Analisi multivariata

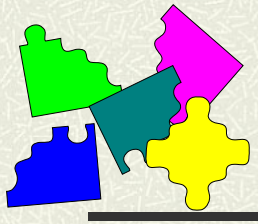
Corso di laurea in Statistica
Carla Rampichini
rampichini@ds.unifi.it

Perchè *Multi*variata?

- Sulla stessa u.s. sono rilevate più di una caratteristica (variabile);
- Necessità di considerare congiuntamente le variabili per capirne le relazioni;
- Necessità di considerare congiuntamente le variabili per capire la struttura dei dati.

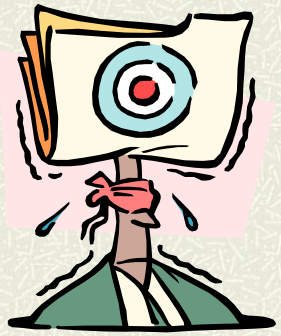
Caratteristiche dell'analisi Multivariata

- Tipicamente esplorativa, non confermativa;
- Riduzione delle dimensioni;
- Ricerca di relazioni tra le variabili;
- Classificazione delle unità.



Preparazione adeguata?

- Corso di Statistica di base
- Conoscenza di SAS o STATA (o altro software statistico)
- Elementi di algebra lineare



Ripasso ...

- Data step in SAS
- Capitolo 1 Zani
- Capitolo 1 Marchetti

- Dispense:
  - algebra delle matrici
  - linear_algebra

Argomenti trattati

- Analisi in componenti principali (PCA)
- Analisi delle corrispondenze
- Distanze e indici di similarità
- Analisi dei gruppi (Cluster Analysis, CA)

Testi consigliati

Zani S. e Cerioli A. (2007) *Analisi dei dati e data mining per le decisioni aziendali*, Milano: Giuffrè, capp. 1, 2, 5, 6, 7, 8, 9

Johnson R.A. e Wichern D.W. (2007). *Applied Multivariate Statistical Analysis*. Sixth Edition. Pearson Education International. Capp.1-2, 8, 12.1-12-5

LA NATURA DEI DATI

- *individui* (unità statistiche):
  da campionamento; intera popolazione
- *caratteri*, misurati sugli individui:
  qualitativi (sconnessi o ordinali)
  quantitativi (discreti o continui)

OBIETTIVO

descrizione e analisi della struttura del collettivo di individui osservati

Esempio

- Nella ricerca sociale

- *Livello di istruzione*
- *Opinioni su welfare*
- *Opinioni sulla sicurezza sociale*
- *Opinioni su ….*

un'unica osservazione

La matrice dei dati

$$\mathbf{X}=\{x_{is}\}$$

- $i=1,\ldots,n$ unità statistiche
- $s=1,\ldots,p$ variabili

Esempio: delinquenza in 16 città americane

7 tipologie di delinquenza
Tasso per 10000 abitanti

Matrice dei dati
$n=16$ unità satistiche, $p=7$ variabili

Obiettivo
- Studio relazione tra variabili;
- Studio delle differenze tra città

⇩

- Riduzione di dimensionalità
- Matrice di correlazione

Esempio: delinquenza in 16 città americane (crimini.sas)

CITTA	OMICIDI	STUPRI	RAPINE	AGGRESS	FURTI	TRUFFE	F_AUTO
Atlanta	16.5	24.8	106	147	1112	905	494
Boston	4.2	13.3	122	90	982	669	954
Chicago	11.6	24.7	340	242	808	609	645
Dallas	18.1	34.2	184	293	1668	901	602
Denver	6.9	41.5	173	191	1534	1368	780
Detroit	13.0	35.7	477	220	1566	1183	788
Hartford	2.5	8.8	68	103	1017	724	468
Honolulu	3.6	12.7	42	28	1457	1102	637
Houston	16.8	16.6	289	186	1509	787	697
Kansas City	10.8	43.2	255	226	1494	955	765
Los Angeles	9.7	51.8	186	355	1902	1386	862
New Orleans	10.3	39.7	266	283	1056	1036	776
New York	9.4	19.4	522	267	1674	1392	848
Portland	5.0	23.0	157	144	1530	1281	488
Tucson	5.1	22.9	85	148	1206	756	483
Washington	12.5	27.6	524	217	1496	1003	739

Vocabolario essenziale

- Varianza ← *Più tipi di varianza multidimensionale…*
- Covarianza
- Correlazione

Queste tre matrici derivano dalla

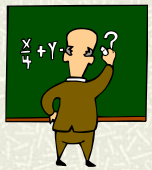
Matrice dei dati ← *Partiamo da qui*

Acquisire confidenza con i concetti di base dell'algebra lineare

- vettore $p \times 1$ e matrice $n \times p$
- operazioni elementari con vettori e matrici
- autovalori e autovettori
- scomposizione spettrale
- cenni alla massimizzazione (minimizzazione) vincolata di forme quadriche.
- proprietà delle matrici partizionate (utili per studio normale multivariata)
- Interpretazione geometrica : angolo tra due vettori, proiezioni, distanze

Fare riferimento a dispense Marchetti, cap. 2 e 2A di Johnson e Wichern

Operazioni sui vettori



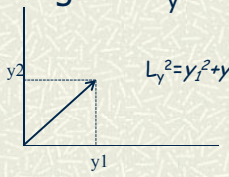
- Come si rappresenta un vettore?
- Come si sommano o sottraggono due vettori ***v*** e ***w***?
- Cosa succede se si moltiplica un vettore per un numero reale?

Operazioni sui vettori

Prodotto scalare e ortogonalità

- Prodotto scalare (interno): $\mathbf{x}'\mathbf{y}=\sum x_i y_i$
- Vettori ortogonali $\Leftrightarrow$ $\mathbf{x}'\mathbf{y}=0$

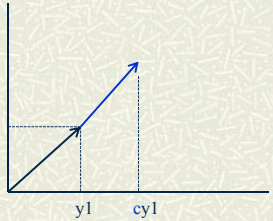
- Ogni vettore ha una direzione, un verso e una lunghezza
- Lunghezza del vettore si trova con teorema di Pitagora: $L_y^2=S(\mathbf{y})=S y_i^2$ ➔ $L_y=\sqrt{\sum y_i^2}$



Lunghezza di un vettore

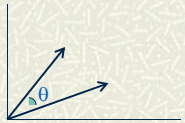
- Ogni vettore può essere allungato o accorciato moltiplicando per un numero reale *c*

- $c\mathbf{y}=[cy_1\ cy_1]'$
  ➔ $L_{cy}^2=c^2y_1^2+c^2y_2^2 \rightarrow L_{cy}=|c|L_y$



- Se $c=L_y^{-1}$ ➔$L_y^{-1}\mathbf{y}$ ha lunghezza unitaria=$[L_y^{-1}y_1\ L_y^{-1}y_2]'$

Angolo tra due vettori

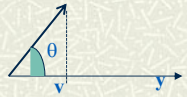


$$\cos(\theta)=\frac{x_1y_1+x_2y_2}{L_xL_y}=\frac{\sum_{i=1}^n x_iy_i}{\sqrt{\sum_{i=1}^n x_i^2}\sqrt{\sum_{i=1}^n y_i^2}}=\frac{\mathbf{x}'\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{x}}\sqrt{\mathbf{y}'\mathbf{y}}}$$

- considerando vettori **x** e **y** di lunghezza unitaria si ha:
  - θ=0 ↔ cos(θ)=1 **x**=[1 0]', **y**=[1 0]'
  - θ=90° ↔ cos(θ)=0 **x**=[1 0]', **y**=[0 1]'
  - θ=180° ↔ cos(θ)=-1 **x**=[1 0]', **y**=[-1 0]'
  - θ=270° ↔ cos(θ)=0 **x**=[1 0]', **y**=[0 -1]'

  - Per θ=90,270 ➔**x** e **y** sono ortogonali

Proiezione di **x** su **y**



$$\mathbf{v}=\frac{\mathbf{x}'\mathbf{y}}{L_yL_y}\mathbf{y}=\frac{\cos(\theta)L_x}{L_y}\mathbf{y}$$

$$L_v=\sqrt{\mathbf{v}'\mathbf{v}}=cL_y=\cos(\theta)L_x=\frac{\mathbf{x}'\mathbf{y}}{L_xL_y}L_x$$

$$c=\frac{\mathbf{x}'\mathbf{y}}{L_y}\frac{1}{L_y}$$

Nb: la proiezione di **x** su **y** è un vettore **v** che ha la stessa direzione di **y** e lunghezza che dipende dalla lunghezza di **x** e dall'angolo tra **x** e **y**

Applicazione del prodotto scalare in statistica

Vettore unitario **u**'=(1,...,1)

- Scarti dalla media $\tilde{\mathbf{x}}=(\mathbf{x}-\mu_x\mathbf{u}),\quad \tilde{\mathbf{y}}=(\mathbf{y}-\mu_y\mathbf{u})$
  $(\mathbf{x}-\mu_x\mathbf{u})'\mathbf{u}=\sum(x_i-\mu)=0$
- Codevianza
  - $(\mathbf{x}-\mu_x\mathbf{u})'(\mathbf{y}-\mu_y\mathbf{u})=\Sigma_i(y_i-\mu_y)(x_i-\mu_x)$
- Devianza
  - $Dev(\mathbf{y})=S(\mathbf{y}-\mu_y\mathbf{u})=\Sigma_i(y_i-\mu_y)^2$
- Coefficiente di correlazione
  $r(x,y)=\cos(\theta)=\tilde{\mathbf{x}}'\tilde{\mathbf{y}}/L_xL_y$

Vedi esempio covarianza

Spazi vettoriali e combinazioni lineari

- Spazio vettoriale: insieme di tutti i vettori reali di dimensione *n* per cui valgono le operazioni di prodotto scalare $\mathbf{x}'\mathbf{y}=\Sigma x_i y_i$ e somma $\mathbf{x}+\mathbf{y}$
- Combinazione lineare $\mathbf{y}=a_1\mathbf{x}_1+a_2\mathbf{x}_2+\ldots+a_k\mathbf{x}_k$
- Dipendenza lineare un insieme di vettori si dice linearmente dipendente se esistono *k* numeri reali non tutti nulli, tali che

  $$a_1\mathbf{x}_1+a_2\mathbf{x}_2+\ldots+a_k\mathbf{x}_k = \mathbf{0}$$

  0 è il vettore nullo

  NB se uno dei vettori dell'insieme ($\mathbf{x}_1, \mathbf{x}_2,\ldots,\mathbf{x}_k$) è il vettore nullo ➔l'insieme di vettori è linearmente dipendente
- Inipendenza lineare se $a_1\mathbf{x}_1+a_2\mathbf{x}_2+\ldots+a_k\mathbf{x}_k = \mathbf{0}$ solo per $a_1=a_2=\ldots=a_k = 0$ i vettori sono linearmente indipendenti
- Vedi esempio vettori lin, indipendenti e lin. dipendenti

Base dello spazio vettoriale

- qualsiasi insieme di *n* vettori linearmente indipendenti è detto base dello spazio vettoriale di ordine *n*

  ⇩

  dati gli *n* vettori della base, utilizzando la somma e il prodotto scalare si possono generare tutti gli altri vettori di ordine *n*

  ⇩
- Ogni vettore può essere espresso come combinazione lineare unica di una base vettoriale data
- Per esempio per n=3 se si sceglie come base l'insieme di vettori

  $\mathbf{x}_1=[1\ 0\ 0]'$ $\mathbf{x}_2=[0\ 1\ 0]'$ $\mathbf{x}_3=[0\ 0\ 1]'$
- $\forall\ \mathbf{y}=[y_1\ y_2\ y_3]'$ può essere generato come $\mathbf{y}= y_1\ \mathbf{x}_1+ y_2\ \mathbf{x}_2+ y_3\ \mathbf{x}_3$

(vedi esempio base n=3)

Vettori della base

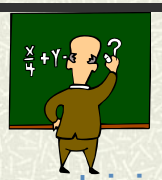
- I vettori che formano la base sono di lunghezza unitaria e sono tra loro perpendicolari
- Per esempio per n=3 i vettori della base sono

  $\mathbf{v}_1=[1\ 0\ 0]'$ $\mathbf{v}_2=[0\ 1\ 0]'$ $\mathbf{v}_3=[0\ 0\ 1]'$

  ⇩
  - $L^2(\mathbf{v}_1)=\mathbf{v}_1'\,\mathbf{v}_1=1$
  - $\mathbf{v}_1'\,\mathbf{v}_2=0$

(vedi esempio vettori_base)

Operazioni sulle matrici



- Operazioni sulle matrici
- Cos'è la traccia di una matrice?
- Che relazione c'è tra gli autovalori e la traccia della matrice?
- Che realzione c'è tra la traccia della matrice e la variabilità?

TRACCIA

- La **traccia** di una matrice quadrata è la somma degli elementi della diagonale principale

$$tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & \dots & a_{1i} & \dots & a_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{i1} & \dots & a_{ii} & \dots & a_{in} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \dots & a_{ni} & \dots & a_{nn} \end{bmatrix}$$

- tr(c**A**)=c×tr(**A**)
- tr(**A**+**B**)=tr(**A**)+tr(**B**) idem per la differenza
- tr(**AB**)=tr(**BA**)
- Per **A** qualsiasi (anche non quadrata): $tr(\mathbf{AA}') = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2$

Principali matrici utilizzate nell'analisi multivariata

- Matrice dei dati $\mathbf{X}_{n\times p}$
- Matrice $p\times p$ dei quadrati $\mathbf{X'X}$
- Matrice degli scarti $\tilde{\mathbf{X}}_{n\times p} = \mathbf{X}_{n\times p} - \bar{\mathbf{X}}_{n\times p}$
- Matrice $p\times p$ delle devianze $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$
- Matrice di covarianza $\mathbf{S}_{p\times p} = \frac{1}{n-1}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$
- Matrice dei dati standardizzati

  $\mathbf{Z}_{n\times p} = \tilde{\mathbf{X}}\mathbf{D}, \mathbf{D}_{p\times p} = diag\left\{1/\sqrt{s_{ii}}\right\}$ s_{ii}=var($\mathbf{X}_j$)
- Matrice di correlazione $\mathbf{R}_{p\times p}$=(1/(*n*-1)) **Z'Z**

Un esempio …

Esempio: matrice di correlazione

	omicidi	stupri	rapine	aggress	furti	truffe
stupri	0.346					
rapine	0.437	0.207				
aggress	0.556	0.758	0.526			
furti	0.232	0.460	0.267	0.420		
truffe	-0.068	0.489	0.258	0.343	0.759	
f_auto	0.050	0.365	0.458	0.378	0.269	0.312

Quasi tutti i pacchetti statistici svolgono operazioni tra matrici.
Operazioni tra matrici: (i) in STATA, (ii) in SAS IML (iii) in R

Alcune definizioni utili

- Rango di una matrice ***A***

Se ***A*** matrice quadrata *p*×*p*

- Traccia: tr(***A***)=a_{11}+ a_{22} +…+ a_{pp}
- Determinante: det(***A***)
- Matrice inversa: $\mathbf{A}^{-1}$
- Autovalore λ_j: soluzione di p(λ)=det(***A***-λ**I**-)
- Autovettore ***x***: ***Ax***=λ***x*** **autovalori.sas**
- det(*A*)=$\Pi\lambda_j$ j=1,…,p

vedremo in ACP

Rango

- Il rango, $r(\mathbf{A})$, di una matrice $\mathbf{A}_{n\times m}$ è definito come il numero massimo di righe (colonne) linearmente indipendenti
- Un insieme di $k \le n$ righe $\mathbf{a}_i$ di $\mathbf{A}$ è linearmente indipendente se nessuna riga può essere espressa come combinazione lineare delle restanti $(k\text{-}1)$ righe, cioè $\sum_{j=1}^{k} c_j a_{ij} = 0 \rightarrow c_j = 0, \forall j$.
- Se $m \le n$, $r(\mathbf{A}) \le m$, e se $r(\mathbf{A}) = r = m$, la matrice $\mathbf{A}$ è detta a pieno rango

$$0 \le r(\mathbf{A}) \le \min(n,m)$$

Determinante

- Per la matrice quadrata $\mathbf{A}$

$$\det(\mathbf{A}) = |\mathbf{A}| = \sum (-1)^{|\tau|} a_{1\tau(1)} \dots a_{n\tau(n)}$$

- La sommatoria è fatta rispetto a tutte le $n!$ permutazioni τ degli indici $\{1, 2, \dots, n\}$
  - $|\tau| = 0$ se la permutazione può essere scritta come prodotto di un numero pari di trasposizioni
  - $|\tau| = 1$ altrimenti.

Proprietà del determinante

$\mathbf{A}(n \times n)$, $\mathbf{B}(n \times n)$, $c \in \mathrm{R}$

- $\det(\mathbf{A}) = |\mathbf{A}| \neq 0$ sse $\mathbf{A}$ è a pieno rango
- $|c\mathbf{A}| = c^n |\mathbf{A}|$
- $|\mathbf{AB}| = |\mathbf{BA}| = |\mathbf{A}||\mathbf{B}|$
- $\det(\mathbf{D}) = d_1 \times \dots \times d_p$, $\mathbf{D} = \mathrm{diag}\{d_j\}$
- $|\mathbf{A}| = |\mathbf{A}'|$

Matrice inversa

- Per la matrice quadrata $\mathbf{A}_{p\times p}$, se $\det(\mathbf{A}) \neq 0$, allora esiste $\mathbf{A}^{-1}$:

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

- Se p è piccolo $\mathbf{A}^{-1}$ può essere calcolata come $\mathbf{A}^{-1} = \frac{\mathbf{C}}{\det(\mathbf{A})}$, dove $\mathbf{C}$ è la matrice aggiunta di $\mathbf{A}$

NB: se $\det(\mathbf{A}) = 0 \rightarrow \mathbf{A}^{-1}$ non esiste!

- Noi useremo computer (p.e. IML in SAS)!

Una matrice di cui esiste l'inversa è detta NON SINGOLARE

Proprietà matrice inversa

- $\mathbf{D}=\text{diag}\{d_j\}$, $\mathbf{D}^{-1}=\text{diag}\{1/d_j\}$
- $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$
- $(\mathbf{A}^{-1})^{-1}=\mathbf{A}$
- $(\mathbf{A}^{-1})'=(\mathbf{A}')^{-1}$
- Se $\mathbf{A}$ è simmetrica anche $\mathbf{A}^{-1}$ è simmetrica
- $(\mathbf{AB})^{-1}=\mathbf{B}^{-1}\mathbf{A}^{-1}$

Autovalori e autovettori

- Consideriamo la matrice quadrata $\mathbf{A}_{p\times p}$. Se esistono uno scalare λ e un vettore $\mathbf{x}$ tali che

$$\mathbf{Ax} = \lambda\mathbf{x}$$

allora:

- λ è un autovalore (eigenvalue) di $\mathbf{A}$
- $\mathbf{x}$ è un autovettore (eigenvector) di $\mathbf{A}$
- considereremo $\mathbf{e}=(1/L_x)\mathbf{x}$ autovettore normalizzato

Come si trovano gli autovalori

- Si dimostra che gli autovalori sono le radici del polinomio di grado p in λ

$$\det(\mathbf{A}-\lambda\mathbf{I}) = 0$$

- Ci sono al più p autovalori distinti non nulli di $\mathbf{A}$.
- Una matrice simmetrica $p\times p$ ha p radici reali distinte
- Esempio

$$\mathbf{A}=\begin{pmatrix}1 & 0\\ 1 & 3\end{pmatrix},\quad \mathbf{A}-\lambda\mathbf{I}=\begin{pmatrix}1-\lambda & 0\\ 1 & 3-\lambda\end{pmatrix}$$

$$\det(\mathbf{A}-\lambda\mathbf{I}) = (1-\lambda)(3-\lambda) = 0 \Rightarrow \lambda^2-4\lambda+3=0$$

$$\lambda_1 = 1,\quad \lambda_2 = 3$$

Come si trovano gli autovettori

- Per ogni autovalore λ_j esiste un autovettore $\mathbf{x}_j$ dato dall'equazione

$$\mathbf{Ax} = \lambda\mathbf{x},\quad \mathbf{x}\neq\mathbf{0}$$

- Esempio

$$\mathbf{A}=\begin{pmatrix}1 & 0\\ 1 & 3\end{pmatrix},\quad \lambda_1=1,\quad \lambda_2=3,\quad \mathbf{Ax}=\lambda_1\mathbf{x}$$

$$\Rightarrow\begin{pmatrix}1 & 0\\ 1 & 3\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}=1\begin{pmatrix}x_1\\ x_2\end{pmatrix}\Rightarrow\begin{pmatrix}x_1\\ x_1+3x_2\end{pmatrix}=\begin{pmatrix}x_1\\ x_2\end{pmatrix}$$

$$x_1 = -2x_2$$

- Esistono più soluzioni che soddisfano l'ultima eguaglianza! Per esempio x_1=-2, x_2=1

quale soluzione si sceglie?

- Si considera il vettore di lunghezza unitaria (normalizzato):

$$L_x = \sqrt{\mathbf{x}'\mathbf{x}} = \sqrt{x_1^2 + x_2^2} = \sqrt{4x_2^2 + 1x_2^2} = x_2\sqrt{5}$$

$$\Rightarrow \mathbf{e} = \frac{\mathbf{x}}{L_x} = \frac{1}{x_2\sqrt{5}}\begin{pmatrix} -2x_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{pmatrix}$$

Alcune proprietà degli autovalori

- Supponiamo che la matrice quadrata **A** ammetta autovalori $\lambda_1, \ldots, \lambda_p$.
- Sia $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_p)$.
- det(**A**) e tr(**A**) possono essere espressi in termini degli autovalori:

$$\det(\mathbf{A}) = \det(\mathbf{\Lambda}) = \lambda_1 \times \ldots \times \lambda_p$$

$$\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{\Lambda}) = \lambda_1 + \ldots + \lambda_p$$

Alcune proprietà degli autovalori

- Se la matrice $\mathbf{A}_{p\times p}$ è simmetrica
  - ammette autovalori distinti $\lambda_1, \ldots, \lambda_p$
  - i *p* autovettori corrispondenti sono ortogonali
  - poiché per $j=1,\ldots, p$ $\mathbf{e}_j'\mathbf{e}_k=0$ e $\mathbf{e}_j'\mathbf{e}_j=1$, $\mathbf{E}'\mathbf{E}=\mathbf{I}$, dove $\mathbf{E}=[\mathbf{e}_1 \ldots \mathbf{e}_p]$.
  - $\mathbf{E}'\mathbf{E}=\mathbf{I}$ (ortonormale) ➔ $\mathbf{E}'=\mathbf{E}^{-1}$

(NB: se A non è simmetrica, non è detto che le radici dell'equazione caratteristica siano reali)

Scomposizione spettrale

- **A** simmetrica➔ $\mathbf{E}'\mathbf{E}=\mathbf{I}$ (ortonormale)
- Possiamo scrivere l'insieme di tutte le equazioni caratteristiche $\mathbf{A}\mathbf{e}_j=\lambda_j\mathbf{e}_j$ come $\mathbf{AE}=\mathbf{E\Lambda}$, $\mathbf{\Lambda}=\text{diag}(\lambda_j)$
- Pre-moltiplicando per $\mathbf{E}'$: $\mathbf{E}'\mathbf{AE}=\mathbf{E}'\mathbf{E\Lambda}=\mathbf{I\Lambda}=\mathbf{\Lambda}$
  ➔ogni matrice simmetrica può essere diagonalizzata
- Post-moltiplicando per $\mathbf{E}'$: $\mathbf{AEE}'=\mathbf{E\Lambda E}'$ ➔ $\mathbf{A}=\mathbf{E\Lambda E}'$
  ➔Ogni matrice simmetrica può essere ricostruita dai suoi autovalori e autovettori

$$\mathbf{A} = \sum_{j=1}^{p} \lambda_j \mathbf{e}_j \mathbf{e}_j'$$

Esempio scomposizione spettrale

$$\mathbf{A}=\begin{pmatrix}1 & -5\\ -5 & 1\end{pmatrix},\quad \det(\mathbf{A}-\lambda\mathbf{I})=(1-\lambda)^2-25=0\Rightarrow \lambda_1=-4,\quad \lambda_2=6$$

$$\mathbf{Ax}=\lambda_1\mathbf{x}\Rightarrow x_1=x_2,\quad \text{per } x_1=1\quad L_x=\sqrt{2}\Rightarrow \mathbf{e}_1=\begin{pmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{pmatrix}$$

$$\mathbf{Ax}=\lambda_2\mathbf{x}\Rightarrow -x_1=x_2,\text{per } x_1=1\quad L_x=\sqrt{2}\Rightarrow \mathbf{e}_2=\begin{pmatrix}1/\sqrt{2}\\ -1/\sqrt{2}\end{pmatrix}$$

$$\lambda_1\mathbf{e}_1\mathbf{e}_1'+\lambda_2\mathbf{e}_2\mathbf{e}_2'=\mathbf{A}$$

Esempio scomposizione spettrale

$$\mathbf{A}=\begin{pmatrix}1 & -5\\ -5 & 1\end{pmatrix},\quad \det(\mathbf{A}-\lambda\mathbf{I})=(1-\lambda)^2-25=0\Rightarrow \lambda_1=-4,\quad \lambda_2=6$$

$$\mathbf{Ax}=\lambda_1\mathbf{x}\Rightarrow x_1=x_2,\quad \text{per } x_1=1\quad L_x=\sqrt{2}\Rightarrow \mathbf{e}_1=\begin{pmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{pmatrix}$$

$$\mathbf{Ax}=\lambda_2\mathbf{x}\Rightarrow -x_1=x_2,\text{per } x_1=1\quad L_x=\sqrt{2}\Rightarrow \mathbf{e}_2=\begin{pmatrix}1/\sqrt{2}\\ -1/\sqrt{2}\end{pmatrix}$$

$$\lambda_1\mathbf{e}_1\mathbf{e}_1'+\lambda_2\mathbf{e}_2\mathbf{e}_2'=\mathbf{A}$$

Rango di una matrice simmetrica

- Il rango di una matrice simmetrica **A** corrisponde al numero di autovalori diversi da zero
- Per trovare questo risultato sfruttiamo la scomposizione spettrale e il teorema seguente
- TEOREMA rango di un prodotto: per ogni matrice **A** e date le matrici non singolari **B** e **C**, il rango di **BAC** è uguale al rango di **A**
- Poiché **E** e **E**' sono non singolari, applicando il teorema a **E'AE=Λ** si ottiene rango(**A**)= rango(**Λ**)
- Il rango della matrice diagonale **Λ** corrisponde al numero di valori non nulli sulla diagonale!

Forme quadratiche

- Data una matrice simmetrica **A**, di dimensione $p\times p$, la doppia sommatoria degli elementi di un vettore **x** di dimensione p con coefficienti pari agli elementi di **A**, può essere scritta in forma matriciale come:

$$Q(\mathbf{x})=\mathbf{x}'\mathbf{Ax}=\sum_{i=1}^{p}\sum_{j=1}^{p}a_{ij}x_ix_j$$

- $Q(\mathbf{x})$ è detta forma quadratica.
- In generale $Q(\mathbf{x})$ può essere positiva, negativa o nulla, dipende da **x** e **A**.

Matrici definite

Per alcune matrici $Q(\mathbf{x})$ è sempre positiva, qualunque sia $\mathbf{x}$, e per altre $Q(\mathbf{x})$ è sempre negativa. Data la matrice simmetrica $\mathbf{A}$,

- $\mathbf{A}$ è definita positiva se vale $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ per $\forall \mathbf{x} \neq \mathbf{0}$
- $\mathbf{A}$ è definita negativa se vale $\mathbf{x}'\mathbf{A}\mathbf{x} < 0$ per $\forall \mathbf{x} \neq \mathbf{0}$
- $\mathbf{A}$ è definita non negativa (semidefinita positiva) se vale $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ per $\forall \mathbf{x} \neq \mathbf{0}$
- $\mathbf{A}$ è definita non positiva (semidefinita negativa) se vale $\mathbf{x}'\mathbf{A}\mathbf{x} \leq 0$ per $\forall \mathbf{x} \neq \mathbf{0}$

Come stabilire se A è definita positiva (o negativa)?

- Ricordiamo che una matrice simmetrica può essere scomposta come $\mathbf{A} = \mathbf{E}\Lambda\mathbf{E}' = \sum_{j=1}^{p} \lambda_j \mathbf{e}_j \mathbf{e}_j'$ con $\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_p), \Lambda = diag(\lambda_j)$

posto $\mathbf{y}=\mathbf{E}'\mathbf{x}$, possiamo scrivere

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{E}\Lambda\mathbf{E}'\mathbf{x} = \mathbf{y}'\Lambda\mathbf{y} = \sum_{j=1}^{p} \lambda_j y_j^2$$

- Se $\lambda_j>0$ per $\forall j$, allora $Q(\mathbf{x})$ sarà positiva qualunque sia $\mathbf{x}$.

Matrici definite

- Sia $\mathbf{A}$ una matrice simmetrica.
  - Se tutte le radici caratteristiche (autovalori) di $\mathbf{A}$ sono positive (negative) allora $\mathbf{A}$ è **definita positiva** (negativa definita).
  - Se alcune delle radici sono nulle, allora $\mathbf{A}$ è **definita non negativa** (non positiva) se le radici rimanenti sono positive (negative).
  - Se $\mathbf{A}$ ha radici sia positive che negative è **indefinita**.

Esempio

$$Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 3x_1^2 + 2x_2^2 - 2\sqrt{2}x_1x_2$$

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det \begin{pmatrix} 3-\lambda & -\sqrt{2} \\ -\sqrt{2} & 2-\lambda \end{pmatrix} = (3-\lambda)(2-\lambda) - 2$$

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \Rightarrow \lambda_1 = 1, \quad \lambda_2 = 4$$

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1 \mathbf{x}'\mathbf{e}_1\mathbf{e}_1'\mathbf{x} + \lambda_2 \mathbf{x}'\mathbf{e}_2\mathbf{e}_2'\mathbf{x} = \lambda_1 y_1^2 + \lambda_2 y_2^2 = y_1^2 + 4y_2^2 > 0, \forall \mathbf{y} \neq \mathbf{0}$$

$$\Rightarrow \mathbf{A} = \begin{pmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{pmatrix}$$ è definita positiva

Proprietà delle matrici definite

- La matrice identità è definita positiva
- Se la matrice simmetrica **A** è
  - definita non negativa, cioè se vale $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$, $\forall \mathbf{x}$ allora $\det(\mathbf{A}) \geq 0$
  - definita positiva, cioè se vale $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$, $\forall \mathbf{x}$ allora anche $\mathbf{A}^{-1}$ è definita positiva

Principali matrici utilizzate nell'analisi multivariata

- Matrice dei dati $\mathbf{X}_{n\times p}$
- Matrice $p\times p$ dei quadrati $\mathbf{X}'\mathbf{X}$
- Matrice degli scarti $\tilde{\mathbf{X}}_{n\times p} = \mathbf{X}_{n\times p} - \bar{\mathbf{X}}_{n\times p}$
- Matrice $p\times p$ delle devianze $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$
- Matrice di covarianza $\mathbf{S}_{p\times p} = \frac{1}{n-1}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$
- Matrice dei dati standardizzati

$$\mathbf{Z}_{n\times p} = \tilde{\mathbf{X}}\mathbf{D},\ \mathbf{D}_{p\times p} = diag\left\{1/\sqrt{s_{ii}}\right\}_i = \text{var}(\mathbf{X}_j)$$

- Matrice di correlazione $\mathbf{R}_{p\times p} = (1/(n-1))\ \mathbf{Z}'\mathbf{Z}$

Esempio: matrice di correlazione

	omicidi	stupri	rapine	aggress	furti	truffe
stupri	0.346					
rapine	0.437	0.207				
aggress	0.556	0.758	0.526			
furti	0.232	0.460	0.267	0.420		
truffe	-0.068	0.489	0.258	0.343	0.759	
f_auto	0.050	0.365	0.458	0.378	0.269	0.312

Quasi tutti i pacchetti statistici svolgono operazioni tra matrici.
Operazioni tra matrici: (i) in STATA, (ii) in SAS IML (iii) in R

Relazioni tra variabili quantitative

- Diagramma di dispersione
- Covarianza
- Correlazione

Forni a micro-onde di 8 marche
(Fonte: Altroconsumo, n.104, aprile 1998)

marca	altezza	capacita	prezzo	timer	cottura
candy	20	15.7	295	1	9
delonghi	17	10.2	260	1	10
electrolux	19	12.5	259	2	7
moulinex	16	10.8	280	2	8
ocean	20	10.5	279	1	4
panasonic	19	11.5	240	2	6
samsung	17	11.8	259	2	6
sharp	17	10.8	339	2	7

microonde.txt; microonde.sas

Matrice di covarianza campionaria

$$\mathbf{S}_n=\{\mathrm{cov}(X_i,X_k)\}=(1/n)\mathbf{X}'(\mathbf{I}-\mathbf{uu}')\mathbf{X} \qquad i,k=1,2,\ldots,p$$

La codevianza può essere calcolata in vari modi …

Matrice di correlazione

$$\mathbf{R}_n=\{r_{sv}\}=(1/n))\ \mathbf{Z}'\mathbf{Z}=\mathbf{D}\,\mathbf{S}_n\,\mathbf{D}$$

Z matrice degli scarti standardizzati $\mathbf{Z}_{n\times p}=\tilde{\mathbf{X}}\mathbf{D}$

D matrice diagonale dei reciproci delle dev. std. $\mathbf{D}_{p\times p}=diag\left\{1/\sqrt{s_{ii}}\right\}$

$r_{ik}=[\mathrm{cov}(X_i,X_k)]/[V(X_i)V(X_k)]^{1/2}$

Vedremo esempio in SAS: microonde2.sas

Forni a microonde di 8 marche

Variabile	Media	Std Dev	Minimo	Massimo
altezza	18.125	1.55265	16.00	20.00
capacita	11.725	1.77261	10.20	15.70
prezzo	276.375	30.43465	240.00	339.00

centroide

	altezza	capacita	prezzo
altezza	**2.4107143**	1.4678571	-6.7678571
capacita	1.4678571	**3.1421429**	3.9892857
prezzo	-6.7678571	3.9892857	**926.2678571**

Matrice di covarianza

	altezza	capacita	prezzo
altezza	1.00000	0.53333	-0.14322
capacita	0.53333	1.00000	0.07395
prezzo	-0.14322	0.07395	1.00000

Matrice di correlazione

Misure di variabilità multi-dimensionale: varianza totale

$$VAR_T=\mathrm{tr}\{\mathbf{S}\}=\Sigma V(\mathbf{x}_j)$$

- problemi interpretativi se le variabili hanno diversa unità di misura
- non tiene conto della correlazione tra variabili
- $\mathrm{tr}\{\mathbf{R}\}=\Sigma 1=p$

Misure di variabilità multidimensionale: varianza generalizzata di Wilks

$$VAR_G = \det(\mathbf{S}) = |\mathbf{S}|$$

- $VAR_G = 0 \Leftrightarrow$ rango($\mathbf{S}$)$<p$
  - almeno una variabile costante
  - almeno una variabile combinazione lineare di altre
- max $VAR_G = \Pi var(X_s)$

Misure di variabilità multidimensionale: variabilità relativa

$$VAR_R = |\mathbf{S}| / \Pi var(X_s) = |\mathbf{R}|$$

$0 \leq VAR_R \leq 1$ (variabili incorrelate)

Una var costante o combinazione lineare di altre

Voto riportato da 100 studenti a 3 esami (dati fittizi)

Variabile	Media	Varianza	Minimo	Massimo
esame1	23.600	7.1515	18.0	30.0
esame2	23.090	8.2241	18.0	30.0
esame3	22.740	8.7600	18.0	30.0

$VAR_T = tr\{\mathbf{S}\} = 24.1357$

Matrice dei dati: n=100, p=3

studenti.sas

Voto riportato da 100 studenti a 3 esami

S

	esame1	esame2	esame3
esame1	7.151515152	7.157575758	6.814141414
esame2	7.157575758	8.224141414	7.983232323
esame3	6.814141414	7.98323232	8.760000000

$VAR_G = \det(\mathbf{S}) = |\mathbf{S}| = 7.517562$

R

	esame1	esame2	esame3
esame1	1.00000	0.93330	0.86091
esame2	0.93330	1.00000	0.94055
esame3	0.86091	0.94055	1.00000

$VAR_R = |\mathbf{S}| / \Pi var(X_s) = |\mathbf{R}| = 0.014591$

Rappresentazioni grafiche di dati multivariati

- Utilissimi per mettere in evidenza caratteristiche e struttura dei dati
- Attenzione alle illusioni ottiche!
- Utili in fase di:
  - Analisi preliminare dei dati: valori anomali, relazioni tra variabili, somiglianza tra unità statistiche
  - Presentazione dei risultati

Rappresentazioni grafiche

- Bibliografia essenziale:
  - Capitolo 5 Zani
  - Capitolo 1 Johnson e Wichern
  - Cap. 2 Everitt
- Per approfondire
  - Everitt B. (1978). Graphical techniques for multivariate data. New York: North-Holland.
  - Tukey J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison Wesley.
  - Friendly M. (1991), Statistical Graphics for Multivariate Data, *SAS SUGI 16* Conference, http://www.math.yorku.ca/SCS/sugi/sugi16-paper.html

Rappresentazioni grafiche

- Dotplot e boxplot marginali
- Diagrammi di dispersione: relazione tra coppie di variabili
- *Convex hull* e boxplot bivariato (con R)
- Matrice dei diagrammi di dispersione
- Le "stelle"
- Facce di Chernoff

DOTPLOT

- mostra i singoli casi osservati come punti
- dal dotplot possiamo vedere la forma, il centro e la dispersione dei dati

Il dot plot è utile quando:

- si hanno pochi casi
- si vogliono vedere i singoli valori
- si vuole vedere qual è la forma della distribuzione
- si vogliono confrontare pochi gruppi

Quando leggete un dot plot, tenete presente che software diversi fanno dot plot diversi: a volte 1 punto rappresenta 1 singolo caso, a volte 2 o più casi, a volte i valori vengono arrotondati

dotplot_auto.sas

Box-plot

Sintesi della distribuzione univariata attraverso 5 numeri:

- **minimo:** il più piccolo valore osservato
- ***Q1*:** la mediana della prima metà dei valori
- **Mediana:** il valore che divide i dati in due parti
- ***Q3*:** la mediana della metà superiore dei valori
- **massimo:** il valore più grande osservato

dotplot_auto.sas

Diagrammi di dispersione

Consentono di evidenziare:

- Outlier uni- e bi-dimensionali;
- Relazione tra coppie di variabili.

Alimetari.sas:

- Outlier Plasmon;
- Relazione non lineare tra ROE e Economic Profit;
- Rifare grafico senza Plasmon per meglio evidenziare posizione dei punti.

Indicatori economici di 10 aziende alimentari '97

(Fonte: Le 5000 Societa' leader, Milano Finanza, 1998)
Zani, 2000, p. 7

Obs	azienda	econ	cash	lavor	roe	inde	fatt
1	Barilla	-25.4	7.39	59.54	4.20	0.83	2867
2	Eridania	-141.0	4.00	68.99	0.84	1.86	1693
3	Ferrero	65.8	9.61	53.70	21.12	-0.02	3031
4	Galbani	-71.9	8.40	56.32	2.66	-0.02	2136
5	Kraft	-32.0	5.88	72.11	3.20	0.35	1563
6	Lavazza	-28.9	4.96	39.08	5.29	-0.05	1117
7	Nestle	-98.8	2.72	81.25	0.00	1.69	3463
8	Parmalat	-145.1	5.96	38.51	2.23	2.91	1664
9	Plasmon	31.7	27.76	31.35	24.60	1.35	858
10	Star	2.4	6.47	62.49	10.60	0.00	811
							=====
							19203

scatterplot_aziende.sas

Indicatori economici di 10 aziende alimentari 1997

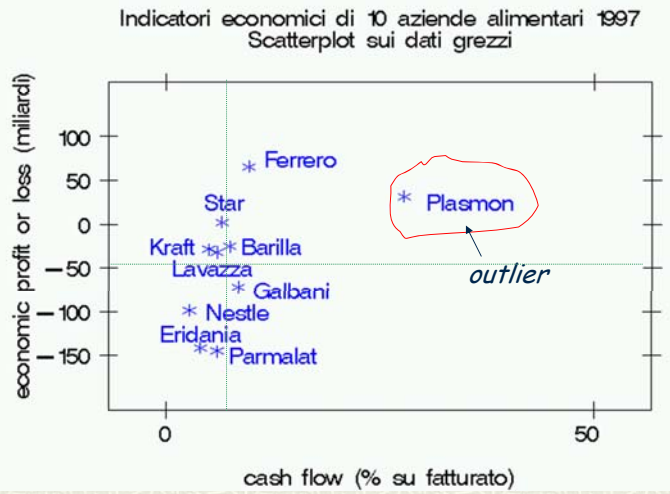
Statistiche descrittive ponderate per il fatturato

Variabile	Descrizione	Media	Std Dev
econ	economic profit or loss (miliardi)	-46.9935	68.2710
cash	cash flow (% su fatturato)	7.1948	5.0081
lavor	costo del lavoro (% su VA)	59.8840	14.5154
roe	utile netto sul patrimonio in %	6.6388	8.0306
inde	indebitamento sul capitale proprio	0.9253	0.9445

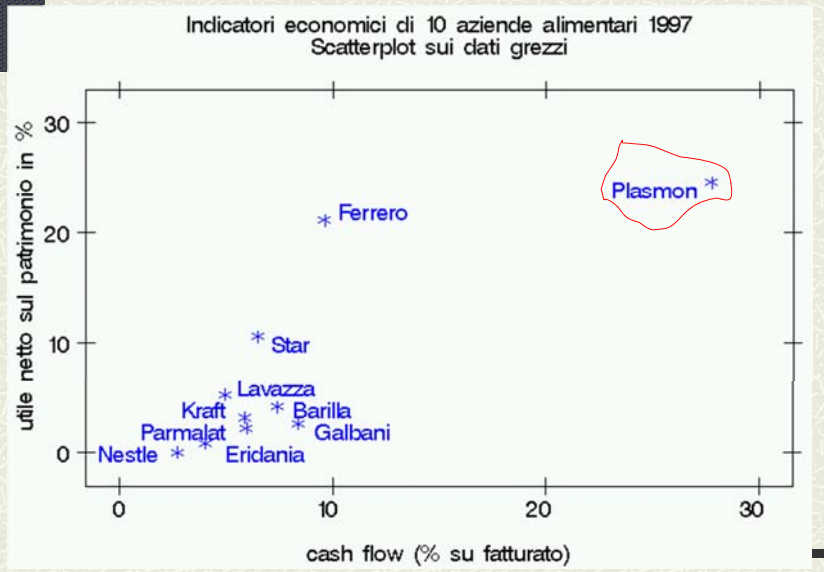
Matrice di correlazione

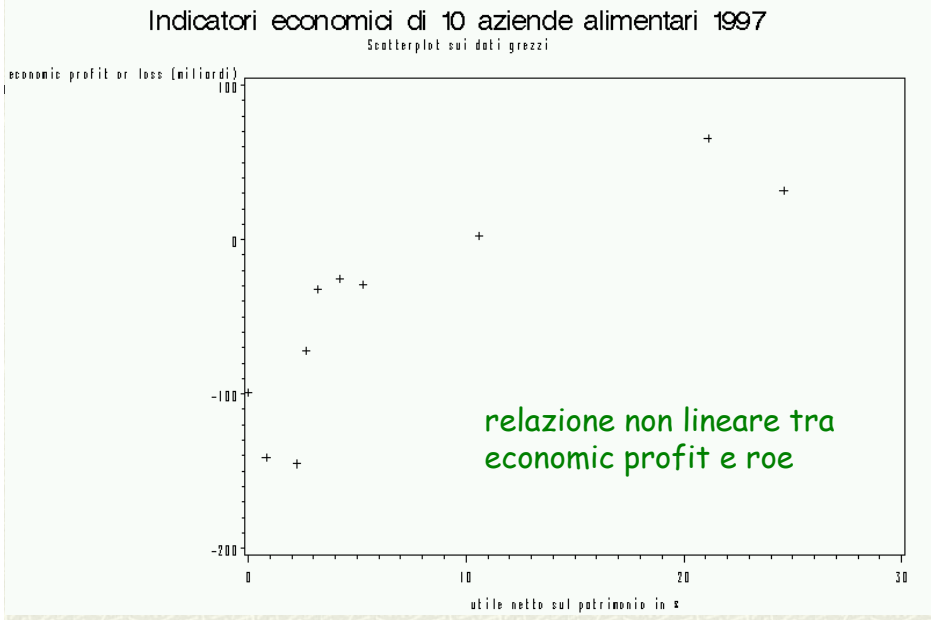
	econ	cash	lavor	roe	inde	fatt
econ	1.00000					
cash	0.53327	1.00000				
lavor	-0.26648	**-0.61737**	1.00000			
roe	**0.82819**	**0.80784**	-0.52861	1.00000		
inde	**-0.68767**	0.00663	-0.07301	-0.26741	1.00000	
fatt	-0.08754	-0.36027	0.50608	-0.22119	0.07883	1.00000

Outliers bivariati dal diagramma di dispersione



Outliers bivariati dal diagramma di dispersione (2)





VETTORI E MATRICI CASUALI

- VETTORI E MATRICI CASUALI: vettori e matrici i cui elementi sono variabili casuali (v.c.)
- Il valore atteso di una matrice casuale **X** si indica con

$$E(\mathbf{X}) = \begin{bmatrix} E(X_{11}) & \dots & E(X_{1p}) \\ \vdots & \ddots & \vdots \\ E(X_{n1}) & \dots & E(X_{np}) \end{bmatrix}$$

$$E(X_{kj}) = \begin{cases} \int_{-\infty}^{+\infty} x_{kj} f_{kj}(x_{kj}) dx_{kj} & \text{se } X_{kj} \text{ è continua} \\ \sum_{\text{per tutti gli } x_{kj}} x_{kj} f_{kj}(x_{kj}) & \text{se } X_{kj} \text{ è discreta} \end{cases}$$

Proprietà dei valori attesi

- **X** e **Y** due matrici casuali della stessa dimensione
- **A** e **B** due matrici di costanti conformabili

$$E(\mathbf{X}+\mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y})$$

$$E(\mathbf{AXB}) = \mathbf{A}E(\mathbf{X})\mathbf{B}$$

Medie e varianze marginali

- $\mathbf{X}'=(X_1, \ldots, X_p)$ vettore casuale di dimensione $p\times 1$
- Ogni elemento di **X** è una v.c. avente la sua distribuzione di probabilità marginale $f_k(x_k)$ (densità se X_k v.c. continua)

$$\mu_k = E(X_k) = \begin{cases} \int_{-\infty}^{+\infty} x_k f_k(x_k) dx_k & \text{se } X_k \text{ è continua} \\ \sum_{\text{per tutti gli } x_k} x_k f_k(x_k) & \text{se } X_k \text{ è discreta} \end{cases}$$

$$\sigma_k^2 = \sigma_{kk} = E(X_k - \mu_k)^2 = \begin{cases} \int_{-\infty}^{+\infty} (x_k - \mu_k)^2 f_k(x_k) dx_k & \text{se } X_k \text{ è continua} \\ \sum_{\text{per tutti gli } x_k} (x_k - \mu_k)^2 f_k(x_k) & \text{se } X_k \text{ è discreta} \end{cases}$$

Covarianza

$$\sigma_{jk} = E\left[(X_j - \mu_j)(X_k - \mu_k)\right]$$

$$\sigma_{jk} = \begin{cases} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x_j - \mu_j)(x_k - \mu_k) f_{jk}(x_j, x_k) dx_j dx_k & \text{se } X_j, X_k \text{ continue} \\ \sum_{\text{per tutti gli } x_j} \sum_{\text{per tutti gli } x_k} (x_j - \mu_j)(x_k - \mu_k) f_{jk}(x_j, x_k) & \text{se } X_j, X_k \text{ discrete} \end{cases}$$

$$F_{jk}(x_j, x_k) = P(X_j \le x_j, X_k \le x_k)$$

Covarianza e indipendenza

- Se $P(X_j \le x_j, X_k \le x_k) = P(X_j \le x_j) P(X_k \le x_k)$

➔ allora X_j *e* X_k *sono indipendenti*

- se X_j *e* X_k sono indipendenti

➔ $Cov(X_j, X_k) = 0$

In generale se $P(X_1 \le x_1, \ldots, X_p \le x_p) = P(X_1 \le x_1) \ldots P(X_p \le x_p)$
allora $X_1 \ldots X_p$ *sono mutualmente indipendenti*

Matrice di covarianza teorica

$$\boldsymbol{\Sigma} = E\left[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\right] = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix}$$

$\boldsymbol{\mu}' = \left[\mu_1, \ldots, \mu_p\right]$ Vettore delle medie della popolazione

Campionamento casuale

- Nel seguito supporremo che i dati siano relativi ad un campione casuale semplice (c.c.s.) di osservazioni da una data popolazione
- Questa assunzione implica: che
  - le *n* osservazioni siano indipendenti
  - la distribuzione congiunta delle *p* variabili sia la stessa per tutte le osservazioni

Media e varianza campionarie

- $\mathbf{X}_1, \ldots, \mathbf{X}_n$ c.c.s. da una distribuzione multivariata con vettore delle medie $\boldsymbol{\mu}$ e matrice di covarianza $\boldsymbol{\Sigma}$
- Allora:
  - il vettore delle medie campionarie è uno stimatore corretto di $\boldsymbol{\mu}$ con matrice di covarianza $(1/n)\,\boldsymbol{\Sigma}$.
  - $E(\mathbf{S}_n)=\boldsymbol{\Sigma}-(1/n)\,\boldsymbol{\Sigma} \rightarrow E(n/(n-1)\mathbf{S}_n)=\boldsymbol{\Sigma}$
  - Matrice di covarianza 'corretta' $\mathbf{S}=(n/(n-1))\mathbf{S}_n$

Analisi in componenti principali

- acp2012

...segue rappresentazioni grafiche

- Giovedì 29 marzo 2012 in LAB C

Esempio
misure di qualità della carta

- A causa dell'orientamento delle fibre di cellulosa rispetto al foglio, la resistenza della carta è differente se misurata nella direzione di produzione della macchina rispetto alla direzione opposta.
- Il file **carta.txt** contiene 41 misure relative a:
  - X1 densità della carta (grammi/cm cubi)
  - X2 resistenza nella direzione della macchina (pounds)
  - X3 resistenza nella direzione opposta (pounds)

carta_scatterplot.sas

Matrice dei diagrammi di dispersione

- Lungo la diagonale principale della matrice riportiamo i box-plot relativi a ogni variabile
- Costruiamo tanti scatter plot quante sono le possibili coppie di variabili **carta_scatterplot.sas**
- Osserviamo:
  - nel box-plot: un outlier per la densità del 25° foglio
  - Alcuni scatter plots hanno un andamento che suggerisce la presenza di due gruppi di osservazioni

Convex hull (guscio convesso) e box-plot bidimensionale

- Poligono convesso che si ottiene congiungendo tra loro i punti più esterni nelle diverse direzioni: è la più piccola figura convessa che contiene tutti i punti della 'nuvola' delle *n* unità statistiche
- Utile per individuare valori anomali
- Si possono tracciare 'gusci' via via più interni eliminando le osservazioni che giacciono sul guscio precedente.
- Il 'guscio' che racchiude almeno il 50% dei dati ha un ruolo analogo alla scatola del box-plot nel caso unidimensionale.

scatterplot_matrice.sas, eepels.sas

Rappresentazioni per icone delle unità statistiche

- Facce di Chernoff
- Diagrammi a stella, o stelle
  - permettono una chiara rappresentazione simultanea di più variabili (10 e oltre)
  - La superficie del poligono fornisce un'impressione immediata dell'ordine di grandezza dei valori delle variabili per una data u.s.
  - Facile confrontare diverse unità statistiche

Diagrammi a stella

- Supponiamo di avere n osservazioni non negative su $p \geq 2$ variabili. Possiamo costruire una circonferenza con p raggi equispaziati per ogni osservazione:
  - La lunghezza di ciascun raggio rappresenta il valore della variabile
  - I raggi possono essere connessi da una linea spezzata ottenendo una 'stella'
  - Ogni osservazione è rappresentata da una stella
  - Le stelle possono essere raggruppate in base alla somiglianza della loro forma

stelle.sas

Facce di Chernoff

- Chernoff (1973) suggerì di rappresentare le dimensioni p-dimesionali con delle 'facce' in due dimensioni, le cui caratteristiche (forma del viso, curva delle labbra, lunghezza del naso, dimensione degli occhi, posizione delle pupille ecc.) sono determinate per ogni unità statistica dal valore delle p variabili
- L'assegnazione delle variabili agli elementi della faccia è arbitraria: assegnazioni diverse producono risultati differenti. Di solito è necessario fare più prove prima di trovare una assegnazione soddisfacente

facce_utility.sas, chernoff.sas

Cosa abbiamo visto in questa introduzione

- Definizione di analisi multivariata
- La matrice dei dati
- Ripasso di algebra lineare
- Le matrici di covarianza e correlazione
- La varianza totale e la varianza generalizzata
- Alcune rappresentazioni grafiche di dati multidimensionali
- Studiare su ZANI cap. 1 e 2 (fino a §4. compreso), 5 (fino a §5. compreso