PSP6075525 - Testing psicologico (matr. dispari)

Caso studio del 10-06-22

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 all'analisi della dimensionalità del test AEO2x di solito utilizzato per la valutazione del dolore post-operatorio (*functional interference of pain*), in particolare della sua interferenza con le attività quotidiane dei pazienti. Il test è composto da due dimensioni (I: *intensity*; F: *interference*) quantificate mediante 8 item di tipo dicotomico (Sì/No). I dati si riferiscono ad uno studio che ha coinvolto un campione di 250 partecipanti (di cui 125 di genere femminile) rappresentativo della popolazione veneta tra 40 e 50 anni. L'obiettivo dell'analisi è quello di studiare la dimensionalità complessiva del test AEO2x, nelle sue proprietà di attendibilità nella rilevazione del fenomeno *functional interference of pain* e di invarianza rispetto al genere.

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 $n = 250$ (maschi: $n = 125$; femmine: $n = 125$). I dati a disposizione consistono nelle risposte su scala ordinale ai $p = 8$ item del test. Le variabili riferite agli item sono di tipo dicotomico.

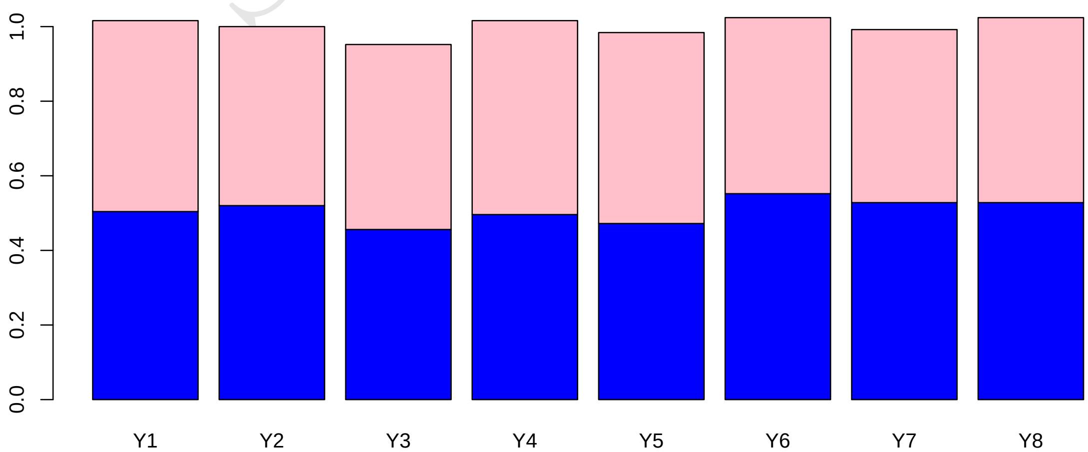
2. Si calcolino e si rappresentino graficamente delle statistiche opportune per sintetizzare l'informazione sugli item in entrambi i gruppi.
   Considerata la natura degli item, una statistica ragionevole per la sintesi dell'informazione in entrambi i gruppi consiste nella proporzione di risposte positive $\overline{y} = \sum_{i \in \mathcal{G}} y_i$ (dove $\mathcal{G}$ è il gruppo delle osservazioni di riferimento). Una possibile rappresentazione grafica consiste nell'utilizzare un grafico a barre di tipo *stacked* da cui è possibile evincere la proporzione di risposte affermative (i.e., di presenza del dolore) in entrambi i gruppi e per ciascun item.

```
yM = apply(datax[datax$group=="M",1:8],2,sum)/125
yF = apply(datax[datax$group=="F",1:8],2,sum)/125
```

```
plot_data = rbind(yM,yF); rownames(plot_data)=c("M","F")
print(plot_data)

      Y1   Y2    Y3    Y4    Y5    Y6    Y7    Y8
  M 0.504 0.52 0.456 0.496 0.472 0.552 0.528 0.528
  F 0.512 0.48 0.496 0.520 0.512 0.472 0.464 0.496

barplot(plot_data,col=c("blue","pink"))
```



```
yM_ov = mean(yM)
yF_ov = mean(yF)
print(c(yM_ov,yF_ov))

  [1] 0.507 0.494
```

Complessivamente possiamo notare come sia maschi che femmine esprimono la stessa proporzione di presenza del dolore.

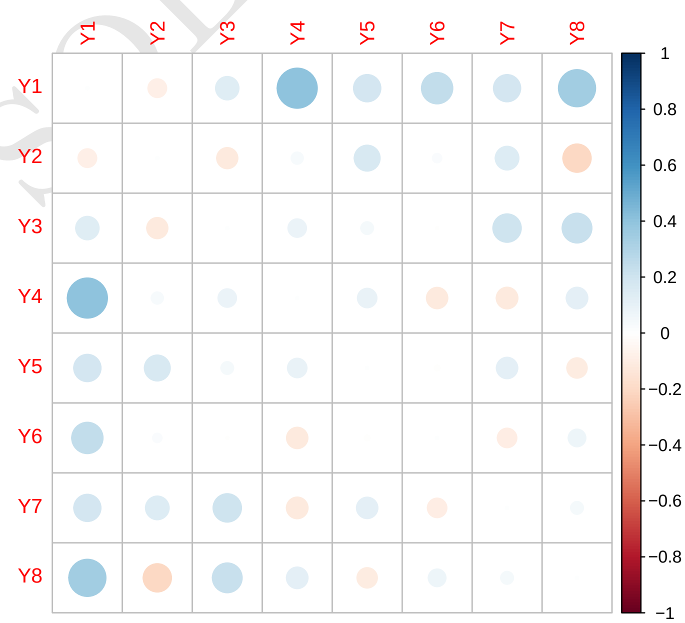
3. Si stimino le matrici di correlazione policorica tra le variabili indicatrici differenziate per la variabile gruppo. Successivamente si valuti mediante un opportuno indice descrittivo se le due strutture di correlazione sono dissimili. Nota: per la stima della matrice di correlazione policorica si può utilizzare la funzione **`psych::polychoric()`** mentre un indice ragionevole per la comparazione tra le due matrici è l'indice di entanglement (**`dendextend::entanglement()`**).
   Le due strutture di correlazione tra maschi e femmine appaiono complessivamente differenti (l'indice di entanglement è, in questo caso, relativamente basso).

```
R_M = psych::polychoric(datax[datax$group=="M",1:8])$rho
R_F = psych::polychoric(datax[datax$group=="F",1:8])$rho
hclust.matA = hclust(d = dist(R_M),method = "ward.D2")
hclust.matB = hclust(d = dist(R_F),method = "ward.D2")
dendA = as.dendrogram(hclust.matA); dendB = as.dendrogram(hclust.matB)
dendextend::entanglement(dendA,dendB)

  [1] 0.2717497
```

4. Sulla base del punto precedente, si considerino le matrici di correlazione policorica per maschi $\mathbf{R}_M$ e femmine $\mathbf{R}_F$. Si calcoli la matrice $\mathbf{R}_{\texttt{diff}} = \mathbf{R}_M - \mathbf{R}_F$ e la si rappresenti graficamente. Si interpreti il risultato ottenuto.

```
R_d = R_M - R_F
corrplot::corrplot(R_d)
```



Il grafico di correlazione evidenzia graficamente il risultato riprodotto dall'indice di entanglement. I due gruppi evidenziano simili correlazioni su alcune coppie di item (ad esempio, `Y2-Y1`, `Y6-Y3`) ma alti valori di correlazione su altre coppie di variabili (ad esempio, `Y4-Y1`). Sulla base di tali considerazioni, a livello del campione raccolto, maschi e femmine riportano strutture di correlazione differenti tra loro.

5. Si definisca e si adatti un modello fattoriale confermativo appropriato rispetto ai dati a disposizione. Per l'adattamento ai dati si consideri la seguente assegnazione:

```
I: Y1 Y2 Y3 Y4
F: Y5 Y6 Y7 Y8
```

Si ricordi che il modello deve essere adattato tenendo conto della variabile categoriale `group` mediante il comando `cfa(..,group="group")`.

```
for(j in 1:8){
  datax[,j] = factor(datax[,j],ordered = TRUE)
}

mod1 = "I =~ Y1+Y2+Y3+Y4 \n F =~ Y5+Y6+Y7+Y8"
mod1_fit = lavaan::cfa(model = mod1,data = datax,group = "group",
                       estimator="DWLS",ordered = colnames(datax)[1:8])
```

Il modello è composto da $q = 2$ variabili latenti per $p = 8$ variabili osservate secondo l'equazione lineare

$$\boldsymbol{\Sigma}_{y_{8\times8}} = \boldsymbol{\Lambda}_{8\times2}\boldsymbol{\Phi}_{2\times2}\boldsymbol{\Lambda}^T_{8\times2} + \boldsymbol{\Theta}_{\delta_{8\times8}}$$

L'adattamento ai dati $\mathbf{S}_{y_{8\times8}}$ deve essere fatto mediante stimatori DWLS per dati ordinali.

6. Si interpreti il risultato del modello adattato al punto precedente anche mediante l'utilizzo di indici di adattamento complessivo.

```
print(mod1_fit)

  lavaan 0.6-7 ended normally after 34 iterations

    Estimator                                       DWLS
    Optimization method                           NLMINB
    Number of free parameters                         34

    Number of observations per group:
      M                                              125
      F                                              125

  Model Test User Model:

    Test statistic                                26.958
    Degrees of freedom                                38
    P-value (Chi-square)                           0.909
    Test statistic for each group:
      M                                           17.314
      F                                            9.644
```

```
res1 = lavaan::inspect(mod1_fit,what="std.all")
Xout_M = cbind(res1$M$lambda,diag(res1$M$theta))
colnames(Xout_M)=c("lambda_I","lambda_F","diag(ThetaDelta)")
print(Xout_M)
```

```
      lambda_I  lambda_F diag(ThetaDelta)
  Y1 0.7868151 0.0000000        0.3809219
  Y2 0.8050793 0.0000000        0.3518473
  Y3 0.7430229 0.0000000        0.4479170
  Y4 0.7183871 0.0000000        0.4839200
  Y5 0.0000000 0.7252080        0.4740733
  Y6 0.0000000 0.7047498        0.5033277
  Y7 0.0000000 0.7686074        0.4092426
  Y8 0.0000000 0.6800304        0.5375586
```

```
Xout_F = cbind(res1$F$lambda,diag(res1$F$theta))
colnames(Xout_F)=c("lambda_I","lambda_F","diag(ThetaDelta)")
print(Xout_F)
```

```
      lambda_I  lambda_F diag(ThetaDelta)
  Y1 0.5645922 0.0000000        0.6812357
  Y2 0.9181854 0.0000000        0.1569356
  Y3 0.7092949 0.0000000        0.4969008
  Y4 0.6817445 0.0000000        0.5352244
  Y5 0.0000000 0.7059057        0.5016971
  Y6 0.0000000 0.7560276        0.4284222
  Y7 0.0000000 0.7030376        0.5057381
  Y8 0.0000000 0.6786618        0.5394181
```

```
res1$M$psi
```

```
    I     F
  I 1.000
  F 0.971 1.000
```

```
res1$F$psi
```

```
    I    F
  I 1.00
  F 0.83 1.00
```

```
fitMeasures(mod1_fit,fit.measures = c("RMSEA","CFI","chisq","df","npar"))
```

```
   rmsea    cfi  chisq     df   npar
   0.000  1.000 26.958 38.000 34.000
```

Globalmente il modello adattato evidenzia ottimi indici di CFI e RMSEA. La struttura fattoriale delle scale I e F è ben formata in entrambi i gruppi. La matrice $\boldsymbol{\Phi}_{2\times 2}$ in entrambi i gruppi evidenzia che le variabili latenti sono correlate tra loro. Ad eccezione di qualche item (es.: `Y1`), i gruppi presentano coefficienti fattoriali relativamente simili.

7. Si valuti il livello di invarianza che il modello adattato al punto 5 può raggiungere. Nota: l'invarianza deve essere valutata rispetto alla variabile categoriale `group`.

```
mod2_fit = lavaan::cfa(model = mod1,data = datax,group = "group",
                       estimator="DWLS",ordered = colnames(datax)[1:8],
                       group.equal=c("loadings"))

anova(mod1_fit,mod2_fit)
```

```
  Chi-Squared Difference Test
```

```
          Df AIC BIC  Chisq Chisq diff Df diff Pr(>Chisq)
  mod1_fit 38         26.959
  mod2_fit 44         34.061     7.1027       6     0.3114
```

```
mod3_fit = lavaan::cfa(model = mod1,data = datax,group = "group",
                       estimator="DWLS",ordered = colnames(datax)[1:8],
                       group.equal=c("loadings","intercepts"))

anova(mod2_fit,mod3_fit)
```

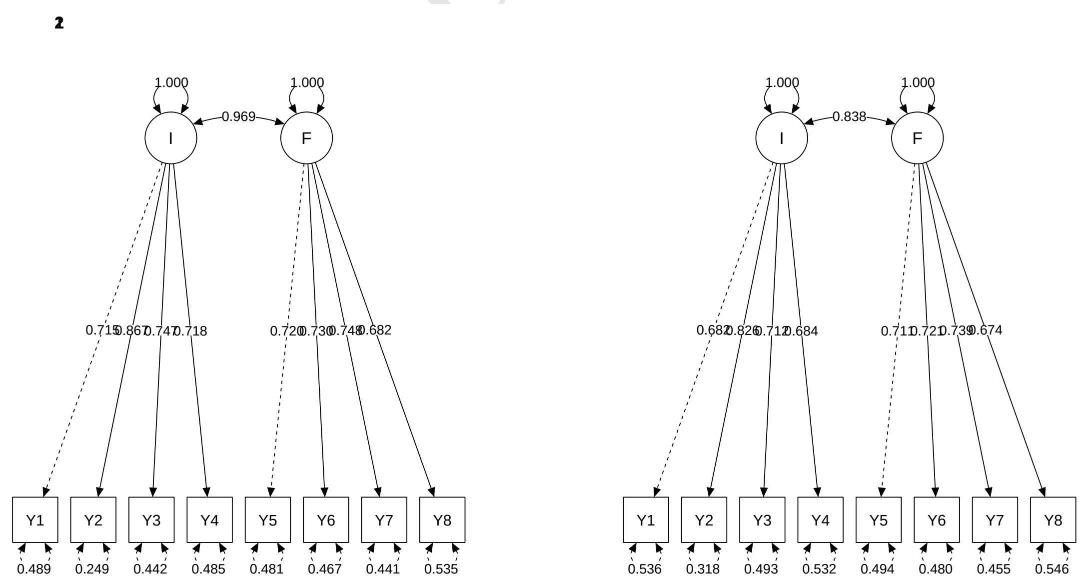
```
  Chi-Squared Difference Test

          Df AIC BIC  Chisq Chisq diff Df diff Pr(>Chisq)
  mod3_fit 42         27.629
  mod2_fit 44         34.061     6.4325       2    0.04011 *
  ---
  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Il modello adattato al punto 5 raggiunge il livello di invarianza debole, ossia quella in cui i due gruppi presentano medesimi coefficienti fattoriali $\boldsymbol{\Lambda}_M = \boldsymbol{\Lambda}_F$. Ciò indica che gli indicatori osservati rispondono ad una struttura a due fattori per il *functional interference of pain* in entrambi i gruppi.

8. Si rappresenti graficamente il modello finale scelto al punto precedente. Nota: per la visualizzazione dei due grafici si ricordi di utilizzare nel comando `semPlot::semPaths(...,panelGroups=TRUE)`.

```
semPlot::semPaths(object = mod2_fit, whatLabels = "std.all",edge.label.cex = 0.95,
                  edge.color = "black",sizeMan = 7,sizeLat=8, style = "openmx",
                  nDigits = 3, intercepts = FALSE,thresholds = FALSE,
                  panelGroups = TRUE)
```



9. Si calcoli l'attendibilità del modello adattato al punto 5 si commenti il risultato ottenuto.

```
semTools::reliability(object = mod2_fit,what = "omega",return.total = TRUE)

  $M
                I         F     total
  omega 0.7250681 0.6794841 0.8227972

  $F
                I         F     total
  omega 0.6883014 0.6706649 0.7945843
```

I risultati indicano che complessivamente il test presenta una buona attendibilità nei due gruppi, sebbene i coefficienti ω parziali siano più bassi rispetto all'indice complessivo.

10. Si stimino i punteggi veri per entrambi i gruppi applicando la seguente formula:

$$y^{M}_{\texttt{tot}_i} = \omega_M \times \underbrace{\left(\frac{1}{p}\sum_{j=1}^{p} y_j^{(i)}\right)}_{\text{media attraverso item}} + (1-\omega_M) \times \underbrace{\left(\frac{1}{p \cdot n_M}\sum_{i=1}^{n_M}\sum_{j=1}^{p} y_{ij}\right)}_{\text{media complessiva}}$$

$$y^{F}_{\texttt{tot}_i} = \omega_F \times \underbrace{\left(\frac{1}{p}\sum_{j=1}^{p} y_j^{(i)}\right)}_{\text{media attraverso item}} + (1-\omega_F) \times \underbrace{\left(\frac{1}{p \cdot n_F}\sum_{i=1}^{n_F}\sum_{j=1}^{p} y_{ij}\right)}_{\text{media complessiva}}$$

dove ω_M e ω_F sono gli indici di attendibilità calcolati al punto precedente per entrambi i gruppi. Si dia una spiegazione degli indici utilizzati e dei risultati ottenuti.

```
res = semTools::reliability(object = mod2_fit,what = "omega",return.total = TRUE)
str(res)

  List of 2
   $ M: num [1, 1:3] 0.725 0.679 0.823
    ..- attr(*, "dimnames")=List of 2
    .. ..$ : chr "omega"
    .. ..$ : chr [1:3] "I" "F" "total"
   $ F: num [1, 1:3] 0.688 0.671 0.795
    ..- attr(*, "dimnames")=List of 2
    .. ..$ : chr "omega"
    .. ..$ : chr [1:3] "I" "F" "total"

ym_M = apply(datax[datax$group=="M",1:8],1,mean)
ym_F = apply(datax[datax$group=="F",1:8],1,mean)
omega_M = res[[1]][3] #males
omega_F = res[[2]][3] #females

yt_M = ym_M*omega_M + mean(ym_M)*(1-omega_M)
yt_F = ym_F*omega_F + mean(ym_F)*(1-omega_F)
```

Gli indicatori permettono di calcolare i punteggi veri al test per entrambi i gruppi. Poiché gli item sono dicotomici, il calcolo della media restituisce la proporzione per ciascun soggetto (attraverso gli item) di risposte Sì, vale a dire la proporzione media di dolore post-operatorio autoriferito. Rispetto alla semplice media attraverso gli item, questi indicatori sono corretti per l'attendibilità ω del test. Valori alti di attendibilità pesano maggiormente le medie calcolate sul campione (prima parte della formula) mentre valori bassi di attendibilità pesano in maniera maggiore la media complessiva al test (minor peso al punteggio individuale).