ISTITUZIONI DI STATISTICA – A. A. 2006/2007
Marco Minozzo
Laurea in Economia del Commercio Internazionale
Laurea in Economia e Amministrazione delle Imprese
Università degli Studi di Verona (sede di Vicenza)

Statistica Descrittiva
Soluzioni 7. Interpolazione: minimi quadrati

Esercizio A.
a) Sulla base della seguente tabella

Regione	x_i	y_i	x_i^2	x_iy_i	y_i^2	y'_i	$(y'_i - m_y)^2$	$(y'_i - y_i)^2$
Piemonte	2,52	17,18	6,350	43,294	295,152	17,007	3,112	0,030
Lombardia	3,78	19,14	14,288	72,349	366,340	18,688	0,007	0,205
Trentino A. A.	6,96	22,55	48,442	156,948	508,503	22,928	17,280	0,143
Veneto	3,61	17,14	13,032	61,875	293,780	18,461	0,096	1,745
Liguria	2,25	15,45	5,063	34,763	238,703	16,647	4,512	1,433
Emilia-Romagna	4,38	19,72	19,184	86,374	388,878	19,488	0,513	0,054
Toscana	3,40	20,22	11,560	68,748	408,848	18,181	0,349	4,158
Totale	26,90	131,40	117,919	524,350	2500,203	131,400	25,870	7,768

si ha che $m_x = 3,843$, $m_y = 18,771$, $\sigma_x^2 = 117,919/7 - (3,843)^2 = 2,077$ e $\sigma_{xy} = 524,35/7 - 3,843 \cdot 18,771 = 2,770$ da cui $b = 2,770/2,077 = 1,334$ e $a = 18,771 - 3,843 \cdot 1,334 = 13,647$.
b) Sulla base della tabella precedente si ha che

$$\begin{aligned} \text{Devianza totale} &= 7 \cdot \sigma_y^2 = 2500,203 - 7 \cdot (18,771)^2 = 33,638, \\ \text{Devianza di regressione} &= 25,870, \\ \text{Devianza residua} &= 7,768, \end{aligned}$$

e quindi la scomposizione della devianza totale è verificata. L'indice r^2 risulta pari a $25,870/33,638 = 0,769$ che indica un buon accostamento della retta ai punti.
c) La percentuale teorica di persone soddisfatte del proprio tempo libero per la regione Umbria è $13,647 + 1,334 \cdot 3,01 = 17,661$.
d) Nel caso in cui si volesse interpolare x rispetto a y, b avrebbe lo stesso segno anche se un diverso valore in quanto il segno di b è lo stesso di σ_{xy}. Infatti per questa nuova interpolazione risulta che $b = 2,770/4,805 = 0,576$.

Esercizio B.
a) Si utilizza la trasformazione $\ln(y) = \ln(\alpha_0) + \alpha_1 \ln(x)$ da cui

Reddito	x_i	y_i	$z_i = \ln(x_i)$	$u_i = \ln(y_i)$	z_i^2	z_iu_i	u_i^2	u'_i	$(u'_i - u_i)^2$
0-5	2,50	4241	0,916	8,353	0,840	7,653	69,765	9,052	0,490
5-6	5,50	2623	1,705	7,872	2,906	13,420	61,970	7,556	0,100
6-7,5	6,75	1682	1,910	7,428	3,646	14,184	55,171	7,168	0,067
7,5-11	9,25	1013	2,225	6,921	4,949	15,396	47,896	6,570	0,123
11-15	13,00	430	2,565	6,064	6,579	15,553	36,769	5,925	0,019
15-19	17,00	219	2,833	5,389	8,027	15,268	29,042	5,416	0,001
19-25	22,00	125	3,091	4,828	9,555	14,925	23,313	4,927	0,010
25-50	37,50	59	3,624	4,078	13,136	14,778	16,626	3,915	0,026
50-100	75,00	9	4,317	2,197	18,641	9,486	4,828	2,600	0,162
Totale	188,500	10401	23,186	53,129	68,278	120,664	345,380	53,129	0,998

Quindi $m_z = 2,576$, $m_u = 5,903$, $\sigma_z^2 = 8,545/9$ e $\sigma_{zu} = -16,210/9$ e di conseguenza $\alpha_1 = -1,897$ e $\ln(\alpha_0) = 10,790$. Si ha quindi che $u' = \ln(y') = 10,790 - 1,897\ln(x)$ o in modo equivalente $y' = 48555,5\ x^{-1,897}$. Il coefficiente di determinazione per l'interpolazione di u rispetto a z è pari a $0,969$.
b) Interpolando i punti originari (x_i, y_i), $i = 1, \ldots, 9$, con una retta si ottiene un valore più elevato della devianza residua ($11.064.157$) a cui corrisponde un valore del coefficiente di determinazione notevolmente più basso ($0,347$).

Esercizio C.
a) Considerando la trasformazione $z = x^2$ si ottiene

Mese	x_i	y_i	$z_i = x_i^2$	z_i^2	$z_i y_i$	y_i^2	y_i'	$(y_i' - y_i)^2$
gennaio	-2	107,2	4	16	428,8	11491,840	109,723	6,365
febbraio	-1	112,5	1	1	112,5	12656,250	109,659	8,074
marzo	0	115,2	0	0	0,0	13271,040	109,637	30,945
aprile	1	99,4	1	1	99,4	9880,360	109,659	105,238
maggio	2	114,1	4	16	456,4	13018,810	109,723	19,159
	0	548,4	10	34	1097	60318,300	548,400	169,782

da cui $m_z = 2$, $m_y = 109,68$, $\sigma_z^2 = 14/5$ e $\sigma_{zy} = 0,3/5$ e quindi $\beta_2 = 0,021$, $\beta_0 = 109,637$ e $r^2 = 0,0000379$ che indica un adattamento dell'interpolazione praticamente nullo.
b) I valori teorici per giugno e luglio sono rispettivamente $y_6' = 109,637 + 0,021 \cdot (3)^2 = 109,826$ e $y_7' = 109,637 + 0,021 \cdot (4)^2 = 109,973$ e risultano entrambi lontani dai valori osservati.

Esercizio D.
a) Per interpolare il numero di nati vivi per 1000 ab. (Y) rispetto al numero di matrimoni per 1000 ab. (X) con una retta mediante il metodo dei minimi quadrati è utile la seguente tabella

Regioni	x_i	y_i	x_i^2	$x_i y_i$	y_i^2	y_i'	$(y_i' - m_y)^2$	$(y_i - y_i')^2$	$(y_i - m_y)^2$
Piemonte	4,7	7,4	22,09	34,78	54,76	8,259	0,0260	0,7375	1,0404
Lombardia	4,6	8,4	21,16	38,64	70,56	7,989	0,1854	0,1686	0,0004
Trentino	5,3	10,6	28,09	56,18	112,36	9,875	2,1176	0,5253	4,7524
Veneto	5,0	8,4	25,00	42,00	70,56	9,067	0,4186	0,4449	0,0004
Friuli	4,2	7,3	17,64	30,66	53,29	6,912	2,2747	0,1507	1,2544
Totale	23,8	42,1	113,98	202,26	361,53	42,102	5,0223	2,0271	7,0480

Si ottiene $m_x = 4,76$, $m_y = 8,42$, $\sigma_x^2 = 113,98/5 - (4,76)^2 = 0,1384$, $\sigma_{xy} = 202,26/5 - 4,76 \cdot 8,42 = 0,3728$, $b = 2,694$, $a = -4,403$. Per ottenere il grafico richiesto, basta tracciare la retta che passa per il punto $(0;\ a) = (0;\ -4,40)$ e per il punto $(m_x; m_y) = (4,76;\ 8,42)$.
b) Per la verifica della scomposizione della devianza totale si ottiene:

$$\begin{aligned} \text{Devianza totale} &= \sum_{i=1}^{N}(y_i - m_y)^2 = 7,0480, \\ \text{Devianza residua} &= \sum_{i=1}^{N}(y_i - y_i')^2 = 2,0271, \\ \text{Devianza di regressione} &= \sum_{i=1}^{N}(y_i' - m_y)^2 = 5,0223, \end{aligned}$$

mentre per il coefficiente di determinazione si ha

$$r^2 = \frac{\text{Devianza di regressione}}{\text{Devianza totale}} = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} = 0,712,$$

il che indica un discreto accostamento.
c) Il valore teorico del numero di nati vivi per 1000 ab. per la regione Valle d'Aosta è pari a $y' = -4,403 + 2,694 \cdot 5 = 9,067$.

Esercizio E.
a) Per interpolare gli Investimenti Fissi Lordi (Y) rispetto al Prodotto Interno Lordo Totale (X) è utile ottenere la seguente tabella:

Paesi	PIL	Investimenti			
	x_i	y_i	x_i^2	$x_i y_i$	y_i^2
Belgio	190,2	33,6	36.176,04	6.390,72	1.128,86
Danimarca	99,6	14,9	9.920,16	1.484,04	222,01
Germania	1.492,1	337,9	2.226.362,41	504.180,59	114.176,41
...	...	...	...	...	...
Austria	76,8	12,1	5.898,24	929,28	146,41
Svezia	142,5	20,6	20.306,25	2.935,50	424,36
	6.188,9	1.174,8	5.649.200	1.092.316	216.835

da cui, essendo $N = 15$, si ricava $m_x = 412,59$, $m_y = 78,32$, $\sigma_x^2 = 206382,83$ e $\sigma_{xy} = 40507,02$ con cui si ottengono il coefficiente angolare $b = 0,1963$ e l'intercetta $a = -2,9182$; si ottiene inoltre $r^2 = 0,962$.
b) La classificazione richiesta fornisce la tabella

	Y				
X	0-50	50-150	150-250	250-350	
0-300	9	1	0	0	10
300-600	0	1	0	0	1
600-1200	0	1	2	0	3
1200-2000	0	0	0	1	1
	9	3	2	1	15

a cui corrispondono i valori centrali di classe $x_1 = 150$, $x_2 = 450$, $x_3 = 900$, $x_4 = 1.600$ e $y_1 = 25$, $y_2 = 100$, $y_3 = 200$, $y_4 = 300$. Da questa si ricavano le tabelle

x_i	$f_{i\cdot}$	$x_i f_{i\cdot}$	$x_i^2 f_{i\cdot}$	y_j	$f_{\cdot j}$	$y_j f_{\cdot j}$	$y_j^2 f_{\cdot j}$
150	10	1.500	225.000	25	9	225	5.625
450	1	450	202.500	100	3	300	30.000
900	3	2.700	2.430.000	200	2	400	80.000
1.600	1	1.600	2.560.000	300	1	300	90.000
	15	6.250	5.417.500		15	1.225	205.625

da cui si possono ricavare le quantità $m_x = 416,67$, $m_y = 81,67$, $\sigma_x^2 = 2.813.291,67/15$, $\sigma_y^2 = 105.575,17/15$, e la tabella

$x_i y_j f_{ij}$	25	100	200	300	
150	33.750	15.000	0	0	
450	0	45.000	0	0	
900	0	90.000	360.000	0	
1.600	0	0	0	480.000	
					1.023.750

da cui si ricava $\sigma_{xy} = 1.023.750/15 - 416,67 \cdot 81,67 = 513.308,42/15$, $b = 0,1825$, $a = 5,6277$ e $r^2 = 0,887$.

Esercizio F.
a) Riportando su di un grafico cartesiano il numero delle tratte protestate (Y) verso il numero di cambiali ordinarie (X) si può notare che all'aumentare dei valori della X aumentano anche i valori della Y. Per il calcolo del coefficiente di correlazione lineare, si consideri la tabella

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
Toscana	128	72	16384	5184	9216
Umbria	30	15	900	225	450
Marche	44	35	1936	1225	1540
Lazio	368	113	135424	12769	41584
Abruzzo	92	46	8464	2116	4232
Molise	18	7	324	49	126
	680	288	163432	21568	57148

da cui si ricava $m_x = 113,33$, $m_y = 48$, $\sigma_x^2 = 163432/6 - (113,33)^2 = 14394,98$, $\sigma_y^2 = 21568/6 - (48)^2 = 1290,67$, $\sigma_{xy} = 57148/6 - 113,33 \cdot 48 = 4084,83$ e quindi $r = 4084,83/\sqrt{14394,98 \cdot 1290,67} = 0,948$.