CdL in Scienze Statistiche ed Economiche - Università degli Studi di Milano-Bicocca

Esercitazione : Varianza totale e generalizzata.

Esercitatrice: Chiara Gaia Magnani

Example 0.1. *Siano S e $\tilde{S}$ due matrici di varianze e covarianze definite come segue:*

$$S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

$$\tilde{S} = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix} \tag{2}$$

Si calcolino per S e $\tilde{S}$ la varianza totale e la varianza generalizzata e si confrontino i risultati.

Dimostrazione. La varianza totale e la varianza generalizzata sono definite rispettivamente come la traccia e
il determinante delle matrici di varianze e covarianze.
In particolare si ha:

$$tr(S) = tr(\tilde{S}) = 3$$

Osservando che le colonne della matrice $\tilde{S}$ sono linearmente dipendenti si ottiene

$$det(S) = 1 \qquad det(\tilde{S}) = 0$$

Se considerassimo la sola varianza totale potremmo pensare che i dati si disperdono in modo simile intorno al vettore delle medie. Tuttavia la varianza totale non dá alcuna informazione sulla relazione tra le variabili.
La varianza generalizzata, al contrario, tiene conto della correlazione tra le variabili assumendo valori piú piccoli nel caso in cui queste siano fortemente correlate tra loro.

□

Example 0.2. *Quando la varianza generalizzata é zero sono le colonne della matrice dei dati centrati ad essere linearmente dipendenti e non necessariamente le colonne della matrice dei dati. Si consideri la seguente matrice*

$$X = \begin{bmatrix} 3 & 1 & 0 \\ 6 & 4 & 6 \\ 4 & 2 & 2 \\ 7 & 0 & 3 \\ 5 & 3 & 4 \end{bmatrix} \tag{3}$$

1. Mostrare che le colonne di X risultano linearmente indipendenti.
2. Calcolare la matrice dei dati centrati $\tilde{X}$ e si verificare che le colonne sono linearmente dipendenti specificando il vettore a che stabilisce la dipendenza.
3. Si calcoli la matrice di varianze e covarianze S e la varianze generalizzata $det(S)$.

Dimostrazione. 1. Si vuole mostrare che $ax_{i1} + bx_{i2} + cx_{i3} = 0$ per ogni $i = 1, \ldots, n$ se e solo se $a = b = c = 0$

Se si considera il sistema :

$$\begin{cases} 3a + b = 0 \\ 6a + 4b + 6c = 0 \\ 4a + 2b + 2c = 0 \\ 7a + 3c = 0 \\ 5a + 3b + 4c = 0 \end{cases} \tag{4}$$

Si osserva che $b = -3a$ e $c = -\frac{7}{3}a$. Sostituendo si ottiene che $a = 0$ e quindi $b = 0$ e $c = 0$.

2. Si ha

$$\tilde{X} = \begin{bmatrix} 3-5 & 1-2 & 0-3 \\ 6-5 & 4-2 & 6-3 \\ 4-5 & 2-2 & 2-3 \\ 7-5 & 0-2 & 3-3 \\ 5-5 & 3-2 & 4-3 \end{bmatrix} = \begin{bmatrix} -2 & -1 & -3 \\ 1 & 2 & 3 \\ -1 & 0 & -1 \\ 2 & -2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \tag{5}$$

E' facile verificare che $\tilde{x}_{i1} + \tilde{x}_{i2} - \tilde{x}_{i3} = 0$ per ogni i e quindi $a' = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}$.

3. Si ottiene che la matrice di varianze e covarianze S ha la seguente forma:

$$S = \begin{pmatrix} 2 & 0 & 2 \\ 0 & 2 & 2 \\ 2 & 2 & 4 \end{pmatrix} \tag{6}$$

Calcolando il determinante della matrice sopra riportata si ottiene $\det(S) = 0$.

□

Example 0.3. *Si individui l'ellisse* $(x - \bar{x})S^{-1}(x - \bar{x}) \leq 1$ *definita dalle tre matrici*

$$S_1 = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix} \quad S_2 = \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix} \quad S_3 = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \tag{7}$$

Dimostrazione. L'equazione $(x - \bar{x})S^{-1}(x - \bar{x}) = 1$ definisce un'ellisse centrata in $\bar{x}$ e il cui j-simo asse é orientato secondo il j-simo autovettore di S e ha una lunghezza proporzionale al j-simo autovalore di S.

Notiamo che la varianza generalizzata S assume lo stesso valore, i.e. $det(S) = 9$, in tutti e tre i casi ovvero che la nuvola di punti 2-dimensionale formata dalle unitá statistiche ricopre la medesima area. Tuttavia la varianza generalizzata non fornisce informazioni sull'orientamento dei punti.

Gli autovalori di S_1 soddisfano l'equazione caratteristica

$$(\lambda - 5)^2 - 16 = (\lambda - 9)(\lambda - 1) = 0$$

dalla quale si ottengono le seguenti coppie di autovalori e autovettori: $\lambda_1 = 9$ e $v_1' = [1/\sqrt{2}, 1/\sqrt{2}]$ e $\lambda_2 = 1$ e $v_2' = [1/\sqrt{2}, -1/\sqrt{2}]$.
Gli autovalori di S_2 soddisfano la medesima equazione caratteristica e le coppie di autovalori e autovettori associate risultano: $\lambda_1 = 9$ e $v_1' = [1/\sqrt{2}, -1/\sqrt{2}]$ e $\lambda_2 = 1$ e $v_2' = [1/\sqrt{2}, 1/\sqrt{2}]$.

Infine gli autovalori di S_3 soddisfano $(\lambda - 3)^2 = 0$ e si scelgono $\lambda_1 = 3$ e $v_1' = [1, 0]$ e $\lambda_2 = 3$ e $v_2' = [0, 1]$.

I risultati ottenuti,relativi alla direzione e lunghezza degli assi delle ellissi, indicano la presenza di diverse strutture di correlazione che la sola varianza generalizzata non é in grado di catturare. $\square$