Contents

List of Figures

1 avvertenza

Gli appunti di questa sezione relativi alla normale bivariata e, in particolare, quelli sulla normale multivariata, sono molto abbondanti rispetto a quanto effettivamente riportato sulla scheda trasparenza e a quanto fatto in aula (*e a quanto sarà utile ai fini dell'esame*).

Tuttavia, almeno ad oggi , ho preferito lasciare tutto il materiale, anche quello non finalizzato al nostro corso. Chiariremo in aula, e in una successiva redazione di questi appunti, le parti essenziali.

Aggiungo con sincerità che ogni anno mi chiedo se è il caso di eliminare del tutto questa parte sulla normale bivariata e multivariata da questo corso e ogni anno, fino ad ora, mi sono sempre risposto di no!

2 La distribuzione normale multivariata

La distribuzione normale multivariata può essere introdotta in numerosi modi, ed espressa con diverse caratterizzazioni.

Qui viene presentata come la distribuzione congiunta di *combinazioni lineari di variabili normali indipendenti e standardizzate*.

Uso adesso una **notazione matriciale compatta** perché non sarebbe conveniente lavorare con una distribuzione normale trivariata, poi con una normale quattro variabili, etc. etc.; alcuni risultati della normale trivariata saranno poi ricavati come casi particolari per qualche esercizio.

Sia $\mathbf{X}$ un vettore di variabili casuali a p componenti indipendenti:

$$\mathbf{X} = \{X_1, X_2, \ldots, X_i, \ldots, X_p\}^{\mathsf{T}}$$

ciascuna distribuita secondo una normale standardizzata.

La densità di tale distribuzione, data l'indipendenza, è data da:

Densità congiunta di p variabili normali standardizzate e indipendenti.

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \prod_{i=1}^{p} f(x_i) = (2\pi)^{-p/2}\exp\left[-\sum_{i=1}^{p}\frac{x_i^2}{2}\right] = \\
&= \frac{1}{\sqrt{(2\pi)^p}}\exp[-\frac{\mathbf{x}^{\mathsf{T}}\mathbf{x}}{2}]
\end{aligned}
\tag{1}
$$

La funzione caratteristica è:

$$
\phi_{\mathbf{X}}(\mathbf{t}) = \exp\left(-\frac{1}{2}\mathbf{t}^{\mathsf{T}}\mathbf{t}\right)
$$

Ovviamente i primi due momenti di $\mathbf{X}$, per le ipotesi fatte, sono:

$$
\begin{aligned}
\mathrm{E}[\mathbf{X}] &= \mathbf{0}_p \quad \text{vettore nullo di } p \text{ componenti} \\
\mathrm{V}[\mathbf{X}] &= \mathbf{I}_p \quad \text{matrice identità } p \times p
\end{aligned}
$$

E' noto, ed è facile comunque vederlo attraverso la funzione caratteristica, che una singola combinazione lineare Z del vettore aleatorio $\mathbf{X}$ si distribuisce secondo una normale univariata, con speranza matematica e varianza ricavabili dalle relazioni per i momenti di combinazioni lineari di vettori aleatori qualsiasi.

Infatti se: $Z = \mathbf{b}^{\mathsf{T}}\mathbf{X} + c$, allora i primi due momenti di Z sono dati da:

$$\begin{aligned} \mathrm{E}[Z] &= \mathbf{b}^{\mathsf{T}}\mathrm{E}[\mathbf{X}] + c = c \\ \mathrm{V}[Z] &= \mathbf{b}^{\mathsf{T}}\mathrm{V}[\mathbf{X}]\,\mathbf{b} = \mathbf{b}^{\mathsf{T}}\mathbf{b} = b_1^2 + b_2^2 + \ldots + b_i^2 + \ldots + b_p^2 \end{aligned}$$

e si ha anche:

$$Z \sim N(\mathrm{E}[Z], \mathrm{V}[Z]).$$

Adesso occorre però studiare la distribuzione congiunta di p combinazioni lineari di variabili normali indipendenti.

---

4 Densità della distribuzione congiunta di p combinazioni lineari di p variabili normali indipendenti

Consideriamo allora il vettore aleatorio $\mathbf{Y}$, trasformazione lineare del vettore aleatorio $\mathbf{X}$, definito dalla relazione:

$$\mathbf{Y} = \mathbf{A}^{\mathsf{T}}\mathbf{X} + \boldsymbol{\mu}$$

essendo:

$\mathbf{A}$ una matrice quadrata di dimensione p e rango pieno;

$\boldsymbol{\mu}$ un vettore di p elementi;

dimostrazione saltata nella versione base

4.1 Densità della distribuzione normale multivariata

In conclusione, si ottiene la densità di $\mathbf{y}$:

abbiamo:

Densità della distribuzione normale multivariata *non singolare* di parametri $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$

$$f_{\mathbf{Y}}(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}}(2\pi)^{-p/2}\exp\left\{-\frac{1}{2}[\mathbf{y}-\boldsymbol{\mu}]^{\intercal}\boldsymbol{\Sigma}^{-1}[\mathbf{y}-\boldsymbol{\mu}]\right\} \tag{2}$$

o anche:

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\sqrt{|\boldsymbol{\Sigma}|\,(2\pi)^{p}}}e^{-\frac{1}{2}[\mathbf{y}-\boldsymbol{\mu}]^{\intercal}\boldsymbol{\Sigma}^{-1}[\mathbf{y}-\boldsymbol{\mu}]}$$

Dalle espressioni della densità riportate sopra, è evidente l'analogia con l'espressione della densità della distribuzione normale univariata.

I primi due momenti multivariati sono (come già visto prima e senza alcun bisogno di effettuare integrazioni p-dimensionali):

$$\mathrm{E}\left[\mathbf{Y}\right] = \boldsymbol{\mu}$$
$$\mathrm{V}\left[\mathbf{Y}\right] = \boldsymbol{\Sigma}$$

Si vede quindi, come nel caso univariato, che la distribuzione normale multivariata dipende soltanto dai primi due momenti (multivariati) di $\mathbf{Y}$.

La funzione caratteristica (applicando la regola per le trasformazioni lineari di variabili aleatorie) è data da:

$$\phi_{\mathbf{Y}}(\mathbf{t}) = \exp[i\mathbf{t}^{\top}\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^{\top}\boldsymbol{\Sigma}\mathbf{t}] \tag{3}$$

Ricordo che i momenti possono essere eventualmente ricavati dalle opportune derivate di $\phi_{\mathbf{Y}}(\mathbf{t})$, valutate in $\mathbf{t} = \mathbf{0}$.

Sottolineo ancora che se A non ha rango p, la distribuzione di Y è sempre una normale multivariata, ma non è dotata di densità . Si intuisce facilmente questa proprietà, senza darne una dimostrazione, perchè in effetti alla funzione caratteristica (3) si arriva anche senza alcuna restrizione sul rango di $\mathbf{A}$.

Inoltre è possibile far vedere, rifacendo a ritroso i passaggi precedenti, che, se $\mathbf{C}$ è una matrice simmetrica definita positiva di rango p, qualsiasi vettore aleatorio $\mathbf{Y}$ la cui densità è data da:

$$f_{\mathbf{Y}}(\mathbf{y}) = \|\mathbf{C}|^{\frac{1}{2}}(2\pi)^{-p/2} \exp\left(-\frac{1}{2}[\mathbf{y}-\boldsymbol{\mu}]^{\mathsf{T}}\mathbf{C}[\mathbf{y}-\boldsymbol{\mu}]\right) \tag{4}$$

è distribuito secondo una normale multivariata di parametri $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma} = \mathbf{C}^{-1}$. Esiste inoltre una trasformazione lineare di $\mathbf{Y}$ che conduce ad un vettore aleatorio $\mathbf{X}$ a componenti standardizzate e indipendenti:

$$\mathbf{X} = \mathbf{B}^{\mathsf{T}}[\mathbf{Y}-\boldsymbol{\mu}], \quad \text{in cui } \mathbf{B} \text{ è tale che:} \quad \mathbf{B}^{\mathsf{T}}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{I}$$

La distribuzione marginale di un qualsiasi sottoinsieme di componenti di un vettore aleatorio distribuito secondo una normale multivariata è ancora distribuito secondo una normale multivariata con parametri uguali ai corrispondenti sottoinsiemi di $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$: il risultato si dimostra facilmente, sfruttando la funzione caratteristica.[1].

Infatti se il vettore $\mathbf{Y}$ è suddiviso in due sottovettori $[\mathbf{Y}_A, \mathbf{Y}_B]$, corrispondentemente suddividiamo il vettore delle medie e la matrice di varianza e covarianza:

$$\boldsymbol{\mu} = \left(\begin{array}{c} \boldsymbol{\mu}_A \\ \hline \boldsymbol{\mu}_B \end{array}\right)$$

$$\boldsymbol{\Sigma} = \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \hline \boldsymbol{\Sigma}_{AB}^{\mathsf{T}} & \boldsymbol{\Sigma}_{BB} \end{array}\right)$$

Posta ora, corrispondentemente alla partizione di $\mathbf{Y}$, una partizione $\mathbf{t} = \{\mathbf{t}_A, \mathbf{t}_B\}$, come si sa la funzione caratteristica della distribuzione marginale di $\mathbf{Y}_1$ si ottiene da quella di $\mathbf{Y}$ ponendo $\mathbf{t}_B = \mathbf{0}$:

$$\phi_{\mathbf{Y}_A}(\mathbf{t}_A) = \phi_{\mathbf{Y}}(\mathbf{t}_A, \mathbf{0}) = \exp[i\mathbf{t}_A^{\mathsf{T}}\boldsymbol{\mu}_A - \frac{1}{2}\mathbf{t}_A^{\mathsf{T}}\boldsymbol{\Sigma}_{AA}\mathbf{t}_A]$$

[1]ottenere lo stesso risultato integrando la densità (2) rispetto alle componenti di mY_2 sarebbe molto più complicato

che è la funzione caratteristica di una normale di parametri μ_A e $\mathbf{\Sigma}_{AA}$.

In particolare tutte le distribuzioni marginali delle singole componenti sono normali univariate.

Come corollario è facile vedere che $\mathbf{Y}_A$ e $\mathbf{Y}_B$ (vettori aleatori normali) sono indipendenti se e solo se $\mathbf{\Sigma}_{AB} = \mathbf{0}$.

E' evidente che l'indipendenza fra tutte le componenti di $\mathbf{Y}$ si può avere solo quando la $f_{\mathbf{Y}}(\mathbf{y})$ è fattorizzabile nelle rispettive densità marginali, il che può avvenire se (*e solo se*) $\mathbf{\Sigma}$ è diagonale, ossia con covarianze nulle, e quindi correlazioni lineari semplici nulle, il che porta un'altra fondamentale proprietà della normale multivariata:

Indipendenza e assenza di correlazione nella normale multivariata

Assenza di correlazione $\Longleftrightarrow$ Indipendenza (nella normale multivariata)

Un vettore aleatorio $\mathbf{Y}$ con distribuzione normale multivariata, è a componenti indipendenti se *(e solo se)* le correlazioni lineari fra le sue componenti prese a due a due sono nulle, ossia se la matrice di varianza e covarianza è diagonale.
Quindi, se due variabili sono congiuntamente normali, l'assenza di correlazione implica l'indipendenza stocastica.

Indipendenza a coppie e a gruppi nella normale multivariata

Nella normale Multivariata vale una proprietà ancora più forte:

L'indipendenza a coppie implica l'indipendenza a gruppi di variabili

5 Distribuzione di combinazioni lineari di variabili normali qualsiasi.

Qualsiasi combinazione lineare (con un numero qualsiasi di componenti) di un vettore aleatorio distribuito secondo una qualsiasi normale multivariata si distribuisce ancora secondo una distribuzione normale multivariata:

Distribuzione congiunta di combinazioni lineari di normali multivariate: $\mathbf{Y} = \mathbf{A}\mathbf{Z}$

Se $\mathbf{Y} = \mathbf{A}^{\mathsf{T}}\mathbf{Z}$, con $\mathbf{A}$ $(p \times k)$, e $\mathbf{Z}(\mathcal{N}_p(\boldsymbol{\mu}_{\mathbf{Z}}, \boldsymbol{\Sigma}_{\mathbf{Z}})$ allora:

$$\mathbf{Y} \sim \mathcal{N}_k(\mathbf{A}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{Z}}, \mathbf{A}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathbf{Z}}\mathbf{A})$$

6 Caratterizzazione della distribuzione normale multivariata.

Le proprietà viste prima sulla distribuzione congiunta di combinazioni lineari di variabili normali costituiscono una caratterizzazione della distribuzione normale multivariata.

Infatti ricordo una importante proprietà che caratterizza la distribuzione normale multivariata (di cui non fornisco la dimostrazione) (Mardia, 1970): citazione

Caratterizzazione della distribuzione normale multivariata.

$\mathbf{X}$, vettore aleatorio a p componenti, è distribuito secondo una normale multivariata *se e solo se* $\mathbf{b}^\top\mathbf{X}$ è distribuito secondo una normale (univariata) per qualsiasi vettore $\mathbf{b}$ di p componenti.

7 Assi principali degli ellissoidi di equiprobabilità

E' immediato vedere che le curve con densità $f(\mathbf{y})$ costante per la normale multivariata di parametri $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ sono, in uno spazio p -dimensionale, degli ellissoidi di centro in $\boldsymbol{\mu}$; infatti le curve a densità costante hanno equazione:

$$\|\boldsymbol{\Sigma}\|^{-\frac{1}{2}}(2\pi)^{-p/2}\exp(-[\mathbf{y}-\boldsymbol{\mu}]^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}[\mathbf{y}-\boldsymbol{\mu}]/2) = k_0$$

essendo k_0 una costante positiva $(0 < k_0 \leq |\boldsymbol{\Sigma}\|^{-\frac{1}{2}}(2\pi)^{-p/2})$, e quindi:

$$[\mathbf{y}-\boldsymbol{\mu}]^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}[\mathbf{y}-\boldsymbol{\mu}] = k_1$$

che è palesemente l'equazione di un ellissoide di centro in $\boldsymbol{\mu}$ e in cui k_1 è un termine costante (rispetto a $\mathbf{x}$), dato da:

$$k_1 = 2\log\left(\|\boldsymbol{\Sigma}\|^{-\frac{1}{2}}(2\pi)^{-p/2}\right) - 2\log k_0$$

- E' facile vedere quindi che al variare del livello costante k_1 , cambia solo il volume dell'ellissoide, ma le proporzioni fra gli assi e le coordinate del centro dell'ellissoide restano inalterate;
- le equazioni degli assi principali di tali ellissoidi sono date dagli autovettori di $\boldsymbol{\Sigma}$, che forniscono l'orientamento degli assi.

- i quadrati delle lunghezze degli assi principali di tali ellissoidi sono proporzionali agli autovalori di $\mathbf{\Sigma}$.
- Se $\mathbf{\Sigma}$ è diagonale, gli ellissoidi hanno assi paralleli agli assi coordinati e lunghezza proporzionale agli scarti quadratici medi delle singole componenti.
- Si può fare vedere che gli autovettori danno le direzioni degli assi principali impostando ancora un problema di massimo, ossia cercando i due punti sulla superficie dell'ellissoide che hanno distanza massima.

Ellissoidi e Normali multivariate

Fissato un qualsiasi valore positivo di k_1, esiste una corrispondenza biunivoca fra ellissoidi in $\Re_p$ e distribuzioni normali multivariate non singolari.

8 Distribuzioni condizionate nella normale multivariata

Una proprietà fondamentale della normale, che oltretutto la caratterizza, riguarda le distribuzioni di un gruppo di componenti condizionatamente ai valori di un altro gruppo di componenti. Questo argomento viene trattato adesso, esponendo i risultati fondamentali per tre ordini di ragioni:

1. La peculiarità delle caratteristiche delle distribuzioni condizionate nella normale multivariata, che ne rappresentano un aspetto fondamentale;
2. Come premessa indispensabile ai modelli lineari, che costituiscono il punto centrale di questo testo;
3. La possibilità di dare un significato statistico autonomo agli elementi dell'inversa della matrice di correlazione di una variabile multipla normale;

Si è già visto molto semplicemente che nella normale bivariata la distribuzione condizionata di una variabile è normale, con varianza costante e speranza matematica che varia linearmente in funzione dei valori assunti dalla variabile condizionante.

Prevedibilmente, il risultato per la normale multivariata generalizza quello della normale bivariata, fatte le opportune modifiche per il passaggio ad una situazione multivariata; come si vedrà nelle pagine successive, la distribuzione di un gruppo di variabili $\mathbf{Y_A}$ condizionata ad un particolare valore $\mathbf{y_B}$ assunto da un altro gruppo di variabili $\mathbf{Y_B}$ è:

1. ancora normale multivariata;
2. La distribuzione ha una matrice di varianze e covarianze costante rispetto a $\mathbf{y_B}$, ossia che non dipende dai valori della componente condizionante (omoscedasticità);
3. La funzione di regressione di una componente $\mathbf{y}_A$ rispetto alle altre componenti è lineare;

I risultati esposti in queste pagine generalizzano le proprietà note per distribuzioni normali bivariate, viste nella (**??**).

8.1 Distribuzione condizionata nel caso generale di un gruppo di componenti rispetto ad un altro gruppo di componenti.

Supponiamo di avere un vettore $\mathbf{Y}$ di p componenti, con distribuzione normale multivariata, suddiviso nel caso più generale in due sottovettori $[\mathbf{Y_A}, \mathbf{Y_B}]$, con corrispondente suddivisione del vettore delle medie e della matrice di varianze e covarianze:

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_A \\ \mathbf{Y}_B \end{pmatrix} \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{AB}^{\mathsf{T}} & \boldsymbol{\Sigma}_{BB} \end{pmatrix}$$

I due insiemi di indici A e B costituiscono una partizione dell'insieme di indici $I = 1, 2, \ldots, p$ così che:

$$A \cup B = I \qquad A \cap B = \emptyset \qquad A \neq \emptyset \qquad B \neq \emptyset$$

per il resto A e B sono costituiti da sottoinsiemi di indici qualsiasi (con la restrizione che esistano le inverse delle matrici di varianze e covarianze che si richiederanno nel seguito).

In effetti i casi più rilevanti, che tratteremo specificatamente, sono quelli in cui:

- $A = \{i\}$, **per lo studio della distribuzione della variabile i-esima condizionatamente alle altre $p-1$;**
- $A = \{i, j\}$, **per lo studio della distribuzione condizionata di due variabili (la i-esima e la j-esima), condizionatamente alle altre $p-2$,** in particolare per lo studio della indipendenza condizionata.

Ci chiediamo qual è la funzione di regressione di $\mathbf{Y}_A$ su $\mathbf{Y}_B$, ossia la speranza matematica di $\mathbf{Y}_A$ condizionata ad un particolare valore $\mathbf{y}_B$ di $\mathbf{Y}_B$: in simboli $\mathrm{E}\left[\mathbf{Y}_A \| \mathbf{Y}_B = \mathbf{y}_B\right]$.

Più in generale ci chiediamo direttamente qual è la distribuzione di $\mathbf{Y}_A$ condizionata ad un particolare valore $\mathbf{y}_B$ di $\mathbf{Y}_B$.

dimostrazione saltata

Distribuzioni condizionate nel caso generale di vettori aleatori normali ($\mathbf{Y}_A$ condiz. a $\mathbf{Y}_B = \mathbf{y}_B$)

$$\mathbf{Y}_A \| \{\mathbf{Y}_B = \mathbf{y}_B\} \sim \mathcal{N}_{n_A}\left[\boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}(\mathbf{y}_B - \boldsymbol{\mu}_B); \boldsymbol{\Sigma}_{AA.B}\right]$$

1. La distribuzione condizionata è normale multivariata con parametri:

$$\mathrm{E}\left[\mathbf{Y}_A \| \{\mathbf{Y}_B = \mathbf{y}_B\}\right] = \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}(\mathbf{y}_B - \boldsymbol{\mu}_B); \quad (5)$$

$$\mathrm{V}\left[\mathbf{Y}_A \| \{\mathbf{Y}_B = \mathbf{y}_B\}\right] = \boldsymbol{\Sigma}_{AA.B}. \quad (6)$$

2. la funzione di regressione (speranza matematica condizionata) è lineare in $\mathbf{y}_B$;

3. la matrice di varianze e covarianze condizionate non dipende da $\mathbf{y}_B$ (omoscedasticità) ed è espressa da:

$$\mathrm{V}\left[\mathbf{Y}_A \| \{\mathbf{Y}_B = \mathbf{y}_B\}\right] = \boldsymbol{\Sigma}_{AA.B} = \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{AB}^{\mathsf{T}} = = (\boldsymbol{\Sigma}^{AA})^{-1}$$

$\boldsymbol{\Sigma}^{AA}$ è il blocco di $\boldsymbol{\Sigma}^{-1}$ corrispondente agli indici AA

I vettori aleatori:

$$\mathbf{Y}_A - \boldsymbol{\mu}_A - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}[\mathbf{Y}_B - \boldsymbol{\mu}_B] \qquad \text{e} \qquad \mathbf{Y}_B$$

(oppure $\mathbf{Y}_A - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\mathbf{Y}_B$ e $\mathbf{Y}_B$)
risultano indipendenti (si verifica subito calcolando $E(\mathbf{Y}_A\mathbf{Y}_B^{\top})$)

Esempio 8.1 *Esempio numerico: Si consideri la matrice 3×3 di varianza e covarianza relativa ad una distribuzione normale multivariata a tre componenti:*

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad \textit{e quindi correlazioni:}$$

$$\mathbf{R} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 1 \end{pmatrix} = \begin{pmatrix} 1.00 & 0.50 & 0.71 \\ 0.50 & 1.00 & 0.71 \\ 0.71 & 0.71 & 1.00 \end{pmatrix}$$

Consideriamo la distribuzione della variabile 1 condizionatamente alla 2 e alla 3. La matrice di varianze e covarianze va quindi partizionata nel seguente *modo:*

$$\mathbf{\Sigma} = \left(\begin{array}{c|cc} 2 & 1 & 1 \\ \hline 1 & 2 & 1 \\ 1 & 1 & 1 \end{array}\right).$$

Con riferimento alla notazione fin qui adottata l'insieme A è costituito dal solo indice 1 e l'insieme B dagli altri due indici (2 e 3); per cui si ha:

$$A = 1 \qquad B = 2,3 \qquad \textit{e quindi:}$$

$$\mathbf{\Sigma}_{BB} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \qquad e \qquad \mathbf{\Sigma}^{-1} = \left(\begin{array}{r|rr} 1 & 0 & -1 \\ \hline 0 & 1 & -1 \\ -1 & -1 & 3 \end{array}\right)$$

$$\mathbf{\Sigma}^{AA} = 1$$

8.2 Significato degli elementi dell'inversa della matrice di varianze e covarianze.

E' possibile dare anche un significato agli elementi dell'inversa di $\mathbf{\Sigma}$, in termini di distribuzioni condizionate, essendo al solito $\mathbf{\Sigma}$ la matrice di varianze e covarianze di una distribuzione normale multivariata; si vedrà in altri capitoli se e in che misura tali concetti possano essere estesi al caso di variabili aleatorie non normali o, anche, nell'analisi di dati multivariati, al caso di variabili statistiche osservate. `link con correlazione parziale per variabili statistiche`

8.2.1 Gli elementi diagonali dell'inversa: la correlazione multipla

Anche gli elementi sulla diagonale principale di $\mathbf{\Sigma}^{-1}$ sono interpretabili tenendo conto delle distribuzioni condizionate, ma in termini di variabilità di una variabile spiegata da tutte le altre, concetto che rivedremo poi nel caso di modelli lineari generali.

Infatti consideriamo ora l'insieme $\mathbf{Y}_A$ costituito da una sola variabile Y_i; nella notazione finora adottata A è uguale all'indice i e B all'insieme degli altri $p-1$ indici, ossia:

$$A = \{i\} \qquad B = \{1, 2, \ldots, i-1, i+1, \ldots, p\}$$

possiamo calcolare la varianza di $\mathbf{y}_i$ condizionata ai valori delle altre $p-1$ variabili, applicando la 6:

$$\mathrm{V}\left[\mathbf{Y}_A \| \{\mathbf{Y}_B = \mathbf{y}_B\}\right] = \mathrm{V}\left[\mathbf{Y}_i \| \{\mathbf{Y}_B = \mathbf{y}_B\}\right] = \left(\mathbf{\Sigma}^{i\,i}\right)^{-1}$$

Tenendo conto che in questo caso $\mathbf{\Sigma}^{AA} = \mathbf{\Sigma}^{i\,i} = c_{ii}$ si ha:

Varianza della distribuzione condizionata di una componente

$$\mathrm{V}\left[\mathbf{y}_i \| \mathbf{y}_B\right] = (\mathbf{\Sigma}^{AA})^{-1} = \frac{1}{c_{ii}} = \frac{|\mathbf{\Sigma}|}{\sigma^{ii}}$$

Quindi l'inverso di un elemento diagonale dell'inversa della matrice di varianze e covarianze esprime la varianza della variabile di posto corrispondente condizionatamente alle altre $p-1$ variabili.

$$\max\left(\frac{1}{c_{ii}}\right) = \sigma_i^2 \qquad \min(c_{ii}) = \frac{1}{\sigma_i^2}$$

Il massimo di questa quantità è proprio la varianza della componente i -esima, ossia σ_i^2

Se $\mathbf{Z}$ è una matrice di correlazione, allora $1/c_{ii}$ indica la variabilità di Y_i non spiegata dalle altre $p-1$ variabili, per cui si può costruire il coefficiente di determinazione multipla:

$$R^2_{i.B} = 1 - \frac{|\mathbf{Z}|}{z^{ii}} = 1 - \frac{\mathrm{V}\left[\mathbf{y}_i\right] \|\mathbf{Y}_B)}{\mathrm{V}\left[\mathbf{y}_i\right]}$$

Misura quanta parte della variabilità di $\mathbf{Y}_i$ è spiegata dalle altre p-1 variabili del vettore aleatorio $\mathbf{y_B}$ chiarire

In generale l'indice di correlazione lineare multipla è dato da:

$$R_{i.B} = \sqrt{1 - \frac{|\mathbf{\Sigma}|}{\sigma_i^2 \sigma^{ii}}} = \sqrt{1 - \frac{1}{\sigma_i^2 c_{ii}}}$$

codice R

8.2.2 Gli elementi non diagonali dell'inversa: la correlazione parziale

Intanto, con riferimento ad una distribuzione normale multivariata con matrice di varianze e covarianze $\mathbf{\Sigma}$, si può dimostrare che, indicata per brevità con $\mathbf{C}$ la sua inversa, $\mathbf{C} = \mathbf{\Sigma}^{-1}$, allora:

$c_{ij} = 0$ è condizione necessaria e sufficiente perché le variabili $\mathbf{Y}_i$ e $\mathbf{Y}_j$ siano indipendenti condizionatamente alle altre $p - 2$ variabili $\mathbf{Y}_B$.

Si può giungere al risultato in due modi:

1) Dalla densità normale multivariata si vede direttamente che:
se e solo se $c_{ij} = 0$ si ha la fattorizzazione:

$$f(\mathbf{y}) = f(\mathbf{y}_i, \mathbf{y}_B) f(\mathbf{y}_j, \mathbf{y}_B)$$

che è una condizione necessaria e sufficiente per l'indipendenza condizionata di due variabili aleatorie qualsiasi dotate di densità. citazione

Infatti, ponendo $\mathbf{Y}_A = (\mathbf{y}_i, \mathbf{y}_j)^T$ e indicando con $\mathbf{Y}_B$ tutte le altre componenti, avendo indicato con $\mathbf{C}$ l'inversa della matrice di varianze e covarianze opportunamente partizionata:

$$\mathbf{C} = \left(\begin{array}{cc|c} c_{ii} & c_{ij} & \mathbf{c}_{iB}^{\mathsf{T}} \\ c_{ij} & c_{jj} & \mathbf{c}_{jB}^{\mathsf{T}} \\ \hline \mathbf{c}_{iB} & \mathbf{c}_{jB} & \mathbf{C}_{BB} \end{array}\right)$$

(per comodità di lettura ho portato gli indici (i, j) in prima e seconda posizione); si ha allora:

$$\begin{aligned} f(\mathbf{y}) &= f(\mathbf{y}_i, \mathbf{y}_j, \mathbf{y}_B) = K \exp[-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\mathbf{C}\mathbf{y}] = \\ &= K \exp[-\frac{1}{2}(c_{ii}\mathbf{y}_i^2 + c_{jj}\mathbf{y}_j^2 + 2c_{ij}\mathbf{y}_i\mathbf{y}_j + 2y_i\mathbf{c}_{iB}^{\mathsf{T}}\mathbf{y}_B + 2y_j\mathbf{c}_{jB}^{\mathsf{T}}\mathbf{y}_B + \mathbf{y}_B^T\mathbf{C}_{BB}\mathbf{y}_B)] \end{aligned}$$

Se ora $c_{ij} = 0$ allora si può facilmente operare su $f(\mathbf{y})$:

$$\begin{aligned} f(\mathbf{y}) &= K \exp[-(c_{ii}\mathbf{y}_i^2 + c_{jj}\mathbf{y}_j^2 + 2y_i\mathbf{c}_{iB}^{\mathsf{T}}\mathbf{y}_B + 2y_j vecc_{jB}^{\mathsf{T}}\mathbf{y}_B + \mathbf{y}_B^T\mathbf{C}_{BB}\mathbf{y}_B)/2] = \\ &= K \underbrace{\exp[-\frac{1}{2}(c_{ii}\mathbf{y}_i^2 + 2y_i\mathbf{c}_{iB}^{\mathsf{T}}\mathbf{y}_B + \mathbf{y}_B^{\mathsf{T}}\mathbf{C}_{BB}\mathbf{y}_B)] \times \exp[-(c_{jj}\mathbf{y}_j^2 + 2y_j vecc_{jB}^{\mathsf{T}}\mathbf{y}_B)/2]}_{g(\mathbf{y}_i,\mathbf{y}_B)\times g(\mathbf{y}_j,\mathbf{y}_B)} \end{aligned}$$

in modo da ottenere la fattorizzazione desiderata in due funzioni, in cui non compaiono simultaneamente termini in $\mathbf{y}_i$ e $\mathbf{y}_j$

Per una interpretazione in generale del significato dei termini dell'inversa, e non solo per il caso estremo $c_{ij} = 0$, *conviene riferirsi alle distribuzioni condizionate.*

Dalla distribuzione di $\mathbf{Y}_A$ condizionata a $\mathbf{Y}_B = \mathbf{y}_B$, ponendo $\mathbf{Y}_A = (Y_i, Y_j)^{\mathsf{T}}$ (e quindi nella notazione della sezione precedente A è uguale alla coppia di indici i, j e B all'insieme degli altri $p - 2$ indici) si ricava che essendo la distribuzione condizionata di $\mathbf{Y}_A$ ancora normale, l'indipendenza condizionata si ha se e solo se $\mathbf{y}_i$ e $\mathbf{y}_j$ risultano non correlati, condizionatamente a $\mathbf{Y}_B = \mathbf{y}_B$.

Si è visto che:

$$\mathrm{V}\left[(\mathbf{Y}_A \| \mathbf{y}_B)\right] = (\mathbf{\Sigma}_{AA.B})^{-1} = \mathbf{\Sigma}_{AA} - \mathbf{\Sigma}_{AB}\mathbf{\Sigma}_{BB}^{-1}\mathbf{\Sigma}_{AB}^{\mathsf{T}} = (\mathbf{\Sigma}^{AA})^{-1}$$

cioè la varianza condizionata di $\mathbf{Y}_A$ è uguale all'inversa del blocco di elementi corrispondenti ad $\mathbf{Y}_A$ nell'inversa di $\mathbf{\Sigma}$.

Nel caso di due variabili i e j , occorre invertire la matrice 2×2 $\mathbf{\Sigma}^{AA}$ di elementi:

$$\mathbf{\Sigma}^{AA} = \begin{pmatrix} c_{ii} & c_{ij} \\ c_{ij} & c_{jj} \end{pmatrix}$$

e quindi:

$$(\mathbf{\Sigma}^{AA})^{-1} = \begin{pmatrix} c_{jj} & -c_{ij} \\ -c_{ij} & c_{ii} \end{pmatrix} / (c_{ii}c_{jj} - c_{ij}^2)$$

pertanto Y_i *e* Y_j *sono non correlati condizionatamente alle altre* $p-2$ *variabili (e quindi indipendenti data la normalità della distribuzione condizionata) se e solo se* $c_{ij} = 0$.

Ricordando la relazione che lega gli elementi c_{hk} dell'inversa di $\mathbf{\Sigma}$, ai cofattori

σ^{hk} di σ_{hk} in $\boldsymbol{\Sigma}$, ossia:

$$c_{hk} = \frac{\sigma^{hk}}{|\boldsymbol{\Sigma}|}$$

dagli elementi di $(\boldsymbol{\Sigma}^{AA})^{-1}$ è possibile calcolare l'indice di correlazione lineare fra Y_i e Y_j condizionatamente a $\mathbf{Y}_B$:

Correlazione fra $\mathbf{Y}_i$ e $\mathbf{Y}_j$ (condizionatamente alle altre $p-2$ variabili)

$$corr(\mathbf{y}_i, \mathbf{y}_j \| \mathbf{Y}_B = \mathbf{y}_B) = \rho_{ij.B} = \frac{-c_{ij}}{\sqrt{c_{ii}c_{jj}}} = \frac{\frac{-\sigma^{ij}}{|\boldsymbol{\Sigma}|}}{\sqrt{\frac{\sigma^{ii}}{|\boldsymbol{\Sigma}|}\frac{\sigma^{jj}}{|\boldsymbol{\Sigma}|}}} = \frac{-\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}}$$

indice di correlazione lineare parziale ossia correlazione fra due variabili tenute costanti le altre $p-2$ variabili

L'indipendenza condizionata e l'indipendenza marginale sono due concetti diversi, e nessuno dei due implica l'altro.———————————————-
verificare

FIGURA DA FARE

Figure 1: distribuzioni condizionate in una normale multivariata

FIGURA DA FARE

Figure 2: distribuzioni condizionate in una normale multivariatamatrice di varianze e covarianze e inversa

---

link o riferimento
(vedere anche →)(figure varie)

`controllare`
`inserire lucidi manuali ed esercizio`
`completare`

8.2.3 Esempio di variabili condizionatamente non correlate

```
\begin{fig}

in ese2000_correlaz1.nb
```

```
\end{fig}
```

Inserire grafici sulle distribuzioni normali condizionate

```
\figura{fig2000regr4.png}
{relazione fra due variabili in funzione
 del valore di una terza variabile}{regr4}


\subsection{cenno ai modelli grafici}
```

8.2.4 Esempi sulla differenza fra l'indipendenza condizionata e l'indipendenza marginale in tabelle $2 \times 2 \times 2$.

L'indipendenza condizionata e l'indipendenza marginale sono due concetti diversi, e nessuno dei due implica l'altro. forse va spostato completamente

Per chiarire la differenza fra indipendenza marginale e indipendenza condizionata, ricorro qui ad un esempio relativo alla distribuzione congiunta di tre variabili dicotomiche A,B e C.

Esempio 8.2 *Si ha una tavola $2 \times 2 \times 2$ di tre mutabili A,B, e C. Le due tavole $A \times B$ condizionate ai valori di C sono:*

$\mathbf{C} = c_1$	b_1	b_2	$tot.$	$\mathbf{C} = c_2$	b_1	b_2	$tot.$
a_1	$0,24$	$0,06$	$0,30$	a_1	$0,12$	$0,28$	$0,4$
a_2	$0,56$	$0,14$	$0,70$	a_2	$0,18$	$0,42$	$0,6$
$tot.$	$0,80$	$0,20$	$1,00$	$tot.$	$0,30$	$0,70$	$1,00$

In queste distribuzioni condizionate A e B sono indipendenti; se P(C=c_1)= P(C=c_2)=$\frac{1}{2}$la tavola marginale $A \times B$ è:

$\mathbf{C}tot.$	b_1	b_2	$tot.$
a_1	$0,18$	$0,17$	$0,35$
a_2	$0,37$	$0,28$	$0,65$
$tot.$	$0,55$	$0,45$	$1,00$

Nella distribuzione marginale A e B non sono indipendenti.

Si può presentare il caso opposto, di caratteri indipendenti marginalmente e associati condizionatamente (paradosso di Simpson). Si ha un'altra tavola citazione *$2 \times 2 \times 2$ di tre mutabili A,B, e C. Le due tavole $A \times B$ condizionate ai valori* *di C sono ora:*

$\mathbf{C} = c_1$	b_1	b_2	$tot.$
a_1	$0,5$	0	$0,5$
a_2	0	$0,5$	$0,5$
$tot.$	$0,5$	$0,5$	1

$\mathbf{C} = c_2$	b_1	b_2	$tot.$
a_1	0	$0,5$	$0,5$
a_2	$0,5$	0	$0,5$
$tot.$	$0,5$	$0,5$	1

In queste distribuzioni condizionate A e B sono associati (addirittura sono massimamente associati)

Infatti se P(C $= c_1$*)*$=$ *P(C*$=$*c*$_2$*)*$=\frac{1}{2}$*la tavola marginale* $A \times B$ *è:*

$\mathbf{C}tot.$	b_1	b_2	$tot.$
a_1	$0,25$	$0,25$	$0,5$
a_2	$0,25$	$0,25$	$0,5$
$tot.$	$0,5$	$0,5$	1

Nella distribuzione marginale A e B sono indipendenti (addirittura equidistribuite)

8.2.5 Impiego delle informazioni dell'inversa C nell'analisi di dati multivariati.

Come si è visto, l'analisi degli elementi dell'inversa della matrice di correlazione può fornire degli elementi utili per indagare sulla dipendenza fra variabili sia in termini marginali che in termini condizionati.

9 Utilità della distribuzione normale multivariata

In effetti quanto visto finora riguarda solo il modello teorico della normale multivariata, ossia le caratteristiche delle distribuzioni di vettori aleatori normali multivariati, che riassumo brevemente (e solo per le proprietà più rilevanti)

- è unimodale;
- dipende solo dai primi due momenti multivariati;
- la non correlazione è condizione necessaria e sufficiente per l'indipendenza;
- ha contorni iper-ellissoidali;
- ha distribuzioni marginali normali multivariate;

- ha distribuzioni condizionate (o parziali) normali multivariate omoschedastiche e con funzioni di regressione lineari;
- combinazioni lineari di sue componenti sono ancora normali multivariate;
- si ottiene come distribuzione limite di processi multivariati come teorema limite centrale multivariato

Non ci stiamo per ora ponendo il problema di adattare una tale distribuzione a dati osservati. In effetti in questo corso questo problema verrà affrontato solo marginalmente: l'importanza del modello normale multivariato in questo corso sta nel fatto che è un modello utile per la definizione di relazioni di dipendenza in media *esattamente* lineari ed omoschedastiche, con distribuzione normale attorno al valor medio.

Un altro aspetto rilevante della distribuzione normale multivariata, è che rappresenta, nei cai regolari, la distribuzione limite degli stimatori di massima verosimiglianza: sfrutteremo questa proprietà in capitoli successivi (per esempio nella regressione non lineare o nei modelli lineari generalizzati) quando in assenza di risultati esatti sulla distribuzione degli stiamtori, ricorreremo a risultati asintotici.

10 Stimatori di massima verosimiglianza dei parametri di una normale multivariata

In questo capitolo sulla normale multivariata non ci siamo mai preoccupati della stima dei parametri: abbiamo solo esposto le proprietà formali di una distribuzione normale multivariata [2] di cui conosciamo tutti i parametri, ossia gli elementi di $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$. Non affronterò in modo completo il problema dell'inferenza sui parametri sulla base campioni provenienti da una normale multivariata dei, ma mi limiterò in queste pagine al problema della stima puntuale dei parametri, che è ovviamente basilare.

Supponiamo di avere un campione(multivariato) casuale di ampiezza n estratto da una normale multivariata a p componenti, ossia una matrice $\mathbf{X}$ $(n \times p)$ di dati,

[2] che ritengo di importanza fondamentale come premessa allo studio dei modelli lineari e dei modelli di dipendenza in genere.

le cui righe sono le n determinazioni di una variabile normale multipla:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix}$$

In analogia al caso univariato, i momenti primi e secondi calcolati sul campione multivariato sono le stime di massima verosimiglianza dei corrispondenti parametri della distribuzione di provenienza; in sintesi:

Stimatore di massima verosimiglianza di $\boldsymbol{\mu}$

Lo stimatore di massima verosimiglianza del vettore delle speranze matematiche $\boldsymbol{\mu}$ di una variabile normale multipla è dato dal vettore $M(\mathbf{X})$ delle medie aritmetiche di un campione multivariato $\mathbf{x}$ di n osservazioni i.i.d. estratto dalla corrispondente distribuzione.
Tale stimatore, come nel caso univariato, è non distorto.

Stimatore di massima verosimiglianza di $\mathbf{\Sigma}$

Lo stimatore di massima verosimiglianza della matrice di varianze e covarianze $\mathbf{\Sigma}$ di tale variabile è dato dalla matrice delle varianze e covarianze empiriche calcolata su un campione multivariato di n osservazioni i.i.d. estratto dalla corrispondente distribuzione. Tale stimatore, come nel caso univariato, è invece distorto.
E' possibile costruire uno stimatore corretto moltiplicando sia le varianze che le covarianze empiriche per il fattore correttivo $\frac{n}{n-1}$, ottenendo quindi lo stimatore:

$$\hat{\mathbf{\Sigma}} = \mathrm{V}[\mathbf{X}] = \mathrm{V}[\mathbf{Z}] = \frac{\mathbf{Z}^{\mathsf{T}}\mathbf{Z}}{n-1}$$

In effetti, dal momento che gli unici parametri della distribuzione normale multivariata sono il vettore delle medie e la matrice di varianza e covarianza, per ottenere gli stimatori di massima verosimiglianza (puntuali!) di tutte le quantità necessarie per calcolare le distribuzioni congiunte, marginali, condizionate e per le componenti principali da un campione proveniente da una normale multivariata, si impiegheranno le stesse formule già viste per la distribuzione teorica, sostituendo ai momenti primi e secondi teorici quelli empirici stimati dal campione, dal momento che lo stimatore di massima verosimiglianza di una funzione dei parametri $g(\boldsymbol{\theta})$ è dato dalla stessa funzione dello stimatore di Massima verosimiglianza, $g(\hat{\theta})$.

Per esempio, lo stimatore di massima verosimiglianza degli autovalori della matrice di varianze e covarianze $\boldsymbol{\Sigma}$, ossia del vettore delle varianze delle componenti principali, si ottiene ovviamente dagli autovalori dello stimatore di massima verosimiglianza $\hat{\boldsymbol{\Sigma}}$:

$$\hat{\lambda}(\boldsymbol{\Sigma}) = \lambda\left(\hat{\boldsymbol{\Sigma}}\right)$$

10.1 Dimostrazione

11 Un test di Multinormalità: cenni

Quando si ha a disposizione un campione di dati multivariato, molto spesso è necessario verificare se è plausibile l'ipotesi di provenienza da un universo normale multivariato.

Un modo semplice per verificare la normalità di un campione di osservazioni multivariate, consiste ovviamente nell'effettuare dei test di normalità su ciascuna delle distribuzioni univariate.

Ricordo che la normalità delle distribuzioni marginali è una condizione necessaria ma non sufficiente per la normalità multivariata: pertanto i test sulla normalità delle distribuzioni marginali costituiscono uno sbarramento preliminare, nel senso che se danno esito negativo possiamo senz'altro scartare l'ipotesi di multinormalità, altrimenti occorrerà procedere col saggiare l'ipotesi di normalità multivariata con test basati sulla distribuzione congiunta.

Se l'insieme in esame è costituito da molte variabili non sarà possibile utilizzare i normali test di bontà dell'adattamento; tuttavia è possibile ottenere delle informazioni eventualmente anche grafiche trasformando opportunamente l'insieme di dati multivariato.

Come si è visto prima, la forma quadratica ad esponente della densità normale ha una distribuzione proporzionale a quella di una χ^2 con p gradi di libertà.

Infatti se:

$$\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

si è già visto prima che la variabile casuale

$$\mathbf{Q} = (\mathbf{Y} - \boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \sim \chi^2_p$$

Pertanto se trasformiamo ognuno degli n vettori osservati $\mathbf{x}_i$ a p componenti secondo la stessa relazione, dovremo aspettarci che questi n valori trasformati q_i seguano ciascuno una distribuzione χ^2 con p gradi di libertà:

$$Q_i = (\mathbf{x}_i - \boldsymbol{\mu})^{\mathsf{T}}(^{-1}\mathbf{x}_i - \boldsymbol{\mu})\boldsymbol{\Sigma} \sim \chi^2_p$$

(le n trasformate Q_i risultano indipendenti per l'indipendenza ipotizzata degli n vettori osservati $\mathbf{x}_i$)

Quindi, se è valida l' ipotesi di multinormalità, il vettore delle n trasformate q_i costituisce un campione casuale semplice estratto da una distribuzione χ^2 con p gradi di libertà. In effetti le quantità che si usano effettivamente per il calcolo delle q_i sono gli stimatori di $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$, M e $\mathbf{S}$, e non i parametri veri (usualmente incogniti); questo fa sì che le quantità:

$$\hat{q}_i = (\mathbf{x}_i - M)^{\mathsf{T}}\mathbf{S}^{-1}(\mathbf{x}_i - M)$$

seguano una distribuzione χ^2_p solo approssimativamente; l'approssimazione è soddisfacente per campioni grandi.

In effetti, un'informazione utile si ricava dalla rappresentazione grafica di tali valori trasformati in corrispondenza dei percentili teorici di una variabile χ^2; un altro elemento di cui si potrebbe tenere conto nella costruzione di un test di normalità è dato dagli angoli che i vettori osservati formano con il centroide M del campione; tuttavia adesso per semplicità non vedremo quest'ulteriore possibilità. Questa trasformazione ci darà fra l'altro la possibilità di creare (sebbene in modo non univoco) un ordinamento in un insieme di dati multivariati, cosa molto utile quando nell'analisi esplorativa occorre trovare *outliers* ossia dati lontani rispetto alla massa dei dati `link con outlier e analisi esplorativa` , facendo uso della cosiddetta distanza di MahalaNobis

citazione esempi: uno positivo e uno negativo

citazione codice ed esempio in R

11.0.1 Esempio:

Questo esempio è tratto dall'insieme di dati antropometrici di cui si è fatto cenno in capitoli precedenti (1432 casi x 7 variabili).