

SEM group assignment (data analysis)

Xiaochang Zhao, Joyce Pang, Laura Springer

3/28/2022

Introduction

This is a R-markdown file that explicates the data analysis we did for our group assignment. This document includes part A (fitting the hypothesized model) and part B (testing for equal variances and means)

We are questioning measurement invariance for the need for uniqueness, perceived stress, and political cynicism between males and females. Females might have less strongly articulated opinions as they are expected to be more modest and less assertive than males. Thus, females might respond more conservatively on questionnaires while males might show more extreme answers. Consequently, this can lead to bigger variances for males than females, potentially biasing results if you do not take sex into account. Using 11 items, we will test if the measurement for the need for uniqueness, perceived stress, and political cynicism differs for males and females.

The questions we used to build our model:

Higher need for uniqueness (NU):

Q3: I think I am a special person

Q4: I like to be the center of attention

Q6: I like to look at myself in the mirror

Perceived stress (PS):

Q8: I have difficulty falling asleep

Q9: I often wake up in the middle of the night and cannot fall asleep again

Q13: I am worried about my current sleeping behaviour

Q14: My sleep interferes with my daily functioning e.g., daytime fatigue mood ability to function at work daily chores concentration memory mood etc

Political cynicism (PC):

Q17: Covid 19 regulations are a way for governments to gain more power

Q19: Politicians usually do not tell us about the true motives of their decisions

Q51: I believe that the government is hiding information about the COVID-19 pandemic from me

Q62: the governments' guidelines to prevent the spread of Covid-19 are appropriate

Instructions

please adapt to your folder path after downloading the csv file on OSF "<https://osf.io/3v2rt/>"

```
data <- read.csv("NA_2020_data.csv")
```

Below are some packages needed for running the analysis script

```
if (! require(tidyverse)) install.packages('tidyverse')
if (! require(semTools)) install.packages('semTools')
if (! require(semPlot)) install.packages('semPlot')
```

Part A

```
library(tidyverse)
library(semTools)
library(semPlot)
data <- read.csv("NA_2020_data.csv")
data1 <- data[!data$gender == "Other",]
## As we are using gender with two categories, we exclude those who answered "Other" on gender
data1$gender <- as.factor(data1$gender)
levels(data1$gender)
```

```
## [1] "Female" "Male"
```

```
Model1 <- '
  NU  =~ Q3 + Q4 + Q6
  PS  =~ Q8 + Q9 + Q13 + Q14
  PC  =~ Q17 + Q19 + Q51 + Q62
'
fit1 <- cfa(Model1, data1)
fit1
```

```
## lavaan 0.6-10 ended normally after 44 iterations
```

```
##
##      Estimator                      ML
##      Optimization method          NLMINB
##      Number of model parameters      25
##
##                                     Used      Total
##      Number of observations          480       491
##
## Model Test User Model:
##
##      Test statistic                  160.917
##      Degrees of freedom              41
##      P-value (Chi-square)            0.000
```

```
fitMeasures(fit1)
```

```
##              npar              fmin              chisq              df
##              25.000             0.168             160.917             41.000
##              pvalue      baseline.chisq      baseline.df      baseline.pvalue
##              0.000             1325.706             55.000             0.000
##              cfi              tli              nnfi              rfi
##              0.906             0.873             0.873             0.837
##              nfi              pnfi              ifi              rni
##              0.879             0.655             0.907             0.906
```

```
##          logl      unrestricted.logl          aic          bic
##      -9687.865      -9607.407      19425.730      19530.075
##          ntotal          bic2          rmsea      rmsea.ci.lower
##      480.000      19450.727      0.078      0.066
##      rmsea.ci.upper      rmsea.pvalue      rmr      rmr_nomean
##          0.091          0.000      0.198      0.198
##          srmr      srmr_bentler srmr_bentler_nomean      crmr
##          0.068          0.068          0.068          0.074
##      crmr_nomean      srmr_mplus      srmr_mplus_nomean      cn_05
##          0.074          0.068          0.068      170.854
##          cn_01          gfi          agfi          pgfi
##      194.740          0.942          0.907          0.585
##          mfi          ecvi
##          0.883          0.439
```

```
#rmsea = 0.078
#df = 41, chisq = 160.917, p.value <0.01. The exact fit for Model 1 is rejected.
#cfi, rni and tli are equal or lower than 0.9.
#This indicates that the fit of Model 1 is not good.
#Therefore, we check modification indices to add possible paths for a better fit.
```

```
#modification indices
modificationindices(fit1) %>% arrange(-mi) %>% head(10)
```

```
##      lhs op rhs      mi      epc sepc.lv sepc.all sepc.nox
## 1   Q8 ~~ Q9 62.555  0.880   0.880   0.415   0.415
## 2  Q13 ~~ Q14 50.664  1.601   1.601   1.568   1.568
## 3   PS =~ Q19 18.587  0.274   0.336   0.194   0.194
## 4   Q8 ~~ Q14 17.690 -0.592 -0.592 -0.334 -0.334
## 5   Q9 ~~ Q13 15.529 -0.461 -0.461 -0.378 -0.378
## 6   NU =~ Q62 15.270  0.384   0.348   0.215   0.215
## 7   PS =~ Q62 12.853  0.230   0.282   0.175   0.175
## 8   PS =~ Q6  11.379 -0.194 -0.238 -0.154 -0.154
## 9   Q3 ~~ Q62 10.864  0.316   0.316   0.168   0.168
## 10  PC =~ Q9   7.249 -0.225 -0.222 -0.128 -0.128
```

```
#Modification indices imply that Q8 and Q9 have shared variance.
#Q8 ~~ Q9 is added in Model 2 to see whether Model 2 is significantly better than Model 1
```

```
Model2 <- '
  NU  =~ Q3 + Q4 + Q6
  PS  =~ Q8 + Q9 + Q13 + Q14
  PC  =~ Q17 + Q19 + Q51 + Q62
  Q8  ~~ Q9
'
fit2 <- cfa(Model2, data1)

anova(fit1,fit2)
```

```
## Chi-Squared Difference Test
##
##      Df    AIC    BIC   Chisq Chisq diff Df diff Pr(>Chisq)
## fit2 40 19366 19475   99.162
```

```
## fit1 41 19426 19530 160.917      61.755      1 3.889e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fitMeasures(fit2)
```

```
##          npar          fmin          chisq          df
##          26.000          0.103          99.162          40.000
##          pvalue      baseline.chisq      baseline.df      baseline.pvalue
##          0.000          1325.706          55.000          0.000
##          cfi          tli          nnfi          rfi
##          0.953          0.936          0.936          0.897
##          nfi          pnfi          ifi          rni
##          0.925          0.673          0.954          0.953
##          logl      unrestricted.logl          aic          bic
##          -9656.987          -9607.407          19365.975          19474.493
##          ntotal          bic2          rmsea      rmsea.ci.lower
##          480.000          19391.972          0.056          0.042
##          rmsea.ci.upper      rmsea.pvalue          rmr          rmr_nomean
##          0.069          0.240          0.176          0.176
##          srmr          srmr_bentler srmr_bentler_nomean          crmr
##          0.062          0.062          0.062          0.068
##          crmr_nomean          srmr_mplus srmr_mplus_nomean          cn_05
##          0.068          0.062          0.062          270.904
##          cn_01          gfi          agfi          pgfi
##          309.300          0.965          0.942          0.585
##          mfi          ecvi
##          0.940          0.315
```

```
#fit 2 has a lower AIC and BIC and a significant the chi-square value.
#Therefore, Model 2 is significantly better than Model 1.
#Furthermore, fit indices such as RMSEA, CFI, RNI indicate a good fit for Model 2.
#Chi-square = 99.16, DF = 40, p < 0.001. The exact fit of Model 2 is rejected.
#However, the test for exact fit may be too sensitive to minor misspecifications with large sample size
#Overall, Model 2 fits the dataset better than Model 1.
```

```
modificationindices(fit1) %>% arrange(-mi) %>% head(10)
```

```
##    lhs op rhs      mi      epc sepc.lv sepc.all sepc.nox
## 1  Q8 ~~ Q9 62.555  0.880   0.880   0.415   0.415
## 2 Q13 ~~ Q14 50.664  1.601   1.601   1.568   1.568
## 3  PS =~ Q19 18.587  0.274   0.336   0.194   0.194
## 4  Q8 ~~ Q14 17.690 -0.592 -0.592 -0.334 -0.334
## 5  Q9 ~~ Q13 15.529 -0.461 -0.461 -0.378 -0.378
## 6  NU =~ Q62 15.270  0.384   0.348   0.215   0.215
## 7  PS =~ Q62 12.853  0.230   0.282   0.175   0.175
## 8  PS =~ Q6  11.379 -0.194 -0.238 -0.154 -0.154
## 9  Q3 ~~ Q62 10.864  0.316   0.316   0.168   0.168
## 10 PC =~ Q9   7.249 -0.225 -0.222 -0.128 -0.128
```

```
#Modification indices imply that Q29 and Q33 have shared variance.
#We also go on to add Q13 ~~ Q14 in Model 3 and test whether Model 3 is significantly better than Model
```

```
Model3 <- '
  NU =~ Q3 + Q4 + Q6
  PS =~ Q8 + Q9 + Q13 + Q14
  PC =~ Q17 + Q19 + Q51 + Q62
  Q8 ~~ Q9
  Q13 ~~ Q14
'
fit3 <- cfa(Model3, data1)
anova(fit1,fit2,fit3)
```

```
## Chi-Squared Difference Test
##
##      Df   AIC   BIC   Chisq Chisq diff Df diff Pr(>Chisq)
## fit3 39 19365 19478  96.383
## fit2 40 19366 19475  99.162      2.778      1   0.09556 .
## fit1 41 19426 19530 160.917     61.755      1  3.889e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

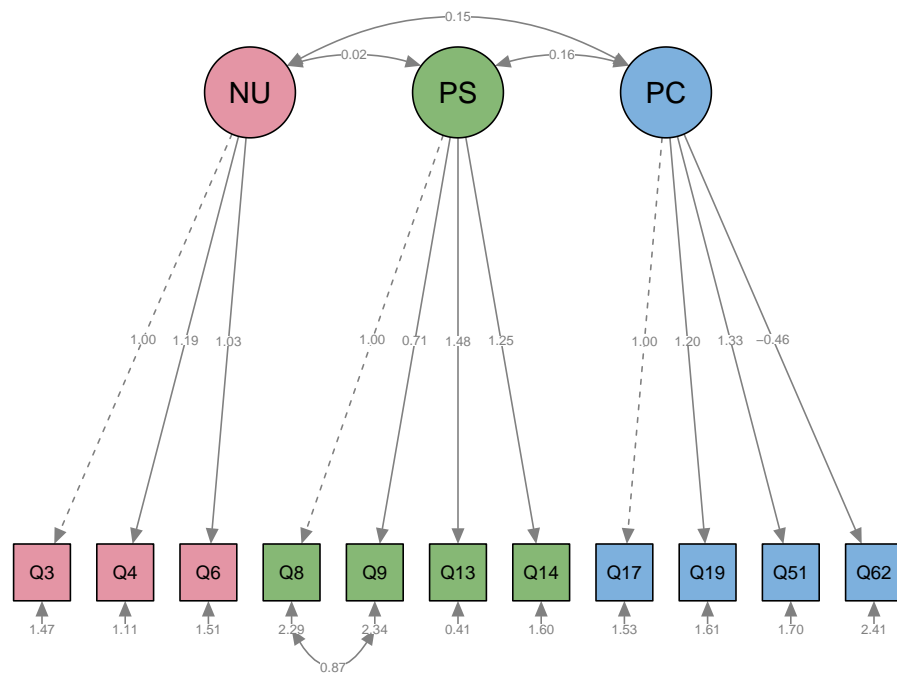
```
fitMeasures(fit3)
```

```
##          npar          fmin          chisq          df
##          27.000          0.100          96.384         39.000
##          pvalue      baseline.chisq      baseline.df      baseline.pvalue
##          0.000          1325.706          55.000          0.000
##          cfi          tli          nnfi          rfi
##          0.955          0.936          0.936          0.897
##          nfi          pnfi          ifi          rni
##          0.927          0.658          0.955          0.955
##          logl      unrestricted.logl          aic          bic
##          -9655.598          -9607.407          19365.197          19477.889
##          ntotal          bic2          rmsea      rmsea.ci.lower
##          480.000          19392.194          0.055          0.042
##          rmsea.ci.upper      rmsea.pvalue          rmr          rmr_nomean
##          0.069          0.248          0.170          0.170
##          srmr          srmr_bentler      srmr_bentler_nomean          crmr
##          0.060          0.060          0.060          0.065
##          crmr_nomean          srmr_mplus      srmr_mplus_nomean          cn_05
##          0.065          0.060          0.060          272.775
##          cn_01          gfi          agfi          pgfi
##          311.899          0.966          0.942          0.571
##          mfi          ecvi
##          0.942          0.313
```

*#p-value = 0.0956, indicating that Model 3 is not significantly better than Model 2
 #AIC is smaller by one unit (not meaningful) and the BIC is higher.
 #Model 2 fits the data better than Model 3.*

*#Conclusion
 #Model 2 is better than Model 1 and Model 3.
 #We will use model 2 for the analysis.*

```
#Plot for Model 2:
semPaths(fit2, whatLabels="est",
  sizeMan = 5,
  groups = "latent",
  layout = "tree",
  edge.label.cex = .5,
  style = "lisrel",
  pastel = TRUE)
```



Part B

Part B includes data analysis needed for testing equal variances and means for the two groups.

We first test for measurement invariance

```
conf <- cfa(Model2, data1, group = "gender", std.lv=TRUE)
fitMeasures(conf, c("chisq", "pvalue", "rmsea", "cfi", "rni", "ifi"))
```

```
##   chisq  pvalue  rmsea   cfi    rni    ifi
## 155.020   0.000   0.063   0.942  0.942  0.943
```

*#The configural invariance model has a significant p-value which rejects the exact fit.
 #But other incremental fit indices indicate a good fit (rmsea = 0.0625 and cfi, rni, ifi are close to 0
 #So we will go on and test for other models.*

```
weak <- cfa(Model2, data1, group = "gender",
            group.equal="loadings", std.lv=TRUE)
anova(conf, weak)
```

```
## Chi-Squared Difference Test
##
##      Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## conf  80 19414 19722 155.02
## weak  88 19400 19676 157.84      2.8233      8      0.9449
```

*#The weak invariance model has a lower AIC and BIC and a non-significant p-value when compared to the c
 #So, weak invariance is accepted.*

```
strong <- cfa(Model2, data1, group="gender",
              group.equal = c("loadings", "intercepts"), std.lv=TRUE)
anova(conf, weak, strong)
```

```
## Chi-Squared Difference Test
##
##      Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## conf  80 19414 19722 155.02
## weak  88 19400 19676 157.84      2.8233      8      0.9449
## strong 96 19392 19634 165.84      8.0004      8      0.4334
```

*#The strong invariance model has a lower AIC and BIC and a non-significant p-value when compared to the
 #So, strong invariance is accepted.*

```
strict <- cfa(Model2, data1, group="gender",
              group.equal= c("loadings", "intercepts", "residuals",
                           "residual.covariances"))
anova(strong,strict)
```

```
## Chi-Squared Difference Test
##
##      Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## strong  96 19392 19634 165.84
## strict 108 19393 19585 190.91      25.071      12      0.01449 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

*#The strict invariance model has a smaller BIC but it also has a similar AIC and a significant p-value.
 #Though it can be debatable, strict invariance is not accepted.*

```
table <- anova(conf,weak,strong,strict)
knitr::kable(table[,1:4])
```

	Df	AIC	BIC	Chisq
conf	80	19413.53	19722.39	155.0201
weak	88	19400.36	19675.83	157.8434
strong	96	19392.36	19634.44	165.8438
strict	108	19393.43	19585.42	190.9149

#The above table shows all chi-square, df, AIC and BIC for all models.

We will test for equal variance and means with the strong model

```
eqvars <- cfa(Model2, data1, group="gender",
              group.equal = c("loadings", "intercepts", "lv.variances"))

anova(strong, eqvars)
```

```
## Chi-Squared Difference Test
##
##           Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## strong  96 19392 19634 165.84
## eqvars  99 19393 19622 172.32      6.4792      3    0.09049 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

*#The model assuming equal variance is non-significant.
 #It also has a similar AIC but lower BIC.
 #So, this model is accepted.
 #The assumption of equal variances between groups holds
 #We do not reject the hypothesis that the variances of both groups are equal.*

```
eqmeans <- cfa(Model2, data1, group="gender",
               group.equal = c("loadings", "intercepts", "lv.variances",
                              "means"))

anova(eqvars, eqmeans)
```

```
## Chi-Squared Difference Test
##
##           Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## eqvars   99 19393 19622 172.32
## eqmeans 102 19391 19608 176.05      3.7287      3    0.2923
```

*#The model assuming equal variance and equal means is non-significant.
 #It also has a lower AIC and BIC compared to the equal variance model.
 #So, the model of equal variance and means is accepted.
 #The assumption of equal means between groups holds.
 #We do not reject the hypothesis that the means of both groups are equal.
 #In our case, this means that we do not have reason to believe that*

*#the questions regarding need for uniqueness, perceived stress and political
#cynicism are answered differently by male and female.
#Hence, we can assume measurement invariance.*