
PSP6075525 - Testing psicologico (matr. dispari)

Caso studio del 25-01-21

Istruzioni iniziali

- Si avvii una nuova sessione di R (o RStudio).
- Si crei un nuovo script di R e lo si salvi come `cognome_nome.R`.
- Si effettui il download del file di dati dell'esame `dati_esame.Rdata` disponibile presso la pagina moodle del corso e lo si carichi nell'ambiente di lavoro di R.
- Si crei un nuovo documento di testo (mediante LibreOffice Writer, Microsoft Word o software analogo) e lo si salvi come `cognome_nome.doc`. Il file dovrà contenere le risposte ai quesiti d'esame accompagnati dai comandi di R, dai risultati ottenuti e dai grafici prodotti. Le risposte dovranno essere inserite in ordine, rispettando il numero del quesito a cui si riferiscono. Alla fine, il file dovrà essere convertito in formato non modificabile (PDF: `cognome_nome.pdf`) ed inviato al docente utilizzando la procedura "Consegna documento" disponibile presso la pagina Moodle del corso. Nel caso di utilizzo di **R-markdown** per la compilazione dinamica di documenti di testo, sarà necessario inviare il file sorgente `.Rnw` unitamente al file PDF generato.
- La valutazione della prova sarà effettuata utilizzando primariamente il file `cognome_nome.pdf`: si raccomanda pertanto la chiarezza nella scrittura delle risposte e la correttezza nel riportare i comandi e gli output di R. Il file `cognome_nome.R` dovrà essere allegato al file `cognome_nome.pdf` solo per un controllo aggiuntivo (pertanto non verrà primariamente valutato).

Caso studio

Il dataset complessivo contiene 12 variabili riferite ad un test di personalità somministrato ad un campione di $n = 1500$ partecipanti. Le variabili originarie, rilevate mediante scala di rating a cinque punti, sono state pretrattate mediante un'adeguata procedura di quantificazione. Le variabili sono riferite ad item che indagano la struttura del sé (**SELF**), aspetti depressivi (**DEPRESS**) e di impulsività (**IMPULS**). L'obiettivo dell'analisi è quello di definire e adattare uno o più modelli CFA per lo studio della dimensionalità del test in oggetto. Considerata la numerosità delle unità statistiche a disposizione, il dataset originario è stato casualmente diviso in due sottoinsiemi rispettivamente di $n = 900$ (**Y0**) e $n = 600$ (**Y1**) unità. Questo consente di poter utilizzare il dataset **Y0** per le analisi esplorative e di utilizzare invece **Y1** per quelle confermative.

1. Si individuino il numero di unità statistiche e di variabili a disposizione, indicando per queste ultime il tipo di variabili coinvolte.
2. Utilizzando il dataset **Y0**, si rappresenti graficamente la struttura di correlazione delle variabili e la loro aggregazione gerarchica.
3. Si ottenga lo stesso risultato del punto precedente mediante clustering gerarchico tipo Ward applicato sulla matrice di correlazione delle variabili osservate. Si scelgano ragionevolmente due diverse soluzioni di aggregazione sulla base dei risultati ottenuti.¹
4. Utilizzando il dataset **Y1**, si definisca un modello CFA ortogonale relativo alla soluzione 1 ottenuta al punto precedente, lo si adatti ai dati a disposizione e se ne valuti l'adattamento complessivo.
5. Utilizzando il dataset **Y1**, si definisca un modello CFA relativo alla soluzione 2 ottenuta al punto precedente, lo si adatti ai dati a disposizione e se ne valuti l'adattamento complessivo. Si confronti inoltre l'adattamento ottenuto rispetto a quello del punto 4.
6. Si utilizzi una procedura razionale per migliorare il modello finale scelto al punto 5 e si individui il modello che meglio si adatta ai dati rispetto a quest'ultimo.
7. Si rappresenti graficamente il modello finale scelto al punto 6 e lo si interpreti.
8. Si calcoli mediante un opportuno indice l'attendibilità delle scale derivanti dal modello finale scelto al punto 6 e se ne interpreti il risultato.
9. Si consideri la sola scala **SELF** del dataset **Y1** e si valuti, mediante opportuni modelli CFA, se essa sia un modello di misura parallelo, τ -equivalente o congenerico.
10. Per ogni sottomodello calcolato al punto 9, si calcoli la statistica $s = \|\Sigma_y - \hat{\Sigma}_y\|^2$ che rappresenta una distanza tra la matrice delle covarianze osservata Σ_y e quella riprodotta dal modello $\hat{\Sigma}_y$.² Si interpreti il risultato anche alla luce di quanto ottenuto al punto 9.

¹ Si suggerisce di ricercare una prima soluzione a due cluster ed una seconda soluzione a tre cluster.

² In generale, $\|\mathbf{X}\|$ con $\mathbf{X}_{p \times p}$ matrice quadrata può essere calcolata in R mediante la funzione `norm()`.