

**Scuola di Scienze Politiche**  
**Corso di Laurea LM-88 SOCIOLOGIA E RICERCA SOCIALE**  
**Compito di Metodi Statistici per la Ricerca Sociale**  
**09 Febbraio 2016**  
**Parte applicativa**

**Esercizio 1**

La seguente tabella mostra alcune statistiche descrittive sull'età al primo figlio calcolate su un campione di donne italiane con almeno un figlio classificate per titolo di studio:

Titolo di studio	Dimensione campionaria	Età al primo figlio	
		Media	Varianza
Titolo inferiore alla laurea	20	28.5	9.61
Laurea o più	16	30.5	10.24

Rispondere ai seguenti quesiti nell'ipotesi che la distribuzione dell'età al primo figlio nelle due popolazioni di donne definite dal titolo di studio sia Normale con uguale varianza.

- Stimare la differenza media tra l'età al primo figlio di donne con laurea o più e con titolo di studio inferiore alla laurea. Stimare l'errore standard della stimatore della differenza tra le medie.
- Verificare l'ipotesi che l'età al primo figlio di donne con laurea o più sia maggiore rispetto l'età al primo figlio di donne con titolo di studio inferiore alla laurea: (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 5%; (iii) Prendere una decisione
- Per il test al punto (b) – (i) scrivere la regione critica al livello di significatività del 1% e prendere una decisione. Come varia la regione critica? (Giustificare la risposta)
- Il  $p$ -value per tale test è 0.033. Commentare tale valore anche sulla base della decisione presa al punto (b) e al punto (c)

**Esercizio 2**

I dati che seguono sono relativi a un campione di  $n = 7$  donne con almeno un figlio su cui si è osservato il numero di anni di istruzione (variabile  $X$ ) e l'età al primo figlio (variabile  $Y$ ).

Donna	1	2	3	4	5	6	7
Anni di istruzione ( $X$ )	13	10	18	15	19	24	20
Età al primo figlio ( $Y$ )	27	25	31	33	32	34	28

- Stimare la covarianza tra numero di anni di istruzione e età al primo figlio
- Stimare i coefficienti del modello di regressione che pone l'età al primo figlio in funzione del numero di anni di istruzione. Riportare le stime dei coefficienti approssimate a due cifre decimali.
- Stimare la varianza degli errori del modello di regressione sapendo che la somma dei quadrati dei residui è 32.97
- Utilizzando il modello di regressione stimato, stimare età al primo figlio attesa per donne con 19 anni di istruzione. Costruire quindi un intervallo di confidenza al 90% per l'età al primo figlio atteso per donne con 19 anni di istruzione

**Esercizio 3**

Su una campione di  $n = 30$  donne italiane con almeno un figlio vengono osservate le seguenti variabili.  $Y$  = Età al primo figlio;  $X_1$  = Zona di residenza (1 = Sud-isole; 2 = Centro; 3 = Nord) e  $X_2$  = Titolo di studio (1 = Nessun titolo o scuola dell'obbligo; 2 = Diploma; 3 = Laurea o più). Si supponga di adattare un modello di analisi della varianza che pone l'età al primo figlio in funzione della zona di residenza e del titolo di studio.

- Completare la tavola di analisi della varianza che segue relativa a tale modello:

Fonte di variabilità	Somma dei quadrati	Gradi di libertà	Media dei quadrati	$F$ -value	$p$ -value
Regressione					0.0044
Residua	72.16			—	—
Totale	129.4			—	—

Specificare l'ipotesi nulla e l'ipotesi alternativa a cui si riferisce il valore della statistica  $F$  e prendere una decisione sulla base del  $p$ -valore riportato nell'ultima colonna

- (b) Confrontare tale modello con il modello ridotto in cui non è inserita la variabile  $X_1 = \text{Zona di residenza}$  sapendo che la somma dei quadrati dei residui del modello ridotto è  $SQE_r = 100.42$ . (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 1%; (iii) Calcolare il valore della statistica test e prendere una decisione
- (c) In base al risultato ottenuto al punto (b), il  $p$ -valore relativo al test condotto al punto (b) sarebbe stato maggiore, minore o uguale a 1%? Giustificare la risposta.

### Domande di Teoria

#### Domanda 1

Si consideri lo studio dell'esercizio 3. La seguente tabella riporta le stime dei coefficienti e degli errori standard del modello di analisi della varianza che pone l'età al primo figlio in funzione della zona di residenza e del titolo di studio (le categorie di riferimento sono tra parentesi).

Variabile	Stima	Errore standard
Costante	26.8	0.724
Zona di residenza (Sud-isole)		
Centro	1.6	0.735
Nord	2.7	0.872
Livello di istruzione (Obbligo o meno)		
Diploma	1.3	0.759
Laurea o +	2.4	0.808

- (a) Scrivere l'equazione di regressione stimata
- (b) Interpretare i coefficienti di regressione
- (c) Scrivere il modello di analisi della varianza in presenza di interazione tra zona di residenza e livello di istruzione.

#### Domanda 2

In un indagine sulle attività del tempo libero a un campione di adulti tra i 30 e i 50 anni si chiede di esprimere una preferenza tra 1 = "Lettura"; 2 = "Musica"; 3 = "Cinema" (Variabile  $Y$ ). La seguente tabella mostra i risultati di un'analisi di regressione logistica multinomiale in cui  $Y$  è posta in funzione dell'età e del titolo di studio (1 = Nessun titolo o scuola dell'obbligo; 2 = Diploma; 3 = Laurea o più).

Variabile	Coefficiente	Errore standard
<u>Logit Lettura versus Cinema</u>		
Costante	1.28	0.40
Età	0.02	0.003
Livello di istruzione (Obbligo o meno)		
Diploma	0.82	0.05
Laurea o +	1.20	0.25
<u>Logit Musica versus Cinema</u>		
Costante	1.62	0.84
Età	0.03	0.004
Livello di istruzione (Obbligo o meno)		
Diploma	0.55	0.02
Laurea o +	0.90	0.03

- (a) Scrivere l'espressione delle tre probabilità stimate
- (b) Calcolare le probabilità di successo stimate per un soggetto laureato di età uguale a 40 anni
- (c) Interpretare i coefficienti dell'età in termini di odds