

AUTOVALUTAZIONE di Regressione lineare e Lisrel (introduzione) - SOLUZIONI

N.B. Fate i calcoli con *almeno* 3 decimali e arrotondate i risultati (non truncate).

Svolgi gli esercizi usando i seguenti dati, che corrispondono all'esercizio 9 dell'Autovalutazione "Ripasso della statistica":

a1=A	a2=B	a3=C	a4=D	a5=E
1	1	4	1	3
1	2	2	1	5
1	2	3	2	1
4	1	1	2	4
2	2	5	4	1
2	1	2	1	4
2	5	1	4	2
5	1	3	3	5
5	4	3	2	3
4	5	4	2	3

1. Usando la formula algebrica di b e di a , calcola i parametri non standardizzati della regressione che spiega B a partire da A;

Per il calcolo di b usiamo la formula:

$$b = \frac{N \sum X_i Y_i - \sum X_i \sum Y_i}{N \sum X_i^2 - (\sum X_i)^2} = \frac{10 \times 70 - 27 \times 24}{10 \times 97 - (27)^2} = 0.216$$

$$a = \bar{Y} - b\bar{X} = \frac{\sum Y}{N} - b \frac{\sum X}{N} = \frac{24}{10} - 0.216 \frac{27}{10} = 1.817$$

2. Usando la formula matriciale ripeti i calcoli precedenti e verifica che i parametri trovati siano uguali;

In questo caso la formula da usare è $\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

$$\left(\begin{array}{c} \left[\begin{array}{cccccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 4 & 2 & 2 & 2 & 5 & 5 & 4 \end{array} \right] \left[\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{array} \right] \\ \left[\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{array} \right] \end{array} \right)^{-1} \left[\begin{array}{c} 1 \\ 2 \\ 2 \\ 1 \\ 2 \\ 2 \\ 5 \\ 5 \\ 4 \end{array} \right]$$

$$\begin{bmatrix} 10 & 27 \\ 27 & 97 \end{bmatrix}^{-1} \begin{bmatrix} 24 \\ 70 \end{bmatrix} = \frac{1}{241} \begin{bmatrix} 97 & -27 \\ -27 & 10 \end{bmatrix} \begin{bmatrix} 24 \\ 70 \end{bmatrix} = \begin{bmatrix} \frac{97 \times 24 - 27 \times 70}{241} \\ \frac{-27 \times 24 + 10 \times 70}{241} \end{bmatrix} = \begin{bmatrix} 1.817 \\ 0.216 \end{bmatrix}$$

3. **Standardizza i parametri precedenti;**

Applichiamo la formula:

$$b^* = b \frac{s_x}{s_y}$$

$$b_0^* = 0$$

$$b_1^* = 0.216 \frac{1.552}{1.562} = 0.215$$

4. **Verifica che coincidono con la correlazione fra A e B;**

$$r_{AB} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} = \frac{10 \times 70 - 27 \times 24}{\sqrt{[10 \times 97 - (27)^2][10 \times 82 - (24)^2]}}$$

$$= \frac{52}{\sqrt{241 \times 244}} = \frac{52}{242.495} = 0.214$$

Se ricalcoliamo b_1 , b_1^* e r_{AB} con 5 decimali, otteniamo 0.21439 e 0.21444, quindi le differenze sono solo dovute agli arrotondamenti.

5. **Calcola la proporzione di varianza spiegata;**

$$r^2 = .214^2 = 0.046$$

6. **Calcola la deviazione standard della stima;**

$$s_{y.x} = s_y \sqrt{1 - r^2} = 1.562 \sqrt{1 - 0.046} = 1.526$$

7. **Usando la formula $C_{xx}^{-1}c_{xy}$ calcola i parametri non standardizzati dell'equazione che spiega C a partire da A e da E;**

Per usare questa formula ci serve la matrice delle varianze/covarianze (che è stata calcolata in un'autovalutazione precedente).

	A	B	C	D	E
A	2,41				
B	0,52	2,44			
C	-0,06	-0,02	1,56		
D	0,36	0,62	0,14	1,16	
E	0,73	-0,74	-0,68	-0,72	1,89

Da questa dobbiamo ricavare C_{xx} e c_{yx}

$$C_{xx} = \begin{bmatrix} 2.41 & 0.73 \\ 0.73 & 1.89 \end{bmatrix} \quad c_{yx} = \begin{bmatrix} -0.06 \\ -0.68 \end{bmatrix}$$

Ora siamo pronti per i calcoli:

$$\begin{bmatrix} 2.41 & 0.73 \\ 0.73 & 1.89 \end{bmatrix}^{-1} \begin{bmatrix} -0.06 \\ -0.68 \end{bmatrix} =$$

$$\frac{1}{2.41 \times 1.89 - 0.73^2} \begin{bmatrix} 1.89 & -0.73 \\ -0.73 & 2.41 \end{bmatrix} \begin{bmatrix} -0.06 \\ -0.68 \end{bmatrix} =$$

$$\begin{bmatrix} \frac{1.89 \times (-0.06) + (-0.73) \times (-0.68)}{4.022} \\ \frac{(-0.73) \times (-0.06) + 2.41 \times (-0.68)}{4.022} \end{bmatrix} = \begin{bmatrix} \frac{0.383}{4.022} \\ \frac{-1.595}{4.022} \end{bmatrix} = \begin{bmatrix} 0.095 \\ -0.397 \end{bmatrix}$$

8. **Standardizza i parametri trovati;**

Applichiamo ancora la formula:

$$b^* = b \frac{s_x}{s_y}$$

$$b_1^* = 0.95 \frac{\sqrt{2.41}}{\sqrt{1.56}} = 0.118$$

$$b_2^* = -0.397 \frac{\sqrt{1.89}}{\sqrt{1.56}} = -0.437$$

9. **Usando la formula $R_{xx}^{-1}r_{xy}$ calcola i parametri standardizzati dell'equazione che spiega C a partire da A e da E;**

In questo caso dovremmo calcolare la matrice di correlazione a partire da quella delle varianze, usando ripetutamente la formula

$$r_{xy} = \frac{cov(xy)}{\sqrt{var(x)var(y)}}$$

Tuttavia non è necessario calcolarle tutte:

$$r_{ac} = \frac{-0.06}{\sqrt{2.41 \times 1.56}} = -0.031$$

$$r_{ae} = \frac{0.73}{\sqrt{2.41 \times 1.89}} = 0.342$$

$$r_{ce} = \frac{-0.68}{\sqrt{1.56 \times 1.89}} = -0.396$$

Ora siamo pronti per i calcoli:

$$\begin{bmatrix} 1 & 0.342 \\ 0.342 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -0.031 \\ -0.396 \end{bmatrix} = \frac{1}{0.883} \begin{bmatrix} 1 & -0.342 \\ -0.342 & 1 \end{bmatrix} \begin{bmatrix} -0.031 \\ -0.396 \end{bmatrix} = \begin{bmatrix} 0.118 \\ -0.436 \end{bmatrix}$$

10. **Verifica che corrispondano a quelli calcolati in precedenza (es. 7);**

Salvo il terzo decimale del secondo parametro, le stime sono uguali.

11. **Quale dei due parametri spiega meglio la variabile?**

Il secondo parametro (-.436).

12. Calcola l' R^2 dell'equazione $C = b_1A + b_2E + e$;

$$.118 \times (-.031) + (-.436) \times (-.396) = .169$$

13. Calcola la statistica di significatività dell'intera equazione;

Per calcolare la statistica F possiamo usare la formula:

$$F = \frac{R_f^2/k}{(1 - R_f^2)/(N - k - 1)} = \frac{.169/2}{(1 - .169)/(10 - 2 - 1)} = \frac{.085}{0.831/7} = \frac{.085}{0.119} = 0.714$$

14. Calcola statistica F per la verifica della significatività di b_1 ;

Ponendo $b_1 = 0$, l'equazione diventa $C = b_2E + e$ e quindi l' R^2 coincide con la correlazione al quadrato fra C ed E. Possiamo allora usare la formula (con $d_f = 10 - 3$ e $d_r = 10 - 2$):

$$F = \frac{(R_f^2 - R_r^2)/(d_r - d_f)}{(1 - R_f^2)/d_f} = \frac{(.169 - (-0.396)^2)/(8 - 7)}{(1 - .169)/7} = \frac{(.169 - .157)/(1)}{(.831)/7} = \frac{.012}{.119} = .101$$

15. Ripeti i calcoli per la verifica della significatività di b_2

In modo assolutamente analogo:

$$F = \frac{(.169 - (-0.031)^2)/(8 - 7)}{(1 - .169)/7} = \frac{(.169 - .001)/(1)}{(.831)/7} = \frac{.165}{.119} = 1.387$$

16. Scrivi il programma Lisrel relativo all'esercizio 1;

DA NI=5 NO=10 MA=CM

RA

1 1 4 1 3

1 2 2 1 5

1 2 3 2 1

4 1 1 2 4

2 2 5 4 1

2 1 2 1 4

2 5 1 4 2

5 1 3 3 5

5 4 3 2 3

4 5 4 2 3

LA; A B C D E

SE; B A /

MO NX=1 NY=1

PD; OU

oppure anche con

CM

2.41

0.52 2.44

-0.06 -0.02 1.56

0.36 0.62 0.14 1.16

0.73 -0.74 -0.68 -0.72 1.89

al posto di RA.

17. **Scrivi il programma Lisrel relativo all'esercizio 7;**

```
DA NI=5 NO=10 MA=CM
CM
  2.41
  0.52  2.44
-0.06 -0.02  1.56
  0.36  0.62  0.14  1.16
  0.73 -0.74 -0.68 -0.72  1.89
LA; A B C D E
SE; C A E /
MO NX=2 NY=1
PD; OU
```

18. **Scrivi il programma Lisrel relativo all'esercizio 9;**

Non è necessario calcolare la matrice di correlazione, perchè basta usa KM

```
DA NI=5 NO=10 MA=KM
CM
  2.41
  0.52  2.44
-0.06 -0.02  1.56
  0.36  0.62  0.14  1.16
  0.73 -0.74 -0.68 -0.72  1.89
LA; A B C D E
SE; C A E /
MO NX=2 NY=1
PD; OU
```

19. **Scrivi il programma Lisrel per verificare che la variabile D sia influenzata da tutte le altre, usando la matrice delle covarianze.**

```
DA NI=5 NO=10 MA=cM
CM
  2.41
  0.52  2.44
-0.06 -0.02  1.56
  0.36  0.62  0.14  1.16
  0.73 -0.74 -0.68 -0.72  1.89
LA; A B C D E
SE; D A B C E
MO NX=4 NY=1
PD; OU
```

20. **Usando il file di output, indica:**

- (a) **quali variabili sono statisticamente non significative;**
Tutte.

(b) qual è l' R^2 del modello;

0.42

(c) se il modello è adeguato ai dati osservati;

Il modello è saturo e quindi si adegua perfettamente.

(d) se ci sono legami da aggiungere e da quale inizieresti;

Non ce ne sono. Se vi fossero stati avrei iniziato con l'indice di modifica più grande.