

METODI STATISTICI PER LA RICERCA SOCIALE

CAPITOLO 9.

REGRESSIONE LINEARE E CORRELAZIONE

ESERCITAZIONE

Alessandra Mattei

Dipartimento di Statistica, Informatica, Applicazioni (DiSIA)
Università degli Studi di Firenze
mattei@disia.unifi.it

LM-88 SOCIOLOGIA E RICERCA SOCIALE

Esercizio 1

Su un campione di 6 lavoratori autonomi di un certo paese si è rilevato il numero di ore lavorate (in media) alla settimana e il numero di anni di attività:

Anni di attività	12	18	8	22	21	15
Ore Lavorate	42	40	56	38	39	49

- (a) Calcolare il coefficiente di correlazione tra gli anni di attività e le ore lavorate
- (b) Stimare i coefficienti di regressione di un modello di regressione che pone le ore lavorate (variabile Y) in funzione degli anni di attività (variabile X)
- (c) Calcolare l'indice di determinazione lineare e interpretarlo
- (d) Stimare la varianza degli errori del modello di regressione
- (e) Stimare il numero medio di ore che in media lavoratori autonomi che hanno iniziato la loro attività da 20 anni dedicano al lavoro alla settimana. Costruire quindi un intervallo di confidenza al livello di confidenza del 90% per il numero medio di ore che in media lavoratori autonomi che hanno iniziato la loro attività da 20 anni dedicano al lavoro alla settimana

Esercizio 2

Sembrerebbe che bere quantità moderate di vino possa prevenire l'infarto. La seguente tabella mostra alcune statistiche descrittive relative al consumo di vino (variabile X) e il numero di morti annui (morti per 100 000 persone) per malattie cardiache (variabile Y) in 19 paesi.

$$\bar{x} = 3.03 \quad \bar{y} = 191.05 \quad s_X^2 = 6.30 \quad s_Y^2 = 4678.053 \quad s_{XY} = -144.674$$

- (a) Calcolare il coefficiente di correlazione e interpretarlo
- (b) Stimare il coefficiente del modello di regressione che pone la mortalità per malattie cardiache in funzione del consumo del vino
- (c) Sottoporre a test l'ipotesi che il coefficiente di regressione relativo al consumo di vino sia significativamente diverso da zero: (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 1%; (iii) Calcolare il valore della statistica test e prendere una decisione
- (d) Costruire un intervallo di confidenza al 99% per il coefficiente di regressione relativo al consumo di vino. Interpretare l'intervallo di confidenza e confrontarlo con il risultato del test ottenuto al punto (c)

Esercizio 3

In uno studio sulla relazione tra Y = tasso di criminalità (numero di crimini ogni 1000 abitanti) e X = tasso di povertà (percentuale di popolazione con reddito sotto il livello di povertà) nelle città di un certo paese si è selezionato un campione di 20 città. Per tale campione si sono osservate le seguenti statistiche:

$$\sum_{i=1}^{20} x_i = 280 \quad \sum_{i=1}^{20} y_i = 170 \quad \sum_{i=1}^{20} x_i^2 = 4851 \quad \sum_{i=1}^{20} y_i^2 = 4181 \quad \sum_{i=1}^{20} x_i \cdot y_i = 3776.5$$

- (a) Calcolare il coefficiente di correlazione lineare tra tasso di criminalità e tasso di povertà
- (b) Sottoporre a test l'ipotesi che il coefficiente di correlazione sia significativamente diverso da zero: (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 5%; (iii) Calcolare il valore della statistica test e prendere una decisione.
- (c) Il p -valore per il test al punto precedente è $p = 0.00000045$. Interpretare il p -valore e confrontarlo con il risultato ottenuto al punto (b)
- (d) Stimare i coefficiente del modello di regressione che pone il tasso di criminalità in funzione del tasso di povertà. Il coefficiente di regressione β_1 è significativamente diverso da zero?
- (e) Stimare la varianza degli errori del modello di regressione
- (f) Per una certa città si è osservato un tasso di povertà $x = 12$ e un tasso di criminalità $y = 6$. Calcolare il tasso di criminalità atteso utilizzando il modello di regressione e il residuo di regressione per tale città. Costruire quindi un intervallo di confidenza al 95% per il tasso di criminalità atteso di città con un tasso di povertà $x = 12$.

Esercizio 4

Ognuna delle seguenti affermazioni contiene un errore. Individuare l'errore, spiegando perché si tratta di un errore

- (a) C'è una forte correlazione fra titolo di studio e reddito.
- (b) La correlazione fra qualità della vita e ore settimanali medie dedicate al tempo libero calcolata su un campione di soggetti è risultata uguale a $r = 0.7$ ore.
- (c) È stata trovata un'alta correlazione ($r = 1.1$) tra la qualità della vita di soggetti anziani e il tempo medio da essi trascorso settimanalmente in compagnia (misurato in ore).