

Statistica Descrittiva

Esercitazione 7. Interpolazione: minimi quadrati

Esercizio A. Si consideri il seguente estratto dalla Tavola 8.1 “Persone di 14 anni e più per livello di soddisfazione su situazione economica, salute, relazioni familiari, relazioni con amici e tempo libero per regione” (“Compendio Statistico Italiano 1998”, ISTAT, pag. 98):

% persone soddisfatte di:	Piemonte	Lombardia	Trentino A.A.	Veneto	Liguria	Emilia-Romagna	Toscana
situazione economica (x)	2,52	3,78	6,96	3,61	2,25	4,38	3,4
tempo libero (y)	17,18	19,14	22,55	17,14	15,45	19,72	20,22

- a) Si interpoli con una retta la percentuale di persone soddisfatte del proprio tempo libero (y) rispetto alla percentuale di persone soddisfatte della propria situazione economica (x).
- b) Si verifichi la relazione

$$\text{Devianza totale} = \text{Devianza residua} + \text{Devianza di regressione},$$

si calcoli il coefficiente di determinazione r^2 per valutare la bontà dell'accostamento e si commenti il risultato ottenuto.

- c) Si calcoli la percentuale teorica di persone soddisfatte del proprio tempo libero per la regione Umbria sapendo che la percentuale di soddisfatti della situazione economica è di 3,01.
- d) Nel caso in cui, diversamente dal punto a) si volesse interpolare (con una retta) la percentuale di persone soddisfatte della propria situazione economica (x) rispetto alla percentuale dei soddisfatti del proprio tempo libero (y), il valore e il segno del coefficiente angolare b cambierebbero?

Esercizio B. Si consideri la distribuzione cumulata dei redditi mensili in ordine discendente (retrocumulata) indicata nella seguente tabella, relativa a $N = 4241$ individui:

Classe di reddito (milioni di lire)	0-5	5-6	6-7,5	7,5-11	11-15	15-19	19-25	25-50	50-100
$(c_i - c_{i+1}]$									
f_i (migliaia)	1618	941	669	583	211	94	66	50	9
F_i	1618	2559	3228	3811	4022	4116	4182	4232	4241
\bar{F}_i	4241	2623	1682	1013	430	219	125	59	9

dove \bar{F}_i indica il numero di individui con reddito superiore a c_i .

- a) Si interpoli la distribuzione retrocumulata con una funzione del tipo $y = \alpha_0 x^{\alpha_1}$ e si calcoli un opportuno indice di adattamento (suggerimento: si considerino come x_i i valori centrali di classe e come y_i le \bar{F}_i , e si cerchi una trasformazione dei dati che consenta di ridurre il problema all'interpolazione di una retta).
- b) Nel caso in cui si volesse utilizzare come funzione interpolante (dei dati originari non trasformati) una retta, come cambierebbe la devianza residua rispetto a quella calcolata al punto precedente?

Esercizio C. La seguente tavola contiene la serie dei numeri indici della produzione media giornaliera di beni di consumo nel 1997 (Base 1990=100):

Mese	gennaio	febbraio	marzo	aprile	maggio	giugno	luglio
Indici della produzione	107,2	112,5	115,2	99,4	114,1	112,5	122,1

a) Considerando solo i mesi da gennaio a maggio, si interpoli la serie di numeri indici in funzione dei mesi con una parabola del tipo $y = \beta_0 + \beta_2 x^2$ (suggerimento: per facilitare i calcoli, si scelga marzo come mese centrale, ossia si ponga $x_1 = -2$ per gennaio, $x_2 = -1$ per febbraio, $x_3 = 0$ per marzo, e così via).

b) Si confrontino i valori teorici e i valori effettivi dei mesi di giugno e luglio.

Esercizio D. Si consideri l'informazione sul valore aggiunto dei prezzi dei prodotti agricoli (in milioni di euro) e sulla produzione di grano (in milioni di euro) per l'anno 2005 riferiti ai paesi europei riportata nella Tavola 11.4 (“Eurostat 2006-2007”, pag. 292). Indicando con Y la variabile dipendente data dal valore aggiunto dei prezzi e con X la variabile indipendente data dal valore della produzione di grano, si calcoli quanto segue (si faccia riferimento solo ai paesi per i quali compaiono entrambe le osservazioni di X e Y):

a) Si interpoli con una retta l'informazione Y rispetto a X .

b) Si calcoli il coefficiente di correlazione lineare di Bravais-Pearson tra X e Y .

c) Si calcoli il coefficiente di determinazione R^2 e si commenti in base a questo l'adattamento della retta di regressione alle osservazioni.

d) Si valuti il valore teorico di Y per l'Italia sulla base della retta di regressione.

e) A partire dal risultato del punto c) (quindi usando il valore R^2), si calcolino la devianza di regressione e la devianza residua.

Esercizio E. Si consideri il seguente estratto della Tavola 9.2 “Conto economico delle risorse e degli impieghi per i Paesi Ue - Anno 1994 (dati in miliardi di S.P.A. a prezzi correnti)” (“Conoscere l'Italia: 1997”, ISTAT)

	PIL	Inv.		PIL	Inv.		PIL	Inv.
Belgio	190,2	33,6	Francia	1.035,7	185,2	Portogallo	112,9	27,9
Danimarca	99,6	14,9	Irlanda	50,6	7,3	Regno Unito	959,6	148,9
Germania	1.492,1	337,9	Italia	995,2	169,2	Finlandia	151,2	39,8
Grecia	109,9	22,5	Lussemburgo	10,9	2,3	Austria	76,8	12,1
Spagna	495,3	99,4	Paesi Bassi	266,4	53,2	Svezia	142,5	20,6

a) Considerando come unità statistiche i singoli Paesi, si interpolino gli Investimenti Fissi Lordi (Y) rispetto al Prodotto Interno Lordo Totale (X) mediante una retta con il metodo dei minimi quadrati.

b) Si costruisca una tabella a doppia entrata raggruppando i valori della X nelle classi 0–300, 300–600, 600–1200, 1200–2000, ed i valori della Y nelle classi 0–50, 50–150, 150–250, 250–350. Si determini la retta di regressione sulla base dei dati raggruppati in classi e si confrontino i risultati ottenuti con i risultati ottenuti al punto precedente (suggerimento: si usino come x_i e y_j i valori centrali delle classi).