

4.1 Analisi in Componenti Principali (ACP)

Sia considerata la **matrice dei dati**, unità \times variabili:

$$\mathbf{X} = \begin{pmatrix} x_{11} & \dots & x_{1s} & \dots & x_{1p} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \dots & x_{is} & \dots & x_{ip} \\ \vdots & & \vdots & & \vdots \\ x_{n1} & \dots & x_{ns} & \dots & x_{np} \end{pmatrix}$$

vettore-variabile

ove x_{is} rappresenta la determinazione della s -esima variabile quantitativa osservata sull' i -esima unità statistica ($i=1,\dots,n$; $s=1,\dots,p$). In forma compatta:

$$\mathbf{X} = \{x_{is} : i = 1, \dots, n; s = 1, \dots, p\}.$$

Quando le p variabili sono **numerose** è molto difficile riuscire a cogliere le **strutture** esistenti nei dati.

Si pone quindi il **PROBLEMA**:

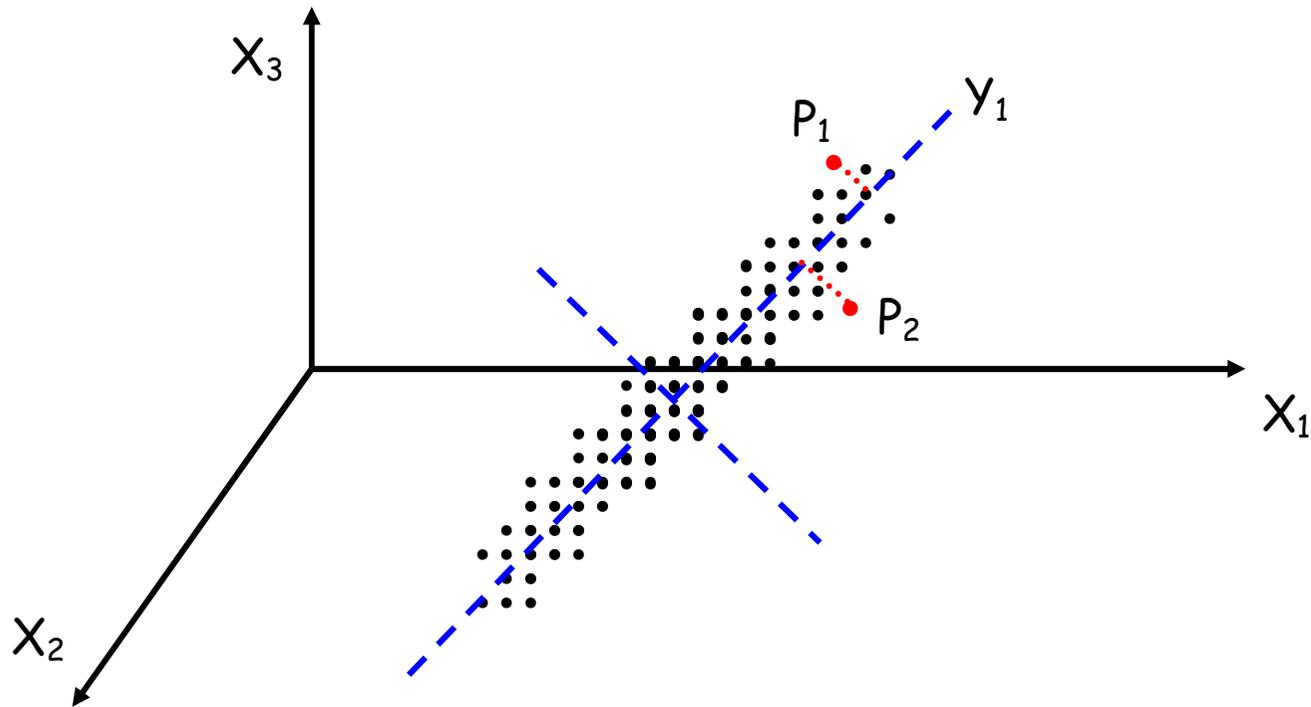
è possibile sostituire le p variabili originarie con un numero minore di variabili "artificiali" ($k \ll p$) (che impareremo a chiamare **COMPONENTI PRINCIPALI**) che garantiscono la **SINTESI** con la **MINOR PERDITA DI INFORMAZIONE POSSIBILE?**

Ossia, in termini geometrici, è possibile rappresentare le osservazioni, anziché nello spazio originario R^p , in uno **spazio di dimensioni ridotte** (R, R^2, R^3, \dots), con una **perdita limitata d'informazione?**

A tal riguardo consideriamo il seguente **ESEMPIO**.

Siano considerate 3 variabili X_1 , X_2 , X_3 osservate su n unità statistiche.

Rappresentiamo quindi le n osservazioni sullo spazio 3-dimensionale seguente:



◆ Se la nuvola dei punti presentasse una forma **ELLISSOIDALE** (simile ad un **SIGARO**) la dimensione più importante è la **LUNGHEZZA**, per cui si possono trascurare le altre dimensioni. Tuttavia, la sostituzione della **retta Y_1** allo spazio originario R^3 comporta una **perdita d'informazione**.

Infatti, ad un generico punto P_1 appartenente alla nuvola si sostituisce la sua **PROIEZIONE** sulla retta Y_1 , con la conseguenza che punti originariamente in posizioni opposte rispetto alla retta (e quindi distanti in R^3) possono avere proiezioni molto vicine o addirittura coincidenti su Y_1 .

◆ Se la nuvola dei punti avesse la forma di **OSSO DI SEPPIA**, occorrerebbe tener conto di 2 dimensioni: **LUNGHEZZA** E **LARGHEZZA**, trascurando lo **SPESSORE** che è molto minore rispetto alle altre 2 grandezze.

◆ Se la nuvola dei punti assumesse la forma **SFERICA** (variabili incorrelate) non sarebbe possibile nessuna riduzione delle dimensioni con limitata perdita d'informazione. Infatti tutte le 3 dimensioni (**LUNGHEZZA**, **LARGHEZZA** e **SPESSORE**) sono importanti.

Il discorso può **generalizzarsi** al caso in cui **$p > 3$** .

Osservazione

Si osserva che se le variabili sono tra loro **CORRELATE**, le dimensioni d'interesse sono in realtà minori di p (spesso molto minori).

È quindi possibile sostituire a R^p un nuovo spazio R^k , con $k \ll p$.

Risulta particolarmente utile il caso in cui $k=2$, poiché in tal caso le unità sono rappresentate sul piano cartesiano e ciò facilita la lettura della **CONFIGURAZIONE DEI DATI**.

Un **metodo statistico multivariato** che permette la **RIDUZIONE DELLE DIMENSIONI** con la **MINORE PERDITA D'INFORMAZIONE** è:

l'ANALISI IN COMPONENTI PRINCIPALI (ACP).

L'ACP è stata introdotta da **HOTELLING** (1933).

Tuttavia una soluzione al problema della riduzione delle dimensioni dello spazio osservazionale era già stata fornita da **K. PEARSON** nel 1901.

Abbiamo visto che l'obiettivo della sintesi può essere perseguito con l'ACP sostituendo le p variabili osservate con k ($k \ll p$) variabili artificiali che garantiscono la minore perdita d'informazione.

PROBLEMA: COME SCEGLIERE LE VARIABILI ARTIFICIALI CHE SINTETIZZANO, con la minor perdita d'informazione, le variabili osservate?

Il problema viene risolto assumendo che le variabili artificiali siano **COMBINAZIONE LINEARE** delle variabili osservate.

PROBLEMA: quale **COMBINAZIONE LINEARE** scegliere?

A tal riguardo osserviamo che il **CONTRIBUTO INFORMATIVO** fornito da una variabile statistica è legato alla sua **VARIABILITA'**.

Infatti una variabile statistica con **ELEVATA VARIABILITA'** fornisce di solito **PIU' INFORMAZIONE** di una con bassa variabilità, poiché tende ad essere "**DISPERSA**", cioè ad assumere modalità molto differenti tra loro.

Quindi, se si vuole costruire una variabile artificiale che sintetizzi, mediante una combinazione lineare, le variabili osservate occorre assicurare che essa fornisca un **CONTRIBUTO INFORMATIVO RILEVANTE** ossia abbia un'**ALTA VARIABILITA'**.

Poiché la misura di variabilità più semplice è la **VARIANZA**, il problema che si pone è quello di costruire una (o più) variabile artificiale che sia una **COMBINAZIONE LINEARE** delle variabili originarie e abbia **VARIANZA MASSIMA**.

Determinazione delle componenti principali

L'ACP è una metodologia statistica multivariata che, partendo da una matrice dei dati $n \times p$ con variabili quantitative, consente di sostituire alle p variabili (tra loro correlate) un nuovo insieme di variabili artificiali dette **COMPONENTI PRINCIPALI (CP)** che:

1. sono tra loro **INCORRELATE (ORTOGONALI)**;
2. sono **elencate in ordine decrescente rispetto alla loro varianza.**

- ◆ La **prima CP** Y_1 è la **COMBINAZIONE LINEARE** delle p variabili di partenza avente **MAX VARIANZA**.
- ◆ La **seconda CP** Y_2 è la **COMBINAZIONE LINEARE** delle p variabili di partenza con **VARIANZA IMMEDIATAMENTE INFERIORE**, soggetta al vincolo di essere **ORTOGONALE** alla CP precedente.
- ◆ La **terza CP** Y_3etc....

Se le p variabili sono **FORTEMENTE CORRELATE**, un numero $k \ll p$ di CP tiene conto di una **ELEVATA QUOTA DI VARIANZA TOTALE**.
Quindi, possiamo considerare solo tali k CP, **trascurando le restanti $p-k$** , ottenendo una **SENSIBILE PARSIMONIA** nella descrizione dei dati.

◆ La **DETERMINAZIONE DELLA 1° CP** y_1 richiede l'individuazione del vettore p-dimensionale \mathbf{a}_1 dei coefficienti della seguente **COMBINAZIONE LINEARE** delle p variabili espresse in termini di scostamento dalle loro medie:

$$\mathbf{y}_1 = \tilde{\mathbf{X}} \mathbf{a}_1$$

ove $\mathbf{a}_1 = (a_{11}, \dots, a_{1s}, \dots, a_{1p})'$; $\tilde{\mathbf{X}}$ = matrice dei dati centrata.

Per definizione la **prima CP** è la **COMBINAZIONE LINEARE** di **MASSIMA VARIANZA**.

Occorre quindi **determinare** il vettore dei coefficienti \mathbf{a}_1 tale che:

$$\max : \text{var}(\tilde{\mathbf{X}} \mathbf{a}_1) = \mathbf{a}_1' \mathbf{S} \mathbf{a}_1$$

ove \mathbf{S} = matrice di varianza e covarianza di $\tilde{\mathbf{X}}$.

Le **soluzioni** di tale problema di massimo sono però **INFINITE PROPORZIONALI**, poiché la combinazione lineare contiene un **FATTORE DI SCALA ARBITRARIO**.

Infatti, ad esempio, se anziché considerare il vettore \mathbf{a}_1 si considera $\mathbf{b}_1 = c\mathbf{a}_1$, $c > 1$, si ha:

$$\text{var}(\mathbf{y}_1) = \mathbf{b}'_1 \mathbf{S} \mathbf{b}_1 = c^2 \mathbf{a}'_1 \mathbf{S} \mathbf{a}_1 .$$

Quindi, per avere varianza max occorrerebbe far crescere indefinitamente c ($c \rightarrow \infty$); in tal caso infatti:

$$\text{var}(\mathbf{y}_1) \rightarrow \infty .$$

Al fine di individuare una **SOLUZIONE FINITA**, occorre quindi porre un **VINCOLO SULLE COMPONENTI DEL VETTORE** \mathbf{a}_1 , in modo da assicurarsi che nessuna di esse possa diventare il valore assoluto infinitamente grande.

Si considera quindi il **VINCOLO DI NORMALIZZAZIONE**:

$$\mathbf{a}'_1 \mathbf{a}_1 = 1.$$

Quindi, per **determinare la 1° CP** si deve risolvere il seguente **PROBLEMA DI MAX VINCOLATO**:

$$\begin{cases} \max : \text{var}(\tilde{\mathbf{X}}\mathbf{a}_1) = \mathbf{a}'_1 \mathbf{S} \mathbf{a}_1 \\ \mathbf{a}'_1 \mathbf{a}_1 = 1. \end{cases}$$

Soluzione del problema

Il problema può essere risolto considerando il **METODO DEI MULTIPLICATORI DI LAGRANGE**.

La **FUNZIONE DI LAGRANGE** è:

$$L = \mathbf{a}'_1 \mathbf{S} \mathbf{a}_1 - \lambda (\mathbf{a}'_1 \mathbf{a}_1 - 1) \quad (1)$$

ove λ = moltiplicatore di Lagrange. Annullando le derivate prime parziali di (1) rispetto ad \mathbf{a}_1 si ottiene:

$$\frac{\partial L}{\partial \mathbf{a}_1} = 2\mathbf{S}\mathbf{a}_1 - 2\lambda\mathbf{a}_1 = 0$$

ossia

$$(\mathbf{S} - \lambda\mathbf{I})\mathbf{a}_1 = \mathbf{0} \quad (2)$$

ove \mathbf{I} =matrice identità $p \times p$; $\mathbf{0}$ =vettore p -dimensionale nullo.

Il sistema lineare di p equazioni e p incognite precedente è detto **SISTEMA DI EQUAZIONI AGLI AUTOVALORI**.

Poiché ci riferiamo alla 1° CP possiamo porre $\lambda = \lambda_1$:

$$(\mathbf{S} - \lambda_1 \mathbf{I}) \mathbf{a}_1 = \mathbf{0} . \quad (2)$$

Il sistema ammette soluzioni non tutte nulle se il suo determinante è uguale a zero:

$$|\mathbf{S} - \lambda_1 \mathbf{I}| = 0 \quad (3)$$

che definisce l'**EQUAZIONE CARATTERISTICA DELLA MATRICE S** con p soluzioni chiamate **AUTOVALORI**.

Determinati gli autovalori da (3) e inseriti in (2) si ottiene così l'**AUTOVETTORE** \mathbf{a}_1 e quindi la 1° CP \mathbf{y}_1 .

Essendo la matrice di varianza e covarianza \mathbf{S} **SEMIDEFINITA POSITIVA**, gli **AUTOVALORI** sono **TUTTI NON NEGATIVI**.

Inoltre poiché \mathbf{S} è **SIMMETRICA** ammetterà solo **SOLUZIONI REALI (NON COMPLESSE)**.

Ora, poiché l'obiettivo è la massimizzazione della varianza della 1° CP si sceglie come λ_1 il massimo degli autovalori, in quanto vale la seguente uguaglianza (ottenuta da (2) pre-moltiplicando per \mathbf{a}'_1 ed essendo $\mathbf{a}'_1 \mathbf{a}_1=1$):

$$\text{var}(\tilde{\mathbf{X}}\mathbf{a}_1) = \mathbf{a}'_1 \mathbf{S} \mathbf{a}_1 = \mathbf{a}'_1 \lambda_1 \mathbf{a}_1 = \lambda_1.$$

Il primo autovalore λ_1 è dunque uguale alla varianza della prima CP.

Quindi:

Definizione

Si definisce 1° CP la combinazione lineare

$$\mathbf{y}_1 = \tilde{\mathbf{X}}\mathbf{a}_1$$

in cui \mathbf{a}_1 è l'autovettore corrispondente all'autovalore più grande λ_1 della matrice \mathbf{S} .

◆ La **DETERMINAZIONE DELLA 2° CP** y_2

$$y_2 = \tilde{X} a_2$$

avviene considerando i **2 vincoli**:

$$a_2' a_2 = 1 \quad (\text{vincolo di normalizzazione})$$

$$a_1' a_2 = 0 \quad (\text{vincolo di ortogonalità})$$

nel problema di max della varianza della 2° CP

$$\text{var}(\tilde{X} a_2) = a_2' S a_2 .$$

Il **problema di max vincolato** è:

$$\left\{ \begin{array}{l} \max : \text{var}(\tilde{X} a_2) = a_2' S a_2 \\ a_2' a_2 = 1 \\ a_1' a_2 = 0 . \end{array} \right.$$

Soluzione del problema

Il problema viene risolto considerando, come in precedenza, il **metodo dei moltiplicatori di Lagrange**.

In tal caso la **funzione di Lagrange** è:

$$L = \mathbf{a}'_2 \mathbf{S} \mathbf{a}_2 - \lambda (\mathbf{a}'_2 \mathbf{a}_2 - 1) - \gamma \mathbf{a}'_1 \mathbf{a}_2 \quad (3)$$

ove λ, γ sono i moltiplicatori di Lagrange.

Considerando le derivate prime parziali di (3) rispetto ad \mathbf{a}_2 ed annullandole, si ottiene:

$$\frac{\partial L}{\partial \mathbf{a}_2} = 2 \mathbf{S} \mathbf{a}_2 - 2 \lambda \mathbf{a}_2 - \gamma \mathbf{a}_1 = 0. \quad (4)$$

Pre-moltiplicando per \mathbf{a}'_1 si ottiene:

$$2\mathbf{a}'_1\mathbf{S}\mathbf{a}_2 - 2\lambda\mathbf{a}'_1\mathbf{a}_2 - \gamma\mathbf{a}'_1\mathbf{a}_1 = 0.$$

Poiché $\mathbf{a}'_1\mathbf{a}_1 = 1$ (condizione di normalizzazione sulla 1° CP) e $\mathbf{a}'_2\mathbf{a}_1 = 0$ (condizione di ortogonalità) si ha:

$$\gamma = 2\mathbf{a}'_1\mathbf{S}\mathbf{a}_2.$$

Ora, poiché dalla $(\mathbf{S} - \lambda\mathbf{I})\mathbf{a}_1 = \mathbf{0}$ (relativa alla 1°CP) ossia da $\mathbf{S}\mathbf{a}_1 = \lambda\mathbf{a}_1$ si ottiene, pre-moltiplicando per \mathbf{a}'_2

$$\mathbf{a}'_2\mathbf{S}\mathbf{a}_1 = \lambda\mathbf{a}'_2\mathbf{a}_1 \text{ ossia } \mathbf{a}'_1\mathbf{S}\mathbf{a}_2 = \lambda\mathbf{a}'_1\mathbf{a}_2$$

e poiché $\mathbf{a}'_1\mathbf{a}_2 = 0$ (condizione ortogonalità) si ottiene

$$\mathbf{a}'_1\mathbf{S}\mathbf{a}_2 = 0 \text{ e quindi } \gamma = 0.$$

In tal modo, il sistema (4) diviene:

$$(\mathbf{S} - \lambda \mathbf{I}) \mathbf{a}_2 = \mathbf{0}$$

per cui si sceglie il secondo autovalore, in ordine decrescente della matrice di varianza e covarianza e lo si indica con λ_2 ; l'autovettore corrispondente è \mathbf{a}_2 .

In tal caso si ottiene (pre-moltiplicando per \mathbf{a}'_2 e poiché $\mathbf{a}'_2 \mathbf{a}_2 = 1$):

$$\text{var}(\tilde{\mathbf{X}} \mathbf{a}_2) = \mathbf{a}'_2 \mathbf{S} \mathbf{a}_2 = \mathbf{a}'_2 \lambda_2 \mathbf{a}_2 = \lambda_2$$

ossia il secondo autovalore λ_2 è uguale alla varianza della seconda CP.

Definizione

Si definisce 2° CP la combinazione lineare

$$\mathbf{y}_2 = \tilde{\mathbf{X}} \mathbf{a}_2$$

in cui \mathbf{a}_2 è l'autovettore corrispondente all'autovalore più grande (dopo λ_1) λ_2 della matrice \mathbf{S} .

In modo analogo si definiscono le successive CP.

Definizione

Si definisce v -esima CP la combinazione lineare

$$\mathbf{y}_v = \tilde{\mathbf{X}} \mathbf{a}_v \quad v=1, \dots, k \leq p$$

in cui \mathbf{a}_v è l'autovettore associato al v -esimo autovalore λ_v , in ordine decrescente, della matrice \mathbf{S} .

Alcune proprietà

Proprietà 1

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_v \geq \dots \geq \lambda_k \geq \dots \geq \lambda_p .$$

Proprietà 2

Poiché $\lambda_v = \text{var}(\mathbf{y}_v)$ per la Proprietà 1 si ha:

$$\text{var}(\mathbf{y}_1) \geq \text{var}(\mathbf{y}_2) \geq \dots \geq \text{var}(\mathbf{y}_v) \geq \dots \geq \text{var}(\mathbf{y}_k) \geq \dots \geq \text{var}(\mathbf{y}_p)$$

Proprietà 3

$$\text{tr}(\mathbf{S}) = \sum_{v=1}^p \lambda_v$$

ossia la somma delle varianze delle CP è uguale alla somma delle varianze delle variabili originarie (=traccia di \mathbf{S}).

(Si dimostra utilizzando il *Teorema di decomposizione spettrale*).

Proprietà 4

$$\frac{\lambda_v}{\text{tr}(\mathbf{S})} 100$$

rappresenta la **PERCENTUALE DI VARIANZA TOTALE SPIEGATA DALLA v-ESIMA CP y_v** .

Proprietà 5

$$\frac{\lambda_1 + \dots + \lambda_v + \dots + \lambda_k}{\text{tr}(\mathbf{S})} 100$$

rappresenta la **PERCENTUALE DI VARIANZA TOTALE SPIEGATA DALLE K CP**.

Proprietà 6

Il **COEFFICIENTE DI CORRELAZIONE LINEARE** tra la v -esima CP e la s -esima **VARIABILE** è:

$$r(Y_v, X_s) = \frac{a_{vs} \sqrt{\lambda_v}}{\text{var}(X_s)}$$

(cfr. dim. in Zani).

Proprietà 7

Se le variabili originarie sono tutte (esattamente) incorrelate esse sono esattamente uguali alle CP.

Punteggi (scores) e pesi (loadings)

La prima CP è definita nel modo seguente:

$$\mathbf{y}_1 = \tilde{\mathbf{X}} \mathbf{a}_1 .$$

Il **PUNTEGGIO (SCORE)** della 1° CP per l'*i*-esima unità statistica è:

$$y_{i1} = a_{11} \tilde{x}_{i1} + \dots + a_{1s} \tilde{x}_{is} + \dots + a_{1p} \tilde{x}_{ip} \quad i = 1, \dots, n$$

ove a_{1s} = **COEFFICIENTE (LOADING)** della 1° CP e dell'*s*-esima variabile (cioè *s*-esimo elemento del 1° autovettore).

Esso fornisce il **PESO** assegnato all'*s*-esima variabile nella definizione della 1° CP.

Il **SEGNO** di tale coefficiente indica il **TIPO DI RELAZIONE** (**DIRETTA** o **INDIRETTA**) tra la 1° CP e la s-esima VARIABILE e il suo **VALORE NUMERICO** indica in quale misura tale variabile concorre alla **DETERMINAZIONE** dei punteggi della 1° CP.

In generale, considerando le prime k CP, la **MATRICE DEGLI SCORES**, di dimensione $n \times k$, è:

$$Y = \tilde{X} A .$$

Osservazione

Poiché le CP sono ottenute in ordine decrescente, i punteggi della 1° CP presentano **MAGGIORE VARIABILITA'** rispetto a quelli della 2° CP, e così via. Quindi, gli SCORES di ogni CP hanno lo **SVANTAGGIO** di **NON ESSERE DIRETTAMENTE COMPARABILI** per CP differenti (alle quali corrispondono autovalori diversi).

Per risolvere tale **INCONVENIENTE** si possono ricavare **PUNTEGGI** delle CP **STANDARDIZZATI** (con media nulla e varianza unitaria), dividendo gli score per la radice quadrata del rispettivo autovalore. Infatti, ad esempio:

$$\text{var}\left(\frac{y_1}{\sqrt{\lambda_1}}\right) = \frac{1}{\lambda_1} \text{var}(y_1) = \frac{1}{\lambda_1} \lambda_1 = 1.$$

Esempio numerico

Data una matrice di dati $\tilde{\mathbf{X}}$, supponiamo che la sua matrice di varianza e covarianza \mathbf{S} sia (2 variabili)

$$\mathbf{S} = \begin{pmatrix} 6 & -2 \\ -2 & 3 \end{pmatrix}.$$

Da $|\mathbf{S} - \lambda \mathbf{I}| = 0$ (equazione caratteristica) si ottiene

$$\left| \begin{pmatrix} 6 & -2 \\ -2 & 3 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = 0$$

ossia

$$\begin{vmatrix} 6 - \lambda & -2 \\ -2 & 3 - \lambda \end{vmatrix} = 0$$

e quindi:

$$(6 - \lambda) \cdot (3 - \lambda) - 4 = 0$$

da cui

$$\lambda^2 - 9\lambda + 14 = 0.$$

Le soluzioni sono:

$$\lambda_{1/2} = \frac{9 \pm \sqrt{81 - 56}}{2} = \frac{9 \pm 5}{2} = \left\langle \begin{array}{l} 7 \\ 2 \end{array} \right.$$

Scegliamo l'autovalore più grande: $\lambda_1 = 7$.

Si noti che $9 = \text{tr}(\mathbf{S}) = \sum_{\nu=1}^2 \lambda_{\nu} = 9$.

Inserendo i valori numerici nel sistema di equazioni agli autovalori

$$(\mathbf{S} - \lambda_1 \mathbf{I}) \mathbf{a}_1 = \mathbf{0}$$

si ottiene

$$\left[\left(\begin{array}{cc} 6 & -2 \\ -2 & 3 \end{array} \right) - 7 \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) \right] \begin{pmatrix} a_{11} \\ a_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

da cui

$$\begin{pmatrix} -1 & -2 \\ -2 & -4 \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

e quindi si ottiene il sistema di equazioni:

$$\begin{cases} -a_{11} - 2a_{12} = 0 \\ -2a_{11} - 4a_{12} = 0 \end{cases} \Leftrightarrow \begin{cases} a_{11} = -2a_{12} \\ a_{12} = -\frac{1}{2}a_{11} \end{cases}$$

che presenta infinite soluzioni.

Consideriamo quindi il vincolo di normalizzazione

$$\mathbf{a}'_1 \mathbf{a}_1 = 1 \quad (a_{11}^2 + a_{12}^2 = 1).$$

Quindi

$$\begin{cases} a_{11} = -2a_{12} \\ a_{11}^2 + a_{12}^2 = 1 \end{cases}$$

da cui

$$4a_{12}^2 + a_{12}^2 = 1$$

e quindi

$$a_{12} = \frac{1}{\sqrt{5}} \quad a_{11} = \frac{-2}{\sqrt{5}}$$

ossia

$$\mathbf{a}_1 = \left(\frac{-2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right)'$$

Analogamente da

$$(\mathbf{S} - \lambda_2 \mathbf{I}) \mathbf{a}_2 = \mathbf{0}$$

si ottiene

$$\begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

e quindi il sistema di equazioni

$$\begin{cases} 4 a_{21} - 2 a_{22} = 0 \\ -2 a_{21} + a_{22} = 0 \end{cases} \Leftrightarrow \begin{cases} a_{21} = \frac{1}{2} a_{22} \\ a_{22} = 2 a_{21} \end{cases}$$

che presenta soluzioni infinite. Quindi col vincolo di normalizzazione

$$\mathbf{a}'_2 \mathbf{a}_2 = 1 \quad (a_{21}^2 + a_{22}^2 = 1)$$

si ha:

$$\begin{cases} a_{21} = \frac{1}{2} a_{22} \\ a_{21}^2 + a_{22}^2 = 1 \end{cases}$$

da cui

$$\frac{1}{4} a_{22}^2 + a_{22}^2 = 1$$

e quindi

$$a_{22} = \frac{2}{\sqrt{5}} \quad a_{21} = \frac{1}{\sqrt{5}}$$

ossia

$$\mathbf{a}_2 = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right)'$$

Si noti che $\mathbf{a}_1 = \left(\frac{-2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right)'$ e $\mathbf{a}_2 = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right)'$ sono ortogonali.

Infatti

$$\mathbf{a}_1' \mathbf{a}_2 = \left(-\frac{2}{\sqrt{5}} \quad \frac{1}{\sqrt{5}} \right) \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} = -\frac{2}{\sqrt{5}} + \frac{2}{\sqrt{5}} = 0.$$

L'ACP basata sulla matrice di correlazione

Abbiamo finora applicato l'ACP sulla matrice di varianza e covarianza S . Ciò ha senso se:

1. le variabili osservate sono espresse nella **medesima unità di misura**;
2. le variabili osservate sono dello **stesso ordine di grandezza** (infatti un cambiamento di scala di una variabile modifica il valore della varianza totale e quindi influenza fortemente i risultati dell'ACP).

Altrimenti occorre basare l'ACP sulla **MATRICE DI CORRELAZIONE**.

Infatti la difficoltà precedente viene superata considerando **VARIABILI STANDARDIZZATE**, che equivale a basare l'ACP sulla **MATRICE DI CORRELAZIONE**.

Nelle **APPLICAZIONI** tale situazione è molto frequente.

Osservazione

Le CP del medesimo dataset ottenute dalla **MATRICE DI VARIANZA E COVARIANZA** o dalla **MATRICE DI CORRELAZIONE** non sono le stesse.

Quindi la scelta della matrice su cui basare l'ACP è molto importante.

La procedura di calcolo delle CP basata sulla matrice di correlazione segue le linee già descritte, tenendo presente che la varianza totale di p variabili standardizzate è uguale a p :

$$\text{var}(\mathbf{Z}) = \text{tr}(\mathbf{R}) = p.$$

Definizione

Si definisce v -esima CP di p variabili standardizzate la combinazione lineare

$$\mathbf{y}_v = \mathbf{Z}\mathbf{a}_v \quad v=1,\dots,k \leq p$$

in cui \mathbf{a}_v è l'autovettore associato al v -esimo autovalore λ_v , in ordine decrescente, della matrice di correlazione \mathbf{R} .

In tal caso si ha, schematicamente, ad esempio per le prime 2 CP:

$$\mathbf{y}_1 = \mathbf{Z}\mathbf{a}_1 \quad \begin{cases} \max : \text{var}(\mathbf{y}_1) = \text{var}(\mathbf{Z}\mathbf{a}_1) = \mathbf{a}_1' \mathbf{R} \mathbf{a}_1 \\ \mathbf{a}_1' \mathbf{a}_1 = 1 \end{cases} \quad \begin{aligned} (\mathbf{R} - \lambda_1 \mathbf{I}) \mathbf{a}_1 &= \mathbf{0} \\ |\mathbf{R} - \lambda_1 \mathbf{I}| &= 0 \end{aligned}$$

$$\mathbf{y}_2 = \mathbf{Z}\mathbf{a}_2 \quad \begin{cases} \max : \text{var}(\mathbf{y}_2) = \text{var}(\mathbf{Z}\mathbf{a}_2) = \mathbf{a}_2' \mathbf{R} \mathbf{a}_2 \\ \mathbf{a}_2' \mathbf{a}_2 = 1 \\ \mathbf{a}_1' \mathbf{a}_2 = 0 \end{cases} \quad \begin{aligned} (\mathbf{R} - \lambda_2 \mathbf{I}) \mathbf{a}_2 &= \mathbf{0} \\ |\mathbf{R} - \lambda_2 \mathbf{I}| &= 0. \end{aligned}$$

Interpretazione delle CP

Distinguiamo **2 CRITERI**:

- ◆ criterio basato sui **COEFFICIENTI** di ogni componente principale a_{vs} .
- ◆ criterio basato sui **COEFFICIENTI DI CORRELAZIONE** $r(Y_v, X_s)$.

◆ Criterio basato sui **COEFFICIENTI** di ogni componente principale a_{vs}

Il generico coefficiente a_{vs} misura il peso assunto dalla s -esima variabile X_s nella determinazione della v -esima CP Y_v .

Infatti quanto più grande è a_{vs} (in valore assoluto) tanto maggiore sarà il peso che z_{is} hanno nella determinare il punteggio y_{iv}

$$y_{iv} = a_{v1}z_{i1} + \dots + a_{vs}z_{is} + \dots + a_{vp}z_{ip} \quad i = 1, \dots, n .$$

Ciò significa che la CP Y_v sarà maggiormente caratterizzata dalla variabile s -esima X_s a cui corrispondono i coefficienti a_{vs} più grandi in valore assoluto. In tal modo sono proprio tali variabili a conferire un significato alla CP Y_v .

◆ **Criterio basato sui COEFFICIENTI DI CORRELAZIONE $r(Y_v, X_s)$**

Abbiamo visto che

$$r(Y_v, X_s) = \frac{a_{vs} \sqrt{\lambda_v}}{\text{var}(X_s)}.$$

Quindi quanto più $r(Y_v, X_s)$ è elevato, in valore assoluto, tanto maggiore è il legame lineare tra la v -esima CP Y_v e la s -esima variabile X_s .

Ciò significa che a determinare il significato di Y_v saranno le variabili X_s , $s=1, \dots, p$, con cui è maggiormente correlata, cioè le variabili X_s , $s=1, \dots, p$, a cui corrispondono i più elevati (in valore assoluto) coefficienti di correlazione $r(Y_v, X_s)$.

Scelta del numero delle CP

Come abbiamo visto, l'obiettivo dell'ACP è quello di ridurre le dimensioni dello spazio originario R_p , rappresentando le unità in un nuovo spazio R_k .

PROBLEMA: Come determinare il numero di CP k da considerare?

Basandoci sull'ACP condotta sulla matrice di correlazione, distinguiamo **3 CRITERI**.

- ◆ Quota di varianza totale spiegata
- ◆ *Scree-graph*
- ◆ *Eigenvalue one* o Regola di Kaiser.

Di solito, nelle applicazioni l'individuazione del numero k delle CP da utilizzare avviene utilizzando congiuntamente i precedenti criteri.

◆ Quota di varianza totale spiegata

Si considera un numero di CP tale che esse tengano conto di una percentuale sufficientemente elevata (ad esempio, almeno l'80%) della varianza totale.

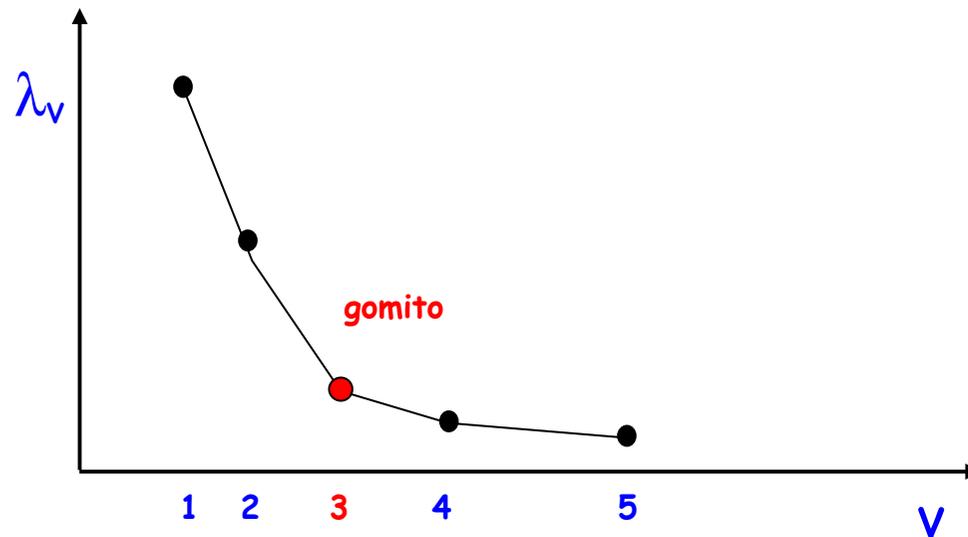
Questo criterio può essere perfezionato variando la soglia della percentuale suddetta in funzione del numero di variabili originarie: al crescere di p aumenta la varianza totale e quindi può essere ragionevole accontentarsi di una percentuale minore di varianza spiegata.

◆ *Scree-graph*

Si costruisce un grafico, chiamato *Scree-graph* o *Scree-plot*, degli autovalori λ_v in funzione del numero v di CP ($v=1, \dots, p$).

Essendo gli autovalori ottenuti in ordine decrescente, tale grafico si presenta nella forma di una spezzata sempre discendente; se k CP sono importanti e le restanti $(p-k)$ trascurabili, tra k e $k+1$ si manifesta una brusca variazione di pendenza (gomito), che segnala che k è il numero opportuno di CP da considerare.

Esempio



Osservazione 1

Non sempre l'andamento del grafico fornisce una risposta univoca, poiché la diminuzione degli autovalori può essere graduale, senza salti evidenti.

Osservazione 2

Alcuni autori suggeriscono di **escludere** tra le CP scelte quelle sul gomito (Harman, 1976). Altri suggeriscono di **includere** tra le CP scelte quelle sul gomito (Cattell, 1966).

◆ *Eigenvalue one* o Regola di Kaiser

Si considerano tutte le CP il cui autovalore è maggiore di 1.

La "ratio" di questo criterio deriva dal fatto che l'autovalore di una CP è uguale alla sua varianza e che operando su variabili standardizzate queste hanno varianza unitaria.

Pertanto, si decide di mantenere una CP solo se essa spiega una quota di varianza totale maggiore di quella di una singola variabile.