CAPITOLO 8 – Valutazione delle performance economico-finanziarie delle imprese
Paragrafo 8.4 - Analisi in componenti principali (ACP)

8

PEARSON

Contenuto

- La logica della riduzione delle variabili mediante l'ACP
- Proprietà algebriche delle componenti principali (CP)
- Esempio numerico
- Le fasi di un'ACP

Analisi multivariata dei ratio di bilancio: la riduzione delle variabili

Per una corretta interpretazione dei dati derivati dal bilancio, è opportuno considerare simultaneamente più indici di bilancio (colonne della matrice dei dati).

Per facilitare l'interpretazione → riduzione delle variabili da p *a* $k<p$

Matrice dei dati

Impresa	Indice 1	Indice 2	…	Indice *j*	…	Indice *p*
U_1	x_{11}	x_{12}	…	x_{1j}	…	x_{1p}
U_2	x_{21}	x_{22}	…	x_{2j}	…	x_{2p}
…	…	…	…	…	…	…
U_i	x_{i1}	x_{i2}	…	x_{ij}	…	x_{ip}
…	…	…	…	…	…	…
U_n	x_{n1}	x_{n2}	…	x_{nj}	…	x_{np}

La riduzione delle variabili: ACP (analisi delle componenti principali)

Obiettivo.

Interpretare la situazione reddituale-finanziaria complessiva tenendo conto di p indici di bilancio attraverso la riduzione del numero di variabili da p indici di bilancio (variabili originarie) *a* $k<p$ **nuove variabili.**

Come sono fatte le nuove variabili.

Sono dette **componenti principali** e sono **combinazione lineare** delle p variabili originarie.

PEARSON

Proprietà algebriche delle CP

Algebricamente le componenti principali (CP) sono combinazioni lineari delle variabili originarie e il valore assunto dalla componente su ogni unità è detto **score**.

Con p variabili correlate si possono al massimo ricavare p CP.

$$C_{ih} = a_{h1}x_{i1} + \ldots + a_{hj}x_{ij} + \ldots + a_{hp}x_{ip} = \sum_{j=1}^{p} a_{hj}x_{ij}$$

è lo score della CP *h-esima* per l'unità *i-esima;*

a_{hj} è il coefficiente da applicare alla variabile *j-esima* per il calcolo nella CP *h-esima*

Sono i dati relativi alle variabili x a determinare i coefficienti a_{hj}

PEARSON

Proprietà algebriche delle CP

I coefficienti a_{hj} sono tali che le CP posseggono le seguenti proprietà:

1. le CP sono fra loro **incorrelate**;

2. le CP sono ordinate **gerarchicamente** rispetto alla **varianza**:

$$var(\mathrm{C}_1) \geq var(\mathrm{C}_2) \geq \ldots \geq var(\mathrm{C}_\mathrm{p})$$

3. la **somma** delle varianze delle CP è uguale alla somma delle varianze delle variabili originarie

$$\sum_{j=1}^{p} var(x_j) = \sum_{h=1}^{p} var(C_h)$$

Se le *p* variabili originarie sono **standardizzate**, ciascuna ha varianza **1** e la somma delle varianze è uguale a *p*

Ancora sulle proprietà algebriche delle CP

La prima componente principale C_1 è la combinazione lineare delle p variabili di partenza, avente la *massima varianza* (v. proprietà 2)

La seconda componente principale C_2 è la combinazione lineare delle p variabili di partenza, avente varianza *immediatamente inferiore* alla varianza di C_1 e ad essa ***incorrelata*** (v. proprietà 2 e 1)

E così via fino alla p-esima componente.

Tutte le p CP riproducono la somma delle varianze delle variabili originarie (v. proprietà 3) ma le **prime** CP tengono conto della **maggior parte** della variabilità nelle variabili originali.

Scegliendo le $k<p$ CP con varianza più elevata si realizza la **riduzione** dei dati.

L'efficacia riduttiva è maggiore quanto **più correlate** sono le variabili.

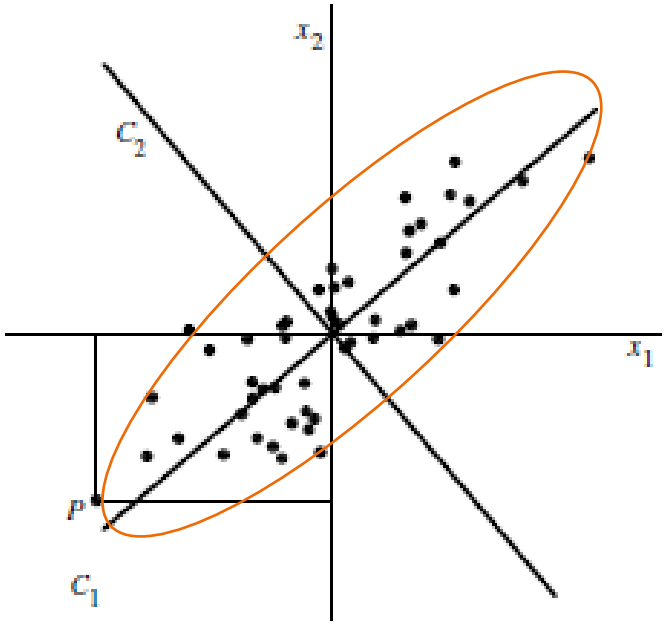
Interpretazione geometrica delle CP

Geometricamente le CP rappresentano un nuovo sistema di coordinate ottenuto ruotando gli assi originali che hanno le variabili originarie come coordinate.

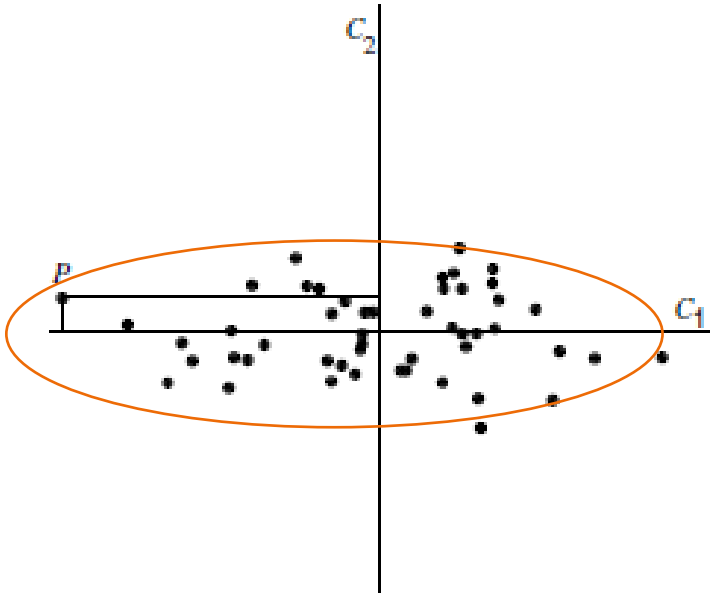
I nuovi assi rappresentano le **direzioni principali** cioè le direzioni di **variabilità massima**.

Caso p=2

Assi originali



Componenti principali



Come si determinano i coefficienti delle CP

Le varianze e i coefficienti delle CP si determinano derivando gli autovalori e gli autovettori associati della matrice di varianza e covarianza **S** delle variabili originarie e cioè risolvendo il seguente sistema di equazioni:

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

S è la matrice *pxp* di varianza e covarianza delle *p* variabili (indici di bilancio)

I: matrice unitaria *pxp*

λ: autovalore di S (varianza della CP)

$\underline{a}$: autovettore di S associato all'autovalore (coefficienti della CP)

Per dettagli si rimanda al testo (esempio *p=2*)

Esempio numerico (p=2 e n=6)

Imprese	BETA (x_1)	CR (x_2)
U_1 (Alfa)	1.502	1.617
U_2	2.953	0.673
U_3	0.957	1.693
U_4	1.631	0.584
U_5	2.205	0.586
U_6	2.016	0.508
Media	1.877	0.944

Matrice di covarianza variabili originarie

	BETA	CR
BETA	0.389	-0.220
CR	-0.220	0.256

BETA (rapporto di indebitamento)

$$\frac{\text{totale debiti}}{\text{patrimonio netto}}$$

CR (*current ratio*)

$$\frac{\text{attivo circolante}}{\text{passività correnti}}$$

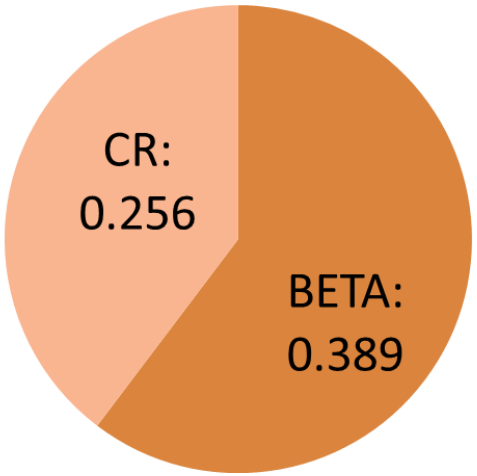
PEARSON

La riduzione dei dati (esempio *continuazione*)

Variabili originarie

Somma varianze:

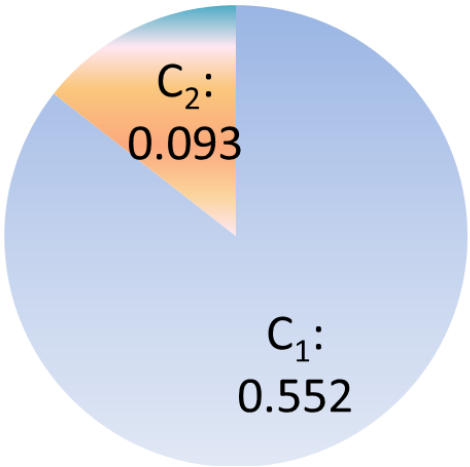
0.389+0.256=**0.645**



Componenti principali

$$\begin{cases} C_1 = -0.803 \times \text{BETA} + 0.596 \times \text{CR} \\ C_2 = -0.596 \times \text{BETA} - 0.803 \times \text{CR} \end{cases}$$

Somma varianze: 0.552+0.093=**0.645**



La prima CP da sola riproduce oltre l'85.7% della varianza totale delle variabili originarie (0.552/0.645).

PEARSON

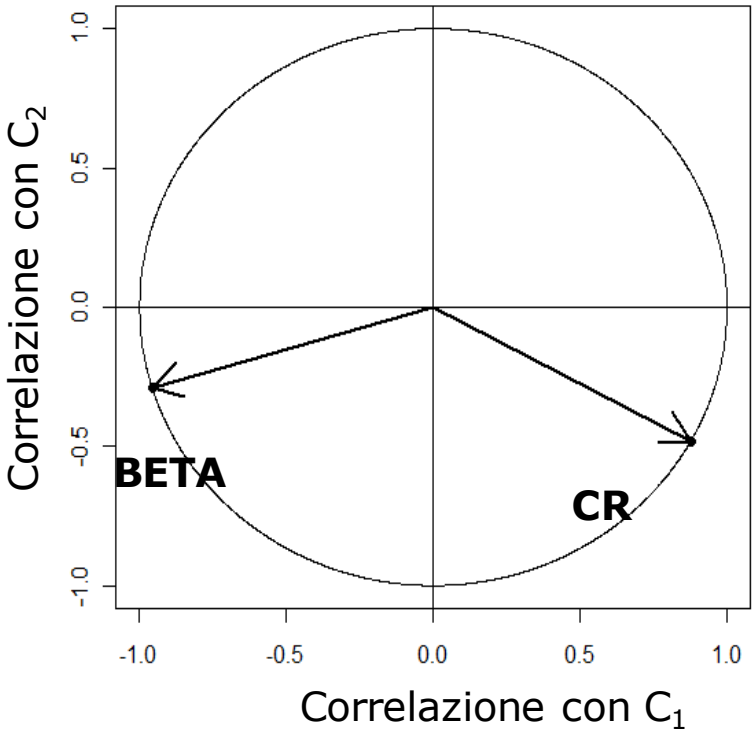
Interpretazione delle CP

In genere, si esaminano segno e ordine di grandezza della **correlazione** fra queste e ognuna delle variabili originarie. Ci si serve del **grafico delle correlazioni**.

Correlazione fra variabili e componenti

	C1	C2
BETA	-0.957	-0.290
CR	0.876	-0.480

All'aumentare di C_1, CR tende ad aumentare mentre BETA tende a diminuire



PEARSON

Calcolo score delle CP

NOTA BENE - Gli score delle CP vengono calcolati sui valori delle variabili originarie espressi come **scarti** dalle rispettive medie

Esempio: calcolo score di C_1 per l'impresa **Alfa**

Valori delle variabili originarie come scarti dalla media:

BETA*=1.502 – 1.877= –0.375

CR*= 1.617–0.944 = 0.673

Calcolo score:

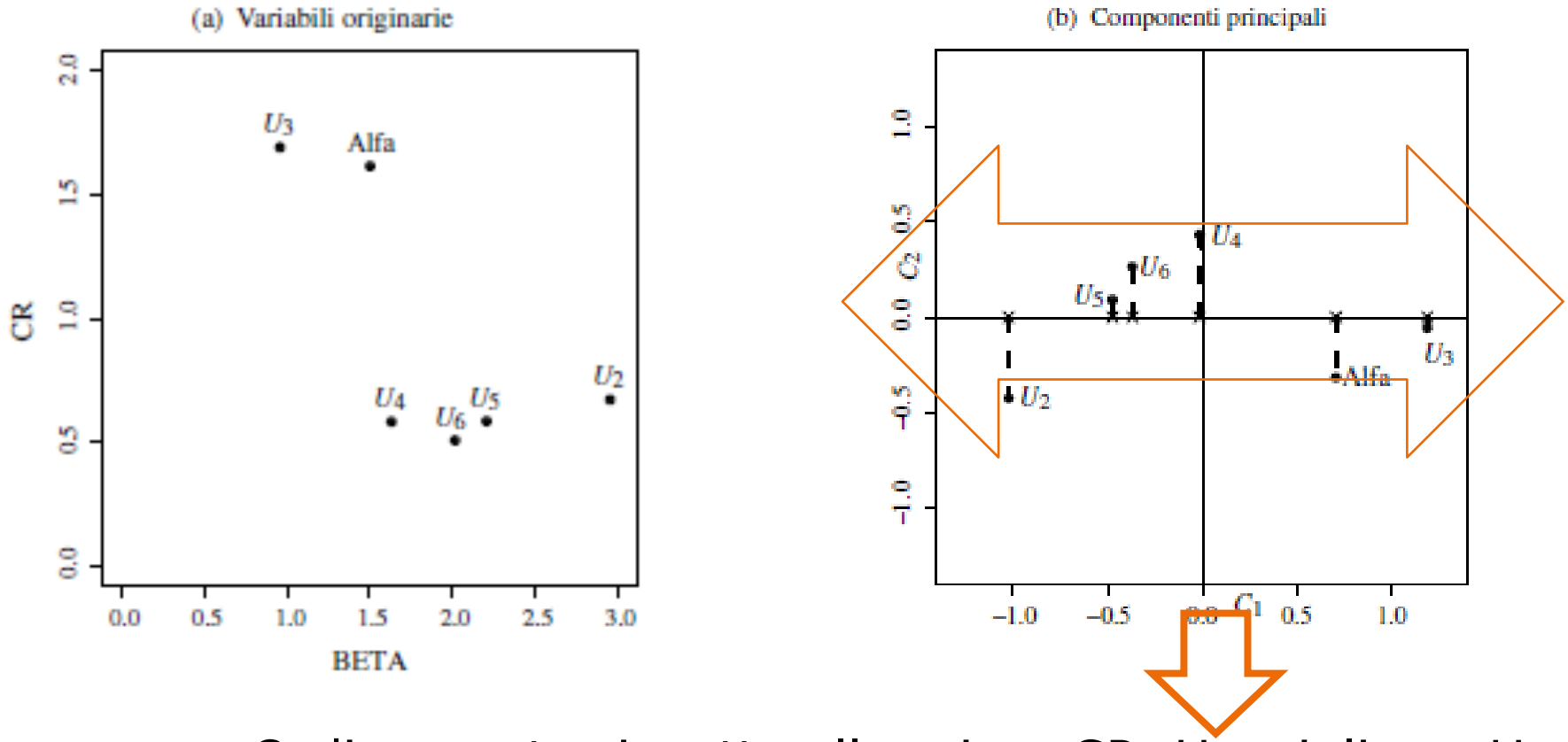
$$C_1 = -0.803 \times (-0.375) + 0.596 \times 0.673$$

PEARSON

Comparazione fra unità sulla base delle CP

Score della prima CP

	Alfa (U_1)	U_2	U_3	U_4	U_5	U_6
C_1	0.703	-1.025	1.186	-0.017	-0.476	-0.371



Ordinamento rispetto alla prima CP: U_3 migliore, U_2 peggiore

PEARSON

Fasi di una ACP

1. Scelta delle variabili

L'obiettivo di studio e il significato e la qualità dei dati sugli indici di bilancio guidano in questo senso.

2. Calcolo delle CP

La questione da risolvere è se usare i valori **originari** o quelli **standardizzati**. Ciò dipende da vari fattori:

i) scala delle variabili (espresse nella stessa unità di misura: es. cm.)
ii) differente unità di misura delle variabili (es. grammi e mm.)

3. Scelta del numero delle CP

Un numero relativamente ridotto di CP dovrebbe catturare un'elevata quota di varianza totale delle variabili originarie.

4. Interpretazione delle CP scelte

Uso del grafico delle correlazioni (già visto).

5. Comparazione delle unità rispetto alle CP scelte.

Uso del diagramma cartesiano con le CP come assi (già visto).

2. Calcolo delle CP: effetto scala (varianza)

Le variabili con varianza maggiore hanno maggiore importanza.

Matrice di covarianza

	X_1	X_2
X_1	80	40
X_2	40	80

$$C_1 = 0.707\,X_1 + 0.707\,X_2$$

$$Var(C_1) = 120\,(75\%\,\text{della varianza totale})$$

Matrice di covarianza

	X_1	X_2
X_1	800	40
X_2	40	80

$$C_1 = 0.998\,X_1 + 0.055\,X_2$$

$$Var(C_1) = 802\,(91\%\,\text{della varianza totale})$$

Se vogliamo attribuire la stessa importanza alle due variabili, si può standardizzare i valori.

ATTENZIONE: i risultati da dati standardizzati sono differenti da quelli ricavati dai dati originari.

2. Calcolo delle CP: variabili con diversa unità di misura

Dal punto di vista **logico** è opportuno procedere alla standardizzazione poiché non ha senso sommare la varianza della LUNGHEZZA (X_1 in mm) e quella del PESO (X_2 in g).

Dal punto di vista **statistico**, però, si modifica l'importanza delle due variabili.

Matrice di covarianza (dati originari)

	X_1	X_2
X_1	800	40
X_2	40	80

Diversa importanza

Matrice di covarianza (dati standardizzati)
→ è la matrice di **correlazione**

	X_1	X_2
X_1	1	0.16
X_2	0.16	1

Stessa importanza

I risultati su dati standardizzati sono differenti da quelli su dati originari.

3. Scelta del numero delle componenti

L'obiettivo è: scegliere **poche** componenti che riassumano **tanta** varianza delle variabili originarie.
Due possibili criteri.

1. Criterio della % di varianza.
Si scelgono le prima *k* componenti che cumulano una prestabilita quota (es. 70%) del totale della varianza delle variabili originarie.

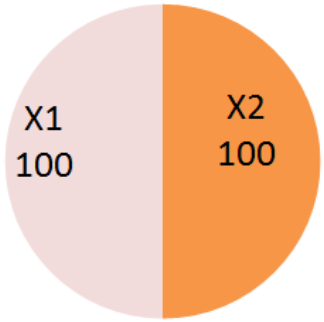
2. Criterio della varianza media.
Si scelgono le componenti la cui varianza è maggiore della varianza media (varianza media = totale varianza/numero di variabili).

Attenzione: più le variabili sono correlate, più efficace è la quota di varianza delle prime CP

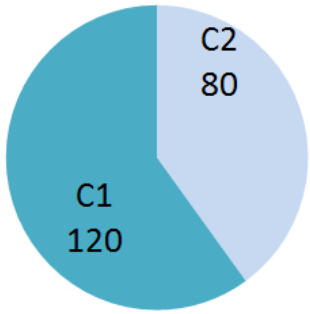
Matrice di covarianza (correlazione 0.2)

	X_1	X_2
X_1	100	20
X_2	20	100

Varianza variabili



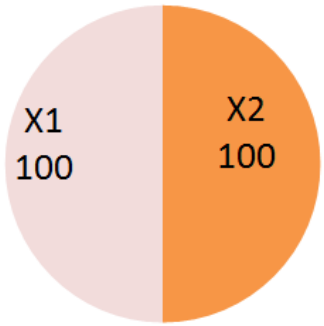
Varianza CP

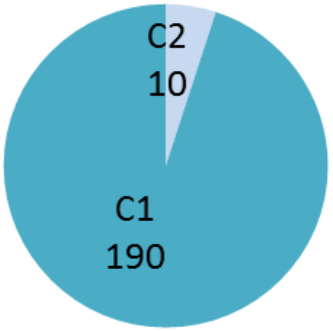


Prima CP:
60% di varianza totale (120/200)

Matrice di covarianza (correlazione 0.9)

	X_1	X_2
X_1	100	90
X_2	90	100





Prima CP:
95% di varianza totale (190/200)

PEARSON

Conclusioni

- L'ACP è utile per individuare gli indici di bilancio maggiormente responsabili delle differenze fra le imprese: si tratta degli indici correlati con le componenti principali più importanti.
- Gli score delle CP danno indicazioni in merito alla *posizione comparativa* di una unità rispetto alle altre. Sarà necessario analizzare i dati originari per capire il significato della posizione individuata attraverso le CP.
- Non sempre si riuscirà a ridurre *p* indici di bilancio in una o due sole componenti principali anche perché *p* può essere relativamente grande.
- Più spesso l'ACP è utile come tecnica preliminare di trasformazione e riduzione dei dati in vista della loro successiva elaborazione mediante altre tecniche statistiche (es. l'analisi cluster).

PEARSON

Aspetti computazionali

L'ACP deve essere condotta mediante *software* apposito.

Per ***Excel*** si rimanda al testo: Bini, Scaffai

Per **R**, **SAS** e altro si rimanda al materiale online