

Scuola di Scienze Politiche
Corso di Laurea LM–88 in SOCIOLOGIA E RICERCA SOCIALE
Compito di Metodi Statistici per la Ricerca Sociale
30 Gennaio 2018
Parte applicativa

Esercizio 1

La seguente tabella è relativa a uno studio sull'occupazione giovanile. Un campione di soggetti in cerca di prima occupazione è stato suddiviso in due gruppi. Un gruppo ha frequentato un corso di formazione di sei mesi e un gruppo no. A sei mesi dalla fine del corso di formazione è stato rilevato lo stato occupazionale di tutti i soggetti: $Y_i = 1$ se il soggetto i è occupato dopo un anno e $Y_i = 0$ altrimenti.

Corso di formazione	Status occupazionale	
	Non Occupato	Occupato
No	225	525
Si	125	375

- (a) Stimare le proporzioni di occupati nella popolazione di soggetti che non hanno frequentato il corso di formazione e nella popolazione di soggetti che hanno frequentato il corso di formazione. Stimare quindi la differenza tra la proporzione di occupati nella popolazione di soggetti che hanno frequentato il corso di formazione e la proporzione di occupati nella popolazione di soggetti che non hanno frequentato il corso di formazione
- (b) Confrontare le proporzioni di occupati nelle due popolazioni usando come ipotesi alternativa l'ipotesi che la proporzione di occupati sia maggiore nella popolazione di soggetti che hanno frequentato il corso di formazione: (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) specificare la regione critica al livello di significatività del 5%; (iii) calcolare il valore della statistica test e prendere una decisione. (iv) Calcolare il p -valore e interpretarlo
- (c) Stimare gli odds a favore di essere occupato nelle due popolazioni
- (d) Stimare il rapporto degli odds e costruire un intervallo di confidenza per il rapporto degli odds al livello di confidenza del 97%

Esercizio 2

La seguente tabella mostra i risultati dell'analisi di regressione su un campione di 25 individui in cui la variabile Y = punteggio a un test attitudinale di accesso a un corso di formazione è posto in funzione dell'età X_1 = età (scartata dalla media uguale a 30) e dagli anni di istruzione scartati dalla media (uguale a 15).

Variabile	Coefficiente	Errore standard
Costante	80.35	9.21
Età - 30 ($X_1 - \bar{X}_1$)	-0.50	0.04
Anni di istruzione - 15 ($X_2 - \bar{X}_2$)	1.25	0.12

Inoltre $s_Y^2 = 81$ e $SSE = 950.4$

- (a) Completare la seguente tavola di analisi della varianza

Fonte di variabilità	Somma dei quadrati	Gradi di libertà	Media dei quadrati	F-value
Regressione				
Residua				
Totale				

- (b) Scrivere l'ipotesi nulla e l'ipotesi alternativa a cui è relativo il valore della statistica F (F-value) riportato nella tavola di analisi della varianza. Costruire la regione critica al livello di significatività del 5% per tale test. Prendere una decisione
- (c) Si consideri il test delle ipotesi $H_0 : \beta_2 = 0$ versus $H_0 : \beta_2 \neq 0$ dove β_2 è il coefficiente relativo alla variabile anni di istruzione. Costruire la regione critica al livello di significatività del 5% per tale test e decidere se rifiutare o non rifiutare l'ipotesi nulla.
- (d) Stimare il punteggio medio per soggetti di 35 anni con 20 anni di studio

Esercizio 3

Di seguito si riportano i risultati di un'analisi in cui un modello di regressione logistica è stato stimato su un campione di 254 uomini anziani. La variabile risposta è una variabile binaria che assume valore 1 per soggetti che presentano sintomi di senilità. Come variabile esplicativa si considera il punteggio ottenuto a un particolare test (il punteggio WAIS = Wechsler Adult Intelligence Scale). Valori maggiori della variabile WAIS indicano capacità intellettive migliori.

Variabile	Coefficiente	SE	p -value
Costante	3.42	1.61	0.0378
WAIS	-0.43	0.16	0.0081

- (a) Interpretare il coefficiente relativo alla variabile WAIS in termini di odds
- (b) Stimare la probabilità di soffrire di senilità per soggetti con WAIS uguale a 5
- (c) Il logaritmo della verosimiglianza per tale modello è -1942.7 . Utilizzare il test del rapporto di verosimiglianza per confrontare tale modello con un modello esteso in cui è inclusa la variabile genere ($\text{logit } \pi_i = \beta_0 + \beta_1 \text{ WAIS} + \beta_2 \text{ Maschio}$) sapendo che logaritmo della verosimiglianza per il modello esteso è -1943.8 : (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% (iii) Calcolare la statistica test e prendere una decisione giustificando la risposta; (iv) 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 finalizzato ad analizzare la relazione tra il reddito annuale in euro e genere (M= Maschio, F= Femmina) e titolo di studio (Obbligo, Diploma, Laurea) si sono ottenuti i seguenti risultati:

```
> A <- factor(A, labels=c("Obbligo", "Diploma", "Laurea"))
> summary(lm(Y~M+A))
```

Call:

```
lm(formula = Y ~ M + A)
```

Residuals:

```
Min      1Q  Median      3Q      Max
-4536.7 -832.7  -77.2   952.9  4148.0
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  35755.0      202.8  176.342 < 2e-16 ***
M              773.2       192.1   4.025 7.58e-05 ***
ADiploma     1590.9       238.9   6.660 1.78e-10 ***
ALaurea     2577.3       238.1  10.822 < 2e-16 ***
```

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

Residual standard error: 1515 on 246 degrees of freedom

Multiple R-squared: 0.3541, Adjusted R-squared: 0.3462

F-statistic: 44.95 on 3 and 246 DF, p-value: < 2.2e-16

- (a) Scrivere l'equazione di regressione stimata
- (b) Interpretare i coefficienti di regressione relativi alla variabile titolo di studio
- (c) Si consideri il valore della statistica test **F-statistic**: (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa a cui è relativo tale valore; (ii) Interpretare il relativo p -valore

Domanda 2

Si consideri uno studio di associazione tra livello di benessere economico e livello di soddisfazione verso un certo servizio

- (a) Completare la seguente tabella nell'ipotesi che nella popolazione le due variabili ' siano indipendenti

Benessere Economico	Livello di Soddisfazione			Totale
	Insoddisfatto	Soddisfatto	Molto Soddisfatto	
Basso				360
Medio				540
Alto				180
Totale	240	720	120	1080

Si consideri il test delle ipotesi:

H_0 : Benessere economico e livello di soddisfazione sono statisticamente indipendenti

H_a : Benessere economico e livello di soddisfazione non sono statisticamente indipendenti

- (b) Definire la statistica test
- (c) Definire la distribuzione campionaria (asintotica) della statistica test sotto l'ipotesi nulla.