

## LM-88 SOCIOLOGIA E RICERCA SOCIALE

### Compito di Metodi Statistici per la Ricerca Sociale 06 Luglio 2015

Studente: \_\_\_\_\_

Matricola: \_\_\_\_\_

#### Esercizi

1. La seguente tabella riporta alcune statistiche di sintesi del numero (medio) di ore impiegate settimanalmente per i lavori domestici dagli uomini e dalle donne

Genere	Numero di osservazioni	Media campionaria	Varianza campionaria
Uomini	20	9.5	4
Donne	27	12.5	9

- (a) Stimare la differenza tra le medie del numero (medio) di ore impiegate settimanalmente per i lavori domestici dalle donne e dagli uomini. Stimare quindi l'errore standard delle differenze tra le medie campionarie.
- (b) Stabilire con un test al livello di significatività del 10% se il numero (medio) di ore impiegate settimanalmente per i lavori domestici è maggiore tra le donne. Specificare le ipotesi, scrivere la regione critica, calcolare il valore della statistica test e prendere una decisione. Il  $p$ -value per tale test è 0.1295. Interpretare il  $p$ -value.
- (c) Costruire un intervallo di confidenza al 90% per la differenza tra le medie del numero (medio) di ore impiegate settimanalmente per i lavori domestici dalle donne e dagli uomini.
2. In uno studio sulla relazione tra fecondità e livello di istruzione su un campione di 5 donne si è rilevato il numero dei figli (variabile  $Y$ ) e il livello di istruzione misurato in anni di scuola (variabile  $X$ ). La seguente tabella riporta i valori osservati:

Variabile	<i>Donna</i>				
	1	2	3	4	5
Livello di istruzione ( $X$ )	20	15	8	18	16
Numero di figli ( $Y$ )	0	1	3	1	2

- (a) Calcolare il coefficiente di correlazione tra livello di istruzione e numero di figli
- (b) Determinare le stime dei minimi quadrati dei coefficienti della retta di regressione che pone il numero dei figli ( $Y$ ) in funzione del livello di istruzione ( $X$ )
- (c) Stimare la varianza degli errori
- (d) Stimare il numero medio di figli per una donna con 10 anni di istruzione e costruire il relativo intervallo di confidenza al livello di confidenza del 90%.

3. In uno studio sul consumo di droghe leggere viene stimato un modello logit in cui l'uso della marijuana ( $0 = \text{No}$ ,  $1 = \text{Si}$ ) è la variabile risposta ( $Y$ ) e l'uso di sigarette ( $X = 0 = \text{No}$ ,  $X = 1 = \text{Si}$ ) è la variabile esplicativa. La seguente tabella mostra i risultati della stima del modello:

Variabile	Coefficiente	Errore standard
Costante	-5.309	0.4752
Uso di sigarette	2.848	0.1638

- Interpretare il coefficiente dell'uso di sigarette in termini di odds
- Costruire un intervallo di confidenza al livello di confidenza del 95% per il rapporto degli odds a favore dell'uso di marijuana per coloro che fumano rispetto a coloro che non fumano
- Calcolare la probabilità stimata di uso di marijuana per soggetti che fumano sigarette

## Domande Teoriche

1. La seguente tabella mostra i risultati di un'analisi di regressione sugli stipendi (in Euro) di 500 neo-laureati alla prima occupazione. Come variabili esplicative sono considerate 2 variabili categoriche, Genere e Tipo di laurea (con categoria di riferimento tra parentesi) e una variabile continua (tempo trascorso dalla laurea in mesi)

Variabile	Stima	Errore standard
Costante	860	12.5
Genere (Femmina)		
Maschio	150	16.2
Tipo di laurea (Umanistica)		
Tecnico/Scientifica	82.8	17.2
Socio-Economica	138.3	27.8
Tempo dalla laurea	-10.5	2.0

- Scrivere l'equazione di regressione stimata e interpretare i coefficienti di regressione
  - Una delle ipotesi sottostanti il modello di regressione è l'ipotesi di omoschedasticità. Spiegare il significato dell'ipotesi di omoschedasticità
2. Ognuna delle seguenti affermazioni contiene un errore. Individuare l'errore spiegando perchè si tratta di un errore.
- La correlazione fra reddito e spese per consumo calcolata su 100 famiglie corrisponde a  $r = 0.76$  Euro.
  - L'indice di determinazione lineare calcolato da una regressione dello stipendio di ricercatori universitari in funzione degli anni di esperienza, il settore di ricerca e il genere suggerisce che le variabili esplicative spiegano il 110% ( $R^2 = 1.1$ ) della variabilità totale dello stipendio
  - In uno studio sull'abbandono degli studi universitari si trova che la differenza tra le proporzioni di abbandoni entro il primo anno in due atenei ( $A$  e  $B$ ) è  $-1.5$ , suggerendo che nell'Ateneo  $A$  la propensione a abbandonare gli studi delle matricole è inferiore.