

## Scuola di Scienze Politiche

## Corso di Laurea LM–88 in SOCIOLOGIA E RICERCA SOCIALE

## Compito di Metodi Statistici per la Ricerca Sociale

27 Settembre 2017 – Appello Straordinario

## Parte applicativa

## Esercizio 1

La seguente tabella riporta per un campione di 5 studenti il punteggio (misurato su scala 0-100) a un test finalizzato a valutare le conoscenze di matematica di matricole di una certa scuola prima e dopo un corso di introduzione alla matematica

Studente	Punteggio al test				
	1	2	3	4	5
Prima	66	62	67	68	75
Dopo	72	62	74	77	83

Si supponga che il punteggio al test prima e dopo il corso abbia distribuzione Normale con medie  $\mu_{pre}$  e  $\mu_{post}$  e varianze  $\sigma_{pre}^2$  e  $\sigma_{post}^2$ .

- (a) Stimare la media e la varianza delle differenze tra il punteggio al test dopo il corso e il punteggio al test prima del corso

Si supponga di voler valutare l'ipotesi che la media del punteggio al test prima del corso sia inferiore alla media del punteggio al test dopo il corso (ipotesi alternativa)

- (b) (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 5%;
- (c) Calcolare il valore della statistica test e prendere una decisione
- (d) Il  $p$ -valore per tale test è  $p = 0.018$ . Interpretare tale valore anche sulla base del risultato ottenuto al punto (c)

## Esercizio 2

La seguente tabella mostra il numero di anni di studio della lingua inglese e il punteggio ottenuto in un test di conoscenza della lingua per un campione di 6 persone

Anni di studio ( $X$ )	6	8	7	4	12	5
Punteggio ( $Y$ )	74	78	79	56	89	71

- (a) Determinare le stime dei minimi quadrati dei coefficienti della retta di regressione che pone il punteggio ( $Y$ ) in funzione del numero di anni di studio ( $X$ ):  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$
- (b) Stimare la varianza degli errori
- (c) Si consideri il test delle ipotesi  $H_0 : \beta_1 = 0$  versus  $H_a : \beta_1 \neq 0$ : (i) scrivere la regione critica (di rifiuto) al livello di significatività del 5%; (ii) calcolare il valore osservato della statistica test e prendere una decisione
- (d) Stimare il punteggio medio per soggetti con 10 anni di studio e costruire il relativo intervallo di confidenza al livello di confidenza del 95%.

## Esercizio 3

In uno studio sulla conoscenza della lingua inglese in un campione di 30 giovani adulti si sono rilevate le seguenti variabili:  $Y$  = punteggio ottenuto in un test di conoscenza della lingua inglese ( $SST = 1075.5$ );  $X_1$  = numero di anni di studio della lingua inglese;  $X_2$  = soggiorni di studio/lavoro in paesi di lingua madre inglese (1 = Mai; 2 = Una volta; 3 = Più di una volta). Si considerino i seguenti modelli di regressione: Modello esteso in cui il punteggio al test è posto in funzione del numero di anni di studio della lingua inglese e della variabile soggiorni di studio in paesi di lingua madre inglese ( $SSE_e = 671.5$ ); modello ridotto in cui il punteggio al test è posto in funzione del numero di anni di studio della lingua inglese. ( $SSE_r = 820.7$ )

- (a) Costruire la tavola di analisi della varianza per il modello esteso e per il modello ridotto
- (b) (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa relativa al confronto tra il modello esteso e il modello ridotto;  
(ii) Scrivere la regione critica al livello di significatività del 5%
- (c) Calcolare la statistica test e prendere una decisione giustificando la risposta. In base al risultato ottenuto, il  $p$ -valore per tale test sarebbe maggiore, uguale o minore di 0.05 (giustificare la risposta)

## Domande di Teoria

### Domanda 1

In uno studio sulla conoscenza della lingue inglese in un campione di 255 giovani adulti si è stimato un modello di regressione (analisi della covarianza) in cui la variabile risposta  $Y$  = punteggio ottenuto in un test di conoscenza della lingua inglese viene posta in funzione delle variabili  $X_1$  = numero di anni di studio della lingua inglese;  $X_2$  = soggiorni di studio/lavoro in paesi di lingua madre inglese (1 = Mai; 2 = Una volta; 3 = Più di una volta) e della loro interazione

Variabile	Coefficiente	
Costante	20.00	1.07
Anni di studio ( $X_1$ )	1.50	0.09
Soggiorni studio ( $X_2$ )		
Una volta	1.75	1.2
Più di una volta	2.25	1.98
Interazione ( $X_1 \cdot X_2$ )		
$X_1 \cdot$ Una volta	0.50	0.12
$X_1 \cdot$ Più di una volta	3.00	0.17

- (a) Scrivere l'equazione di regressione stimata
- (b) Interpretare il coefficiente relativo agli anni di studio
- (c) Determinare il punteggio medio atteso per un soggetto con 15 anni di studio della lingua inglese con ha fatto un solo soggiorno di studio in un paese di lingua madre inglese

### Domanda 2

La seguente tabella mostra le distribuzioni di frequenza marginali delle preferenze per il tempo libero e del titolo di studio su un campione di giovani adulti

- (a) Completare la tabella che segue ipotizzando che preferenze per il tempo libero e del titolo di studio siano variabili statisticamente indipendenti

Titolo di studio	Cinema e TV	Lettura	Musica	Sport	Totale
Scuola dell'Obbligo					360
Diploma					720
Diploma universitario o più					540
Totale	486	162	324	648	1620

- (b) Supporre di fare un test  $\chi^2$  per valutare l'ipotesi che preferenze per il tempo libero e titolo di studio siano statisticamente indipendenti. Specificare la distribuzione (asintotica) della statistica test.
- (c) Scrivere la regione critica del test al punto (b) per un livello di significatività  $\alpha = 0.01$ .