

## Scuola di Scienze Politiche

## Corso di Laurea LM–88 in SOCIOLOGIA E RICERCA SOCIALE

## Compito di Metodi Statistici per la Ricerca Sociale

13 Giugno 2017

## Parte applicativa

## Esercizio 1

In uno studio sull'inquinamento atmosferico si sono rilevati dati relativi al livello di  $\text{PM}_{10}$  ( $\mu\text{g}/\text{m}^3$ ) in un campione di giorni dei mesi estivi (giugno-settembre) in tre città.

Gruppo	Numerosità del gruppo	Media Campionaria	Varianza Campionaria
$j$	$n_j$	$\bar{x}_j$	$s_j^2$
Città A	10	32.8	15
Città B	9	31.5	14.2
Città C	12	33.0	16

- Calcolare la devianza entro i gruppi e la devianza tra gruppi
- Verificare l'ipotesi che il livello di inquinamento medio sia lo stesso nelle tre città. (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
- Il p-valore relativo al test al punto (b) è uguale a  $p = 0.657$ . Interpretare tale valore anche sulla base del risultato ottenuto al punto (b)
- Scrivere il modello di analisi della varianza considerando 'la città A' come gruppo di riferimento e stimare i parametri del modello

## Esercizio 2

Su 5 giorni si sono rilevati i seguenti dati relativi al livello di  $\text{PM}_{10}$  ( $\mu\text{g}/\text{m}^3$ ) e alla temperatura (in gradi celsius)

Temperatura( $X$ )	26	13	24	27	16
$\text{PM}_{10}$ ( $Y$ )	28	20	22	25	18

- Determinare le stime dei minimi quadrati dei coefficienti della retta di regressione che pone il livello di  $\text{PM}_{10}$  ( $Y$ ) in funzione della temperatura ( $X$ ) misurata in gradi celsius:  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$
- Calcolare l'indice di determinazione lineare e interpretarlo
- Stimare la varianza degli errori
- Stimare il livello di  $\text{PM}_{10}$  per giorni con temperatura pari a 25 °C e costruire il relativo intervallo di confidenza al livello di confidenza del 90%.

## Esercizio 3

Su  $n = 30$  giorni si sono rilevate le seguenti variabili:  $Y$  = livello di  $\text{PM}_{10}$  (misurato in  $\mu\text{g}/\text{m}^3$ );  $X_1$  = Temperatura (misurata in gradi celsius);  $X_2$  = Umidità relativa (misura su una scala a tre categorie 1 = Bassa; 2 = Media; 3 = Alta). Si supponga di adattare un modello di regressione che pone il livello di  $\text{PM}_{10}$ ,  $Y$ , in funzione delle variabili esplicative  $X_1$  e  $X_2$ . Si consideri tale modello come modello esteso. Per tale modello la somma dei quadrati dei residui è  $SSE_e = 45.3$ .

- Confrontare il modello esteso con il modello nullo sapendo che la somma dei quadrati totali è  $SST = 182.3$ . (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Definire la statistica test e la sua distribuzione sotto l'ipotesi nulla; (iii) Calcolare il valore osservato della statistica test.

- (b) Prendere una decisione sapendo che il  $p$ -valore per il test al punto è inferiore a 0.0001. (Giustificare la decisione presa)
- (c) Confrontare il modello esteso con un modello ridotto in cui si considera come variabile esplicativa solo la temperatura, sapendo che per il modello ridotto la somma dei quadrati dei residui è  $SSE_r = 60.9$ . (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa relativa al confronto dei due modelli; (ii) Scrivere quindi la regione critica al livello di significatività  $\alpha = 0.05$ ; (iii) Calcolare il valore osservato della statistica test e prendere una decisione.

## Domande di Teoria

### Domanda 1

Su un campione di giorni si sono rilevate le seguenti variabili:  $Y$  = livello di  $PM_{10}$  (misurato in  $\mu g/m^3$ );  $X_1$  = Temperatura (misurata in gradi celsius);  $X_2$  = Umidità relativa (misura su una scala a tre categorie 1 = Bassa; 2 = Media; 3 = Alta). La tabella sottostante mostra i risultati dell'adattamento di un modello di regressione in cui il livello di  $PM_{10}$  ( $Y$ ) è posto in funzione della temperatura, dell'umidità (relativa) e della loro interazione

Variabile	Coefficiente
Costante	21.2
Temperatura ( $X_1$ )	0.6
Umidità ( $X_2$ )	
Media	1.2
Alta	1.7
Temperatura $\times$ Umidità	
Temperatura $\times$ Umidità Media	0.2
Temperatura $\times$ Umidità Alta	0.3

- (a) Scrivere l'equazione di regressione stimata
- (b) Interpretare il coefficiente relativo alla temperatura e il coefficiente relativo all'umidità media
- (c) Stimare il livello di  $PM_{10}$  ( $Y$ ) per giorni con temperatura pari a 20 °C e umidità relativa media

### Domanda 2

Su un campione di giorni si è rilevata una variabile binaria  $Y = 1$  se il livello di  $PM_{10}$  nel giorno è inferiore (o uguale) alla soglia di 40  $\mu g/m^3$ ; e  $Y = 0$  se il livello di  $PM_{10}$  nel giorno è superiore alla soglia di 40  $\mu g/m^3$  e  $X$  = la temperatura (misurata in gradi celsius). La seguente tabella mostra i risultati dell'adattamento di modello di regressione logistica:

Variabile	Coefficiente	Errore standard
Costante	0.5	0.07
Temperatura	-0.2	0.03

- (a) Scrivere l'equazione che definisce le probabilità stimate  $\hat{\pi}_i = \hat{Pr}(Y_i = 1 | X_i = x_i)$ .
- (b) Interpretare il coefficiente relativo alla temperatura in termini di odds.
- (c) Stimare la probabilità che in un giorno con temperatura pari a 20 °C il livello di  $PM_{10}$  nel giorno sia inferiore (o uguale) alla soglia di 40  $\mu g/m^3$