Statistica

Esercitazione VIII

17 Novembre 2022

Esercizio 1 (funzione di regressione e metodo dei minimi quadrati) Si considerino i dati riassunti nella seguente tabella:

$X \backslash Y$	10	20	
0	4	1	5
12	1	4	5
	5	5	10

1. Si rappresentino graficamente le coppie di dati e la funzione di regressione, ipotizzando Y in funzione di X.
2. Si stimino, secondo il criterio dei minimi quadrati, i parametri del modello di regressione $Y = a+bX$.
3. Si calcoli l'indice ρ^2.

Esercizio 2 (funzione di regressione e metodo dei minimi quadrati) La seguente tabella riporta osservazioni di due variabili quantitative X e Y in corrispondenza di 5 individui:

x_i	5	5	3	1	1
y_i	11	25	25	25	39

1. Si rappresentino graficamente le coppie di dati e la funzione di regressione, ipotizzando Y in funzione di X.
2. Si calcolino, secondo il criterio dei minimi quadrati, i parametri e gli indici di adattamento dei modelli
$$Y = a + bX, \qquad Y = c.$$
3. In corrispondenza del modello migliore, si fornisca la previsione relativa a $X = 3$.
4. Si calcoli l'indice di dipendenza in media $\eta^2_{Y|X}$ e lo si confronti con i risultati al punto 2.

Esercizio 3 (funzione di regressione e metodo dei minimi quadrati) Il seguente prospetto riporta i valori rilevati in corrispondenza di 6 unità statistiche con riferimento a due caratteri quantitativi X e Y:

x_i	1	2	3	4	5	6
y_i	100	100	130	133	180	200

1. Si rappresentino graficamente le coppie di dati e la funzione di regressione, ipotizzando Y in funzione di X.
2. Si calcolino, secondo il criterio dei minimi quadrati, i parametri e gli indici di adattamento dei seguenti modelli, individuando quello migliore

$$Y = a + bX, \qquad Y = c + dX^2, \qquad y = hX.$$

3. Si determinino i valori dei rapporti di correlazione $\eta^2_{Y|X}$ e $\eta^2_{X|Y}$.

Esercizio 4 Si completino le seguenti tabelle, riportando modalità e frequenze e motivando teoricamente la risposta, in modo tale che siano soddisfatte le condizioni sotto riportate.

$X \setminus Y$	2	5	8
x_1	2	4	2
x_2		4	4

$\eta^2_{Y|X} = 0, \eta^2_{X|Y} > 0$

$X \setminus Y$	y_1	y_2	y_3
x_1	2	4	2
x_2		4	4

$\rho^2 = 0$

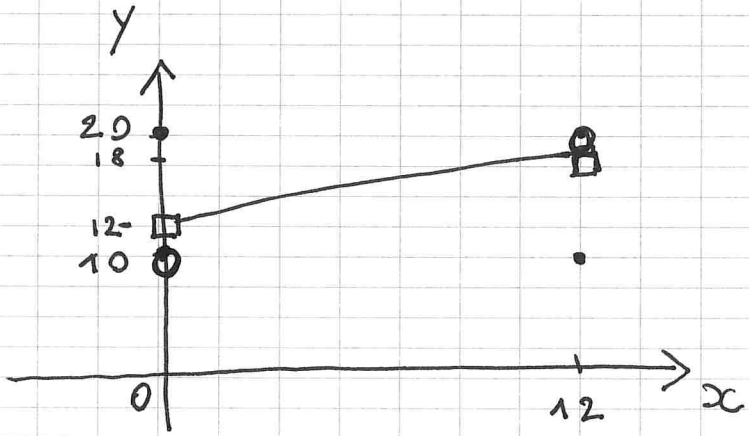
Esercizio 5 Con riferimento a $n = 50$ unità statistiche sono stati raccolti alcuni dati relativi alle variabili (X, Y) nella seguente tabella

$Y \setminus X$	1	2	3
10	8	1	0
20	8	4	a
30	0	1	b

Si determinino i valori (interi) da assegnare alle costanti a e b affinché $\eta^2_{Y|X} = \rho^2$.

ESERCIZIO 1

X \ Y	10	20	
0	4	1	5
12	1	4	5
	5	5	10



1. $(0,10)$; $(0,20)$; $(12,10)$; $(12,20)$

$$M(Y|x=0) = \frac{10\cdot 4 + 20\cdot 1}{5} = 12$$

$$M(Y|x=12) = \frac{10\cdot 1 + 20\cdot 4}{5} = 18$$

2. $Y = a + bX \Rightarrow \hat{b} = Cov(X,Y)/Var(X)$, $\hat{a} = M(Y) - \hat{b}\,M(X)$

$$M(X) = \frac{0\cdot 5 + 12\cdot 5}{10} = 6; \quad M(X^2) = \frac{0^2\cdot 5 + 12^2\cdot 5}{10} = 72;$$

$$Var(X) = 72 - 6^2 = 36; \quad M(Y) = \frac{10\cdot 5 + 20\cdot 5}{10} = 15; \quad M(Y^2) = \frac{10^2\cdot 5 + 20^2\cdot 5}{10} = 250$$

$$Var(Y) = 250 - 15^2 = 25; \quad Cov(X,Y) = M(XY) - M(X)M(Y) =$$

$$= \frac{(0\cdot 10\cdot 4 + 0\cdot 20\cdot 1 + 12\cdot 10\cdot 1 + 12\cdot 20\cdot 4)}{10} - 6\cdot 15$$

$$= 18$$

$$\hat{b} = \frac{18}{36} = \frac{1}{2} \; ; \quad \hat{a} = 15 - \frac{1}{2} \cdot 6 = 12 \quad \Rightarrow \quad \hat{y} = 12 + \frac{1}{2} X$$

2

3.

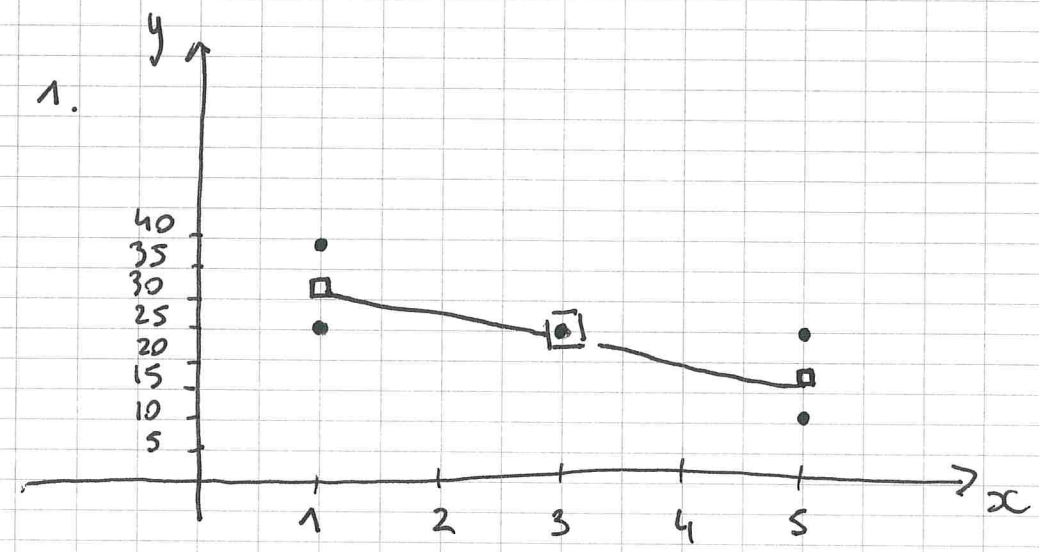
$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{18}{\sqrt{36 \cdot 25}} = \frac{\cancel{18}^{\,3}}{\cancel{6} \cdot 5} = \frac{3}{5} = 0.6$$

$$\rho^2 = 0.6^2 = 0.36 = R^2$$

ESERCIZIO 2

x_i	5	5	3	1	1
y_i	11	25	25	25	39

1.



$$M(Y|x=1) = \frac{25+39}{2} = 32$$

$$M(Y|x=3) = 25$$

$$M(Y|x=5) = \frac{11+25}{2} = 18$$

LE MEDIE CONDIZIONATE SONO ALLINEATE (VI PASSA UNA RETTA)

2. $\boxed{Y = a + bX}$, $\hat{b} = \dfrac{Cov(X,Y)}{Var(X)}$, $\hat{a} = M(Y) - \hat{b}\,M(X)$

$M(X) = 3$; $M(X^2) = \dfrac{5^2+5^2+\dots+1^2}{5} = 12.2$; $Var(X) = 12.2 - 3^2 = 3.2$

$M(Y) = 25$; $M(Y^2) = \dfrac{11^2+25^2+\dots+39^2}{5} = 703.4$; $Var(Y) = 703.4 - 25^2 = 78.4$

$Cov(X,Y) = M(XY) - M(X)M(Y) = \dfrac{5\cdot 11 + 5\cdot 25 + 3\cdot 25 + 1\cdot 25 + 1\cdot 39}{5} - 3\cdot 25 = -11.2$

$\hat{b} = -\dfrac{11.2}{3.2} = -3.5$; $\hat{a} = 25 - (-3.5)3 = 35.5$ $\Rightarrow$ $\boxed{\hat{Y} = 35.5 - 3.5X}$

$$R^2 = \rho^2$$

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = -\frac{11.2}{\sqrt{3.2}\cdot\sqrt{78.4}} = -0.7071 \Rightarrow R^2 = \rho^2 = (-0.7071)^2 \simeq 0.5$$

$$\boxed{Y = c}$$

$$\hat{c} = M(Y) = 25 \Rightarrow \boxed{Y = 25}$$

$R^2 = 0$ (DALLA TEORIA)

3. $R_1^2 = 0.5 > R_2^2 = 0$

IL PRIMO MODELLO É MIGLIORE

$$\hat{Y}\Big|_{X=3} = 35.5 - 3.5\cdot 3 = 25$$

4. $$\eta^2_{Y|X} = \frac{\sigma^2_{BETWEEN,Y}}{\sigma^2_Y} = \frac{39.2}{78.4} = 0.5$$; $\sigma^2_{BETWEEN,Y}$ = VARIANZA DELLE $M(Y|X)$

$$= \{[M(Y|x=1) - M(Y)]^2\cdot 2 + [M(Y|x=3) - M(Y)]^2\cdot 1 + [M(Y|x=5) - M(Y)]^2\cdot 2\}/5 =$$

$$= \frac{(32-25)^2\cdot 2 + (25-25)^2\cdot 1 + (18-25)^2\cdot 2}{5} = 39.2$$

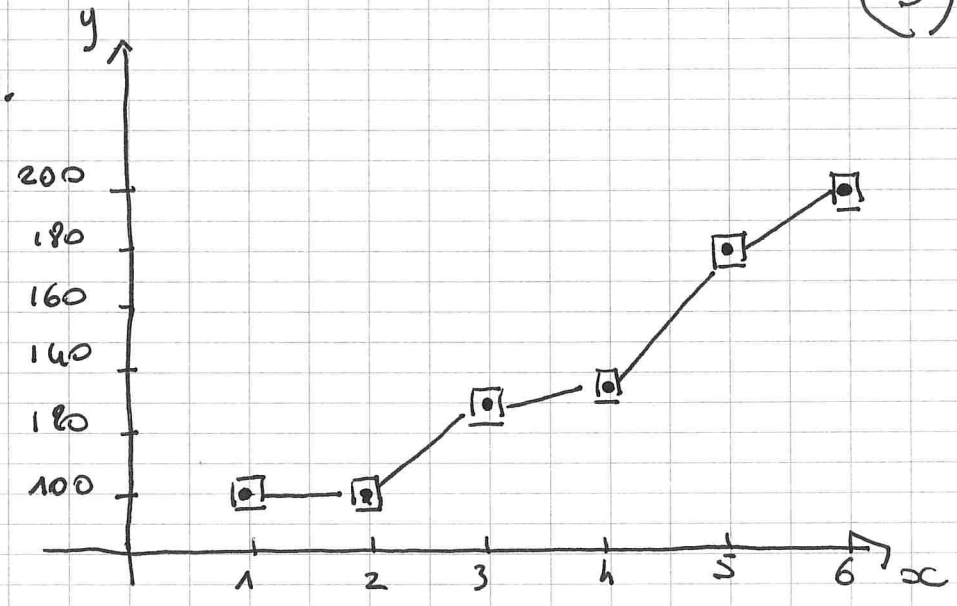
$R_1^2 = \eta^2_{Y|X}$ $\Rightarrow$ MEDIE CONDIZIONATE ALLINEATE, DUNQUE LA FUNZ. DI REGRESSIONE COINCIDE CON LA RETTA DEI MINIMI QUADRATI

$$\hat{Y} = 35.5 - 3.5X$$

ESERCIZIO 3

x_i	y_i	$z_i = x_i^2$
1	100	1
2	100	4
3	130	9
4	133	16
5	180	25
6	200	36

1.



$M(Y|x=1) = 100$

$M(Y|x=2) = 100$

$M(Y|x=3) = 130$

$M(Y|x=4) = 133$

$M(Y|x=5) = 180$

$M(Y|x=6) = 200$

8. $\boxed{Y = a + bX}$ $\quad \hat{b} = \dfrac{Cov(X,Y)}{Var(X)}$; $\quad \hat{a} = M(Y) - \hat{b}\,M(X)$

$M(X) = 3.5$; $\quad M(X^2) = \dfrac{1^2 + 2^2 + \dots + 6^2}{6} = 15.1667$; $\quad Var(X) = 15.1667 - 3.5^2 = 2.9167$

$M(Y) = 140.5$; $\quad M(Y^2) = \dfrac{100^2 + 100^2 + \dots + 200^2}{6} = 21\,164.83$; $\quad Var(Y) = 21\,164.83 - 140.5^2 = 1424.583$

$Cov(X,Y) = M(XY) - M(X)M(Y) = \dfrac{1\cdot 100 + 2\cdot 100 + 3\cdot 130 + \dots + 6\cdot 200}{6} - 3.5\cdot 140.5 = 61.9167$

$\hat{b} = \dfrac{61.9167}{2.9167} = 21.2286$; $\quad \hat{a} = 140.5 - 21.2286\cdot 3.5 = 66.2$

$$\boxed{\hat{Y} = 66.2 + 21.2286\cdot X}$$

$\rho = \dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \dfrac{61.9167}{\sqrt{2.9167\cdot 1424.583}} = 0.9606$

$R^2 = \rho^2 = 0.9606^2 = 0.9227$; $\quad \sigma^2_{RES} = (1-R^2)\,\sigma^2_Y = (1-0.9227)\,1424.583$

$= 110.182$

$$\boxed{Y = c + dX^2}$$

MODELLO LINEARE NEI PARAMETRI c E d, MA NON LINEARE IN X (X^2). $\Rightarrow$ TRASFORMO LA X^2: $Z = X^2$

7

$$Y = c + dZ \quad : \quad \hat{d} = \frac{Cov(Z,Y)}{Var(Z)}, \quad \hat{c} = M(Y) - \hat{d}\,M(Z)$$

$$M(Z) = \frac{1+4+\dots+36}{6} = 15.1667; \quad M(Z^2) = \frac{1^2+4^2+\dots+36^2}{6} = 379.1667$$

$$Var(Z) = 379.1667 - 15.1667^2 = 149.1389$$

$$Cov(Z,Y) = M(ZY) - M(Z)M(Y) = \frac{1\cdot 100 + 4\cdot 100 + \dots + 36\cdot 200}{6} - 15.1667 \cdot 140.5 = 452.08$$

$$\hat{d} = \frac{452.08}{149.1389} = 3.0313, \quad \hat{c} = 140.5 - 3.0313 \cdot 15.1667 = 94.5294$$

$$\boxed{\hat{Y} = 94.5294 + 3.0313\,X^2}$$

$$\rho_{Z,Y} = \frac{Cov(Z,Y)}{\sqrt{Var(Z)Var(Y)}} = \frac{452.08}{\sqrt{149.1389 \cdot 1424.583}} = 0.9808 \Rightarrow R^2_{Z,Y} = \rho^2_{Z,Y} = 0.9808^2 = 0.9620$$

$$\sigma^2_{RES} = (1-R^2)\sigma^2_Y = (1-0.9620)\,1424.583 = 54.187$$

$$\boxed{Y = hX}$$

POLINOMIO IN X DI 1° GRADO, SENZA INTERCETTA, CIOÈ SENZA IL TERMINE DI GRADO 0 $\Rightarrow$ POLINOMIO INCOMPLETO

$$\hat{h} = \frac{M(X,Y)}{M(X^2)} = \frac{553.667}{15.1667} = 36.5054 \qquad \boxed{\hat{Y} = 36.5054\,X}$$

NON VALE IL TEOREMA DI SCOMPOSIZIONE DELLA VARIANZA $\Rightarrow R^2$ NON PUÒ ESSERE CALCOLATO

$$\sigma^2_{RESIDUA} = M\left[(Y - \hat{h}X)^2\right] = M[Y^2] - \hat{h}\,M(XY)$$

$$= 21164.83 - 36.5054 \cdot 553.667 = 952.9726$$

IL 2° MODELLO $(Y = c + dX^2)$ È IL MIGLIORE IN QUANTO HA σ^2_{RES} PIÙ BASSA.

3. $\eta^2_{Y|X} = 1$, IN QUANTO VI È DIP. FUNZIONALE DI Y DA X
(CIOÈ $\forall x \; \exists 1! \; y$)

$\eta^2_{X|Y} < 1$, POICHÉ NON C'È DIP. FUNZ. DI X DA Y (ES. PER $y = 100$, $x = \{1, 2\}$).

$M(X|Y=100) = \frac{1+2}{2} = 1.5$

$M(X|Y=130) = 3$

$M(X|Y=133) = 4$

$M(X|Y=180) = 5$

$M(X|Y=200) = 6$

$M(X) = 3.5$

$\Rightarrow \sigma^2_{BETWEEN,X} = \text{VAR. DELLE } M(X|Y)$

$= \{(1.5-3.5)^2 \cdot 2 + (3-3.5)^2 \cdot 1 + \dots + (6-3.5)^2 \cdot 1\} / 6$

$= 2.8333$

$$\eta^2_{X|Y} = \frac{\sigma^2_{BETWEEN,X}}{\sigma^2_X} = \frac{2.8333}{2.9167} = 0.9714$$

ESERCIZI 4 E 5 PROSSIMA ESERCITAZIONE