

PSP6075525 - Testing psicologico (matr. dispari)

Caso studio del 08-07-22

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 chiaramente 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 bh90210 sovente utilizzato per la quantificazione dell'abilità nell'individuare dissonanze sonore all'interno di una sequenza armonica consonante. In letteratura tale test è spesso associato a due dimensioni sebbene non manchino lavori che riferiscono una sola dimensione. Il test è composto da otto items rilevati su una scala pseudo-continua nell'intervallo $[-4, 4]$ dove valori positivi e di magnitudine elevata indicano forte abilità nell'individuazione delle dissonanze armoniche. I dati si riferiscono ad uno studio sperimentale che ha coinvolto due gruppi sottoposti rispettivamente ad una sorgente sonora a frequenze lente (*slow*: 79 partecipanti) e una a frequenze veloci (*fast*, 101 individui). L'obiettivo dell'analisi è quello di studiare la dimensionalità complessiva del test bh90210 e di valutare se esso sia invariante nella misurazione della dissonanza cognitiva sonora per entrambi i gruppi.

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 $n = 180$ non ripartiti equamente nei due gruppi sperimentali. I dati a disposizione consistono nelle risposte sulle otto scale pseudo-continue. Le variabili riferite agli item sono di tipo continuo.

2. Si calcolino e si rappresentino graficamente delle statistiche opportune per sintetizzare l'informazione sugli item per entrambi i gruppi.

Considerando la natura delle variabili disponibili, alcune statistiche opportune sono quelle notevoli per le variabili reali (es.: minimo, massimo, media, varianza). Il calcolo di esse può essere agevolmente effettuato mediante la funzione `describeBy()` della libreria `psych`.

```
psych::describeBy(x = datax[,1:8], group = datax$group_label)
```

Descriptive statistics by group

group: slow

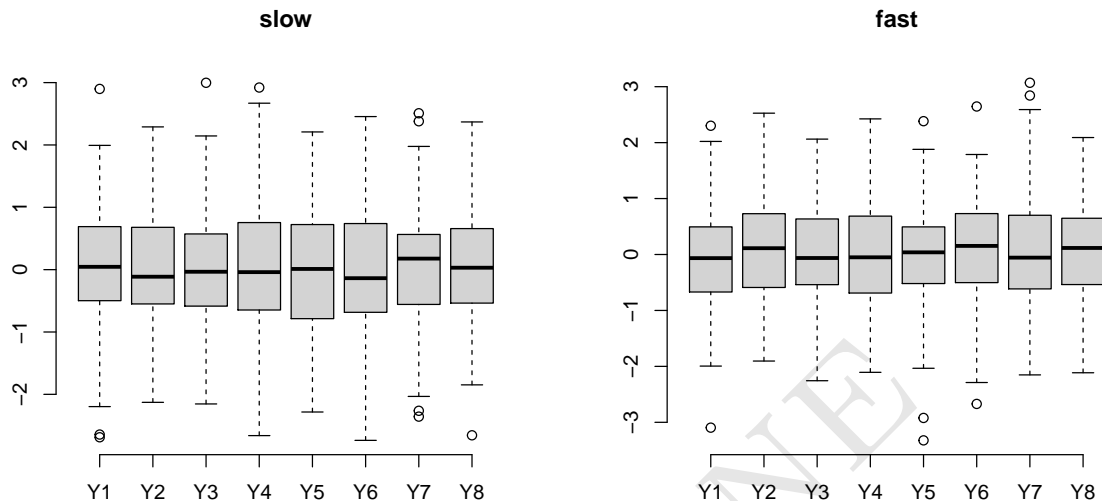
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Y1	1	79	0.08	1.03	0.05	0.08	0.86	-2.69	2.90	5.59	-0.11	0.45	0.12
Y2	2	79	0.08	0.93	-0.11	0.04	0.87	-2.13	2.29	4.42	0.35	-0.17	0.10
Y3	3	79	0.04	0.93	-0.03	0.01	0.86	-2.15	3.00	5.15	0.47	0.43	0.10
Y4	4	79	0.05	1.04	-0.04	0.03	0.95	-2.66	2.92	5.58	0.20	0.57	0.12
Y5	5	79	-0.02	0.91	0.01	-0.02	1.15	-2.28	2.21	4.49	-0.06	-0.73	0.10
Y6	6	79	0.05	1.02	-0.14	0.01	1.00	-2.74	2.46	5.19	0.29	-0.08	0.11
Y7	7	79	0.00	0.97	0.18	0.01	1.02	-2.36	2.51	4.87	-0.04	0.09	0.11
Y8	8	79	0.07	0.94	0.03	0.05	0.90	-2.66	2.37	5.03	0.08	0.10	0.11

group: fast

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Y1	1	101	-0.03	0.95	-0.06	-0.03	0.90	-3.10	2.30	5.40	-0.09	0.22	0.09
Y2	2	101	0.12	0.96	0.11	0.09	1.04	-1.90	2.53	4.43	0.17	-0.54	0.10
Y3	3	101	-0.02	0.85	-0.06	-0.01	0.89	-2.26	2.06	4.32	-0.03	-0.34	0.08
Y4	4	101	0.01	1.02	-0.05	0.01	1.07	-2.11	2.42	4.53	0.10	-0.40	0.10
Y5	5	101	0.01	1.00	0.04	0.03	0.83	-3.32	2.38	5.71	-0.36	0.74	0.10
Y6	6	101	0.05	0.96	0.15	0.08	0.91	-2.67	2.65	5.32	-0.25	0.16	0.10
Y7	7	101	0.05	1.10	-0.06	0.02	0.99	-2.15	3.07	5.22	0.30	-0.06	0.11
Y8	8	101	0.09	0.87	0.12	0.08	0.81	-2.11	2.09	4.21	0.04	-0.09	0.09

3. Si rappresentino le variabili che compongono il test mediante una opportuna rappresentazione grafica. Si interpreti il risultato ottenuto.

```
par(mfrow=c(1,2))
boxplot(datax[datax$group_label=="slow",1:8],frame=FALSE,main="slow")
boxplot(datax[datax$group_label=="fast",1:8],frame=FALSE,main="fast")
```



Complessivamente, in entrambi i gruppi, le variabili sono centrate sul valore neutro della scala e presentano dispersione contenuta. Graficamente, il gruppo *slow* non sembra evidenziare differenze in termini di variabilità o tendenza centrale rispetto al gruppo *fast*.

4. Si calcoli, per entrambi i gruppi, il seguente indicatore sintetico:

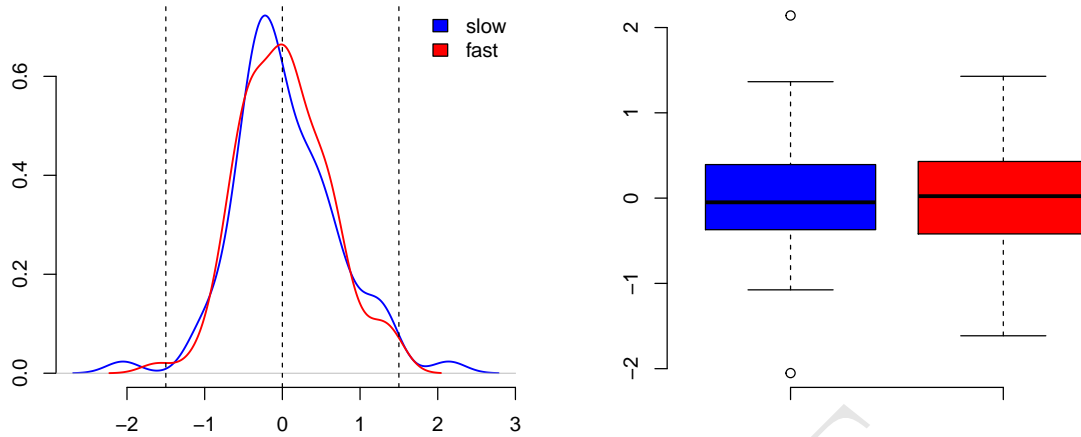
$$h_i^g = \frac{\alpha_g}{8} \sum_{i=1}^{n_g} \mathbf{y}_{ig} \quad g \in \{\text{slow}, \text{fast}\}$$

dove α_g è l'indice α di Chronbach calcolato per il g -esimo gruppo mentre \mathbf{y}_{ig} si riferisce al vettore 1×8 contenente le variabili osservate per l' i -esima unità statistica e il g -esimo gruppo. Si rappresenti il risultato graficamente e lo si interpreti.

```
alpha_slow = coef_alpha(datax[datax$group_label=="slow",1:8])
alpha_fast = coef_alpha(datax[datax$group_label=="fast",1:8])

h_slow = (1/8)*rowSums(x = datax[datax$group_label=="slow",1:8])*alpha_slow
h_fast = (1/8)*rowSums(x = datax[datax$group_label=="fast",1:8])*alpha_fast

par(mfrow=c(1,2))
plot(density(h_slow),bty="n",main="",xlab="",ylab="",col="blue",lwd=1.5)
lines(density(h_fast),bty="n",main="",xlab="",ylab="",col="red",lwd=1.5)
legend(x="topright",fill=c("blue","red"),legend = c("slow","fast"),bty = "n")
abline(v = c(-1.5,0,1.5),lty=2)
boxplot(h_slow,h_fast,frame=FALSE,main="",col=c("blue","red"))
```



```
print(c(alpha_slow,alpha_fast))
[1] 0.8948277 0.8482264
```

I due gruppi presentano indici di attendibilità molto elevati sugli otto item rilevati e presentano, complessivamente, punteggi complessivi al test sovrapponibili sia in termini di variabilità che di tendenza centrale. Entrambi i gruppi mostrano punteggi medi di individuazione della dissonanza sonora intorno al valore neutro.

5. Si definiscano e si adattino due modelli fattoriali confermativi appropriati rispetto ai dati a disposizione. Il primo modello lo si consideri unidimensionale mentre per il secondo si consideri la seguente assegnazione:

η_1 : Y1 Y2 Y3 Y4
 η_2 : Y5 Y6 Y7 Y8

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

```
mod1 = "eta =~ Y1+Y2+Y3+Y4+Y5+Y6+Y7+Y8"
mod1_fit = lavaan::cfa(model = mod1,data = datax,group = "group_label")

mod2 = "eta1 =~ Y1+Y2+Y3+Y4 \n eta2 =~ Y5+Y6+Y7+Y8"
mod2_fit = lavaan::cfa(model = mod2,data = datax,group = "group_label")
```

Il primo modello è composto da $q = 1$ variabili latenti per $p = 8$ variabili osservate secondo l'equazione lineare

$$\Sigma_{y_{8 \times 8}} = \lambda_{8 \times 1} \lambda_{8 \times 1}^T \phi + \Theta_{\delta_{8 \times 8}}$$

mentre il secondo, essendo composto da $q = 2$ variabili latenti sempre per $p = 8$ variabili osservate, risponde all'equazione seguente:

$$\Sigma_{y_{8 \times 8}} = \Lambda_{8 \times 2} \Phi_{2 \times 2} \Lambda_{8 \times 2}^T + \Theta_{\delta_{8 \times 8}}$$

L'adattamento ai dati $\mathbf{Y}_{180 \times 8}$ deve essere fatto, in questo caso, mediante gli stimatori ottenuti per massima verosimiglianza.

6. Si scelga il modello ottimale tra i due adattati al punto precedente e lo si interpreti anche mediante l'utilizzo di indici di adattamento complessivo.

```
lavaan::anova(mod2_fit,mod1_fit,test="Chisq")
```

Chi-Squared Difference Test

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
mod2_fit	38	3535.0	3694.6	41.801			
mod1_fit	40	3531.2	3684.5	42.055	0.25423	2	0.8806

Il modello che meglio si adatta ai dati è quello unidimensionale che, rispetto a quello bidimensionale, presenta meno parametri liberi (maggiore parsimonia).

```
res1 = lavaan::inspect(mod1_fit,what="std.all")
```

```
Xout_s = cbind(res1$slow$lambda,diag(res1$slow$theta))
colnames(Xout_s)=c("lambda_s","diag(ThetaDelta)")
print(Xout_s)
```

	lambda_s	diag(ThetaDelta)
Y1	0.6684987	0.5531095
Y2	0.7004690	0.5093432
Y3	0.7460113	0.4434672
Y4	0.7674974	0.4109478
Y5	0.6742017	0.5454521
Y6	0.7304284	0.4664743
Y7	0.7202169	0.4812876
Y8	0.7443552	0.4459354

```
Xout_f = cbind(res1$fast$lambda,diag(res1$fast$theta))
colnames(Xout_f)=c("lambda","diag(ThetaDelta)")
print(Xout_f)
```

	lambda	diag(ThetaDelta)
Y1	0.6990994	0.5112601
Y2	0.5891378	0.6529166
Y3	0.4841763	0.7655733
Y4	0.5910593	0.6506489
Y5	0.6873442	0.5275580
Y6	0.7486233	0.4395631
Y7	0.7180677	0.4843787
Y8	0.6111800	0.6264590

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

rmsea	cfi	chisq	df	npar
0.024	0.996	42.055	40.000	48.000

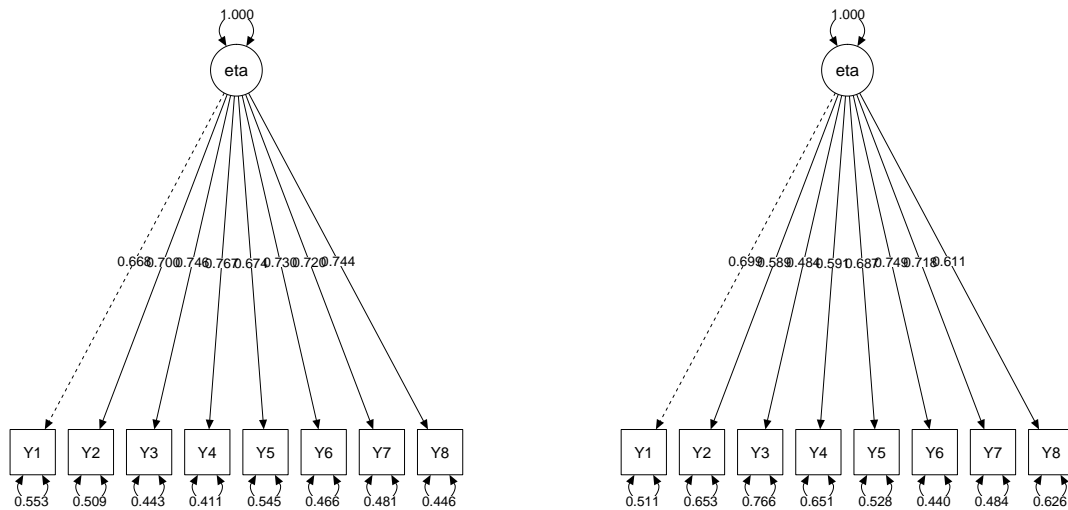
Tale modello presenta buoni indici di adattamento complessivi con coefficienti fattoriali sufficientemente elevanti nel sottomodello *slow* e accettabili nel sottomodello *fast*. In quest'ultimo caso, qualche indicatore osservato presenta bassi coefficienti fattoriali (es.: Y2-Y3).

7. 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 = mod1_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)
```

2



8. 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_label`.

```
mod2_fit = lavaan::cfa(model = mod1,data = datax,group = "group_label",
  group.equal=c("loadings"))

lavaan::anova(mod1_fit,mod2_fit)

Chi-Squared Difference Test

      Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
mod1_fit 40 3531.2 3684.5 42.055
mod2_fit 47 3527.5 3658.4 52.275      10.219      7    0.1765

mod3_fit = lavaan::cfa(model = mod1,data = datax,group = "group_label",
  group.equal=c("loadings","intercepts"))

lavaan::anova(mod2_fit,mod3_fit)

Chi-Squared Difference Test

      Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
mod2_fit 47 3527.5 3658.4 52.275
mod3_fit 54 3515.3 3623.9 54.121      1.8462      7    0.9679

mod4_fit = lavaan::cfa(model = mod1,data = datax,group = "group_label",
  group.equal=c("loadings","intercepts","residuals"))
```

```
lavaan::anova(mod3_fit,mod4_fit)
```

Chi-Squared Difference Test

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
mod3_fit	54	3515.3	3623.9	54.121			
mod4_fit	62	3510.0	3593.0	64.769	10.648	8	0.2225

Il modello scelto al punto 6 raggiunge il livello di invarianza esatta, dove vale anche il vincolo sulle varianze d'errore $\Theta_{\delta_{\text{slow}}} = \Theta_{\delta_{\text{fast}}}$ tra i due gruppi. Ciò indica che la struttura fattoriale complessiva del test bh90210 è la medesima sia per i partecipanti al gruppo *slow* sia per quelli del gruppo *fast*.

9. Si calcoli l'attendibilità del modello adattato scelto al punto 6 e si commenti il risultato ottenuto.

```
semTools::reliability(object = mod1_fit,what = "omega")
```

```
$slow
      eta
omega 0.8956415

$fast
      eta
omega 0.8521887
```

I risultati indicano che complessivamente il test bh90210 presenta una buona attendibilità nei due gruppi.

10. Si calcolino i *factor scores* del modello scelto al punto 6 e li si confrontino graficamente con i punteggi complessivi campionari calcolati invece al punto 4. Nota: per la predizione dei *factor scores* si usi il metodo di Bartlett (`lavaan::lavPredict(...,method="Bartlett")`).

```
Eta_pred = lavaan::lavPredict(object = mod1_fit,type = "lv",method = "Bartlett")
par(mfrow=c(1,2))
boxplot(cbind(Eta_pred$slow,h_slow),frame=FALSE,ylim=c(-3,3),main="slow")
boxplot(cbind(Eta_pred$fast,h_fast),frame=FALSE,ylim=c(-3,3),main="fast")
```

