



Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

Esercitazione 8

Statistica

Alfonso Iodice D'Enza
iodicede@gmail.com

Università degli studi di Cassino



Outline

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

- 1 Studio della dipendenza
- 2 La retta di regressione
- 3 Qualità della soluzione trovata
- 4 Outliers
- 5 Regressione su tabella a doppia entrata



Dipendenza lineare

Esercitazione 8

A. Iodice

Studio della dipendenza

La retta di regressione

Qualità della soluzione trovata

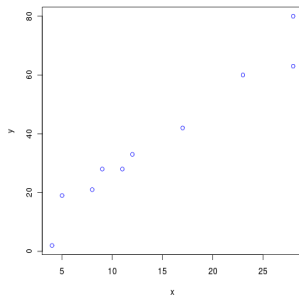
Outliers

Regressione su tabella a doppia entrata

Lo studio della relazione tra caratteri statistici è, nel caso della interdipendenza, di tipo simmetrico: due caratteri quantitativi X e Y hanno lo stesso ruolo e si vuole studiare se essi siano indipendenti o meno. A questo scopo sono stati introdotti gli indici di covarianza σ_{xy} e di correlazione lineare ρ . Si consideri di aver osservato due caratteri quantitativi X ed Y . Si riportano i valori e il grafico di dispersione:

	Y	X
1	28	11
2	21	8
3	63	28
4	42	17
5	28	9
6	2	4
7	80	28
8	19	5
9	33	12
10	60	23
	376	145

Il diagramma di dispersione (scatter plot)





Dipendenza lineare

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

covarianza e coefficiente di correlazione

$$\bullet \mu_x = \frac{\sum_{i=1}^{10} x_i}{10} = 14.5$$

$$\bullet \mu_y = \frac{\sum_{i=1}^{10} y_i}{10} = 37.6$$

$$\bullet \sigma_x = \sqrt{\frac{\sum_{i=1}^{10} (x_i - \mu_x)^2}{10}} = 8.57$$

$$\bullet \sigma_y = \sqrt{\frac{\sum_{i=1}^{10} (y_i - \mu_y)^2}{10}} = 22.49$$

$$\bullet \sigma_{xy} = \frac{\sum_{i=1}^{10} (x_i - \mu_x)(y_i - \mu_y)}{10} = 187.3$$

$$\bullet \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = 0.97$$

Dipendenza funzionale lineare

Essendo il valore del coefficiente di correlazione lineare prossimo ad 1 esiste una forte relazione lineare tra X ed Y . Come confermato dal grafico di dispersione, i dati sono approssimativamente allineati lungo una retta crescente. Ci si può dunque aspettare che sussista una relazione funzionale tra i dati del tipo

$$Y = f(X) = b_0 + b_1 X$$

che rappresenta l'equazione di una retta passante attraverso la nube di punti di coordinate (x_i, y_i) .



La retta di regressione

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

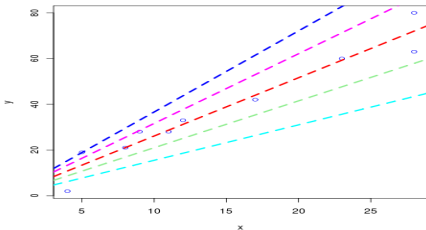
Outliers

Regressione su
tabella a
doppia entrata

La retta di regressione

La retta di regressione fornisce una approssimazione della dipendenza dei valori di Y dai valori di X . La relazione di dipendenza non è esattamente riprodotta dalla retta; i valori $y_i^* = b_0 + b_1 x_i$ sono dunque i valori teorici, ovvero i valori che la variabile Y assume, secondo il modello $Y = b_0 + b_1 X$, in corrispondenza dei valori x_i osservati.

rette passanti per la nube di punti



Determinazione della retta di regressione

L'identificazione della retta avviene attraverso la determinazione dei valori di b_0 , l'intercetta, e b_1 , il coefficiente angolare o pendenza. La retta 'migliore' è quella che passa più 'vicina' ai punti osservati. In altre parole, si vuole trovare la retta per la quale le differenze tra i valori teorici y_i^* e i valori osservati y_i siano minime.



La retta di regressione

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

I residui

le differenze tra i valori teorici y_i^* e i valori osservati y_i vengono definite **residui**. La retta di regressione è tale che la somma dei residui al quadrato sia minima. Formalmente

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - y_i^*)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$

Il problema consiste dunque nel ricercare b_0 e b_1 che minimizzano la precedente espressione. Da un punto di vista operativo bisogna risolvere il seguente sistema di equazioni

$$\frac{\partial}{\partial b_0} \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 = 0$$

$$\frac{\partial}{\partial b_1} \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 = 0$$

Ricerca dei parametri della retta di regressione: (b_0)

$$-2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) =$$

$$\sum_{i=1}^n y_i - n * b_0 - b_1 \sum_{i=1}^n x_i = 0$$

$$b_0 = \mu_y - b_1 \mu_x$$



La retta di regressione

Esercitazione 8

A. Iodice

Studio della dipendenza

La retta di regressione

Qualità della soluzione trovata

Outliers

Regressione su tabella a doppia entrata

I residui

Le differenze tra i valori teorici y_i^* e i valori osservati y_i vengono definite **residui**. La retta di regressione è tale che la somma dei residui al quadrato sia minima. Formalmente

$$\begin{aligned} \sum_{i=1}^n e_i^2 &= \sum_{i=1}^n (y_i - y_i^*)^2 = \\ &= \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 \end{aligned}$$

Il problema consiste dunque nel ricercare b_0 e b_1 che minimizzano la precedente espressione. Da un punto di vista operativo bisogna risolvere il seguente sistema di equazioni

$$\frac{\partial}{\partial b_0} \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 = 0$$

$$\frac{\partial}{\partial b_1} \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 = 0$$

Ricerca dei parametri della retta di regressione: (b_1)

$$-2 \sum_{i=1}^n x_i (y_i - b_0 - b_1 x_i) = 0$$

$$\sum_{i=1}^n x_i y_i - b_0 \sum_{i=1}^n x_i - b_1 \sum_{i=1}^n x_i^2 = 0$$

$$b_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \left(\frac{\sum_{i=1}^n y_i}{n} - b_1 \frac{\sum_{i=1}^n x_i}{n} \right)$$

$$b_1 \left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) = n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i$$

$$b_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{\sigma_{xy}}{\sigma_x^2}$$



Determinazione della retta di regressione

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

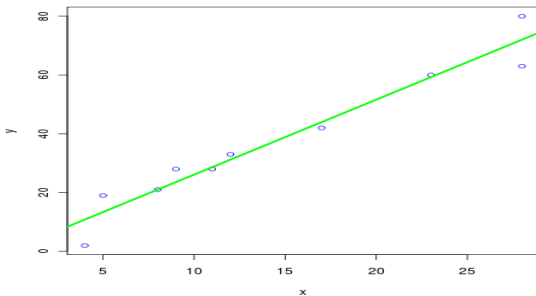
Calcolo dei coefficienti

Richiamando le quantità calcolate in precedenza e le formule per il calcolo dei parametri si ha

$$\bullet \quad b_1 = \frac{\sigma_{xy}}{\sigma_x^2} = 2.55$$

$$\bullet \quad b_0 = \mu_y - b_1 \mu_x = 37.6 - (2.55 * 14.5) = 0.62$$

La retta 'migliore'





Interpretazione dei valori dei coefficienti di regressione

Esercitazione 8

A. Iodice

Studio della dipendenza

La retta di regressione

Qualità della soluzione trovata

Outliers

Regressione su tabella a doppia entrata

- b_0 rappresenta l'intercetta della retta di regressione ed indica il valore della variabile di risposta Y quando il predittore X assume valore 0.
- b_1 rappresenta l'inclinazione della retta di regressione, ovvero la variazione della variabile di risposta Y in conseguenza di un aumento unitario del predittore X .



Bontà di adattamento

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

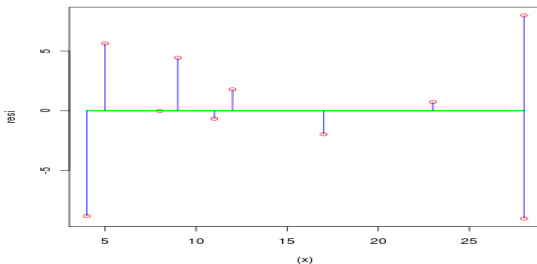
Esistono diversi strumenti grafici ed analitici per valutare la bontà dell'adattamento della retta di regressione ai dati

- Strumenti grafici: **plot dei residui**
- Strumenti analitici: **coefficiente di determinazione lineare**
 R^2

Perchè la retta possa essere considerata una buona approssimazione della relazione che intercorre tra Y ed X è necessario che i residui abbiano un andamento casuale rispetto ai valori della X . Se, ad esempio, all'aumentare dei valori della X aumentassero sistematicamente anche i residui, allora la relazione potrebbe non essere non lineare: la retta di regressione ne sarebbe dunque una cattiva approssimazione.

Plot dei residui

Per verificare che l'andamento dei residui sia effettivamente casuale rispetto ad X , è possibile utilizzare un diagramma di dispersione tra i valori x_i ed i corrispondenti residui $e_i (i = 1, \dots, n)$





coefficiente di determinazione lineare R^2

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

Ricordando che la devianza il numeratore della varianza...

$$\begin{aligned} Dev_y &= \sum_{i=1}^n (y_i - \mu_y)^2 = \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \mu_y)^2 = \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \mu_y)^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \mu_y) \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \mu_y)^2 + 2 \left(\sum_{i=1}^n y_i - \sum_{i=1}^n \hat{y}_i \right) \left(\sum_{i=1}^n \hat{y}_i - n\mu_y \right) \end{aligned}$$

Il metodo dei minimi quadrati assicura che $\sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n y_i$, quindi

$$\begin{aligned} Dev(y) &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \mu_y)^2 + 2 * 0 * \left(\sum_{i=1}^n \hat{y}_i - n\mu_y \right) \\ &= \sum_{i=1}^n (\hat{y}_i - \mu_y)^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 = Dev_r + Dev_e \end{aligned}$$

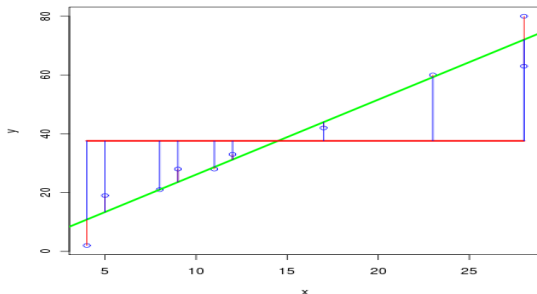


Decomposizione della devianza

La devianza può essere decomposta dunque nelle seguenti quantità $Dev_y = Dev_r + Dev_e$

- $Dev_y = \sum_{i=1}^n (y_i - \mu_y)^2$ devianza totale
- $Dev_r = \sum_{i=1}^n (\hat{y}_i - \mu_y)^2$ devianza di regressione
- $Dev_e = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ devianza dei residui

Interpretazione grafica





Bontà dell'adattamento

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

Intuitivamente, l'adattamento della retta è migliore quanto maggiore sarà la proporzione di variabilità totale che la retta di regressione riesce a spiegare; ovvero, l'adattamento della retta è migliore quanto minore sarà la variabilità residua. Una misura di come il modello approssima i dati osservati è data dal coefficiente di determinazione lineare R^2 , dato da

$$R^2 = \frac{Dev_r}{Dev_y} = \frac{\sum_{i=1}^n (\hat{y}_i - \mu_y)^2}{\sum_{i=1}^n (y_i - \mu_y)^2}$$

ovvero

$$R^2 = 1 - \frac{Dev_e}{Dev_y} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \mu_y)^2}$$

esempio di calcolo R^2

- $Dev_y = \sum_{i=1}^n (y_i - \mu_y)^2 = 5058.4$
- $Dev_r = \sum_{i=1}^n (\hat{y}_i - \mu_y)^2 = 4776.214$
- $Dev_e = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = 282.1862$

$$R^2 = \frac{Dev_r}{Dev_y} = \frac{4776.214}{5058.4} = 0.94$$

ovvero

$$R^2 = 1 - \frac{Dev_e}{Dev_y} = 1 - \frac{282.1862}{5058.4} = 1 - 0.0557 = 0.94$$



Influenza di un outlier sulla soluzione

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

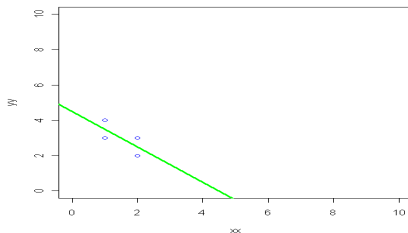
Un piccolo esempio

Si considerino le seguenti
osservazioni

xx	yy
1	4
1	3
2	3
2	2

Retta di regressione

La soluzione induce a concludere che vi sia una relazione di proporzionalità inversa: poichè la retta è decrescente si deduce che all'aumentare di X , la variabile dipendente Y diminuisce.





Influenza di un outlier sulla soluzione

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

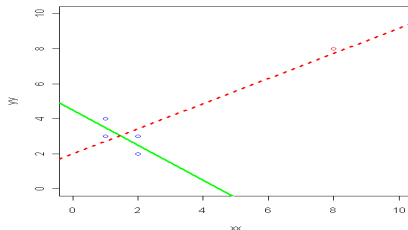
Regressione su
tabella a
doppia entrata

Un (altro) piccolo esempio

Si considerino le osservazioni precedenti a cui è aggiunta un'unica coppia di valori (8, 8). I dati sono

xx2	yy2
1	4
1	3
2	3
2	2
8	8

Retta di regressione



In questo caso, la sola presenza della nuova osservazione conduce all'identificazione di una retta di regressione diversa dalla prima: l'inclinazione positiva della retta indica una relazione di diretta proporzionalità. Tuttavia tale soluzione è unicamente dovuta dalla presenza dell'osservazione (8, 8) che pertanto induce a valutare la relazione di dipendenza tra Y ed X in maniera errata. L'osservazione (8, 8) si definisce pertanto un **outlier**. L'identificazione e la conseguente eliminazione degli eventuali outlier è un elemento molto importante nello studio della dipendenza tra fenomeni.



Esercizio regressione: distribuzione doppia di frequenze

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

Si consideri di aver osservato su 10 rivenditori di componenti informatiche le variabili *numero di punti vendita* e *Fatturato settimanale complessivo*. Si studi la dipendenza del fatturato dal numero di punti vendita.

	fino a 2	tra 2 e 4	tra 4 e 6
fino a 5000	3	2	0
tra 5000 e 1000	1	2	2

- Si stimino i coefficienti della retta di regressione.
- Si valuti la bontà di adattamento della retta ai dati.



Esercizio regressione: distribuzione doppia di frequenze

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

Essendo le modalità delle variabili qualitative espresse in intervalli di valori, è necessario fare riferimento ai centri di ciascun intervallo. La tabella è dunque data da

Y/X	1	3	5	Tot
2500	3	2	0	5
7500	1	2	2	5
Tot	4	4	2	10

- Le medie aritmetiche si ottengono a partire dalle distribuzioni marginali di frequenze:

$$\mu_x = \frac{1}{n} \sum_{j=1}^k x_j n_{.j} = \frac{1}{10} \times (1 \times 4) + (3 \times 4) + (5 \times 2) = \frac{4 + 12 + 10}{10} = 2.6$$

$$\mu_y = \frac{1}{n} \sum_{i=1}^h y_i n_{i.} = \frac{1}{10} \times (2500 \times 5) + (7500 \times 5) = \frac{12500 + 37500}{10} = 5000$$

dove h rappresenta numero di righe della tabella, k il numero di colonne della tabella.



Esercizio regressione: distribuzione doppia di frequenze

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

Per calcolare le varianze si fa riferimento agli scarti dalla media al quadrato

Y/X	$(1 - 2.6)^2$	$(3 - 2.6)^2$	$(5 - 2.6)^2$	Tot
$(2500 - 5000)^2$	3	2	0	5
$(7500 - 5000)^2$	1	2	2	5
Tot	4	4	2	10

- Le varianze si ottengono a partire dalle distribuzioni marginali di frequenze:

$$\sigma_x^2 = \frac{1}{n} \sum_{j=1}^k (x_j - \mu_x)^2 n_{.j} = \frac{1}{10} \times ((1 - 2.6)^2 \times 4) + ((3 - 2.6)^2 \times 4) + ((5 - 2.6)^2 \times 2) = \frac{10.24 + 0.64 + 11.52}{10} = 2.24$$

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^h (y_i - \mu_y)^2 n_{i.} = \frac{1}{10} \times (2500 \times 5)^2 + (7500 \times 5)^2 = \frac{31250000 + 31250000}{10} = 6250000$$

dove h rappresenta numero di righe della tabella, k il numero di colonne della tabella.



Esercizio regressione: distribuzione doppia di frequenze

Per calcolare la covarianza si deve fare riferimento alle distribuzioni condizionate di frequenza.

Y/X	(1 - 2.6)	(3 - 2.6)	(5 - 2.6)	Tot
(2500 - 5000)	3	2	0	5
(7500 - 5000)	1	2	2	5
Tot	4	4	2	10

y_i	x_i	$y_i - \mu_y$	$x_i - \mu_x$
2500	1	(2500-5000)	(1-2.6)
2500	1	(2500-5000)	(1-2.6)
2500	1	(2500-5000)	(1-2.6)
2500	3	(2500-5000)	(3-2.6)
2500	3	(2500-5000)	(3-2.6)
7500	1	(7500-5000)	(1-2.6)
7500	3	(7500-5000)	(3-2.6)
7500	3	(7500-5000)	(3-2.6)
7500	5	(7500-5000)	(5-2.6)
7500	5	(7500-5000)	(5-2.6)

$$\begin{aligned} \sigma_{xy} &= \frac{1}{n} \sum_{i=1}^h \sum_{j=1}^k (y_i - \mu_y) \times (x_j - \mu_x) \times n_{ij} = \\ &= \frac{1}{10} ((2500 - 5000)(1 - 2.6) \times 3 + (2500 - 5000)(3 - 2.6) \times 2 + \\ &+ (7500 - 5000)(1 - 2.6) \times 1 + (7500 - 5000)(3 - 2.6) \times 2 + \\ &+ (7500 - 5000)(5 - 2.6) \times 2) = \frac{12000 - 2000 - 4000 + 2000 + 12000}{10} = 2000 \end{aligned}$$



Esercizio regressione: distribuzione doppia di frequenze

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

Avendo calcolato le quantità $\mu_x = 2.6$, $\mu_y = 5000$, $\sigma_x^2 = 2.24$ e $\sigma_{xy} = 2000$, è possibile calcolare i coefficienti della retta di regressione

Calcolo dei coefficienti

- $b_1 = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{2000}{2.24} = 892.571$
- $b_0 = \mu_y - b_1\mu_x = 5000 - (892.571 * 2.6) = 2679.315$

quindi l'equazione della retta di regressione è

$$y = b_0 + b_1x = 2679.315 + 892.571x$$

Dunque, il valore stimato \hat{y}_i corrispondente ad un valore x_i assegnato è $\hat{y}_i = b_0 + b_1x$.



Valutazione della bontà di adattamento

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

Ricordando che

$$R^2 = \frac{Dev_r}{Dev_y} = \frac{\sum_{i=1}^n (\hat{y}_i - \mu_y)^2}{\sum_{i=1}^n (y_i - \mu_y)^2}$$

ovvero

$$R^2 = 1 - \frac{Dev_e}{Dev_y} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \mu_y)^2}$$

con $Dev_y = Dev_r + Dev_e$

- $Dev_y = \sum_{i=1}^n (y_i - \mu_y)^2$ devianza totale
- $Dev_r = \sum_{i=1}^n (\hat{y}_i - \mu_y)^2$ devianza di regressione
- $Dev_e = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ devianza dei residui

Per ottenere R^2 , misura della bontà di adattamento, si deve calcolare solo la devianza dei residui, avendo già calcolato σ_y^2 .



Calcolo della devianza dei residui

Esercitazione 8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

- $Dev_e = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ devianza dei residui

in base alla retta di regressione stimata, i valori \hat{y}_i stimati in funzione dei valori x_i sono

- $\hat{y}_1 = b_0 + b_1 x_1 = 2679.315 + 892.571 \times 1 = 3571.886$

- $\hat{y}_2 = b_0 + b_1 x_2 = 2679.315 + 892.571 \times 3 = 5357.028$

- $\hat{y}_3 = b_0 + b_1 x_3 = 2679.315 + 892.571 \times 5 = 7142.17$



Calcolo della devianza dei residui

Esercitazione
8

A. Iodice

Studio della
dipendenza

La retta di
regressione

Qualità della
soluzione
trovata

Outliers

Regressione su
tabella a
doppia entrata

Per calcolare i residui $y_i - \hat{y}_i$ nel caso di tabella a doppia entrata si procede come segue

y_i/\hat{y}_j	$\hat{y}_1 = 3571.886$	$\hat{y}_2 = 5357.028$	$\hat{y}_3 = 7142.17$	Tot
$y_1 = 2500$	3	2	0	5
$y_2 = 7500$	1	2	2	5
Tot	4	4	2	10

$$\bullet \text{Dev}_e = \sum_{i=1}^h \sum_{j=1}^k ((y_i - \hat{y}_j)^2) \times n_{ij} \text{ devianza dei residui per tabella doppia}$$

calcolo della devianza dei residui

$$\begin{aligned} \text{Dev}_e &= \sum_{i=1}^h \sum_{j=1}^k ((y_i - \hat{y}_j)^2) \times n_{ij} = ((2500 - 3571.886)^2) \times 3 + ((2500 - 5357.028)^2) \times 2 + \\ &+ ((7500 - 3571.886)^2) \times 1 + ((7500 - 5357.028)^2) \times 2 + ((7500 - 7142.17)^2) \times 2 = \\ &= 44642859 \end{aligned}$$

$$\text{dev}_y = \sum_{i=1}^n (y_i - \mu_y)^2 = \sigma_y^2 \times n = 6250000 \times 10 = 62500000$$

$$R^2 = 1 - \frac{\text{dev}_e}{\text{dev}_y} = 1 - 0.71 = 0.29$$