

Scuola di Scienze Politiche
Corso di Laurea LM–88 in SOCIOLOGIA E RICERCA SOCIALE
Compito di Metodi Statistici per la Ricerca Sociale
20 Dicembre 2017

Parte applicativa

Esercizio 1

La seguente tabella mostra alcune statistiche descrittive del voto alla laurea magistrale di studenti di Scienze politiche di tre Atenei

Linea	Statistiche descrittive		
	Numero di osservazioni	Media	Varianza
Ateneo A	25	105.0	13.2
Ateneo B	31	103.5	14.5
Ateneo C	27	106.7	16.0

- Calcolare la devianza tra gruppi e la devianza entro gruppi
- Verificare l'ipotesi che non ci sia differenza tra il voto medio alla laurea di studentoi di scienze politiche nei tre Atenei al livello di significatività del 1%: (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa, (ii) specificare la regione critica; (iii) calcolare il valore della statistica test e prendere una decisione.
- Il p -value per tale test è 0.0085. Commentare tale valore anche sulla base della decisione presa al punto (b)
- Scrivere il modello di analisi della varianza che pone il voto alla laurea magistrale in funzione dell'Ateneo usando l'Ateneo A come livello di riferimento. Stimare quindi i parametri del modello e interpretarli.

Esercizio 2

La seguente tabella mostra il numero di anni di istruzione e il punteggio ottenuto a un test attitudinale di accesso a un programma di formazione su un campione di 6 soggetti

Anni di istruzione (X)	14	18	16	10	20	12
Punteggio (Y)	79	83	84	61	94	76

- 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): $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$
- Stimare la varianza degli errori
- Costruire un intervallo di confidenza al livello di confidenza del 95% per il coefficiente di regressione β_1 .
- Stimare il punteggio medio per soggetti con 15 anni di studio e costruire il relativo intervallo di confidenza al livello di confidenza del 95%.

Esercizio 3

Su un campione di 47 giovani adulti (maschi) si sono rilevate le seguenti variabili: Y = punteggio a un test attitudinale di accesso a un corso di formazione ($TSS = 2151$); X_1 = età; X_2 = Titolo di studio (1 = Obbligo o meno; 2 = Diploma; 3 = Laurea o più); X_3 = Stato civile (1 = Single; 2 = Coniugato; 3 = Separato/Divorziato; 4 = Vedovo). Si considerino i seguenti modelli di regressione: Modello esteso in cui il punteggio al test è posto in funzione dell'età; del titolo di studio e dello stato civile ($SSE_e = 1075.2$); Modello ridotto in cui il punteggio al test è posto in funzione dell'età e del titolo di studio ($SSE_r = 1250.5$)

- Specificare il modello esteso e il modello ridotto
- Completare la seguente tavola di analisi della varianza per il modello esteso e per il modello ridotto

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

- (c) Confrontare il modello esteso con il modello ridotto: (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

La seguente tabella mostra i risultati dell'analisi di regressione (analisi della covarianza) per i dati dell'esercizio 3 ottenuti adattando il modello ridotto dove la variabile età è stata scartata dalla media (uguale a 32) e la scuola dell'obbligo o meno è usato come livello di riferimento per il titolo di studio.

Variabile	Coefficiente	Errore standard
Costante	30.00	8.07
Età - 32 ($X_1 - \bar{X}_1$)	-0.30	0.04
Titolo di studio (X_2)		
Diploma	2.54	0.85
Laurea o più	5.25	1.73

- (a) Stimare il punteggio medio al test per un soggetto di 35 anni laureato
- (b) Interpretare il coefficiente relativo diploma
- (c) Come cambierebbe l'interpretazione del coefficiente relativo al diploma se nel modello fosse inserita l'interazione tra età e titolo di studio?

Domanda 2

Si consideri uno studio su un campione di soggetti in cerca di prima occupazione. Alcuni hanno frequentato un corso di formazione di sei mesi e altri no ($X_i = 1$ se il soggetto i ha frequentato il corso di formazione e $X_i = 0$ altrimenti). La variabile risposta è lo stato occupazionale dopo 1 anno ($Y_i = 1$ se il soggetto i è occupato dopo un anno e $Y_i = 0$ altrimenti). I seguenti risultato sono relativi a un confronto tra proporzioni.

```
> prop.test(c(y1, y0), c(n1,n0), correct=FALSE)
```

```
2-sample test for equality of proportions without continuity correction
```

```
data: c(y1, y0) out of c(n1, n0)
X-squared = 7.4405, df = 1, p-value = 0.006377
alternative hypothesis: two.sided
95 percent confidence interval:
0.01453061 0.08546939
sample estimates:
prop 1 prop 2
0.75 0.70
```

dove $n1$ = numero di soggetti che hanno frequentato il corso; $y1$ = numero di occupati tra coloro che hanno frequentato il corso; $n0$ = numero di soggetti che non hanno frequentato il corso; $y0$ = numero di occupati tra coloro che non hanno frequentato il corso. Si consideri il risultato del test delle ipotesi.

- (a) Definire l'ipotesi nulla e l'ipotesi alternativa. Definire la statistica test usata dal comando `prop.test` e la sua distribuzione (approssimata) sotto l'ipotesi nulla. Interpretare il risultato del test sulla base del p -valore
- (b) Costruire la regione critica al livello del 1% utilizzando la statistica test riportata nell'output del comando `prop.test`. Decidere se rifiutare o non rifiutare l'ipotesi nulla sulla base della regione critica e del valore osservato della statistica test. Confrontare la decisione con l'interpretazione del p -valore
- (c) Per il confronto tra proporzioni è spesso usata anche la statistica test Z . Definire tale statistica; la sua distribuzione (approssimata) sotto l'ipotesi nulla e la sua relazione con la statistica test usata dal comando `prop.test`