di una regressione polinomiale
  - si pone  $k = n$  dove  $k$  è il grado e  $n$  le osservazioni
  - si minimizza il valore  $k$
  - si cerca il valore  $k$  che massimizza l' $R^2$
  - si guarda ad un criterio come  $R^2$  corretto per tener conto dei gradi di libertà
- La normalizzazione di un indicatore  $x$  può essere ottenuta attraverso la seguente formula:
  - $\frac{x_i - x_{max}}{x_i - x_{min}}$
  - $\frac{x_i - x_{min}}{x_{max} - x_{min}}$
  - $\frac{x_{max} - x_i}{x_{min} - x_i}$
  - $\frac{x_i - x_j}{x_{max} - x_{min}}$

7. una variabile omessa  $Z$  in un modello di regressione:

- A è correlata con  $X$  ma non con  $Y$
- B è correlata con  $X$  e determina  $Y$
- C non è correlata né con  $X$  né con  $Y$
- D nessuna delle precedenti è vera

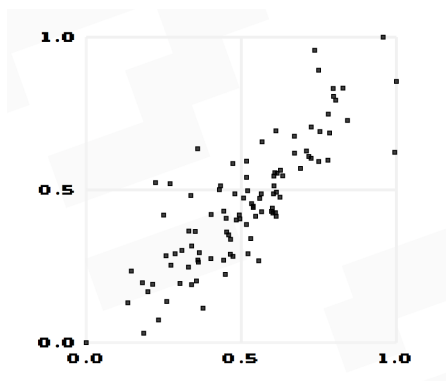
8. in un modello con regressori interagiti:  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 \times X_2$  la variazione di  $Y$  dovuta ad una variazione unitaria di  $X_2$  è pari a:

- A  $\beta_3$
- B  $\beta_3 + \beta_2$
- C  $\beta_3 - \beta_2$  se  $\beta_2 < 0$
- D  $\beta_2 + \beta_3 \times X_1$

9. in una regressione con dati longitudinale (panel) gli effetti fissi temporali

- A colgono la componente di  $Y$  che non varia nel tempo
- B depurano la variabilità di  $Y$  da andamenti costanti nel tempo
- C colgono andamenti nel tempo comuni a tutte le osservazioni
- 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

Principi di Econometria  
 prof. Brunori  
 16/01/2017

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.  $R^2$  corretto
  - A è pari al numero di osservazioni  $n$  meno il numero di regressori  $k$
  - B misura la % della variabilità della  $Y$  catturata dalla variabilità dei regressori tenendo conto dei gradi di libertà
  - C è sempre maggiore o al minimo uguale all' $R^2$
  - D è pari a  $R^2 + \frac{k}{n-k-1}$
2. nel modello  $\log Y = \beta_0 + \beta_1 \log X + u$ 
  - A l'effetto di  $X$  su  $Y$  è lineare
  - B  $\beta_1$  misura l'elasticità di  $X$  a  $Y$
  - C  $\beta_1$  misura la variazione percentuale di  $Y$  dovuta ad una variazione percentuale di  $X$
  - D nessuna delle precedenti è vera
3. se una variabile esplicativa  $X$  è una variabile dicotomica (dummy):
  - A il suo coefficiente si interpreta come la variazione dell'inclinazione della retta di regressione rispetto al coefficiente
  - B il suo coefficiente non può essere stimato per via della multicollinearità perfetta
  - C il suo coefficiente risulta distorto sempre verso il basso
  - D il suo coefficiente è interpretabile come la traslazione dell'intercetta rispetto alla costante
4. La normalizzazione di un indicatore  $x$  può essere ottenuta attraverso la seguente formula:
  - A  $\frac{x_i - x_{min}}{x_{max} - x_{min}}$
  - B  $\frac{x_i - x_{max}}{x_i - x_{min}}$
  - C  $\frac{x_i - x_j}{x_{max} - x_{min}}$
  - D  $\frac{x_{max} - x_i}{x_{min} - x_i}$
5. Per determinare il miglior grado del polinomio di una regressione polinomiale
  - A si pone  $k = n$  dove  $k$  è il grado e  $n$  le osservazioni
  - B si minimizza il valore  $k$
  - C si cerca il valore  $k$  che massimizza l' $R^2$
  - D si guarda ad un criterio come  $R^2$  corretto per tener conto dei gradi di libertà
6. In una regressione lineare l'esistenza di una variabile omessa
  - A implica che le stime dei coefficienti sono distorte
  - B viola una delle assunzioni necessarie alla correttezza dello stimatore OLS
  - C implica una distorsione della stima degli altri coefficienti

D tutte le precedenti sono vere

7. una variabile omessa  $Z$  in un modello di regressione:

A è correlata con  $X$  ma non con  $Y$

B è correlata con  $X$  e determina  $Y$

C non è correlata né con  $X$  né con  $Y$

D nessuna delle precedenti è vera

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

A  $\beta_2 + \beta_3 \times X_1$

B  $\beta_3 + \beta_2$

C  $\beta_3 - \beta_2$  se  $\beta_2 < 0$

D  $\beta_3$

9. in una regressione con dati longitudinale (panel) che spiega la variazione del PIL, gli effetti fissi di paese

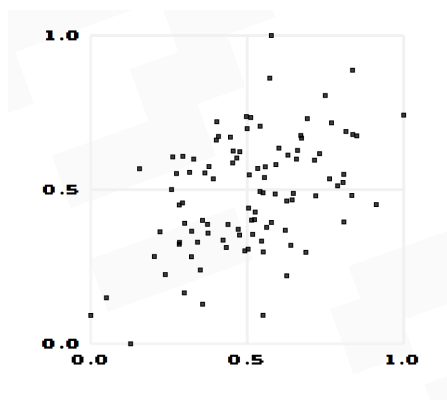
A servono come variabili di controllo per tutti quei fenomeni fissi a livello di paese che influenzano il PIL

B depurano la variabilità di  $Y$  da tutte le variazioni nel tempo

C colgono andamenti nel tempo comuni a tutte le osservazioni

D nessuna delle precedenti è vera

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



C 0.3 - 0.5

D 0.5 - 0.7

A 0.7 - 1

B è minore di 0.3

**Risposte aperte**

1. Quali vantaggi presenta l'utilizzo di dati longitudinali (panel) nella stima di un modello di regressione?

Il fatto che i dati longitudinali raccolgano osservazioni ripetute nel tempo rende possibile:

1) l'identificazioni di andamenti nel tempo della variabile dipendente che riguardano tutte le osservazioni (stati, imprese, persone).

2) l'identificazione di effetti fissi a livello di osservazione sulla variabile dipendente.

Entrambi possono inoltre essere utili nell'identificazione dei coefficienti degli altri regressori in quanto, catturando tutta la variabilità costante a livello individuale e comune a tutte le osservazioni ad ogni momento nel tempo possono fungere da variabili di controllo. In pratica fungendo da variabili di controllo rendono migliore la stima anche degli altri coefficienti.

2. Discutete le stime del modello OLS sotto riportato che spiega la probabilità di iscriversi all'università come funzione di tre variabili: VOTOMAT=voto del diploma, LICEO=1 se ha frequentato il liceo e =0 altrimenti, SUD=1 se residente al sud e =0 altrimenti.

	coefficiente	SE	t	p - value
$\beta_{VOTOMAT}$	0.00399	0.002073	1.242	0.0313
$\beta_{LICEO}$	0.3459	0.03228	10.717	0.0000
$\beta_{SUD}$	-0.2821	0.01332	-21.16	0.0000
$\beta_{SUD \times VOTOMAT}$	-0.0901	0.0384	-2.723	0.0081

La probabilità di iscriversi all'università è funzione del voto di maturità un voto in più aumenta dell' 0,3% la probabilità di iscriversi (significatività del 5%), dipende inoltre dalla scuola di provenienza (chi viene dal liceo ha circa un terzo di probabilità in più di iscriversi). I residenti al sud hanno una probabilità inferiore di iscriversi (-28%). Questi due coefficienti sono entrambi molto significativi (p-value<0.0001). Si noti però che il regressore ottenuto interagendo sud e votomat chiarisce che il significato di un voto elevato è diverso al sud rispetto che nel resto del paese. Il coefficiente è negativo e significativo (al 99%) il che rappresenta un risultato fortemente controintuitivo perché suggerisce che al sud, un voto in più alla maturità abbassi la probabilità di andare all'università, a parità di altre condizioni. Si noti infine che questo modello non prevede la stima di un'intercetta. (nell'altra versione ha segni/p-value leggermente diversi).

3. Cosa si intende per regressione polinomiale? Come si determina il suo grado?

Una regressione polinomiale include fra i controlli uno o più dei regressori elevato a diverso grado. Un caso visto a lezione è quello della stima dell'effetto dell'esperienza sul reddito. In questo caso gli anni di esperienza sono inclusi sia linearmente che al quadrato. Questo serve a cogliere una relazione non lineare fra i regressori e la variabile dipendente. L'effetto della variabile indipendente infatti non è costante ma varia a seconda del suo valore. L'inclusione di regressori in forma polinomiale non è però sempre opportuna. Occorre bilanciare la necessità di tenere conto di possibili effetti non lineari e la necessità di rendere affidabili le stime evitando di inserire troppi termini che, utilizzando gradi di libertà, rendono meno precise le stime.