

L'analisi fattoriale

Stabilire il numero di fattori da estrarre

Lezioni di Psicometria
Giovanni Battista Flebus

Il problema del **numero** di fattori da estrarre

Criteri validi e forti:

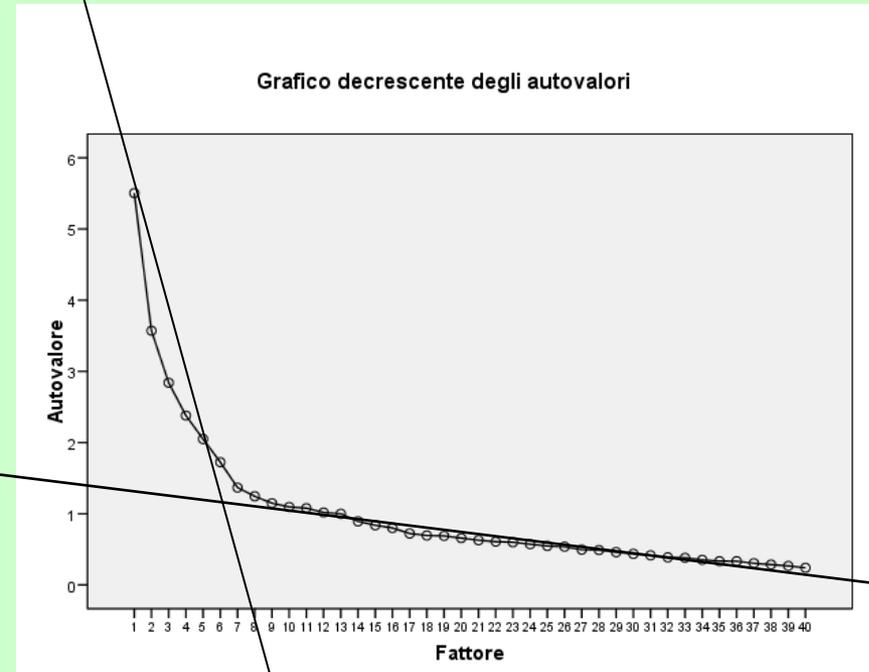
1. Scree test
2. Analisi parallela
3. Calcolo dei punteggi fattoriali su due subcampioni –
4. Interpretabilità dei fattori

Criteri deboli:

5. Autovalore superiore a 1
6. Percentuale di varianza spiegata

Lo scree-test

- È la rappresentazione grafica degli autovalori in forma canonica (i primi, più importanti, sono sulla sinistra del grafico)
- Secondo Cattell, gli autovalori connessi con i fattori reali e non casuali hanno un andamento caratteristico a caduta
- Quelli casuali degradano lentamente



Lo scree-test

- Si estraggono solo quei fattori che sono sulla linea di caduta, e si trascurano quelli che degradano lentamente

Validità dello scree-test

Vantaggi:

- facile da utilizzare
- molto usato

Limiti

- non sempre è efficace e veritiero
- A volte il pendìo o la sua interruzione non è individuabile con facilità

L'analisi parallela

- Consiste nel generare dei numeri casuali, (per esempio punti zeta), uno per ogni variabile osservata e per ogni partecipante.
- Si estraggono tutti gli autovalori
- Si ripete la simulazione molte volte (per es. 100)
- La media del primo, secondo ... kappesimo autovalore servono da confronto per gli autovalori della matrice reale

Esempio con cinque simulazioni

- Si estraggono i 14 autovalori da una matrice di 14 variabili osservate, e si riportano sul grafico, insieme a quelli corrispondenti di cinque simulazioni, ottenute sostituendo a ciascuna risposta R di un soggetto S un dato casuale (per esempio un punto zeta casuale).
- Si estraggono i fattori e si tiene conto degli autovalori.

Esempio con **cinque** simulazioni

Autovalori fuori scala nel grafico, perché sicuramente da conservare

Ambito dell'incertezza

Autovalori reali più piccoli di quelli casuali, sicuramente da scartare

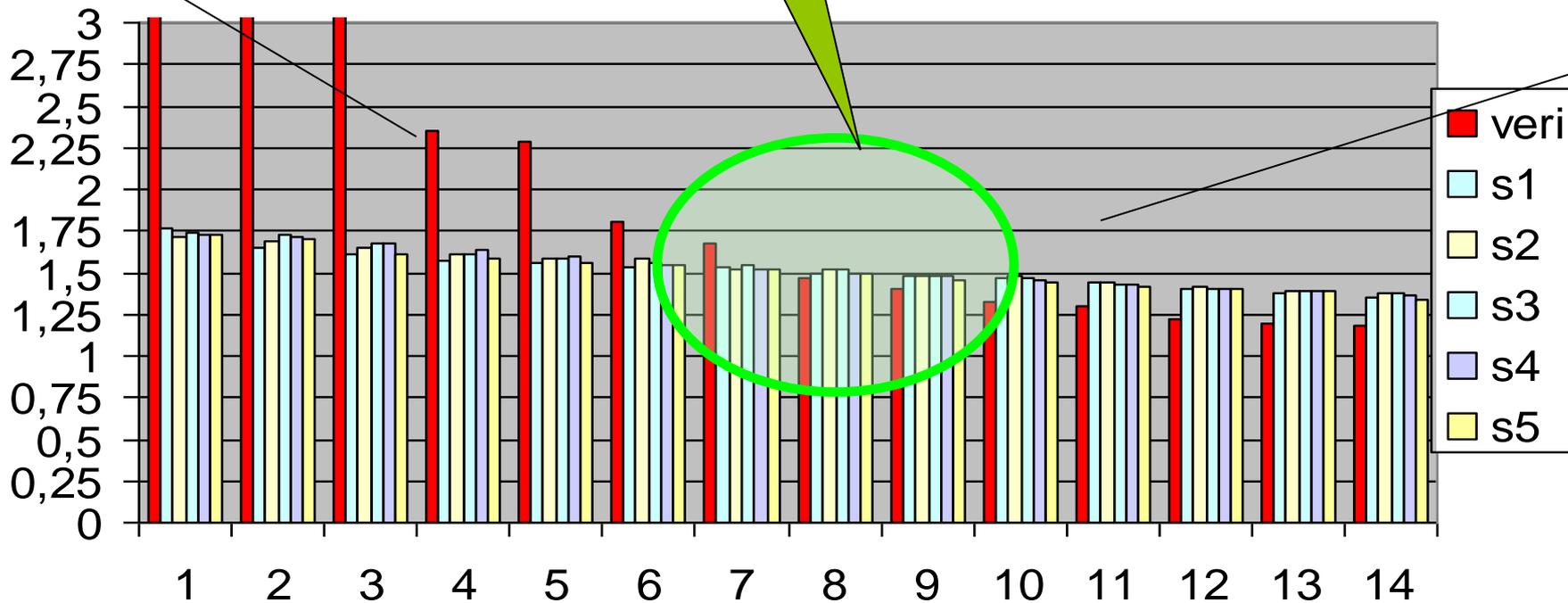
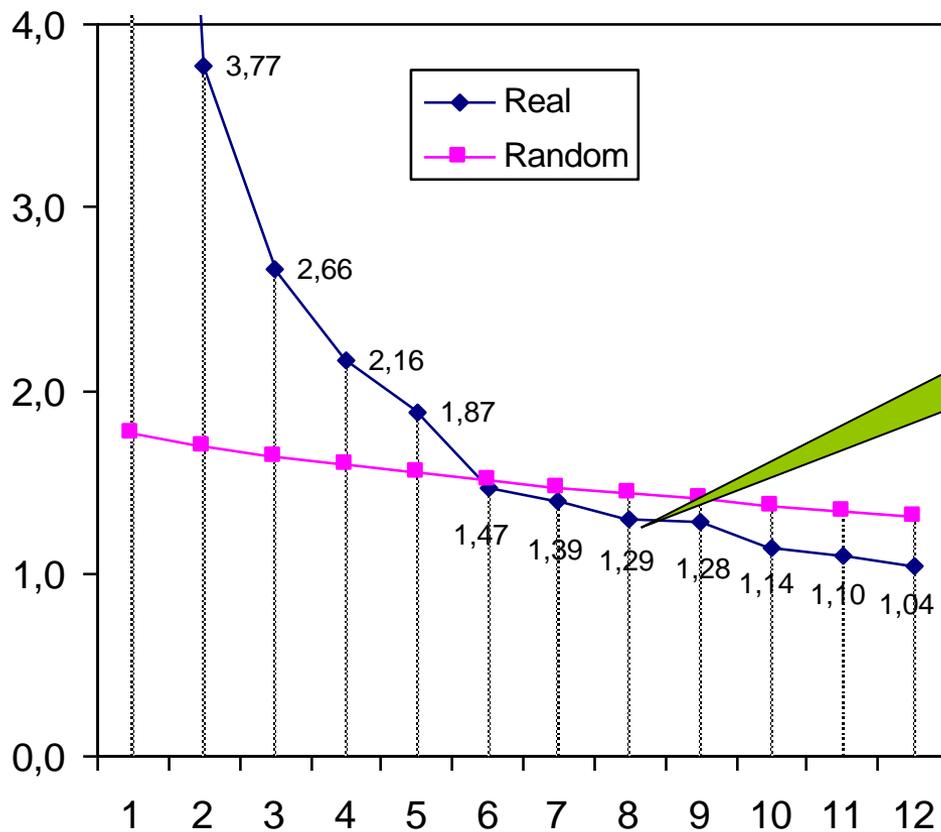


Grafico degli autovalori dei dati reali e simulati

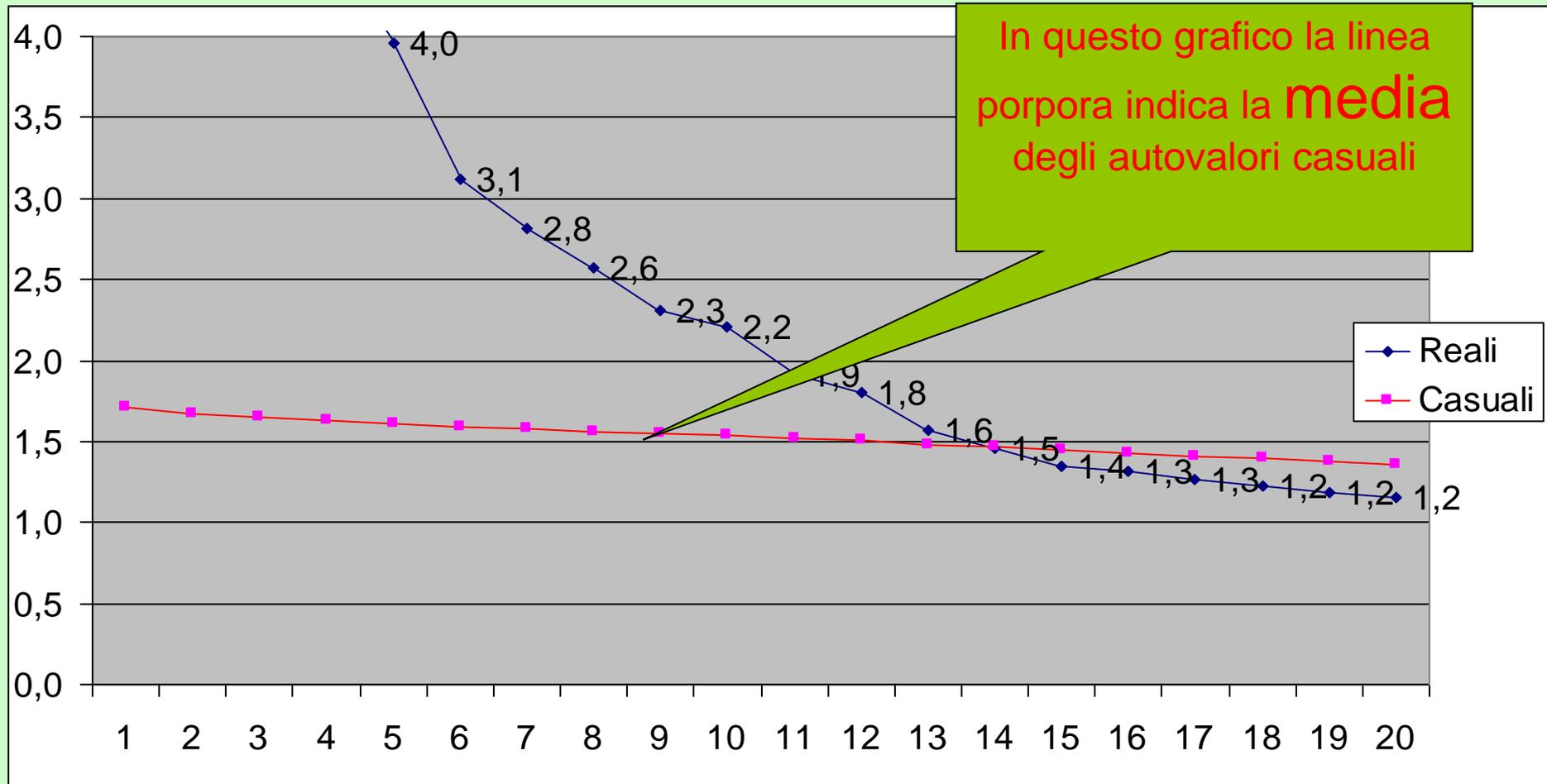
Real and Randomly generated eigenvalues



In questo grafico la linea
porpora indica la
media degli autovalori
casuali

Secondo esempio

Valori reale e **medie** degli autovalori casuali



Analisi parallela

- Anziché la media, si prende in considerazione il 95° percentile, facendo un paragone con la procedura di verifica di ipotesi.
- L'ipotesi nulla è sostanzziata qui con la distribuzione dei k-esimi autovalori: se il 95° percentile del k-esimo autovalore è inferiore al k-esimo autovalore, allora lo si considera non casuale ma reale, e si conserva il k-fattore.
- Se il k-esimo autovalore reale è inferiore al 95° percentile della distribuzione dei k-esimi autovalori casuali, allora tale autovalore viene considerato frutto del caso e il corrispondente fattore viene ignorato

Validità dell'analisi parallela

- Funziona molto bene, anche se non è entrata completamente nell'uso.
- Un software (**Monte Carlo PCA** di Marley Watkins) è disponibile gratuitamente in rete.
- Disponibile con il software Jamovi e Mplus
- SPSS non lo ha implementato, ma esistono degli script (insieme di istruzioni in formato di sintassi, per i bisogni individuali)

Il 95° percentile

- Il software di Witkins permette di confrontare il k-esimo autovalore tratto dai dati con il k-esimo autovalore dei dati simulati. L'autovalore dei dati reali dovrebbe essere superiore al 95% degli autovalori casuali per essere considerato rappresentativo di una dimensione latente da prendere in considerazione.

Output del programmino di Watkins

- Monte Carlo PCA for Parallel Analysis
- Version .

- 08/11/2011 15.57.03
- Number of variables: 5
- Number of subjects: 1559
- Number of replications: 100

```

+++++
+
Eigenvalue #   Random Eigenvalue   Standard Dev
+++++

```

Eigenvalue #	Random Eigenvalue	Standard Dev
1	1,0723	,0157
2	1,0323	,0147
3	0,9984	,0102
4	0,9681	,0112
5	0,9289	,0176

```

+++++
+
08/12/2011 15.57.04

```

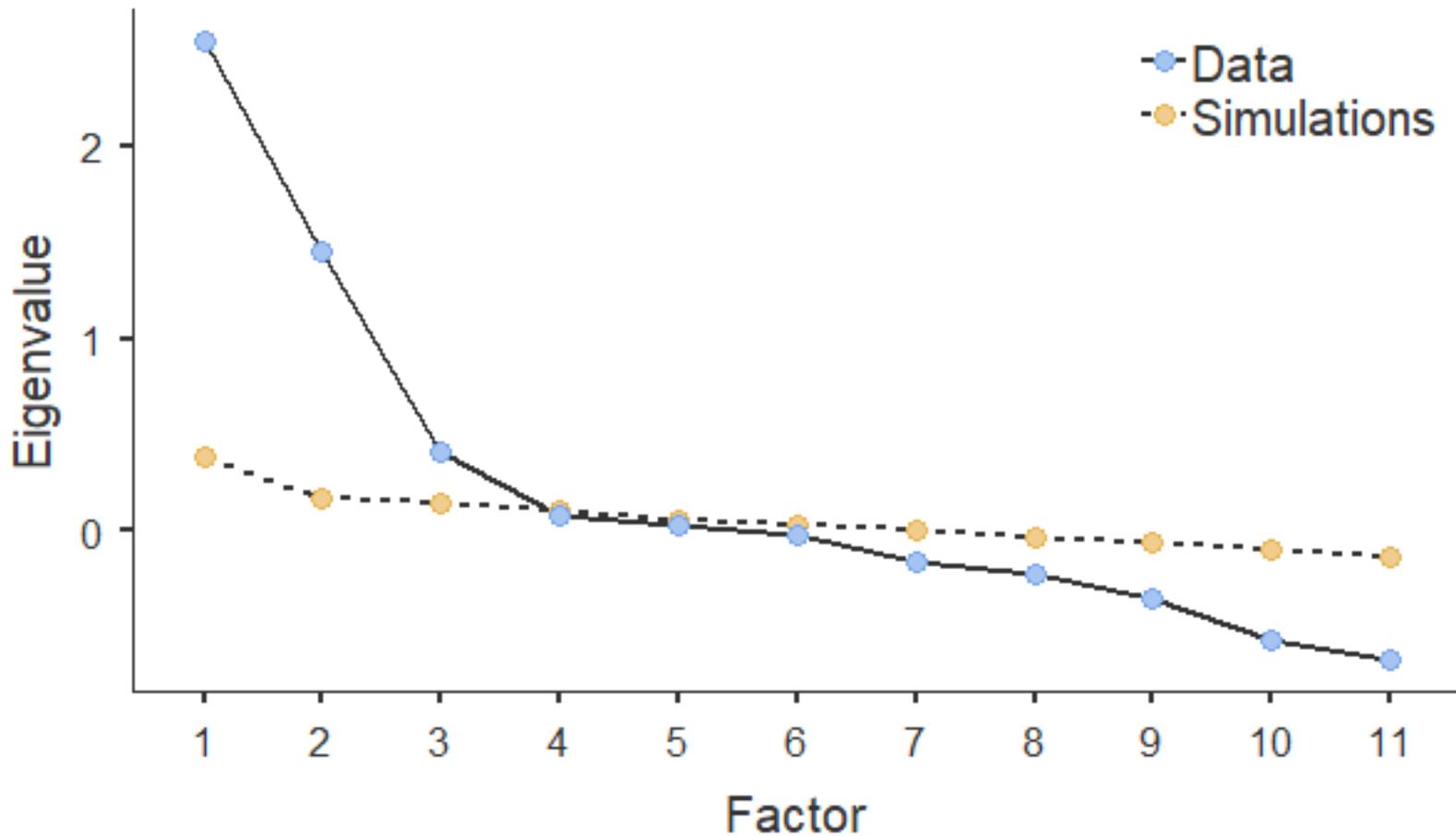
I primi due autovalori dell'esempio sono superiori ai due autovalori dei dati casuali (1,78 e 1,36) ma il terzo è inferiore (0,90), Si sceglie la soluzione a due fattori (da estrarre e ruotare)

Istruzione del software

- Select the number of variables (3-300), subjects (100-2500), and replications (1-1000). The program then: (1) generates random normal numbers for the quantity of variables and subjects selected, (2) computes the correlation matrix, (3) performs Principal Components Analyses and calculates the eigenvalues for those variables, (4) repeats the process as many times as specified in the replications field, and (5) calculates the average and standard deviation of the eigenvalues across all replications.
- For stable results, replicate at least 50-100 times. Use these eigenvalues as the criteria for Horn's Parallel Analysis for the number of factors or components to retain for rotation.

Esempio di output di Jamovi

Scree Plot



Il metodo dei punteggi fattoriali su due sottocampioni

La procedura consiste in questi passi:

Suddividere il campione in subcampione A e B

Sul subcampione A, applicare l'analisi fattoriale con il numero ipotizzato di fattori, e calcolare i punteggi fattoriali, sia sul subcampione A sia sul subcampione B

Seguire la stessa procedura con il subcampione B

Calcolare la correlazione fra i punteggi AB e BA (subcampione A ma con i punteggi ottenuti dal subcampione B e viceversa)

- Se i fattori estratti producono correlazioni elevate allora sono stabili e confermati, altrimenti vanno scartati
- Si conclude ripetendo l'analisi fattoriale sull'intero campione, con il numero di fattori individuato dalla procedura.

Interpretabilità dei fattori

- E' un criterio soggettivo, dipendente dalla natura degli item, dalla teoria soggiacente e dalle conoscenze precedenti: si accettano i fattori che sono stati previsti e risultano comprensibili e interpretabili
- Pertanto il criterio è molto variabile e poco affidabile in situazioni di autentica esplorazione

Autovalore superiore a 1

- Il criterio dovuto a Kaiser stabilisce che si possono accettare quei fattori che hanno una varianza almeno uguale a una variabile standardizzata
- Tuttavia il criterio è molto debole e impreciso, perché tende a dichiarare buoni troppi fattori casuali.
- E' il criterio prestabilito (default) di Spss

Percentuale di varianza spiegata

- Non è un vero criterio, ma un modo molto impreciso per decidere di estrarre almeno, per esempio, 40 o 60 o 80 % di varianza delle variabili osservate
- Si usa raramente e presenta il limite di non avere una vera giustificazione teorica

Concludendo

- Il metodo migliore è l'analisi parallela, disponibile col software aggiuntivo gratuito (jamovi oppure MacParallel di Watkins)
- In mancanza, fare riferimento allo scree-test, disponibile con SPSS
- Non fare affidamento al criterio di Kaiser (autovalore maggiore di 1).