Confirmatory Factor Analysis (CFA)

A Course in `MplusAutomation`

Adam Garber

_____

Load packages

```
library(MplusAutomation) # A conduit between R & Mplus
library(tidyverse);library(here);library(gt) # For tidiness & tables.
library(semPlot);library(DiagrammeR) # For creating path diagrams.
```

_____

Load data example

Data source. This tutorial utilizes the NCES public-use data called the Education Longitudinal Study of 2002 (Lauff & Ingels, 2014) This data can be found on the [NCES website](). Note that all examples used are for purposes of illustration only and are not intended to be interpreted substantively.

```
cfa_data <- read_csv("https://garberadamc.github.io/project-site/data/els_sub5_data.csv")
```

_____

Prepare `data.frame` for analysis (select & reorder columns)

```
school_trouble <- cfa_data %>%
  select(stolen, t_hurt, p_fight, hit, damaged, bullied,  # factor 1 (indicators)
         safe, disrupt, gangs, rac_fght)                  # factor 2 (indicators)
```

Look at variables for CFA example

Applied Example: School Trouble

Name	Variable Description
Factor 1	
stolen	Had something stolen at school

t_hurt	Someone threatened to hurt 10th grader at school
p_fight	Got into a physical fight at school
hit	Someone hit 10th grader
damaged	Someone damaged belongings
bullied	Someone bullied or picked on 10th grader
Factor 2	
safe	Does not feel safe at this school
disrupt	Disruptions get in way of learning
gangs	There are gangs in school
rac_fght	Racial groups often fight

**Make a simple CFA path diagram using package {`DiagrammeR`}**

```
grViz(" digraph cfa_model {

graph [layout = dot, overlap = true]

node [shape=box]
stolen t_hurt p_fight hit damaged bullied safe disrupt gangs rac_fght;

node [shape=circle, width = 0.9]
factor1 factor2;

edge []
factor1 -> {stolen t_hurt p_fight hit damaged bullied}
factor2 -> {safe disrupt gangs rac_fght}
factor1 -> factor2 [dir = both, color = red]
}")
```
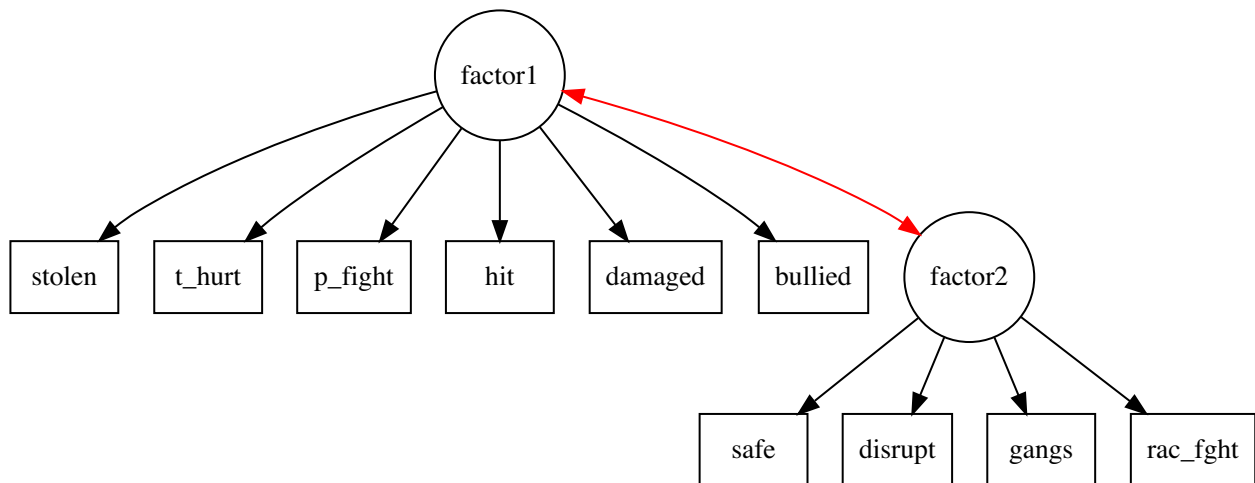
Estimate CFA model 1 with default Unit Loading Identification (ULI)

```r
cfa_m1  <- mplusObject(

  TITLE = "CFA - ULI",

  VARIABLE =
    "usevar = stolen-rac_fght;",

  ANALYSIS =
    "estimator = mlr;",

  MODEL =
    "factor1 by stolen t_hurt p_fight hit damaged bullied;

     factor2 by safe disrupt gangs rac_fght;" ,

  PLOT =
    "type = plot3;",

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

  usevariables = colnames(school_trouble),
  rdata = school_trouble)

cfa_m1_fit <- mplusModeler(cfa_m1,
              dataout = here("05-cfa","cfa_mplus","cfa_data.dat"),
              modelout = here("05-cfa","cfa_mplus","cfa_m1_uli.inp"),
              check=TRUE, run = TRUE, hashfilename = FALSE)
```

Create a path diagram with `semPlot` package

```r
semPaths(cfa_m1_fit$results)
```
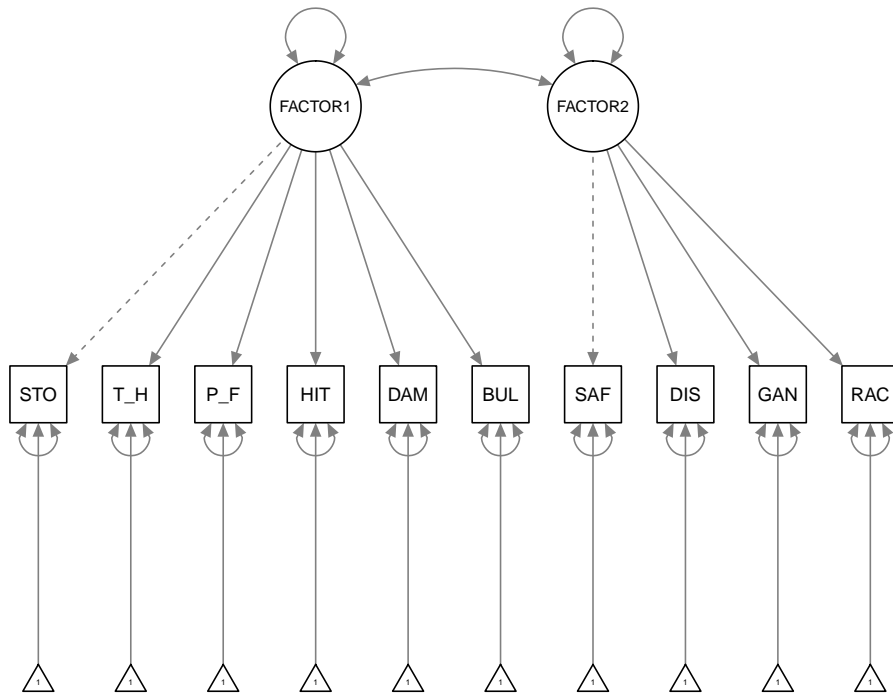
Create table summarizing model fit

```r
model_fit <- LatexSummaryTable(cfa_m1_fit,
  keepCols=c("Title", "Parameters", "LL",
            "ChiSqM_Value","ChiSqM_DF", "ChiSqM_PValue",
            "RMSEA_Estimate","RMSEA_90CI_LB","RMSEA_90CI_UB",
            "CFI", "TLI","SRMR")) %>%
  mutate(RMSEA_90CI_LB = str_pad(RMSEA_90CI_LB, width=6, side="left", pad="("),
         RMSEA_90CI_UB = str_pad(RMSEA_90CI_UB, width=6, side="right", pad=")")) %>%
  unite(CI, RMSEA_90CI_LB:RMSEA_90CI_UB, sep=", ", remove = TRUE) %>%
  unite(RMSEA, RMSEA_Estimate:CI, sep=" ", remove = TRUE)

model_fit %>%
  gt() %>%
  tab_header(
    title = md("**Table 1**"),
    subtitle = md("*Summary of Model Fit Statistics*")) %>%
  cols_label(
    Title = "Model",
    Parameters = md("Par"),
    LL = md("*LL*"),
    ChiSqM_Value = md("Chi^2"),
    ChiSqM_PValue = md("*p-value*"),
    ChiSqM_DF = md("*df*"),
    RMSEA = "RMSEA (90% CI)" ) %>%
  tab_options(column_labels.font.weight = "bold") %>%
  fmt(c(6), fns = function(x) ifelse(x<0.001,"<.001", scales::number(x, accuracy = 0.01)))
```

Table 1

Summary of Model Fit Statistics

Model	Par	LL	Chi^2	df	p-value	RMSEA (90% CI)	CFI	TLI	SRMR
CFA - ULI	31	-6197.667	121.46	34	<.001	0.06 (0.048, 0.071)	0.898	0.865	0.043

Create table of CFA loading & correltation estimates

```
loadings_stdyx <- cfa_m1_fit$results$parameters$stdyx.standardized %>%
  filter(grepl("FACTOR",paramHeader)) %>%
  select(param,est,se) %>%
  mutate(param = str_replace(param, "FACTOR1", "F1 with F2"),
         param = str_to_title(param))

loadings_stdyx %>%
  gt() %>%
  tab_header(
    title = md("**Table 2**"),
    subtitle = md("*Standardized Factor Loading and Factor Correlation Estimates*")) %>%
  cols_label(param = "Item", est = "Loading", se = "SE") %>%
  tab_row_group(group = "Factor 1", rows = 1:6) %>%
  tab_row_group(group = "Factor 2", rows = 7:10) %>%
  tab_row_group(group = "Factor Correlation", rows = 11) %>%
  row_group_order(groups = c("Factor 1","Factor 2","Factor Correlation")) %>%
  tab_options(column_labels.font.weight = "bold")
```

Table 2

Standardized Factor Loading and Factor Correlation Estimates

Item	Loading	SE
Factor 1		
Stolen	0.492	0.042
T_hurt	0.665	0.045
P_fight	0.510	0.059
Hit	0.677	0.043
Damaged	0.634	0.046
Bullied	0.481	0.054
Factor 2		
Safe	0.553	0.052
Disrupt	0.370	0.049
Gangs	0.655	0.045
Rac_fght	0.639	0.042
Factor Correlation		
F1 With F2	-0.350	0.061

Estimate CFA model (UVI; Unit Variance Identification)

```
cfa_m2  <- mplusObject(

  TITLE = "CFA - Unit Variance Identification",

  VARIABLE =
    "usevar = stolen-rac_fght;",

  ANALYSIS =
    "estimator = mlr;",

  MODEL =
    "!!! estimate first variable loading & fix variance at 1 !!!

     factor1 by stolen* t_hurt p_fight hit damaged bullied;
     factor1@1;

     factor2 BY safe* disrupt gangs rac_fght;
     factor2@1;",

  PLOT =
    "type = plot3;",

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

  usevariables = colnames(school_trouble),
  rdata = school_trouble)

cfa_m2_fit <- mplusModeler(cfa_m1,
              dataout = here("05-cfa","cfa_mplus","cfa_data.dat"),
              modelout = here("05-cfa","cfa_mplus","cfa_m2_uvi.inp"),
              check=TRUE, run = TRUE, hashfilename = FALSE)
```
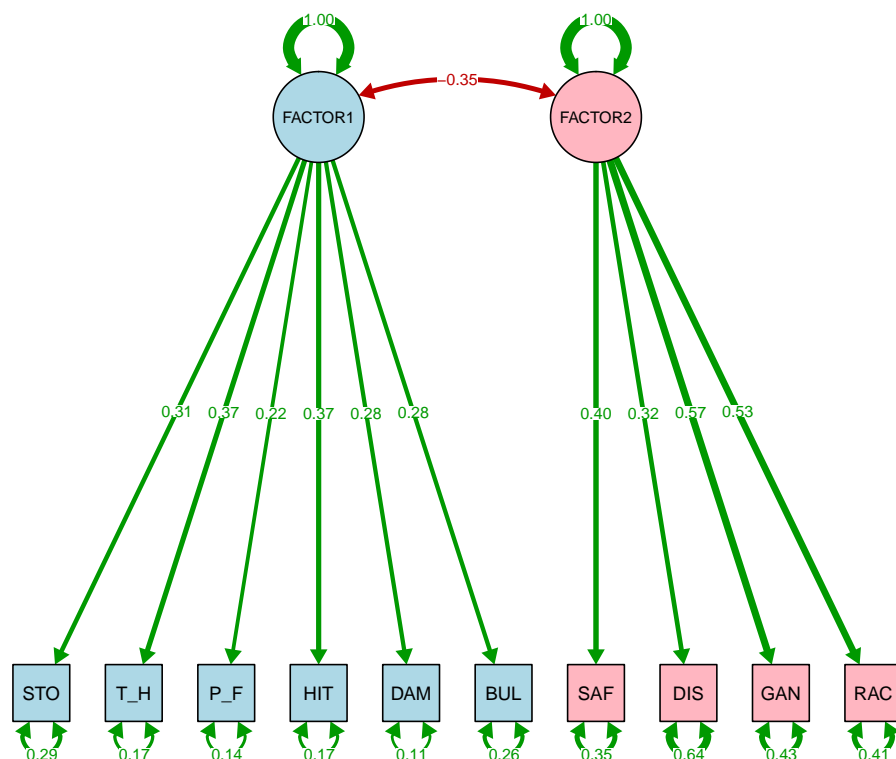
Create a path diagram including parameter estimates

```
semPaths(cfa_m2_fit$results, "std", intercepts=FALSE,
        fixedStyle = c(1),
        fade = FALSE,
        color= list(lat = c("light blue"," light pink")),
)
```

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.

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

_____

Further resources & examples here:

https://garberadamc.github.io/project-site/

https://www.adam-garber.com/

_____