

Scuola di Scienze Politiche
Corso di Laurea LM–88 in SOCIOLOGIA E RICERCA SOCIALE
Compito di Metodi Statistici per la Ricerca Sociale
20 Febbraio 2018

Parte applicativa

Esercizio 1

A un campione di neo-laureati magistrali di una certa scuola viene chiesto un giudizio (Negativo, Neutro, Positivo) sul corso di laurea appena terminato e il voto di laurea (< 100 , $100 - 104$, ≥ 105). La seguente tabella mostra la distribuzione di frequenza congiunta osservata:

Giudizio	Voto di laurea		
	< 100	$100 - 104$	≥ 105
Negativo	35	27	16
Neutro	75	80	104
Positivo	20	55	80

- Costruire la distribuzione di frequenza relativa condizionata del giudizio per studenti che si sono laureati con un voto maggiore o uguale a 105.
- Calcolare la frequenza teorica di indipendenza corrispondente alla frequenza osservata di studenti che hanno espresso un giudizio positivo e si sono laureati con un voto maggiore o uguale a 105. Calcolare quindi il corrispondente residuo.
- Condurre un test statistico per valutare se giudizio e voto di laurea sono statisticamente indipendenti: (i) Scrivere la regione critica al livello di significatività del 5%; (ii) Prendere una decisione sapendo che il valore osservato della statistica test è 34.481.
- Definire il p -valore relativo al test al punto (c). In base al risultato ottenuto al punto (c) il p -valore sarebbe maggiore, minore o uguale di 0.05? Giustificare la risposta.

Esercizio 2

La seguente tabella mostra alcune statistiche descrittive di una misura di benessere socio-economico (misurata su scala 0-100) e della qualità della vita (misurata su scala 0-100) calcolate su un campione di $n = 16$ soggetti di oltre 50 anni di un certo paese.

Variabile	Media	Varianza
Misura di benessere socio-economico	62	144
Qualità della vita	80	49

La covarianza tra la misura di benessere socio-economico e la qualità della vita è risultata pari a $s_{x,y} = 75.6$

- Calcolare il coefficiente di correlazione lineare tra la misura di benessere socio-economico e la qualità della vita
- Stimare i coefficienti del modello di regressione che pone la qualità della vita in funzione della misura di benessere socio-economico
- Stimare la varianza degli errori del modello di regressione
- Stimare la qualità della vita media utilizzando il modello di regressione per soggetti con benessere socio-economico pari a 65. Calcolare il residuo di regressione per un soggetto per cui si osserva un valore della misura di benessere socio-economico pari a 65 e un valore della qualità della vita pari a 78. Costruire quindi un intervallo di confidenza al 95% per la qualità della vita media per soggetti con benessere socio-economico pari a 65.

Esercizio 3

Si consideri lo studio dell'esercizio 2. Si supponga di introdurre nel modello di regressione anche la variabile titolo di studio (Obbligo o meno, Diploma, Laurea o più). Si supponga che la somma dei quadrati degli errori per tale modello sia $SSE_e = 95$

- Completare la seguente tavola di analisi della varianza per il modello esteso (in cui sono incluse entrambe le variabili esplicative: misura di benessere socio-economico e titolo di studio)

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

- Confrontare il modello esteso (in cui sono incluse entrambe le variabili esplicative: misura di benessere socio-economico e titolo di studio) con il modello ridotto (in cui è inclusa solo la variabile esplicativa misura di benessere socio-economico): (i) Scrivere l'ipotesi nulla e l'ipotesi; (ii) Costruire la regione critica al livello di significatività del 5% per tale test; (iii) Calcolare il valore osservato della statistica test e prendere una decisione
- Definire il p -valore per il test al punto (b). Il p -valore per tale test è uguale a 0.0991. Interpretare tale valore.

Domande di Teoria

Domanda 1

La seguente tabella mostra i risultati di un'analisi di regressione condotta su un campione di 100 soggetti di oltre 50 anni in cui la qualità della vita (misurata su scala 0-100) è posta in funzione di una misura di benessere socio-economico (misurata su scala 0-100) e del titolo di studio (Obbligo o meno, Diploma, Laurea o più).

Variabile	Coefficiente	SE
Costante	60.8	5.6
Benessere socio-economico	5.7	1.2
Titolo di studio (Obbligo o meno)		
Diploma	2.4	0.9
Laurea o più	3.1	0.7

- Scrivere l'equazione di regressione stimata
- Interpretare il coefficiente di regressione relativo al benessere socio-economico e il coefficiente relativo al diploma
- Come cambierebbe l'interpretazione del coefficiente di regressione relativo al benessere socio-economico e del coefficiente relativo al diploma se nel modello fosse presente l'interazione tra titolo di studio e benessere socio-economico?

Domanda 2

Su un campione di $n = 350$ soggetti si osservano la variabile binaria Y con $Y = 1$ per soggetti obesi (ossia con indice di massa corporea superiore a 30) e $Y = 0$ per soggetti non obesi (ossia con indice di massa corporea minore o uguale a 30) e X , il numero di ore dedicate in media alla settimana all'attività fisica. Su tale campione di soggetti viene stimato un modello di regressione logistica, ottenendo i seguenti risultati:

```
> summary(glm(Y~X, family=binomial))

Call:
glm(formula = Y ~ X, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6295 -1.2067  0.7849  1.0186  2.1261

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.01966    0.19531   5.221 1.78e-07 ***
X            -0.31697    0.06383  -4.966 6.84e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 481.07  on 349  degrees of freedom
Residual deviance: 453.38  on 348  degrees of freedom
AIC: 457.38
```

Number of Fisher Scoring iterations: 4

- Scrivere l'espressione delle probabilità stimate
- Interpretare il coefficiente relativo alla variabile X in termini di odds
- Si considerino i valori **z value** e $\Pr(> |z|)$ relativi al coefficiente della variabile X . Tali valori sono relativi a un test delle ipotesi. Definire l'ipotesi nulla e l'ipotesi alternativa di tale test, definire il valore **z value** (ossia come viene calcolato) e interpretare il valore $\Pr(> |z|)$