PSP6075525 - Testing psicologico (matr. dispari)

Caso studio del 30-08-21

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 .Rmd unitamente al file PDF generato. Si ricorda di riportare <u>chiaramente</u> Nome, Cognome e Matricola all'interno dei file contenenti le soluzioni finali (.pdf, .R, .Rmd).
- 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 caso studio si riferisce alla valutazione dei test ridotti SWLS-III (*Satisfaction With Life Scale*) e HILS-III (*Harmonic in Life*) utilizzati rispettivamente per la valutazione delle componenti cognitive e affettive del benessere soggettivo (*subjective well-being*). Le versioni abbreviate di entrambi i test comprendono tre item ciascuno. I dati si riferiscono ad uno studio[1] che ha coinvolto 299 partecipanti (di cui 214 di genere femminile, 84 di genere maschile, 1 non dichiarato) di nazionalità britannica. Gli item sono stati rilevati su scale ordinali a 7 livelli (*1*: "Strongly Disagree",...,*7*: "Strongly Agree") e sono descritti dalle seguenti assegnazioni semantiche: (1) *My lifestyle allows me to be in harmony*, (2) *Most aspects of my life are in balance*, (3) *I am in harmony* (HILS-III); (1) *In most ways my life is close to my ideal*, (2) *The conditions of my life are excellent*, (3) *I am satisfied with my life* (SWLS-III). Entrambi i test sono stati somministrati allo stesso campione in due tempi, il secondo dei quali a distanza di quattordici giorni in media dal primo. Per entrambe le somministrazioni è stato anche rilevato il tempo (in minuti) necessario al completamento di emtrambi i test (`CompleteTime`).

L'obiettivo dell'analisi è quello di (i) studiare la dimensionalità complessiva del test HILS-III pe entrambi i tempi di somministrazione; (ii) valutare se il costrutto HILS-III sia invariante rispetto ad alcune dimensioni rilevanti dell'indagine.

1. Si individuino il numero di unità statistiche e si commenti il tipo di dato a disposizione.
   Il numero di unità statistiche è pari a $n = 299$ non equamente raggruppate per la variabile genere (maschi: $n = 84$; femmine: $n = 214$; altro: $n = 1$). I dati a disposizione consistono nelle risposte su scala ordinale ai tre item del test in entrambe le somministrazioni.

2. Si crei una nuova variabile indicatrice `Group` che assuma i seguenti valori:

$$\begin{aligned} \texttt{Group} = 0 \quad &\text{se} \quad \texttt{CompleteTime\_t1} < \text{median}(\texttt{CompleteTime\_t1}) \\ \texttt{Group} = 1 \quad &\text{se} \quad \texttt{CompleteTime\_t1} \geq \text{median}(\texttt{CompleteTime\_t1}) \end{aligned}$$

   La condizione `Group=0` indica quelle unità statistiche particolarmente veloci nel completare il test mentre `Group=1` indica quelle unità che sono più lente nel completare il test.

```
group = rep(0,NROW(datax))
iid = datax$CompleteTime_t1 >= median(datax$CompleteTime_t1)
group[iid] = 1
datax$Group = group
```

3. Si valuti la coerenza interna del test HILS complessivo (tempo `t1` e `t2` congiuntamente) mediante indice α di Cronbach[2] rispetto alla variabile `Group` e si commenti il risultato ottenuto.

```
coef_alpha(datax[datax$Group==0,c(4:6,11:13)]) #HILS t1 + t2 quando Group=0

  [1] 0.948145

coef_alpha(datax[datax$Group==1,c(4:6,11:13)]) #HILS t1 + t2 quando Group=1

  [1] 0.9304861
```

   Per il test complessivo HILS-III l'attendibilità è buona sia nel gruppo più veloce che in quello più lento

[1] Kjell, O. N., & Diener, E. (2021). Abbreviated three-item versions of the satisfaction with life scale and the harmony in life scale yield as strong psychometric properties as the original scales. *Journal of Personality Assessment*, 103(2), 183-194.

[2] L'indice può essere calcolato, ad esempio, mediante la funzione `alpha(x=...,)` della libreria `psych`. In alternativa può essere utilizzata la funzione `coef_alpha()` disponibile nel file `reliability.R` nella cartella "Utilities" alla pagina Moodle del corso.

rispetto al tempo di completamento del test.

4. Si definisca un modello fattoriale confermativo ad una sola variabile latente avente come item quelli del test HILS-III rilevati in entrambi i tempi.
La scala è composta da sei item secondo il modello CFA è definito dall'equazione lineare

$$\boldsymbol{\Sigma}_{y_{6\times 6}} = \boldsymbol{\lambda}_{6\times 1}\boldsymbol{\lambda}^{T}_{6\times 1}\phi + \boldsymbol{\Theta}_{\delta_{6\times 6}}$$

L'adattamento ai dati $\mathbf{S}_{y_{6\times 6}}$ può essere fatto mediante stimatori DWLS per dati ordinali. Il modello necessita di 12 parametri da stimare (5 coefficienti fattoriali, 6 varianze d'errore, 1 varianza della variabile latente) su un totale di $p(p+1)/2 = 21$ parametri totali. Poiché la stima è effettuata mediante DWLS, vi sono parametri aggiuntivi da stimare (c.d. *thresholds* degli item) e che si riferiscono alle soglie continue associate alle categorie di risposta.

```
#ricodifichiamo ciascuna variabile osservata per t1 e t2 come variabile ordinale
for(j in c(4:6,11:13)){
  datax[,j] = factor(datax[,j],ordered=TRUE)
}

model1 = "eta=~HILS1_t1+HILS2_t1+HILS3_t1+HILS1_t2+HILS2_t2+HILS3_t2"
fit1 = lavaan::cfa(model = model1,data = datax,
                   ordered = c("HILS1_t1","HILS2_t1","HILS3_t1",
                               "HILS1_t2","HILS2_t2","HILS3_t2"),
                   estimator="DWLS")
```

5. Si interpreti il risultato del modello adattati al punto 4 anche mediante l'utilizzo di indici di adattamento complessivo (si suggerisce l'utilizzo dei coefficienti standardizzati nell'interpretazione della soluzione fattoriale).

```
print(fit1)

  lavaan 0.6-7 ended normally after 21 iterations

    Estimator                                         DWLS
    Optimization method                             NLMINB
    Number of free parameters                           42

    Number of observations                             299

  Model Test User Model:

    Test statistic                                 284.612
    Degrees of freedom                                   9
    P-value (Chi-square)                             0.000
```

```
res1 = lavaan::inspect(fit1,what="std.all")
Xout = cbind(res1$lambda,diag(res1$theta),rep(res1$psi,6))
colnames(Xout)=c("lambda","diag(ThetaDelta)","phi")
print(Xout)

              lambda diag(ThetaDelta) phi
  HILS1_t1 0.8667906       0.24867406   1
  HILS2_t1 0.8663390       0.24945682   1
```

```
  HILS3_t1 0.8945937       0.19970216   1
  HILS1_t2 0.9048304       0.18128199   1
  HILS2_t2 0.8930909       0.20238867   1
  HILS3_t2 0.9497440       0.09798633   1
```

```
fitMeasures(fit1,fit.measures = c("RMSEA","CFI","chisq","df","npar"))
```

```
    rmsea     cfi   chisq      df    npar
    0.321   0.990 284.612   9.000  42.000
```

Globalmente il modello adattato evidenzia un buon indice CFI ma uno scarso RMSEA. La struttura fattoriale della scala è ben formata dagli item a disposizione, con coefficienti fattoriali di magnitudine sufficientemente elevata. Le varianze d'errore per ciascun item sono contenute.

6. Si definisca un secondo modello fattoriale confermativo a due variabili latenti per gli item del test HILS-III rilevati in entrambi i tempi (una variabile latente per gli item al tempo `t1`, un'altra variabile latente per gli item al tempo `t2`). Il modello inoltre deve assumere che (i) la correlazione tra le due variabili latenti sia zero, (ii) le varianze residue degli item per le due scale siano correlate a coppie (es.: item 1 HILS-III a tempo `t1` deve correlare con item 1 HILS-III a tempo `t2` e così via). Successivamente si adatti il modello ai dati a disposizione mediante opportuno metodo di stima.

   Il modello CFA è definito dall'equazione lineare

$$\boldsymbol{\Sigma}_{y_{6\times 6}} = \boldsymbol{\Lambda}_{6\times 2}\boldsymbol{\Phi}_{2\times 2}\boldsymbol{\Lambda}^T_{6\times 2} + \boldsymbol{\Theta}_{\delta_{6\times 6}}$$

   dove

$$\boldsymbol{\Phi}_{2\times 2} = \begin{bmatrix} \sigma_{\eta_1} & \\ 0 & \sigma_{\eta_2} \end{bmatrix}$$

   è diagonale mentre

$$\boldsymbol{\Theta}_{\delta_{6\times 6}} = \begin{bmatrix} \theta_{\delta_{11}} & & & & & \\ 0 & \theta_{\delta_{22}} & & & & \\ 0 & 0 & \theta_{\delta_{33}} & & & \\ \theta_{\delta_{41}} & 0 & 0 & \theta_{\delta_{44}} & & \\ 0 & \theta_{\delta_{52}} & 0 & 0 & \theta_{\delta_{55}} & \\ 0 & 0 & \theta_{\delta_{63}} & 0 & 0 & 0\,\theta_{\delta_{66}} \end{bmatrix}$$

   contiene ora i parametri di correlazione tra le varianze residue d'errore per gli item del tempo `t1` e `t2`. Rispetto al modello unidimensionale precedente, questo modello necessita della stima di tre parametri aggiuntivi, vale a dire le correlazioni a coppie sulle varianze d'errore. L'adattamento del modello ai dati è fatto mediante metodo DWLS per dati ordinali.

```
model2 = "eta1=~HILS1_t1+HILS2_t1+HILS3_t1
eta2=~HILS1_t2+HILS2_t2+HILS3_t2
HILS1_t1~~HILS1_t2
HILS2_t1~~HILS2_t2
HILS3_t1~~HILS3_t2
eta1~~0*eta2"
fit2 = lavaan::cfa(model = model2,data = datax,
                   ordered = c("HILS1_t1","HILS2_t1","HILS3_t1",
                               "HILS1_t2","HILS2_t2","HILS3_t2"),
                   estimator="DWLS")
```

7. Si interpreti il risultato del modello adattato al punto 6. Si valuti infine, mediante l'utilizzo di indici di adattamento complessivo, se la soluzione a due variabili latenti (punto 6) sia superiore o meno a quella a una singola variabile latente (punto 4). Si scelga, dopo opportune argomentazioni, il modello fattoriale

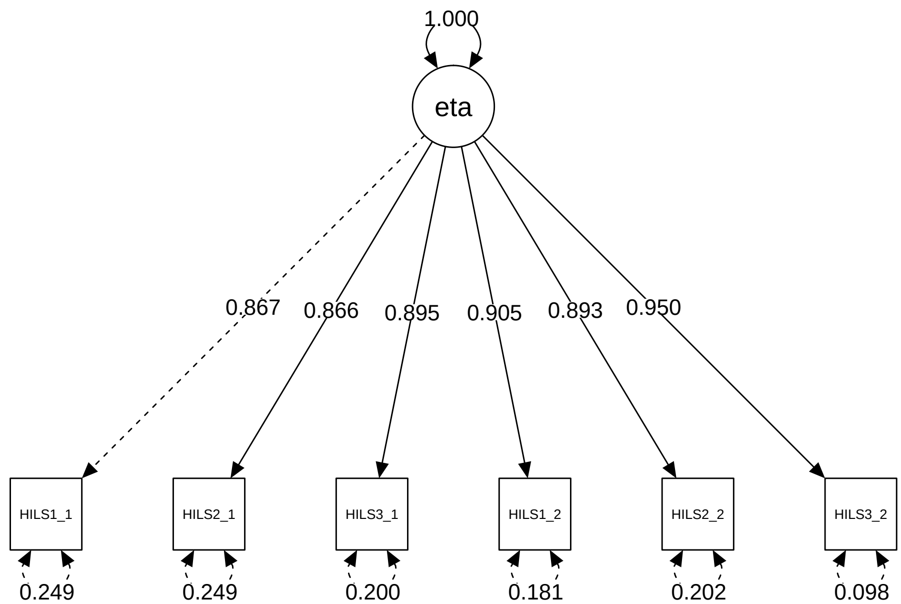
finale che meglio si adatta ai dati.

```
print(fit2)

  lavaan 0.6-7 ended normally after 22 iterations

    Estimator                                       DWLS
    Optimization method                           NLMINB
    Number of free parameters                         45

    Number of observations                           299

  Model Test User Model:

    Test statistic                              3841.843
    Degrees of freedom                                 6
    P-value (Chi-square)                           0.000

res2 = lavaan::inspect(fit2,what="std.all")
Xout = cbind(res2$lambda,diag(res2$theta),res2$psi[2,1],res1$lambda,diag(res1$theta),0)
colnames(Xout)=c("M2_lambda1","M2_lambda2","M2_diag(ThetaDelta)","M2_phi12",
                 "M1_lambda","M1_diag(ThetaDelta)","M1_phi12")
print(Xout)

           M2_lambda1 M2_lambda2 M2_diag(ThetaDelta) M2_phi12 M1_lambda
  HILS1_t1  0.8990754  0.0000000          0.19166337        0 0.8667906
  HILS2_t1  0.9046961  0.0000000          0.18152498        0 0.8663390
  HILS3_t1  0.9538134  0.0000000          0.09024000        0 0.8945937
  HILS1_t2  0.0000000  0.9295306          0.13597295        0 0.9048304
  HILS2_t2  0.0000000  0.9138753          0.16483185        0 0.8930909
  HILS3_t2  0.0000000  0.9706159          0.05790469        0 0.9497440
           M1_diag(ThetaDelta) M1_phi12
  HILS1_t1          0.24867406        0
  HILS2_t1          0.24945682        0
  HILS3_t1          0.19970216        0
  HILS1_t2          0.18128199        0
  HILS2_t2          0.20238867        0
  HILS3_t2          0.09798633        0

Yout = rbind(fitMeasures(fit2,fit.measures = c("RMSEA","CFI","chisq","df","npar")),
             fitMeasures(fit1,fit.measures = c("RMSEA","CFI","chisq","df","npar")))
rownames(Yout)=c("mod2","mod1")
print(Yout)

           rmsea       cfi     chisq df npar
  mod2 1.4646932 0.8640988 3841.8433  6   45
  mod1 0.3205674 0.9902353  284.6115  9   42
```

Il modello a due variabili latenti presenta un adattamento complessivo inferiore rispetto a quello del modello unidimensionale. Ciò è evidente anche se la struttura fattoriale del modello a due fattori sia ben formata, come evidenziato anche dalle basse varianze residue (più basse rispetto al modello unidimensionale). Ciononostante, considerando gli indici complessivi del modello, la soluzione a due fattori risulta più scarsa della precedente e pertanto si sceglie di considerare il modello unidimensionale per le analisi successive.

8. Sulla base dei risultati ottenuti al punto 7, si rappresenti graficamente il modello finale scelto.

```
semPlot::semPaths(object = fit1, whatLabels = "std.all",edge.label.cex = 0.95,
                  edge.color = "black",sizeMan = 7,sizeLat=8, style = "openmx",
                  nDigits = 3, intercepts = FALSE,thresholds = FALSE)
```



9. Si valuti se il modello finale scelto al punto 7 sia invariante in senso forte rispetto alla variabile `Group`. In particolare, si vuole indagare se la struttura fattoriale del test HILS-III complessivo (per entrambi i tempi congiuntamente) differisca tra coloro che presentano tempi di completamento del test più veloci e più lenti.

   Un modello CFA unidimensionale è invariante in senso forte all'interno di $g = (1,2)$ gruppi quando è possibile scriverlo come segue:

$$\boldsymbol{\Sigma}^{(1)}_{y_{3\times 3}} = \boldsymbol{\tau} + \boldsymbol{\lambda}_{3\times 1}\phi^{(1)}_{1\times 1}\boldsymbol{\lambda}^{T}_{3\times 1} + \boldsymbol{\Theta}^{(1)}_{\delta_{3\times 3}}$$
$$\boldsymbol{\Sigma}^{(2)}_{y_{3\times 3}} = \boldsymbol{\tau} + \boldsymbol{\lambda}_{3\times 1}\phi^{(2)}_{1\times 1}\boldsymbol{\lambda}^{T}_{3\times 1} + \boldsymbol{\Theta}^{(2)}_{\delta_{3\times 3}}$$

   dove (i) $\boldsymbol{\lambda}^{(1)} = \boldsymbol{\lambda}^{(2)} = \boldsymbol{\lambda}$ e $\boldsymbol{\tau}^{(1)} = \boldsymbol{\tau}^{(2)} = \boldsymbol{\tau}$. Il test del χ^2 per modelli annidati permette di valutare se un tale modello vincolato $\mathcal{M}_{\text{forte}}$ sia superiore al modello $\mathcal{M}_{\text{debole}}$ in cui si ha il solo vincolo delle matrici dei coefficienti fattoriali $\boldsymbol{\Lambda}^{(1)} = \boldsymbol{\Lambda}^{(2)}$ (*modello debole*). Se l'ipotesi nulla

$$H_0 : \chi^2_{\mathcal{M}_{\text{forte}}} - \chi^2_{\mathcal{M}_{\text{debole}}} = 0$$

   non è rigettata allora i gruppi sono invarianti in senso forte (il modello $\mathcal{M}_{\text{forte}}$ è scelto rispetto a $\mathcal{M}_{\text{debole}}$).

```
#modello CFA debole
fit_debole = lavaan::cfa(model = model1,group = "Group",data=datax,
                         group.equal=c("loadings"),
                         ordered = c("HILS1_t1","HILS2_t1","HILS3_t1",
                                     "HILS1_t2","HILS2_t2","HILS3_t2"),
                         estimator="DWLS")

#modello CFA forte
fit_forte = lavaan::cfa(model = model1,group = "Group",data=datax,
```

```
                        group.equal=c("loadings","intercepts"),
                        ordered = c("HILS1_t1","HILS2_t1","HILS3_t1",
                                    "HILS1_t2","HILS2_t2","HILS3_t2"),
                        estimator="DWLS")

#chi-square test
lavaan::anova(fit_debole,fit_forte)

  Chi-Squared Difference Test

             Df AIC BIC  Chisq Chisq diff Df diff Pr(>Chisq)
  fit_debole 23         359.94
  fit_forte  52         374.10     14.167      29     0.9905
```

Dopo aver adattato i due sottomodelli CFA ai dati a disposizione, il test inferenziale evidenzia il fatto che i due gruppi siano invarianti in senso forte (H_0 non è rigettata ad un $\alpha = 0.05$). Il modello fattoriale confermativo invariante in senso forte rappresenta la struttura fattoriale del test complessivo HILS allo stesso modo per il gruppo dei soggetti più veloci e quelli più lenti. Inoltre, le medie degli item (codificate dal parametro $\boldsymbol{\tau}$) non differiscono rispetto alla velocità di completamento del test.

10. Si valuti se il modello scelto al punto 7 sia di tipo τ-equivalente. Un modello fattoriale è di tipo τ-equivalente quando è possibile scriverlo nella forma

$$\boldsymbol{\Sigma}_{y_{6\times 6}} = \lambda \mathbf{1}_6 \mathbf{1}_6^T \phi + \boldsymbol{\Theta}_{\delta_{6\times 6}}$$

ossia quando gli item che formano il modello di misura presentano stesso coefficiente fattoriale λ (le varianze d'errore possono essere diverse per ciascun item). La stima dei parametri di tale modello può essere fatta come per qualsiasi modello CFA mediante la funzione `cfa(,...std.lv=TRUE)` della libreria `lavaan` come di seguito riportato.

```
model3 = "eta =~ l*HILS1_t1 + l*HILS2_t1 + l*HILS3_t1 +
          l*HILS1_t2 + l*HILS2_t2 + l*HILS3_t2"
model3_fit = lavaan::cfa(model = model3,data = datax,std.lv=TRUE,
                         ordered = c("HILS1_t1","HILS2_t1","HILS3_t1",
                               "HILS1_t2","HILS2_t2","HILS3_t2"),
                          estimator="DWLS")
```

```
res1 = lavaan::inspect(model3_fit,what="std.all")
Xout = cbind(res1$lambda,diag(res1$theta),rep(res1$psi,6))
colnames(Xout)=c("lambda","diag(ThetaDelta)","phi")
print(Xout)

              lambda diag(ThetaDelta) phi
  HILS1_t1 0.9016481        0.1870307   1
  HILS2_t1 0.9016481        0.1870307   1
  HILS3_t1 0.9016481        0.1870307   1
  HILS1_t2 0.9016481        0.1870307   1
  HILS2_t2 0.9016481        0.1870307   1
  HILS3_t2 0.9016481        0.1870307   1
```

```
cfa.fits = fitMeasures(model3_fit,
                       fit.measures = c("Chisq","df","pvalue","RMSEA","CFI"))
print(cfa.fits)
```

```
    chisq       df   pvalue    rmsea      cfi
  338.858   14.000    0.000    0.279    0.988
```

Per il sottomodello di misura τ-equivalente, data la numerosità campionaria elevata, il valore della statistica del χ^2 sotto l'ipotesi nulla $\mathcal{H}_0 : \boldsymbol{\Sigma}_y - \hat{\boldsymbol{\Sigma}}_y = \mathbf{0}$ non può essere utilizzato (il test tende ad essere quasi sempre rigettato). Per tale motivo, valutiamo il sottomodello mediante RMSEA e CFI. Sebbene rispetto all'indice CFI il modello τ-equivalente sembra essere soddisfacente, se valutato anche l'indice RMSEA, questo non sembra essere ottimale. Si noti che, rispetto al modello congenerico (adattato al punto 4), il modello τ-equivalente sembra adattarsi meglio ai dati. Tuttavia, sulla base degli indici complessivi, non possiamo affermare che la scala `HILS-III` complessiva presenti una struttura di misurazione di tipo τ-equivalente.