

LM-88 SOCIOLOGIA E RICERCA SOCIALE

Compito di Metodi Statistici per la Ricerca Sociale 12 Giugno 2014

Studente: _____

Matricola: _____

Esercizi

1. Gli impiegati nelle grandi aziende affermano spesso che lo stipendio medio offerto dalle aziende differisce a seconda della zona. Per verificare questa affermazione sono stati raccolti i dati sugli stipendi medi (in migliaia di Euro) nelle aziende del Nord, Centro e Sud Italia.

Gruppo	Numerosità del gruppo	Media Campionaria	Varianza Campionaria
j	n_j	\bar{x}_j	s_j^2
Nord	25	40.0	6.25
Centro	20	38.5	4.84
Sud	18	36.0	4

- (a) Calcolare la devianza tra gruppi e la devianza entro gruppi.
- (b) Verificare l'ipotesi che non ci sia differenza tra il salario medio nelle diverse zone geografiche al livello di significatività del 10%.
2. I dati che seguono sono relativi al numero dei visitatori (in milioni di unità) e agli introiti dal turismo (in miliardi di euro) in sei paesi.

Paese	Visitatori (in milioni di unità)	Introiti dal turismo in miliardi di euro
Francia	60	27.5
Spagna	48	25.0
USA	45	58.5
Italia	30	27.0
Regno Unito	20	17.5
Germania	15	12.0

- (a) Stimare la retta di regressione che pone gli introiti dal turismo (Y) in funzione del numero di visitatori (X)
- (b) Stimare la varianza degli errori
- (c) Costruire un intervallo di confidenza per il coefficiente di regressione β_1 al livello di confidenza del 95%.

3. Un campione di 300 giovani è stato classificato secondo lo condizione lavorativa e la zona geografica di residenza come segue:

Zona	Condizione lavorativa	
	Occupato	Disoccupato
Nord	65	45
Centro-Sud	75	115

- (a) Verificare l'ipotesi che non ci sia differenza tra le probabilità di essere occupato nelle diverse zone geografiche al livello di significatività del 5%.
- (b) Calcolare il rapporto degli odds considerando come successo l'evento 'essere occupato' e costruire un intervallo di confidenza per il rapporto degli odds al livello di confidenza del 95%.

Domande Teoriche

1. In uno studio finalizzato ad analizzare la relazione tra il prezzo di vendita delle case in euro (Price) e la dimensione della casa in metri quadri (Size) e se la casa è nuova (New) si sono ottenuti i seguenti risultati:

Call:

```
lm(formula = Price ~ Size + New, data = house)
```

Residuals:

Min	1Q	Median	3Q	Max
-205102	-34374	-5778	18929	163866

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-40230.867	14696.140	-2.738	0.00737	**
Size	116.132	8.795	13.204	< 2e-16	***
New	57736.283	18653.041	3.095	0.00257	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 53880 on 97 degrees of freedom

Multiple R-squared: 0.7226, Adjusted R-squared: 0.7169

F-statistic: 126.3 on 2 and 97 DF, p-value: < 2.2e-16

- (a) Scrivere l'equazione di stima e interpretare i coefficienti di regressione
- (b) Interpretare i p -values relativi ai singoli coefficienti e il p -value relativo alla statistica F (specificando le ipotesi a cui si riferiscono)
- (c) Interpretare il coefficiente di determinazione lineare (R^2)
2. Descrivere cos'è l'analisi di sopravvivenza, fornendo le definizioni dei concetti principali (evento, durata, popolazione a rischio, funzione di rischio e funzione di sopravvivenza, eventualmente anche con esempi).