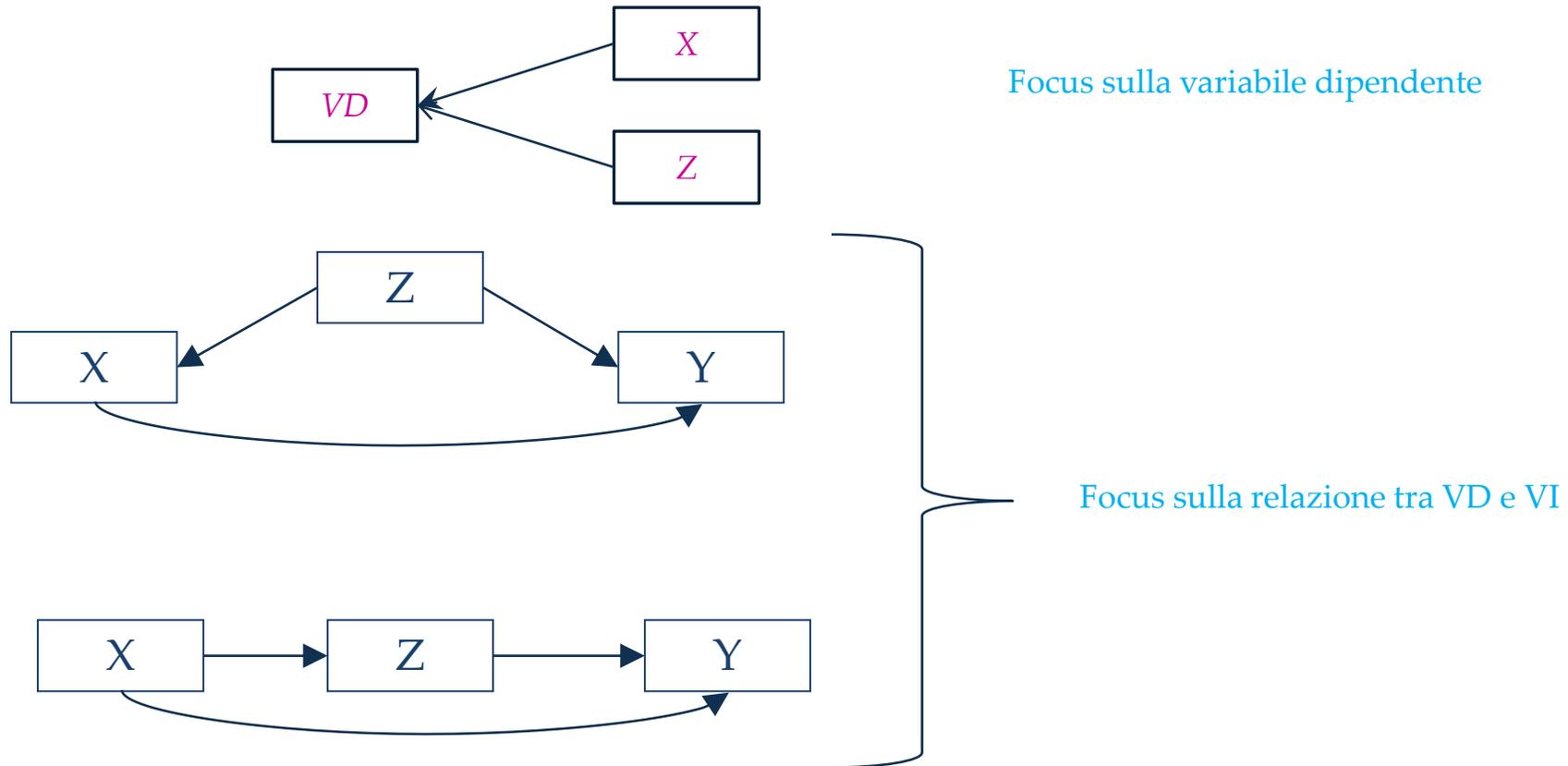


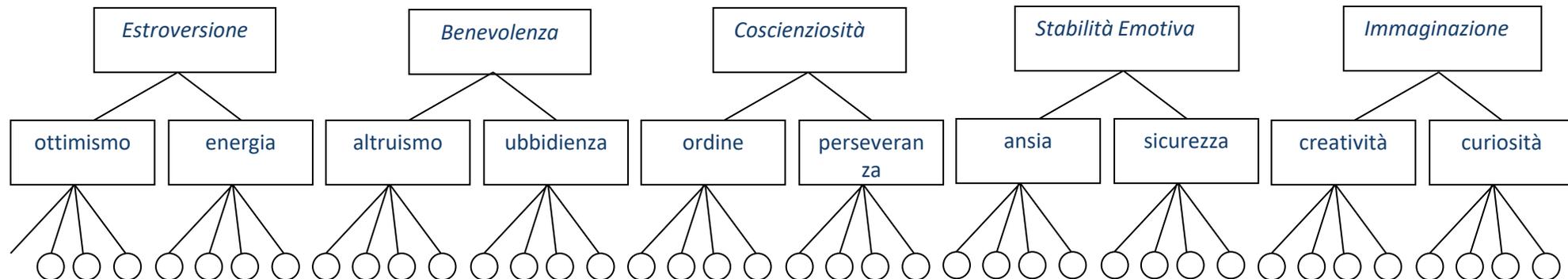
Tecniche di riduzione dei dati

Ci siamo occupati di relazioni come queste



Ora ci occupiamo di relazioni come queste

Relazioni asimmetriche per generalità e semanticamente dipendenti:
organizzazione gerarchica e aggregazione dei punteggi



Tecniche di riduzione dei dati

Obiettivo generale dell'analisi fattoriale è ridurre l'informazione contenuta in una matrice di correlazioni con n variabili in una nuova matrice con $k < n$ nuove variabili

Quando ho bisogno di «comprimere» i miei dati?

- elevata multicollinearità tra stimatori
- raggruppare indicatori in nuove variabili sovra-ordinate per riassumere l'informazione
- individuare fattori latenti per spiegare le intercorrelazioni tra variabili
- verificare la validità «strutturale» di uno strumento di assessment

Tecniche di riduzione dei dati

Obiettivo generale dell'analisi fattoriale è ridurre l'informazione contenuta in una matrice di correlazioni con n variabili in una nuova matrice con $k < n$ nuove variabili

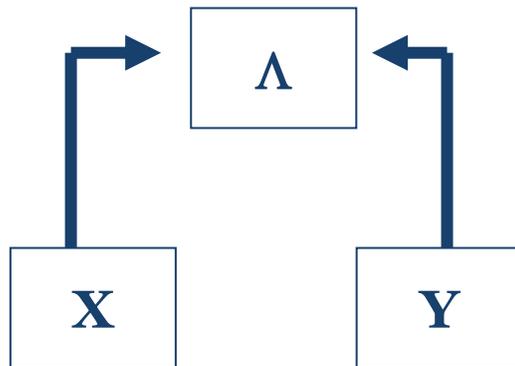
Approcci basilari

- Analisi esplorativa
 - Analisi delle componenti principali (ACP)
 - Analisi dei fattori comuni (AFC)
- Analisi confermativa

Tecniche di riduzione dei dati

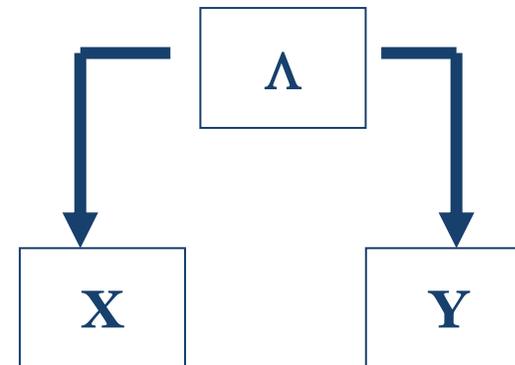
L'ANALISI FATTORIALE INTRODUCE una TERZA VARIABILE,
il fattore o la componente

ACP: relazione di indicazione



*Obiettivo: descrivere e
massimizzare la previsione,
analizzando tutta la varianza
delle variabili osservate*

AFC: relazione di dipendenza



*Obiettivo: spiegare le relazioni tra
variabili analizzandone solo la
varianza comune*

L'analisi delle componenti principali: ridurre una matrice **R** in una nuova matrice **A**

| | x1 | x2 | x3 | x4 | x5 |
|----|-----|-----|-----|-----|-----|
| x1 | 1 | .43 | .18 | .08 | .48 |
| x2 | .43 | 1 | .51 | .49 | .63 |
| x3 | .18 | .51 | 1 | .27 | .55 |
| x4 | .08 | .49 | .27 | 1 | .44 |
| x5 | .48 | .63 | .55 | .44 | 1 |



| I Comp |
|--------|
| .569 |
| .859 |
| .700 |
| .615 |
| .871 |

L'analisi fattoriale e l'analisi delle componenti principali: momenti e decisioni principali nelle tecniche esplorative

Selezione delle variabili, indicatori del costrutto

Selezione del campione, ampio e rappresentativo della popolazione in cui si assume la qualità misurata sia presente con elevata variabilità

Estrazione dei fattori

Determinazione del numero di fattori

Rotazione dei fattori

Interpretazione dei fattori

Stima dei punteggi fattoriali

Eventuale selezione delle variabili o indicatori di un costrutto

Cross-validity

L'analisi delle componenti principali

In termini matematici, obiettivo dell'ACP è riuscire a stimare la matrice delle componenti,

definendo l'*equazione caratteristica della matrice*

$$RV = \lambda V$$

in cui V è il *vettore caratteristico* o *autovettore* della matrice e corrisponde ad una sequenza di pesi applicabili alle variabili analizzate z

λ è l'*autovalore* o *radice caratteristica*, si associa alla componente estratta e esprime la quantità di varianza della matrice R spiegata dalla componente stessa

L'analisi delle componenti principali

In termini matematici, una componente principale è una combinazione lineare di variabili:

$$C_m = b_{1m}x_1 + b_{2m}x_2 + b_{3m}x_3 + \dots + b_{jm}x_j$$

in cui b_{jm} rappresenta il peso che ogni variabile x_j ha nel determinare la componente stessa

L'analisi delle componenti principali: estrazione

| | x1 | x2 | x3 | x4 | x5 |
|----|-----|-----|-----|-----|-----|
| x1 | 1 | .43 | .18 | .08 | .48 |
| x2 | .43 | 1 | .51 | .49 | .63 |
| x3 | .18 | .51 | 1 | .27 | .55 |
| x4 | .08 | .49 | .27 | 1 | .44 |
| x5 | .48 | .63 | .55 | .44 | 1 |



| I Vettore | I Comp. |
|--------------|------------|
| .347 | .569 |
| .524 | .859 |
| .424 | .700 |
| .375 | .615 |
| .531 | .871 |

L'analisi delle componenti principali: estrazione

È possibile verificare l'equazione caratteristica della matrice

$$RV = \lambda V$$

$$\begin{bmatrix} 1.0 & .43 & .18 & .08 & .48 \\ .43 & 1.0 & .51 & .49 & .63 \\ .18 & .51 & 1.0 & .27 & .55 \\ .08 & .49 & .27 & 1.0 & .44 \\ .48 & .63 & .55 & .44 & 1.0 \end{bmatrix} \times \begin{bmatrix} .347 \\ .524 \\ .424 \\ .375 \\ .531 \end{bmatrix} = 2.687 \times \begin{bmatrix} .347 \\ .524 \\ .424 \\ .375 \\ .531 \end{bmatrix}$$

L'analisi delle componenti principali: estrazione

La I componente principale, con i valori b_{ij} si ottiene:

$$\sqrt{2.687} \times \begin{bmatrix} .347 \\ .524 \\ .424 \\ .375 \\ .531 \end{bmatrix} = \begin{bmatrix} .569 \\ .859 \\ .700 \\ .615 \\ .871 \end{bmatrix} \quad V \sqrt{L} = A$$

L'analisi delle componenti principali: estrazione

Estratta la prima componente, si procede con l'estrazione della II componente dalla

MATRICE DEI RESIDUI,

dalla matrice R parzializzata dalla componente già estratta (covarianze parziali)



Per questo le componenti via via estratte sono

ORTOGONALI tra loro

L'analisi delle componenti principali: informazioni salienti

- Saturazione fattoriale b_{ij} che esprime l'intensità della relazione tra variabile e componente; elevata al quadrato ne esprime la % di varianza condivisa
- Autovalore λ_i indica la varianza che la componente estratta spiega della matrice R (se espresso in % si ottiene $(\lambda_i/n) \times 100$)

L'analisi delle componenti principali: esempio estrazione e autovalori

File dati Jamovi «PCA dati»: 10 variabili per valutare i BF

jamovi - PCA dati

Analisi delle Componenti Principali

Metodo: Rotazione **Nessuna**

Numero delle Componenti: **Basato su analisi parallele**

Autovalori maggiori di **1**

Verifiche delle Ipotesi: Test di sfericità di Bartlett, Misura di idoneità al campionamento KMO

Saturazioni fattoriali: Nascondi le saturazioni sotto a **0.2**, Ordina i carichi per dimensione

Risultati Aggiuntivi: Riassunto delle componenti

Spesso una soluzione non ruotata non è interpretabile

Analisi delle Componenti Principali

Caricamenti delle Componenti

| | Componente | | | | Unicità |
|----------|------------|--------|--------|--------|---------|
| | 1 | 2 | 3 | 4 | |
| BFI_NEU1 | -0.672 | 0.391 | | 0.312 | 0.296 |
| BFI_NEU2 | -0.628 | 0.405 | | 0.322 | 0.317 |
| BFI_EX2 | 0.623 | 0.477 | 0.340 | | 0.255 |
| BFI_AG1 | 0.420 | 0.203 | | -0.376 | 0.628 |
| BFI_EX1 | 0.349 | 0.591 | 0.408 | | 0.362 |
| BFI_OP1 | | -0.566 | 0.356 | 0.353 | 0.416 |
| BFI_OP2 | | -0.377 | 0.638 | 0.368 | 0.302 |
| BFI_AG2 | 0.377 | | -0.443 | | 0.646 |
| BFI_CO2 | 0.300 | | -0.249 | 0.550 | 0.542 |
| BFI_CO1 | 0.411 | | -0.478 | 0.495 | 0.341 |

Nota. È stata utilizzata la rotazione 'niente'

Autovalori

Autovalori Iniziali

Matrice di estrazione delle componenti (non ruotata)

Saturazione fattoriale della VARIAB sulla I componente

Unicità: quota di varianza di una VARIAB NON spiegata dalle 4 componenti estratte $1 - U = h^2$ o comunaltà (quota di varianza estratta per ogni variabile)

jamovi - PCA dati

Variabili | Dati | **Analisi** | Modifica

Colorazione | Test t | ANOVA | Regressione | Frequenze | Fattore | Behavior Chang Linear Models | medmod | jpower

Analisi delle Componenti Principali

Variabili

- RSE_3
- RSE_4
- RSE_5
- RSE_6
- RSE_7
- RSE_8
- RSE_9
- RSE_10

Variabili

- BFI_CO1
- BFI_NEU1
- BFI_OP1
- BFI_EX2
- BFI_AG2
- BFI_CO2
- BFI_NEU2
- BFI_OP2

Metodo

Rotazione: Nessuna

Verifiche delle Ipotesi

Test di sfericità di Bartlett

Misura di idoneità al campionamento KMO

Saturazioni fattoriali

Basato su analisi parallele

BFI_CO1 0.411

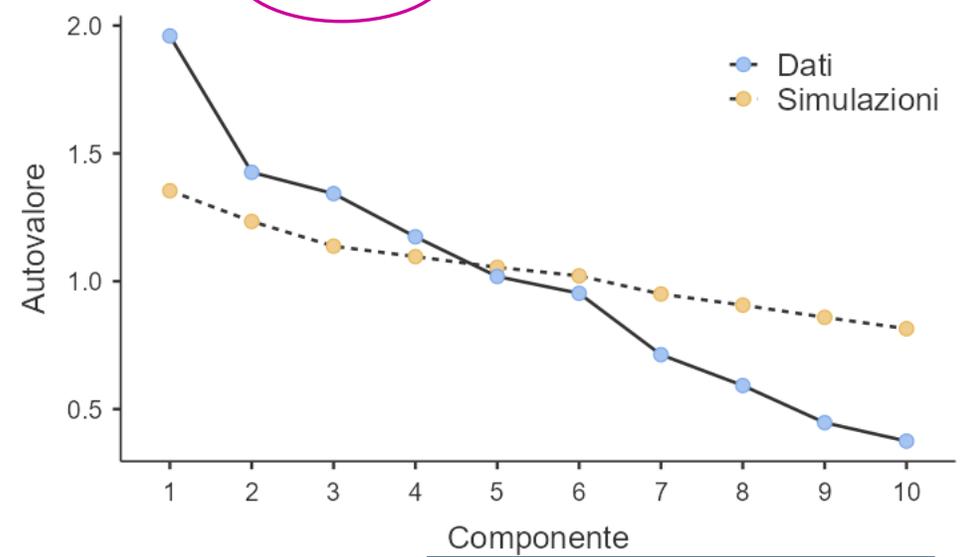
Nota. È stata utilizzata la

Autovalori

Autovalori Iniziali

| Componente | Autovalore | % della Varianza | % Cumulata |
|------------|------------|------------------|------------|
| 1 | 1.960 | 19.60 | 19.6 |
| 2 | 1.426 | 14.26 | 33.9 |
| 3 | 1.343 | 13.43 | 47.3 |
| 4 | 1.174 | 11.74 | 59.0 |
| 5 | 1.019 | 10.19 | 69.2 |
| 6 | 0.953 | 9.53 | 78.7 |
| 7 | 0.713 | 7.13 | 85.9 |
| 8 | 0.592 | 5.92 | 91.8 |
| 9 | 0.447 | 4.47 | 96.3 |
| 10 | 0.375 | 3.75 | 100.0 |

Grafico dei Sedimenti (Scree plot)



Matrice degli autovalori, dove $n = k$

Autovalore per ciascuna componente (corrisponde alla Somma delle saturazioni al quadrato delle 10 variabili sulla I componente)
 Accanto % varianza spiegata dalla componente



Analisi delle Componenti Principali

- RSE_3
- RSE_4
- RSE_5
- RSE_6
- RSE_7
- RSE_8
- RSE_9
- RSE_10

- Variabili
- BFI_EX1
 - BFI_AG1
 - BFI_CO1
 - BFI_NEU1
 - BFI_OP1
 - BFI_EX2
 - BFI_AG2
 - BFI_CO2

Metodo
Rotazione **Nessuna**

Verifiche delle Ipotesi

- Test di sfericità di Bartlett
- Misura di idoneità al campionamento KMC

Numero delle Componenti

Nota. È stata utilizzata la rotazione 'niente'

Verifiche delle Ipotesi

Test di Sfericità di Bartlett

| χ^2 | gdl | p |
|----------|-----|--------|
| 442 | 45 | < .001 |

Testa se N casi è sufficientemente alto per applicare la tecnica di PCA o EFA, se sig allora OK

Autovalori

Autovalori Iniziali

| Componente | Autovalore | % della Varia |
|------------|------------|---------------|
| 1 | 1.888 | 18.88 |
| 2 | 1.000 | 10.00 |
| 3 | 0.500 | 5.00 |
| 4 | 0.250 | 2.50 |
| 5 | 0.125 | 1.25 |
| 6 | 0.062 | 0.62 |
| 7 | 0.031 | 0.31 |
| 8 | 0.016 | 0.16 |
| 9 | 0.008 | 0.08 |
| 10 | 0.004 | 0.04 |

Quanti componenti (fattori) estrarre?

- Criterio di Kaiser
- Scree test
- **Analisi parallela**
- Criteri teorici
- Criteri basati sui punteggi fattoriali (es., metodo definito da Everett)

L'analisi delle componenti principali: Rotazione

Perché?

Una matrice delle componenti principali ottenuta dall'estrazione spesso NON si presta a facile interpretazione, così i contenuti della matrice R rimangono oscuri.

La rotazione serve a produrre una matrice delle componenti che sia interpretabile

L'analisi delle componenti principali: rotazione

Quale?

Matrice estratta e matrici (potenzialmente infinite) ruotate sono matematicamente equivalenti.

Il criterio privilegiato per la rotazione è noto come

STRUTTURA SEMPLICE:

ogni componente deve essere definita da poche saturazioni elevate, mentre le restanti devono approssimarsi a 0 ovvero ogni variabile deve presentare 1 saturazione elevata su una sola componente e saturazioni basse sulle altre componenti

L'analisi delle componenti principali: rotazione

Come?

Metodi analitici di rotazione (ciechi)

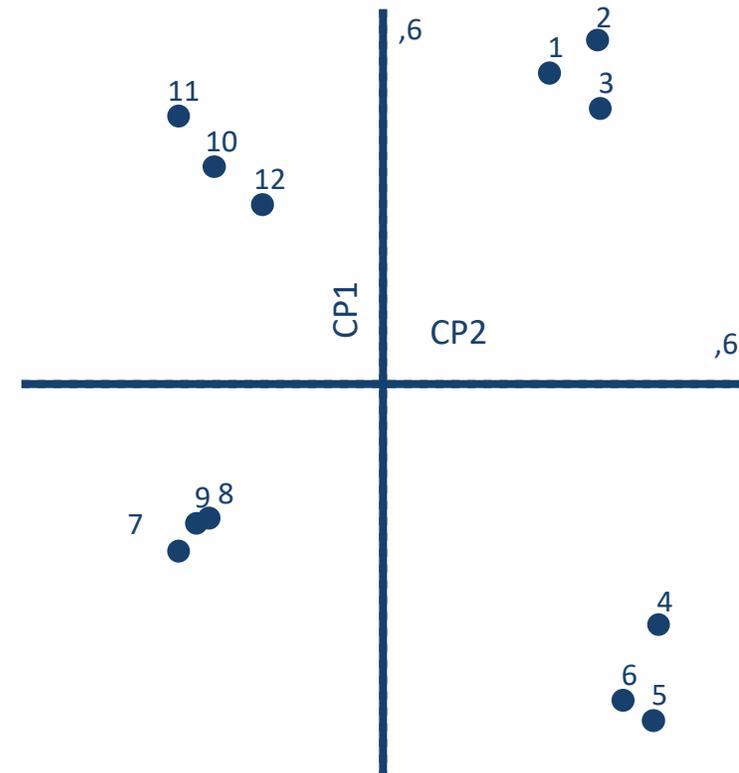
- varimax
- oblimin

Rotazioni grafiche e manuali

L'angolo di rotazione viene scelto da chi analizza i dati, in accordo con una struttura teorica attesa o teoricamente sensata, anche in violazione del principio della struttura semplice

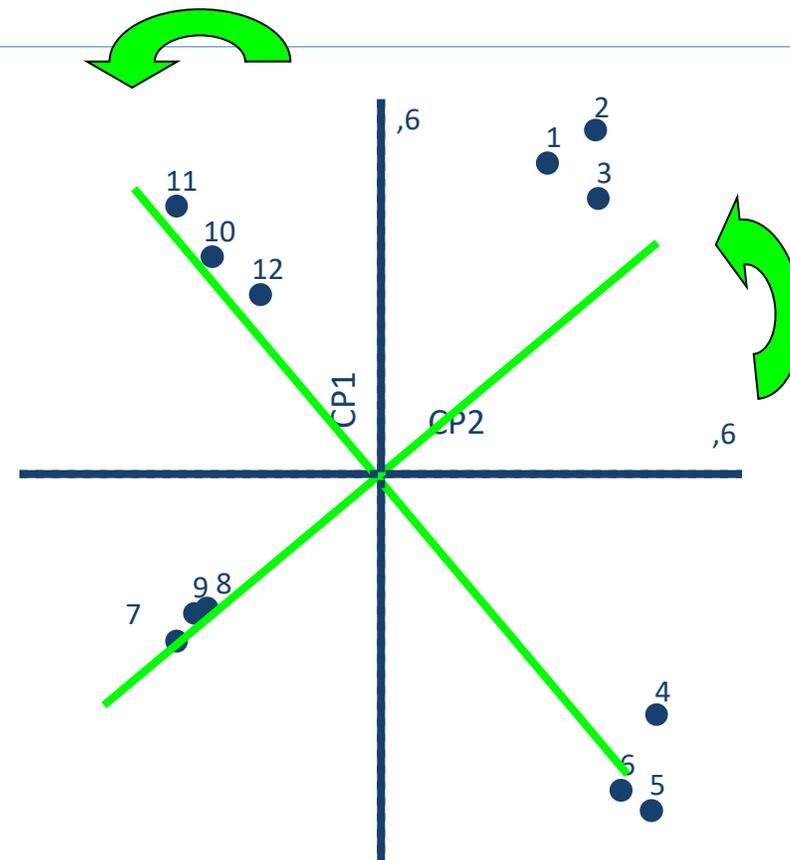
L'analisi delle componenti principali: la rotazione ortogonale

| | I cp (X) | II cp (Y) |
|-------|----------|-----------|
| Soc1 | .50 | .27 |
| Soc2 | .55 | .35 |
| Soc3 | .45 | .36 |
| Dom4 | -.37 | .45 |
| Dom5 | -.52 | .44 |
| Dom6 | -.49 | .39 |
| Tim7 | -.25 | -.34 |
| Tim8 | -.20 | -.29 |
| Tim9 | -.21 | -.31 |
| Sot10 | .36 | -.28 |
| Sot11 | .44 | -.34 |
| Sot12 | .30 | -.20 |



L'analisi delle componenti principali: la rotazione ortogonale

| | I cp (X) | II cp (Y) |
|-------|----------|-----------|
| Soc1 | .50 | .27 |
| Soc2 | .55 | .35 |
| Soc3 | .45 | .36 |
| Dom4 | -.37 | .45 |
| Dom5 | -.52 | .44 |
| Dom6 | -.49 | .39 |
| Tim7 | -.25 | -.34 |
| Tim8 | -.20 | -.29 |
| Tim9 | -.21 | -.31 |
| Sot10 | .36 | -.28 |
| Sot11 | .44 | -.34 |



L'analisi delle componenti principali: la rotazione

Applicando una rotazione si ottiene una nuova matrice delle componenti, la *matrice ruotata*.

Come?

Si definisce una matrice di rotazione in base a

Angolo di rotazione

Direzione della rotazione

Rotazione antioraria

$$\begin{pmatrix} \cos \Phi & \sin \Phi \\ -\sin \Phi & \cos \Phi \end{pmatrix}$$

Rotazione oraria

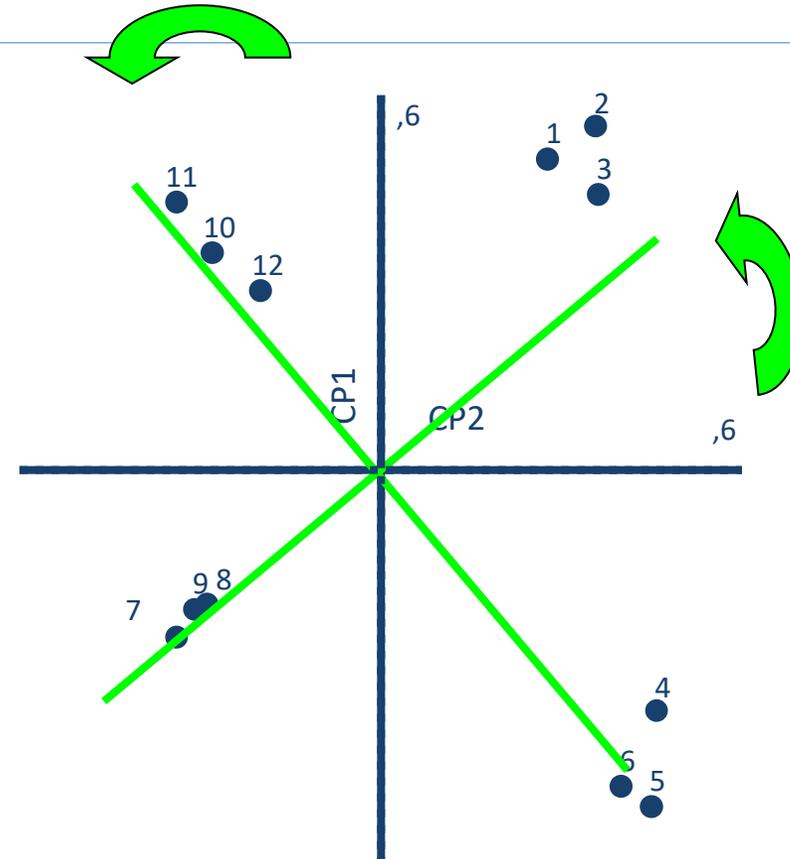
$$\begin{pmatrix} \cos \Phi & -\sin \Phi \\ \sin \Phi & \cos \Phi \end{pmatrix}$$



La matrice di componenti NON Ruotata viene moltiplicata per la matrice di rotazione e si ottiene la matrice RUOTATA

L'analisi delle componenti principali: la rotazione ortogonale

$$\begin{bmatrix} .50 & .27 \\ .55 & .35 \\ .45 & .36 \\ -.37 & .45 \\ -.52 & .44 \\ -.49 & .39 \\ -.25 & -.34 \\ -.20 & -.29 \\ -.21 & -.31 \\ .36 & -.28 \\ .44 & -.34 \\ .30 & -.20 \end{bmatrix} \times \begin{bmatrix} .766 & .643 \\ -.643 & .766 \end{bmatrix} = \begin{bmatrix} .21 & .53 \\ .20 & .62 \\ .11 & .57 \\ -.57 & .11 \\ -.68 & .00 \\ -.63 & -.02 \\ .03 & -.42 \\ .03 & -.35 \\ .04 & -.37 \\ .46 & .02 \\ .56 & .02 \\ .36 & .04 \end{bmatrix}$$



Noi interpretiamo la matrice ruotata

Analisi delle Componenti Principali

- RSE_3
- RSE_4
- RSE_5
- RSE_6
- RSE_7
- RSE_8
- RSE_9
- RSE_10

- Variabili
- BFI_EX1
 - BFI_AG1
 - BFI_CO1
 - BFI_NEU1
 - BFI_OP1
 - BFI_EX2
 - BFI_AG2
 - BFI_CO2

Metodo

Rotazione Varimax

Numero delle Componenti

- Basato su analisi parallele
- Basato sull'autovalore

Autovalori maggiori di 1

Verifiche delle Ipotesi

- Test di sfericità di Bartlett
- Misura di idoneità al campionamento KMO

Saturazioni fattoriali

Nascondi le saturazioni sotto a 0.2

- Ordina i carichi per dimensione

Analisi delle Componenti Principali

Caricamenti delle Componenti

| | Componente | | | | Unicità |
|----------|------------|-------|-------|-------|---------|
| | 1 | 2 | 3 | 4 | |
| BFI_NEU2 | 0.821 | | | | 0.311 |
| BFI_NEU1 | 0.816 | | | | 0.296 |
| BFI_EX2 | | 0.828 | | | 0.255 |
| BFI_EX1 | | 0.794 | | | 0.362 |
| BFI_AG1 | -0.387 | 0.413 | | | 0.628 |
| BFI_CO1 | | | 0.804 | | 0.341 |
| BFI_CO2 | | | 0.658 | | 0.542 |
| BFI_AG2 | -0.321 | | 0.473 | | 0.646 |
| BFI_OP2 | | | | 0.821 | 0.302 |
| BFI_OP1 | | | | 0.741 | 0.416 |

Nota. È stata utilizzata la rotazione 'Varimax'

[3]

Verifiche delle Ipotesi

Test di Sfericità di Bartlett

La quota di varianza spiegata per ogni variabile è invariata ($1 - U$)

La quota di varianza spiegata per ogni variabile viene redistribuita in accordo col principio della struttura semplice per favorire interpretazione

L'analisi delle componenti principali: Interpretazione e punteggi

L'interpretazione dipende
dal contenuto delle componenti,
dal peso delle variabili e dalla teoria di riferimento,
ricordando come la matrice delle componenti richieda una
lettura per colonne o componenti
(dove le componenti sono combinazioni lineari delle variabili
opportunamente ponderate)

$$C_m = b_{m1}x_1 + b_{m2}x_2 + b_{m3}x_3 + \dots + b_{mj}x_j$$

e una lettura per riga o per variabili
(dove le variabili sono combinazioni lineari di componenti)

Analisi delle componenti principali

Riepilogando

Matrice di dati grezzi

Matrice Z dei dati standardizzati

Matrice R delle intercorrelazioni

Matrice A delle componenti Estratte

Matrice A delle componenti Ruotate

Matrice di trasformazione dei componenti

Matrice B dei coefficienti fattoriali

L'analisi delle componenti principali: la rotazione obliqua

Le componenti ruotate possono essere inter-correlate → ROTAZIONE OBLIMIN

→ La quota di variabilità di una variabile può essere simultaneamente condivisa da più componenti

→ la saturazione della variabile sulla componente non è più al contempo un indicatore del peso netto del legame tra componente e variabile

tre nuove matrici:

Matrice P dei Pattern (pesi)

Matrice S di Struttura (saturazioni)

Matrice Φ delle correlazioni tra componenti

$$S = P\Phi$$

Analisi delle Componenti Principali

- RSE_7
- RSE_8
- RSE_9
- RSE_10
- BFI_OP1
- BFI_EX2
- BFI_AG2
- BFI_CO2

Metodo

Rotazione **Oblimin**

Numero delle Componenti

Basato su analisi parallele

Basato sull'autovalore

Autovalori maggiori di

Numero fisso

componente(i)

Verifiche delle Ipotesi

- Test di sfericità di Bartlett
- Misura di idoneità al campionamento KMO

Saturazioni fattoriali

Nascondi le saturazioni sotto a

Ordina i carichi per dimensione

Risultati Aggiuntivi

- Riepilogo delle componenti
- Correlazioni delle componenti

Analisi delle Componenti Principali

Caricamenti delle Componenti

| | Componente | | | | Unicità |
|----------|------------|-------|-------|--------|---------|
| | 1 | 2 | 3 | 4 | |
| BFI_NEU2 | 0.827 | | | | 0.311 |
| BFI_NEU1 | 0.821 | | | | 0.296 |
| BFI_EX2 | | 0.823 | | | 0.255 |
| BFI_EX1 | | 0.802 | | | 0.362 |
| BFI_AG1 | -0.383 | 0.400 | | -0.220 | 0.628 |
| BFI_CO1 | | | 0.799 | | 0.341 |
| BFI_CO2 | | | 0.665 | | 0.542 |
| BFI_AG2 | -0.310 | | 0.449 | | 0.646 |
| BFI_OP2 | | | | 0.820 | 0.302 |
| BFI_OP1 | | | | 0.736 | 0.416 |

Nota. È stata utilizzata la rotazione 'Oblimin'

[3]

Statistiche delle Componenti

Correlazioni Intercomponenti

| | 1 | 2 | 3 | 4 |
|---|---|---------|---------|---------|
| 1 | — | -0.0887 | -0.1261 | -0.0693 |
| 2 | | — | 0.0719 | 0.0191 |
| 3 | | | — | -0.0277 |
| 4 | | | | — |

Matrice delle correlazioni tra le componenti

Per un riepilogo degli argomenti trattati: Obiettivi raggiunti

★ dal punteggi grezzo ai punteggi standardizzati e normalizzati nel testing psicologico;

★ analisi della varianza

- per confronti tra gruppi (ANOVA con 1 VI e con 2 VI)
- per misure ripetute (ANOVA con 1 VD)
- modelli fattoriali misti (ANOVA con 1 VI btw e 1 VI wth)

★ analisi della correlazione semplice e multipla

★ analisi della regressione semplice

★ analisi della regressione multipla (con particolare attenzione ai concetti di parzializzazione, spiegazione e previsione di una variabile, controllo di una relazione)

★ tecniche di analisi di riduzione dei dati (analisi fattoriale esplorativa → solo PCA)

★ elementi psicometrici di base nel testing psicologico (validità e affidabilità)

Contenuti, obiettivi e bibliografia

Del testo di M. Gallucci, L. Leone e M. Berlingeri (2017, seconda ed), “Modelli statistici per le scienze sociali” , Pearson Italia, Milano Torino.

Si confermano i seguenti capitoli

- 2
- 3
- 4 (par 4.1, 4.2, 4.3, 4.5.1, 4.5.6, 4.8.1, 4.8.2)
- 8 (par 8.1, 8.2, 8.3, 8.4.1)
- 9 (par 9.1, 9.2, 9.3.1, 9.3.2, 9.3.3, 9.3.10, 9.3.11)
- (12 non è più in programma d’esame)
- 13 (13.1, 13.2, 13.3, 13.4.1-13.4.4, 13.5)