

# METODI STATISTICI PER LA RICERCA SOCIALE

## CAPITOLO 13.

### COMBINARE REGRESSIONE E ANOVA: PREDITTORI CATEGORIALI E QUANTITATIVI

### 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

Si consideri uno studio in cui su 27 soggetti si osservano le variabili  $Y$  = indice di massa corporea;  $X$  = numero medio di ore dedicate all'attività fisica alla settimana, e  $A$  una variabile binaria uguale a 1 per soggetti che seguono una dieta equilibrata e uguale a 0 altrimenti. Si considerino le seguenti statistiche descrittive:

| Statistica  | Gruppo    |           |
|---|-----------|-----------|
|   | Dieta: Si | Dieta: No |
| $n_\ell$  | 12        | 15        |
| $\bar{x}_\ell$  | 2.8       | 2.26      |
| $\bar{y}_\ell$  | 22.4      | 26        |
| $\sum_{i=1}^{n_\ell} (x_{i\ell} - \bar{x}_\ell)^2$                                | 8.91      | 8.96      |
| $\sum_{i=1}^{n_\ell} (y_{i\ell} - \bar{y}_\ell) \cdot (x_{i\ell} - \bar{x}_\ell)$ | -20.22    | -13.62    |

- (a) Stimare i coefficienti del modello di regressione che pone l'indice di massa corporea in funzione del numero medio di ore dedicate all'attività fisica alla settimana e della variabile binaria  $A$  (modello di analisi della covarianza senza interazione)
- (b) Interpretare i coefficienti di regressione stimati al punto (a)
- (c) Sapendo che la somma dei quadrati totali di  $Y$  è  $SST = 105$  e che la somma dei quadrati dei residui del modello stimato al punto (a) è  $SSE_e = 80$ , confrontare tale modello con il modello nullo: (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 10% (iii) Calcolare la statistica test; (iv) Prendere una decisione giustificando la risposta

## Esercizio 2

La seguente tabella mostra i risultati di un'analisi di regressione che pone il consumo di alcool (misurato come il numero di bevande alcoliche che il soggetto ha bevuto nell'ultimo mese) in funzione delle seguenti variabili esplicative: Sesso (0 = Femmina, 1 = Maschio); Stato civile (1 =single; 2 =sposato; 3 =divorziato/separato; 4 = Vedovo) e età. Per le variabili categoriche la categoria di riferimento è scritta in parentesi.

| Variabile             | Stima           | Errore standard |
|-----------------------|-----------------|-----------------|
| Costante              | 9.40            | 1.25            |
| Genere (Femmina)      |                 |                 |
| Maschio               | 6.54            | 0.74            |
| Stato civile (Single) |                 |                 |
| Sposato               | -3.49           | 0.72            |
| Div./Sep.             | 2.38            | 0.78            |
| Vedovo                | -0.92           | 0.63            |
| Età                   | 0.15            | 0.02            |
| $n = 56$              | $TSS = 1233.09$ | $SSE = 145.39$  |

- (a) Scrivere l'equazione di regressione stimata e interpretare i coefficienti di regressione
- (b) Calcolare una stima della varianza degli errori del modello di regressione
- (c) Sapendo che la somma dei quadrati dei residui di un modello ridotto in cui non è incluso lo stato civile è  $SQE_r = 251.55$ , confrontare il modello esteso (in cui è incluso lo stato civile) con il modello ridotto (in cui non è incluso lo stato civile) utilizzando un livello di significatività pari a 5%. (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.

## Esercizio 3

La seguente tabella mostra i risultati di una analisi di regressione multipla (modello di analisi della covarianza) in cui la qualità della vita (variabile risposta  $Y$ , misurata con un opportuno indice con valore minimo 0 e valore massimo 100) di 29 pazienti in dialisi è posta in funzione del tipo di trattamento (variabile  $X_1$  con  $X_1 = 1$  per il nuovo trattamento; e  $X_1 = 0$  per il trattamento standard) e dell'età del paziente (variabile  $X_2$ ). La variabile età è stata scartata dalla sua media,  $\bar{X}_2 = 55$ , prima di essere introdotta nel modello:

| Variabile                 | Stima | Errore standard |
|---------------------------|-------|-----------------|
| Costante                  | 50.9  | 13.22           |
| Trattamento ( $X_1$ )     |       |                 |
| Nuovo trattamento         | 7.3   | 1.35            |
| Età ( $X_2 - \bar{X}_2$ ) | -0.2  | 0.03            |

- (a) Scrivere l'equazione di regressione stimata
- (b) Interpretare i coefficienti di regressione
- (c) Sottoporre a test l'ipotesi che il coefficiente di regressione relativo alla variabile trattamento 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
- (d) Calcolare l'indice di determinazione lineare sapendo che la somma dei quadrati totale è 3735 e che la somma dei quadrati dei residui è 560.25. Interpretare il valore ottenuto.
- (e) Stimare la qualità della vita media per un paziente di 60 anni esposto al nuovo trattamento

## Esercizio 4

Si consideri lo studio precedente sulla qualità della vita di pazienti in dialisi. La seguente tabella riporta le stime dei coefficienti del modello con interazione tra trattamento e età:

| Variabile                 | Stima |
|---------------------------|-------|
| Costante                  | 49.8  |
| Trattamento ( $X_1$ )     |       |
| Nuovo trattamento         | 7.2   |
| Età ( $X_2 - \bar{X}_2$ ) | -0.18 |
| Interazione               | 0.02  |

- (a) Scrivere l'equazione di regressione stimata
- (b) Interpretare i coefficienti di regressione del modello