

Tecniche di indagine statistica - Prova d'esame

1. Illustrare il significato degli acronimi e indicare il vostro accordo o disaccordo rispetto alle seguenti affermazioni, motivando la risposta:
 - a. *L'elevato numero di rispondenti nelle indagini CAWI assicura buone stime delle caratteristiche della popolazione*
 - b. *L'errore di copertura è il principale problema nelle indagini CAPI sulla popolazione in Italia*
 - c. *L'errore di copertura è il principale problema nelle indagini CATI sulla popolazione in Italia*
2. Con riferimento al disegno di indagine retrospettivo e prospettivo per ottenere informazioni longitudinali:
 - a. descrivere le caratteristiche principali dei due disegni
 - b. fare il confronto tra i due rispetto al problema della non risposta
3. Nella tabella seguente sono riportati per una certa regione alcuni risultati sull'offerta di corsi di informatica nelle scuole superiori, provenienti da fonti amministrative e da una indagine su un campione casuale di 1500 scuole. I dati sono distinti per gestione della scuola (paritetica e statale).

		Scuole paritetiche	Scuole statali
# campione		500	1000
tasso di risposta (%)		50	60
% scuole che offrono corsi di informatica	Scuole rispondenti	5	50
	Scuole non rispondenti	0	35

- a. assumendo che le dimensioni campionarie nei due gruppi (paritetiche/statali) rappresentino le proporzioni nella popolazione, stimare sulla base dei rispondenti la percentuale totale di scuole che offre corsi di informatica e l'errore standard di tale proporzione (tralasciando il termine di correzione per popolazioni finite);
 - b. fornire una stima della distorsione da non risposta nel risultato ottenuto al punto a.
4. Il seguente modello è stato utilizzato per analizzare l'efficacia di un corso di formazione sulle performance di un campione di 32 dipendenti:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \quad i = 1, \dots, 32$$

dove:

$y = 1$ se la performance del dipendente è migliorata nel periodo successivo allo svolgimento del corso, 0 altrimenti;

$x_1 =$ valutazione sul dipendente espressa dal suo dirigente prima dello svolgimento del corso (scala 18-30);

x_2 = punteggio riportato in un test volto ad accertare le conoscenze iniziali sugli argomenti trattati durante il corso (scala 18-30);

$x_3 = 1$ se il dipendente ha seguito il corso, 0 altrimenti.

La stima del modello secondo la formulazione lineare e la formulazione *logit* produce i seguenti risultati (errori standard tra parentesi):

	<i>Modello lineare</i>	<i>Modello logit</i>
costante	-1.498 (2.345)	-13.02 (15.987)
x_1	0.464 (0.043)	2.826 (0.643)
x_2	0.379 (0.087)	2.379 (0.566)
x_3	0.010 (0.001)	0.095 (0.007)
Valore min di \hat{y}_i	-1.03	0.09
Valore max di \hat{y}_i	1.34	0.89

- commentare i valori minimi e massimi previsti per la variabile dipendente nelle due formulazioni del modello
- è possibile interpretare direttamente i coefficienti di x_1 e x_2 come effetti sulla variabile dipendente? Motivare la risposta.
- che cosa si può dire sulla significatività dei coefficienti? E sul modello?
- sulla base delle stime fornite dal modello *logit*, calcolare il valore massimo e minimo di $Prob(y_i = 1)$ e confrontarla con quella ottenuta per un dipendente con valutazione del dirigente e voto al test pari a 18 e che abbia seguito il corso.