Università degli Studi di Cassino

Esercitazione di Statistica 2 dell'8.03.2007

Simona Balzano

Esercizio 1

Da un campione di 7 aziende risultano i seguenti valori del numero totale di dipendenti (X) e del numero di dipendenti laureati (Y):

Dip. totali (X)	5	8	10	11	7	9	6
Dip. Laureati (Y)	3	5	7	6	4	3	2

a) Disegnare la retta di regressione di Y su X;
b) misurare la bontà dell'adattamento;
c) sapendo che $s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = 3,89$ verificare la significatività del modello al livello del 99%

Soluzione

Le quantità utili per rispondere ai 3 quesiti sono riassunte nella seguente tabella:

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
5	3	-3	-1,29	9	1,65	3,86
8	5	0	0,71	0	0,51	0
10	7	2	2,71	4	7,37	5,43
11	6	3	1,71	9	2,94	5,14
7	4	-1	-0,29	1	0,08	0,29
9	3	1	-1,29	1	1,65	-1,29
6	2	-2	-2,29	4	5,22	4,57
56	**30**	**0**	**0**	**28**	**19,43**	**18**

$n = 7$

$\bar{x} = \frac{56}{7} = 8$ $\bar{y} = \frac{30}{7} = 4,29$

$\text{dev}(x) = \sum_{i=1}^{n}(x_i - \bar{x})^2$ $\text{dev}(y) = \sum_{i=1}^{n}(y_i - \bar{y})^2$ $\text{cod}(x,y) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

a)

$$\hat{b} = \frac{\text{cov}(x,y)}{\sigma_x^2} = \frac{\text{cod}(x,y)}{\text{dev}(x)} = \frac{18}{28} = \mathbf{0,64}$$

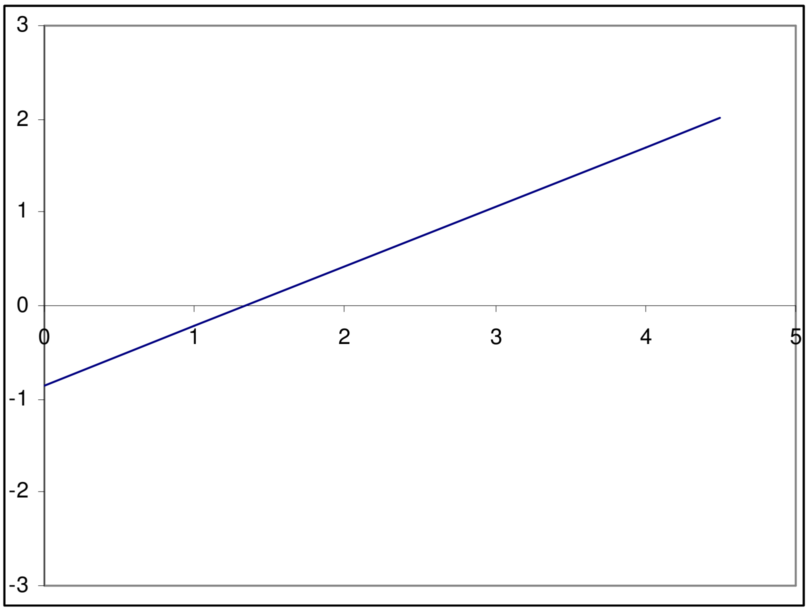
$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 4,29 - 0,64 \cdot 8 = \mathbf{-0,86}$$

Equazione della retta:

$$\hat{y} = -0,86 + 0,64x$$

Per disegnare la retta è necessario determinare almeno 2 punti. Uno, l'intercetta, è già determinato, l'altro va determinato in base all'equazione, ponendo x = 1:

x	y
0	-0,86
1	-0,22



b)

La bontà dell'adattamento si misura con l'indice di determinazione lineare R^2, che può essere calcolato con una delle seguenti espressioni:

$$R^2 = \frac{\text{dev}_{REG}}{\text{dev}(y)} = 1 - \frac{\text{dev}_{RES}}{\text{dev}(y)} = \frac{(\text{cov}(x,y))^2}{\sigma_x^2 \cdot \sigma_y^2} = \frac{(\text{cod}(x,y))^2}{\text{dev}(x) \cdot \text{dev}(y)} = \rho^2$$

$$R^2 = \frac{18^2}{28 \cdot 19,43} = \mathbf{0,6}$$

c)

Test su a, bilaterale

▪ Ipotesi	H_0: a = 0 H_1: a ≠ 0
▪ Statistica test	$X_{test} = \dfrac{\hat{a} - a}{s_a} \sim t_{n-2}$, in cui: $s_a = \sqrt{\dfrac{s^2}{n}\left[1 + \dfrac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]}$
▪ Valori critici	$\pm\, t_{0,005;5} = \pm\, 4,032$
▪ Regola di decisione	$-4,032 \le x_{test} \le 4,032$ ⇒ si accetta H_0 $x_{test} \le -4,032$ oppure $x_{test} \ge 4,032$ ⇒ si rifiuta H_0
▪ Valore test	1. $s_a = \sqrt{\dfrac{3,89}{7}\left[1 + \dfrac{8^2}{28}\right]} = 1,35$ 2. $x_{test} = \dfrac{-0,86}{1,35} = -0,63$
▪ Decisione	$-4,032 \le -0,63 \le 4,032$ **⇒ si accetta H_0**

Test su b, bilaterale

▪ Ipotesi	H_0: b = 0 H_1: b ≠ 0
▪ Statistica test	$X_{test} = \dfrac{\hat{b} - a}{s_b} \sim t_{n-2}$, in cui: $s_b = \sqrt{s^2 \dfrac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$
▪ Valori critici	$\pm\, t_{0,005;5} = \pm\, 4,032$
▪ Regola di decisione	$-4,032 \le x_{test} \le 4,032$ ⇒ si accetta H_0 $x_{test} \le -4,032$ oppure $x_{test} \ge 4,032$ ⇒ si rifiuta H_0
▪ Valore test	1. $s_b = \sqrt{1,57 \dfrac{1}{28}} = 0,37$ 2. $x_{test} = \dfrac{0,64}{0,37} = 1,72$
▪ Decisione	$-4,032 \le 1,72 \le 4,032$ **⇒ si accetta H_0**

Test su R^2, unilaterale

▪ Ipotesi	H_0: $R^2 = 0$ H_1: $R^2 > 0$
▪ Statistica test	$X_{test} = \dfrac{dev_{REG}}{dev_{RES}/n-2} = \dfrac{R^2(n-2)}{1-R^2} \sim F_{1;n-2}$
▪ Valore critico	$F_{0,01;1;5} = 5764$
▪ Regola di decisione	$x_{test} \leq 5764$ $\Rightarrow$ si accetta H_0 $x_{test} \geq 5764$ $\Rightarrow$ si rifiuta H_0
▪ Valore test	$X_{test} = \dfrac{0,6 \cdot 5}{1-0,6} = 7,36$
▪ Decisione	$7,36 \leq 5764$ **$\Rightarrow$ si accetta H_0**

Esercizio 2

Il consumo di carne (X) e vino (Y) in 16 paesi europei fa registrare i seguenti indici:

$\bar{x} = 25,94$ $\bar{y} = 85,78$ $cod_{XY} = 2.121,33$

$dev_X = 6.202,68$ $dev_Y = 4.650,46$

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = 280,35$$

a) Si stimi la retta di regressione tra X e Y;
b) si verifichi la significatività del modello al livello del 95%.

Soluzione

a)

$$\hat{b} = \frac{cod(x,y)}{dev(x)} = \frac{2121,33}{6202,68} = \mathbf{0,34}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 85,78 - 0,34 \cdot 25,94 = \mathbf{76,91}$$

Equazione della retta:

$$\hat{y} = 76,91 + 0,34x$$

$$R^2 = \frac{\left(cod(x,y)\right)^2}{dev(x)\cdot dev(y)} = \frac{2121,33^2}{6202,68 \cdot 4650,46} = 0,156$$

b)

Test su a, bilaterale

▪ Ipotesi	H_0: a = 0 H_1: a ≠ 0
▪ Statistica test	$X_{test} = \dfrac{\hat{a} - a}{s_a} \sim t_{n-2}\,, \qquad \text{in cui: } s_a = \sqrt{\dfrac{s^2}{n}\left[1 + \dfrac{\overline{x}^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right]}$
▪ Valori critici	$\pm\, t_{0,025;14} = \pm\, 2{,}145$
▪ Regola di decisione	$-\,2{,}145 \le x_{test} \le 2{,}145$ ⇒ si accetta H_0 $x_{test} \le -\,2{,}145$ oppure $x_{test} \ge 2{,}145$ ⇒ si rifiuta H_0
▪ Valore test	1. $s_a = \sqrt{\dfrac{280{,}35}{16}\left[1 + \dfrac{25{,}94^2}{6202{,}68}\right]} = 5{,}21$ 2. $x_{test} = \dfrac{76{,}91}{5{,}21} = 14{,}75$
▪ Decisione	14,75 > 2,145 **⇒ si rifiuta H_0**

Test su b, bilaterale

▪ Ipotesi	H_0: b = 0 H_1: b ≠ 0
▪ Statistica test	$X_{test} = \dfrac{\hat{b} - a}{s_b} \sim t_{n-2}\,, \qquad \text{in cui: } s_b = \sqrt{s^2 \dfrac{1}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}$
▪ Valori critici	$\pm\, t_{0,025;14} = \pm\, 2{,}145$
▪ Regola di decisione	$-\,2{,}145 \le x_{test} \le 2{,}145$ ⇒ si accetta H_0 $x_{test} \le -\,2{,}145$ oppure $x_{test} \ge 2{,}145$ ⇒ si rifiuta H_0
▪ Valore test	1. $s_b = \sqrt{\dfrac{280{,}35}{6202{,}68}} = 0{,}2$ 2. $x_{test} = \dfrac{0{,}34}{0{,}2} = 1{,}72$
▪ Decisione	- 2,145 ≤ 1,72 ≤ 2,145 **⇒ si accetta H_0**

Test su R^2, unilaterale

▪ Ipotesi	H_0: $R^2 = 0$ H_1: $R^2 > 0$
▪ Statistica test	$X_{test} = \frac{dev_{REG}}{dev_{RES}/n-2} = \frac{R^2(n-2)}{1-R^2} \sim F_{1;n-2}$
▪ Valore critico	$F_{0,05;1;14} = 245$
▪ Regola di decisione	$x_{test} \leq 245$ $\Rightarrow$ si accetta H_0 $x_{test} \geq 245$ $\Rightarrow$ si rifiuta H_0
▪ Valore test	$X_{test} = \frac{0,156 \cdot 14}{1-0,156} = 2,59$
▪ Decisione	$2,59 \leq 245$ **$\Rightarrow$ si accetta H_0**

Esercizio 3

Su un insieme di 10 aziende sono osservati due indicatori economico-finanziari I_X ed I_Y, i cui risultati possono essere sintetizzati come segue:

$\bar{x} = 12,3$ $\quad$ $\bar{y} = 11,2$ $\quad$ $cod_{XY} = 1.195,71$

$dev_X = 845$ $\quad$ $dev_{REG} = 1.698,87$ $\quad$ $s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = 17,85$

a) Si stimi la retta di regressione tra X e Y;
b) si verifichi la significatività del modello al livello del 95%.

Soluzione