Lab10.2 Factor Analysis - Higher Order Factors

Adam Garber

Factor Analysis ED 216B - Instructor: Karen Nylund-Gibson

March 10, 2020

Contents

1 Getting started: Rprojects, Rmarkdown, Git-Github

- R-studio, Projects, Scripts: Go here
- Rmarkdown basics tutorial: Go here here
- Connect Git-Hithub with R-studio and download Repositories: Go here

---

2 Steps to download repositories from Github and create a version controlled R-project

0. Create a Github account and connect R-Studio with Git
1. Go to the Lab10 repository link to **Fork** and **Clone** (copy address) the repository:
2. Within R-studio create a **New project** and choose the **Version Control** Option (Git)
3. Paste the repository address copied (cloned) from Github and save locally on your computer
4. After making changes in your branch of the repository to update the version on Github follow the following sequence of steps: `Stage`, `Commit` (add commit message), `Pull`, and then `Push`

---

3 Outline

- Prepare data
- EFA (model 0)
- CFA (models 1-3)
- Invariance (models 4-7)
- Generate summary table for model comparison

DATA SOURCE: This lab exercise utilizes a subset of the HSLS public-use dataset: High School Longitudinal Study of 2009 (Ingels et al., 2011) See website: nces.ed.gov

---

4 BEGIN: Higher-Order Factors

```r
# load packages
library(MplusAutomation)
library(rhdf5)
library(tidyverse)
library(here)
library(kableExtra)
library(semPlot)
library(gt)
library(DiagrammeR)
```

---

5 Prepare data

5.1 Read in data

```
data_raw <- read_csv(here("data", "hsls_fa_data_subset.csv"))
```

5.2 Reverse code for factor interpretation

```
hsls_data <- data_raw

cols = c("S1MPERS1", "S1MPERS2", "S1MUSELI", "S1MUSECL", "S1MUSEJO",
         "S1MTESTS", "S1MTEXTB", "S1MSKILL", "S1MASSEX", "S1MENJNG",
         "S1SPERS1", "S1SPERS2", "S1SUSELI", "S1SUSECL", "S1SUSEJO",
         "S1STESTS", "S1STEXTB", "S1SSKILL", "S1SASSEX", "S1SENJNG")

hsls_data[ ,cols] <-  5 - hsls_data[ ,cols]
```

---

6 Run a baseline CFA model with 4 factors (for comparison)

```
m.cfa0  <- mplusObject(
  TITLE = "Higher Order FA Models - HSLS SCIENCE",
  VARIABLE =
 "usevar =
 S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
 S1SUSEJO S1STESTS S1STEXTB S1SSKILL
 S1SASSEX S1SENJNG S1SWASTE S1SBORIN;",

  ANALYSIS =
 "estimator=mlr; ",

     MODEL =
 "SCI_ID BY S1SPERS1* S1SPERS2;
  SCI_ID@1;

  SCI_UT BY S1SUSELI* S1SUSECL S1SUSEJO;
  SCI_UT@1;

  SCI_EFF BY S1STESTS* S1STEXTB S1SSKILL S1SASSEX ;
  SCI_EFF@1;
```

```
  SCI_INT BY S1SENJNG* S1SWASTE S1SBORIN;
  SCI_INT@1; ",

  PLOT = "type = plot3;",
  OUTPUT = "sampstat standardized residual modindices (3.84);",

  usevariables = colnames(hsls_data),
  rdata = hsls_data)

m.cfa0.fit <- mplusModeler(m.cfa,
                            dataout=here("2nd_order_FA", "cfa_baseline.dat"),
                            modelout=here("2nd_order_FA", "cfa_baseline.inp"),
                            check=TRUE, run = TRUE, hashfilename = FALSE)
```

---

7   Make a higher-order model path diagram using package {DiagrammeR}

```
grViz(" digraph higher_order_path_diagram {

graph [overlap = true, fontsize = 10,   # this is the 'graph' statement
       fontname = Times,
       label= '']

  node [shape = box]            # this is the 'node' statement

  ID1; ID2; UT1; UT2; UT3; UT1; UT2;
  UT3; EFF1; EFF2; EFF3; EFF4; INT1; INT2; INT3;

  node [shape = circle, fixedsize = true,
        width = 1.5, label = 'Science Identity']

  F1;

  node [shape = circle, fixedsize = true,
        width = 1.5, label = 'Science Utility']

  F2;

  node [shape = circle, fixedsize = true,
        width = 1.5, label = 'Science Efficacy']

  F3;

  node [shape = circle, fixedsize = true,
        width = 1.5, label = 'Science Interest']
```

```
  F4;

  node [shape = circle, fixedsize = true,
        width = 1.5, label = 'Adaptive Science Orientation']

  G1;

  edge [color = black]    # this is the 'edge' statement

  F1->{ID1 ID2}
  F2->{UT1 UT2 UT3}
  F3->{EFF1 EFF2 EFF3 EFF4}
  F4->{INT1 INT2 INT3}
  G1->{F1 F2 F3 F4}

}")
```
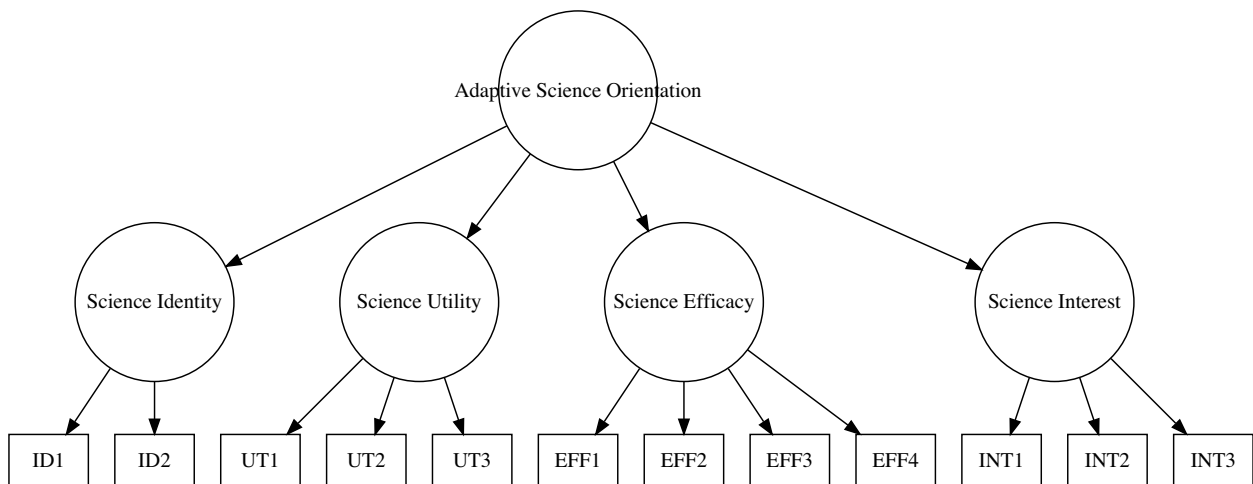

8 Run a higher-order model model with 4 sub-factors

```
m.cfa1 <- mplusObject(
  TITLE = "Higher Order FA Models - HSLS SCIENCE",
  VARIABLE =
"usevar =
 S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
 S1SUSEJO S1STESTS S1STEXTB S1SSKILL
 S1SASSEX S1SENJNG S1SWASTE S1SBORIN;",

  ANALYSIS =
"estimator=mlr; ",

    MODEL =
```

```
 "SCI_ID BY S1SPERS1* S1SPERS2;
  SCI_ID@1;

  SCI_UT BY S1SUSELI* S1SUSECL S1SUSEJO;
  SCI_UT@1;

  SCI_EFF BY S1STESTS* S1STEXTB S1SSKILL S1SASSEX ;
  SCI_EFF@1;

  SCI_INT BY S1SENJNG* S1SWASTE S1SBORIN;
  SCI_INT@1;

  ! Regress the higher-order factor on the 4 sub-factors
  F2NDORDR BY SCI_ID SCI_UT SCI_EFF SCI_INT" ,

  PLOT = "type = plot3;",
  OUTPUT = "sampstat standardized residual modindices (3.84);",

  usevariables = colnames(hsls_data),
  rdata = hsls_data)

m.cfa1.fit <- mplusModeler(m.cfa1,
                           dataout=here("2nd_order_FA", "cfa_baseline.dat"),
                           modelout=here("2nd_order_FA", "cfa_2nd_order.inp"),
                           check=TRUE, run = TRUE, hashfilename = FALSE)
```

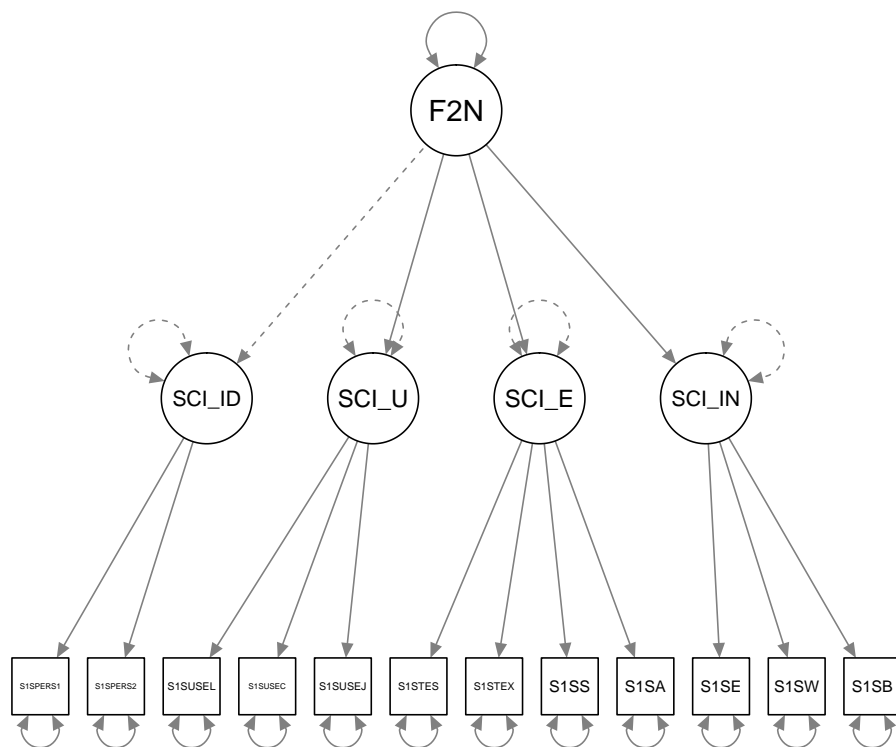# 9 Generate a higher-order model path diagram from Mplus Output with {semPlot}

```
order2_model <- readModels(here("2nd_order_FA",
                                "cfa_2nd_order.out"))
```

```
## Reading model:  /Users/agarber/Desktop/FA_W20/Lab10_FA/2nd_order_FA/cfa_2nd_order.out
```

```
# plot model:
semPaths(order2_model,
         intercepts=FALSE)
```

10   Compare model fit of baseline and higher-order models

10.1   Read into R summary of all models

```
models_2 <- readModels(here("2nd_order_FA"))
```

```
## Reading model:  /Users/agarber/Desktop/FA_W20/Lab10_FA/2nd_order_FA/cfa_2nd_order.out
## Reading model:  /Users/agarber/Desktop/FA_W20/Lab10_FA/2nd_order_FA/cfa_baseline.out
```

10.2   Extract relevant data and generate table

```
order2_table <- LatexSummaryTable(models_2,
  keepCols=c("Filename", "Parameters",
            "ChiSqM_Value", "CFI","TLI",
            "SRMR", "RMSEA_Estimate",
            "RMSEA_90CI_LB", "RMSEA_90CI_UB"),
  sortBy = "Filename")
```

```
order2_table %>%
  kable(booktabs = T, linesep = "",
        col.names = c(
          "Model", "Par", "ChiSq",
          "CFI", "TLI", "SRMR", "RMSEA",
          "Lower CI", "Upper CI")) %>%
  kable_styling(c("striped"),
  full_width = F, position = "left")
```

Model	Par	ChiSq	CFI	TLI	SRMR	RMSEA	Lower CI	Upper CI
cfa__2nd__order.out	40	390.662	0.966	0.956	0.038	0.05	0.045	0.054
cfa__baseline.out	40	390.662	0.966	0.956	0.038	0.05	0.045	0.054

11 References

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. Structural equation modeling: a multidisciplinary journal, 25(4), 621-638.

Horst, A. (2020). Course & Workshop Materials. GitHub Repositories, https://https://allisonhorst.github.io/

Ingels, S. J., Pratt, D. J., Herget, D. R., Burns, L. J., Dever, J. A., Ottem, R., . . . & Leinwand, S. (2011). High School Longitudinal Study of 2009 (HSLS: 09): Base-Year Data File Documentation. NCES 2011-328. National Center for Education Statistics.

Muthén, L.K. and Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

UC **SANTA BARBARA**