

Risposta multipla

- La retta dei minimi quadrati:
 - A si ottiene minimizzando la somma dei valori della variabile dipendente al quadrato
 - B minimizza la somma della differenza fra valore predetto e valore reale
 - C è quella retta che minimizza la somma della distanza fra la retta e le osservazioni elevata al quadrato
 - D misura la distanza fra il valore Y_i e il suo valore predetto
- In una regressione in cui sia la variabile dipendente che quella indipendente sono espressi in logaritmi
 - A il coefficiente misura l'aumento nella Y dovuto ad un aumento unitario della X
 - B il coefficiente si può interpretare come l'elasticità della X alla Y
 - C il coefficiente si può interpretare come l'elasticità della Y alla X
 - D tutte le precedenti sono vere
- Se si utilizza un modello autoregressivo per spiegare l'andamento di una variabile nel tempo:
 - A la presenza di rotture strutturali nell'evoluzione della variabile rappresenta un problema
 - B la presenza di trend nell'andamento della variabile rappresenta un problema
 - C una misurazione del fenomeno con errore rappresenta un problema
 - D tutte le precedenti sono vere
- Una variabile omessa:
 - A è una variabile non correlata con la variabile dipendente
 - B è correlata con la dipendente ma non con la variabile indipendente
 - C è tale solo se è correlata con la dipendente ma non con l'indipendente
 - D nessuna delle precedenti è vera
- in un modello di probabilità lineare che spiega la probabilità di passare un esame:
 - A la costante (β_0) si interpreta come la probabilità di passare di un soggetto per cui tutte le variabili esplicative assumono valore zero
 - B il coefficiente della variabile X è interpretabile come l'elasticità della variabile Y a variazioni della variabile Y
 - C la probabilità predetta è sempre compresa fra 0 e 1
 - D il coefficiente β_1 misura la probabilità di passare l'esame per lo studente con tale caratteristica
- la perfetta multicollinearità:
 - A è il fenomeno opposto all'eteroschedasticità
 - B implica distorsione nelle stime ottenute tramite regressione lineare
 - C impedisce la stima dei coefficienti di una regressione lineare
 - D tutte le precedenti sono vere

Risposte aperte

1. un modello OLS spiega il numero di medaglie vinte alle Olimpiadi di Rio da ogni paese:

$$Y_i = \beta_0 + \beta_1 PIL_i + \beta_2 PIL_i^2 + \beta_3 BRAZIL_i + \beta_4 GUERRA + u_i$$

dove PIL è il Prodotto interno lordo, $BRAZIL$ è una variabile dicotomica che vale 1 solo per il Brasile e $GUERRA$ è una variabile dicotomica se nel paese è presente un conflitto

- cosa significa un coefficiente pari a +3.21 per la variabile $BRAZIL$?

Il Brasile ha ottenuto 3.21 medaglie in più rispetto a quanto atteso tenendo in considerazione le altre variabili esplicative. L'effetto potrebbe essere collegato al fatto che il Brasile ha ospitato i giochi.

- come si interpreta un coefficiente pari a +0.02 per PIL e pari a -0.0001 per PIL^2 ?

I due coefficienti congiuntamente catturano una relazione non lineare che lega PIL di un paese e le medaglie che ottiene. L'aumento delle medaglie dovute a maggior PIL è decrescente via via che il PIL aumenta.

- che segno vi aspettate che abbia il coefficiente per $GUERRA$?

Il segno sarà negativo poiché è improbabile atleti di paesi in guerra possano adeguatamente preparare i giochi olimpici.

2. un modello OLS che spiega il costo di un ombrellone+sdraio per un giorno in Puglia:

variabile	coefficiente	errore standard	t	p-value
costante	3.8522	0.9481	4.0603	0.0000
Agosto	1.0984	0.3504	3.1346	0.0015
Domenica	0.4984	0.2623	1.8997	0.0544
Salento	- 0.2644	0.1065	2.4815	0.0144
Salento × Agosto	1.0043	0.3616	2.7769	-0.0056
Distanza da parcheggio (km)	-0.0893	0.0348	-2.5621	0.0106

- Qual'è la situazione in cui l'ombrellone+sdraio ha il costo massimo?

Una domenica di agosto in Salento in un bagno con il parcheggio accanto (distanza=0).

- come si interpretano i coefficienti di 'Salento' e 'Salento × Agosto'?

Il primo suggerisce che in media nel Salento il costo sia di circa 25 centesimi inferiore ad altre zone della Puglia. Il secondo però chiarisce che nel caso sia Agosto la relazione cambia e il costo è superiore nel Salento di oltre un euro.

3. Sempre riguardo alla stima del modello dell'esercizio 2

- come possiamo interpretare il coefficiente che si riferisce alla distanza dello stabilimento dal parcheggio più vicino?

Ogni Km di distanza dal parcheggio riduce di 8,9 centesimi il costo. Si tratta di un effetto significativo al 5% ma piccolo (10 km a piedi riducono il costo di soli 89 centesimi)

4. Ipotizzate di voler stilare una classifica dei posti dove mangiare panzerotti a Bari

- Indicate 4 dimensioni che secondo voi dovrebbero essere considerate come rilevanti, almeno una deve essere misurata con una variabile continua, almeno uno con una variabile qualitativa e almeno una con una variabile dicotomica.

1) prezzo (variabile continua), 2) qualità da pessima a eccellente (variabile qualitativa), 3) posto a sedere fuori (0 se no, 1 se si = variabile dicotomica), 4) dimensione piccoli, medi o grandi (variabile discreta)

- Spiegate come le 4 dimensioni potrebbero essere aggregate per ottenere un indicatore standardizzato compreso fra 0 e 1. Dove 1 rappresenta il miglior possibile luogo per mangiare panzerotti e 0 il peggiore. Se riuscite scrivete la formula del vostro indicatore.

Il modo più semplice per ottenere l'indicatore consiste nel trasformare tutte le variabili in variabili numeriche. Prezzo = Prezzo - Prezzo_max, qualità = pessima=0, ..., eccellente = 10, posto a sedere fuori=1, dimensioni: piccolo=1, ..., grande=3. Dopo di che si standardizzano tutte le variabili sottraendo il massimo e dividendole per la differenza fra massimo e minimo. Infine si calcola la media di tutte le variabili. Notate però che la variabile prezzo deve essere inserita cambiandogli di segno (tanto più alto il prezzo peggiore è il posto).