

Multiple Indicator, Multiple Cause Models (MIMIC)

A Course in `MplusAutomation`

Adam Garber

Outline

- a. Prepare, wrangle, and explore data
- b. Run an unconditional CFA baseline model
- c. Specify a MIMIC model with a single binary covariate
- d. Specify a MIMIC model and probe for DIF (Differential Item Functioning)
- e. Specify a MIMIC model with a DIF parameter
- f. Specify a MIMIC model with two binary covariates & an interaction
- g. Specify a MIMIC model with three continuous covariates
- h. Experiment with path diagram notation & formatting

loading packages:

```
library(tidyverse)
library(here)
library(semPlot)
library(DiagrammeR)
library(MplusAutomation)
library(stargazer)
library(gtsummary)
library(gt)
library(glue)
```

DATA SOURCE: This exercise utilizes the NCES public-use dataset: Education Longitudinal Study of 2002 (Lauff & Ingels, 2014) [See website: nces.ed.gov](https://nces.ed.gov/ipeds/data/elsa/)

Begin

Read in data

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

Prepare dataframe for analysis

```
mimic_data <- els_subset %>%
  select(bystlang, freelnch, byincome,          # covariates
         stolen, t_hurt, p_fight, hit, damaged, bullied, # factor 1 (indicators)
         safe, disrupt, gangs, rac_fght,         # factor 2 (indicators)
         late, skipped, mth_read, mth_test, rd_test) %>%
  mutate(
    freelnch = case_when( # Grade 10, percent free lunch - transform to binary
      freelnch < 5 ~ 0,   # < 50%
      freelnch >= 5 ~ 1)) # > 50%
```

Take a look at the variables used for this example (same indicators used for 05-cfa lesson)

```
stargazer(as.data.frame(mimic_data), type="text", digits=1)
```

```
##
## =====
## Statistic  N  Mean St. Dev. Min  Pctl(25) Pctl(75) Max
## -----
## bystlang   749 0.8    0.4    0.0    1.0    1.0    1.0
## freelnch   726 0.3    0.4    0.0    0.0    1.0    1.0
## byincome   794 9.2    2.5    1      8     11    13
## stolen     718 1.5    0.6    1.0    1.0    2.0    3.0
## t_hurt     714 1.3    0.6    1.0    1.0    1.0    3.0
## p_fight    715 1.2    0.4    1.0    1.0    1.0    3.0
## hit        711 1.3    0.6    1.0    1.0    1.0    3.0
## damaged    716 1.2    0.4    1.0    1.0    1.0    3.0
## bullied    713 1.3    0.6    1.0    1.0    1.0    3.0
## safe       711 3.3    0.7    1.0    3.0    4.0    4.0
## disrupt    713 2.6    0.9    1.0    2.0    3.0    4.0
## gangs      705 3.0    0.9    1.0    3.0    4.0    4.0
## rac_fght   716 3.0    0.8    1.0    3.0    4.0    4.0
## late       719 2.3    1.2    1.0    1.0    3.0    5.0
## skipped    715 1.5    0.9    1.0    1.0    2.0    5.0
## mth_read   785 50.9   10.0   25.1   44.3   57.6   79.9
## mth_test   785 51.1   10.1   24.4   44.6   57.5   84.8
## rd_test    785 50.6   10.0   22.6   43.4   57.2   78.8
## -----
```

Look for pairwise differences across student language groups (`bystlang`) using the `{gtsummary}` package

```
table_data <- mimic_data %>%
  dplyr::select(bystlang, safe, disrupt, gangs, rac_fght)

table2 <- tbl_summary(table_data,
  by = bystlang,      # split table by group `bystlang` (student language)
  missing = "no"      # don't list missing data separately
) %>%
  add_n() %>%         # add column with total number of non-missing observations
  add_p() %>%         # test if there's difference between groups
  bold_labels()
```

table2

Characteristic	N	0, N = 117	1, N = 632	p-value
safe	711			0.2
1		3 (2.7%)	16 (2.7%)	
2		10 (8.9%)	44 (7.3%)	
3		64 (57%)	285 (48%)	
4		35 (31%)	254 (42%)	
disrupt	713			0.2
1		17 (15%)	68 (11%)	
2		44 (39%)	190 (32%)	
3		38 (33%)	265 (44%)	
4		15 (13%)	76 (13%)	
gangs	705			0.013
1		11 (9.8%)	36 (6.1%)	
2		29 (26%)	98 (17%)	
3		48 (43%)	263 (44%)	
4		24 (21%)	196 (33%)	
rac_fght	716			0.3
1		7 (6.2%)	30 (5.0%)	
2		23 (20%)	98 (16%)	
3		55 (49%)	277 (46%)	
4		28 (25%)	198 (33%)	

Estimate the Unconditional Confirmatory Factor Analysis (CFA) model

```
cfa_m0 <- mplusObject(

  TITLE =
    "CFA model0",

  VARIABLE =
    "usevar = stolen-rac_fght;",

  ANALYSIS =
    "estimator = mlr;",

  MODEL =
    "FACTOR_1 by stolen t_hurt p_fight hit damaged bullied;

    FACTOR_2 by safe disrupt gangs rac_fght;" ,

  PLOT =
    "type = plot3;",

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

  usevariables = colnames(mimic_data),
  rdata = mimic_data)

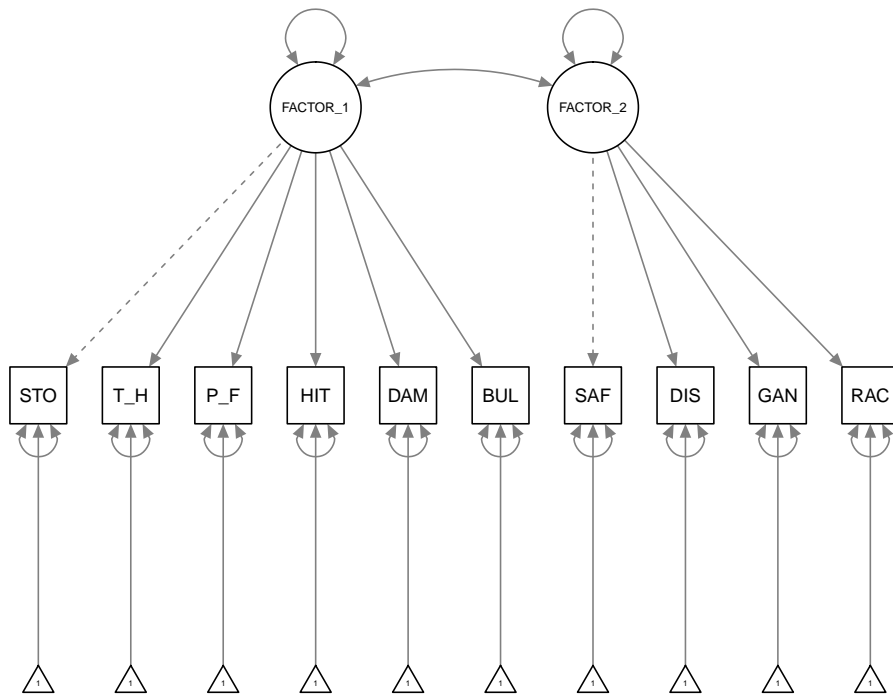
cfa_m0_fit <- mplusModeler(cfa_m0,
  dataout=here("06-MIMIC", "mimic_mplus", "mimic_data.dat"),
  modelout=here("06-MIMIC", "mimic_mplus", "cfa_model0.inp"),
  check=TRUE, run = TRUE, hashfilename = FALSE)
```

Read the model `cfa_model0.out` into R within the `mimic_mplus` sub-folder

```
mimic_output1 <- readModels(here("06-MIMIC", "mimic_mplus", "cfa_model0.out"), quiet = TRUE)
```

Create a path diagram from the model object

```
semPaths(mimic_output1)
```



MIMIC model 1 - single bivariate covariate

Number of parameters for the MIMIC model 1 = 33

- 8 item loadings (10 items; 2 fixed loadings)
- 10 intercepts
- 10 residual variances
- 2 factor variances
- 1 factor co-variance
- 1 covariate mean
- 1 covariate variance

```
grViz(" digraph mimic_path_diagram {

graph [overlap = true, fontsize = 10,    # this is the 'graph' statement
      fontname = Times,
      label=
        'Figure 1: MIMIC model with single covariate.']

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

```

A; B; C; D; E;

node [shape = box,
      label = 'Covariate']

X;

node [shape = circle, fixedsize = true,
      width = 0.9, label = 'Factor 1']

F;

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

F->{A B C D E}
X->F

}"))

```

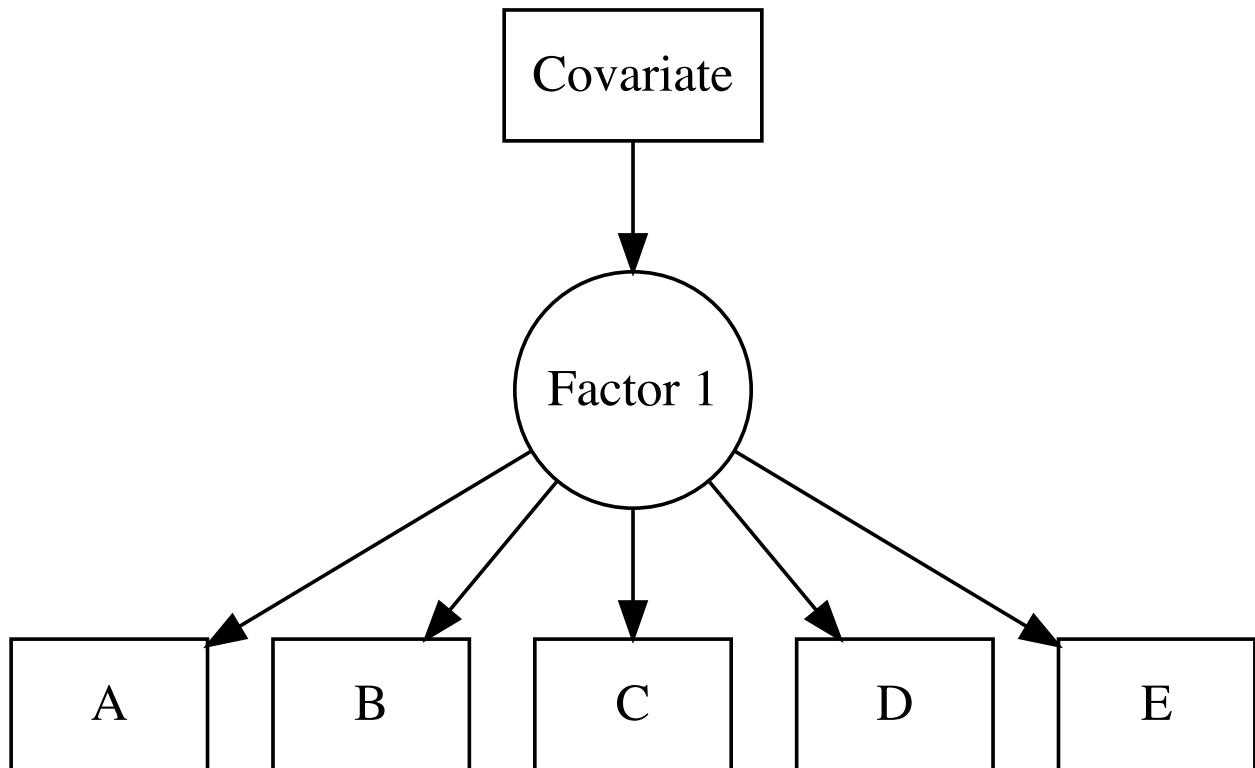


Figure 1: MIMIC model with single covariate.

```

mimic_m1 <- mplusObject(

```

```

TITLE = "MIMIC model1 ",

VARIABLE =
  "usevar = freelnch stolen-rac_fght;",

ANALYSIS =
  "estimator = mlr;",

MODEL =
  "FACTOR_1 by stolen t_hurt p_fight hit damaged bullied;

  FACTOR_2 by safe disrupt gangs rac_fght;

  FACTOR_1 on freelnch;

  FACTOR_2 on freelnch;" ,

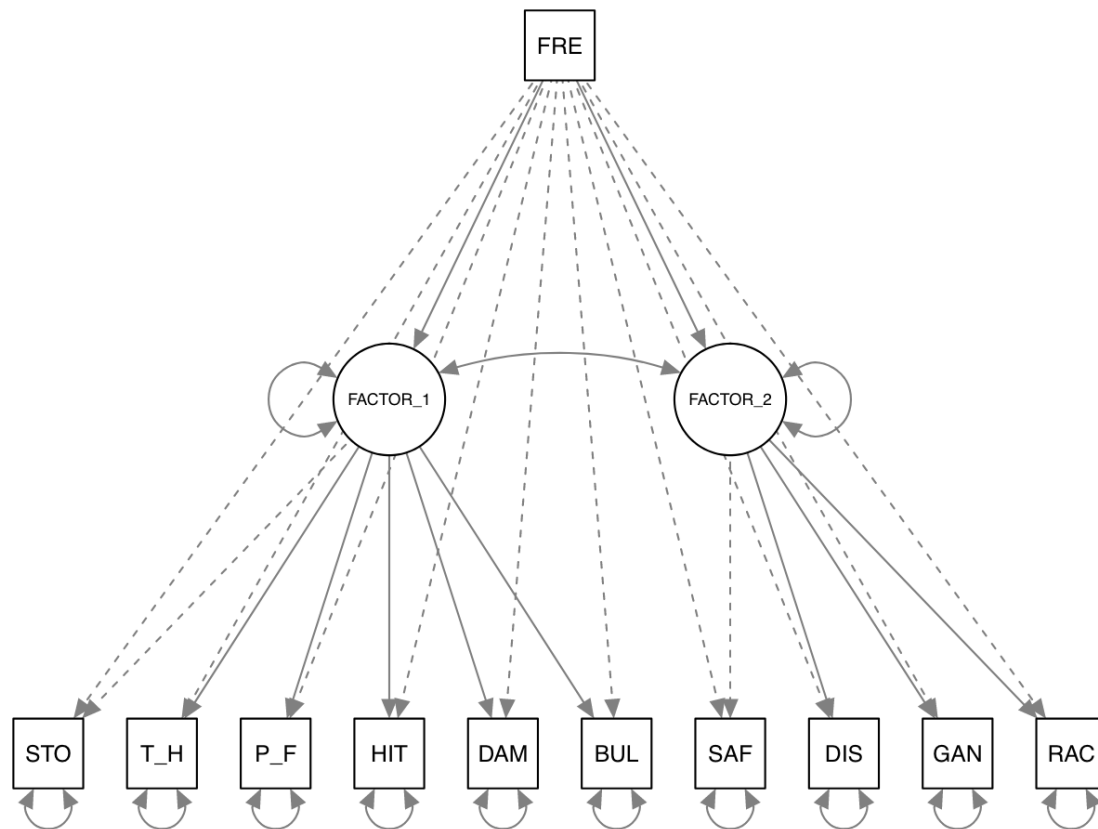
PLOT = "type = plot3;",
OUTPUT = "sampstat standardized residual modindices (3.84);",

usevariables = colnames(mimic_data),
rdata = mimic_data)

mimic_m1_fit <- mplusModeler(mimic_m1,
  dataout=here("06-MIMIC", "mimic_mplus", "mimic_data.dat"),
  modelout=here("06-MIMIC", "mimic_mplus", "mimic_model1.inp"),
  check=TRUE, run = TRUE, hashfilename = FALSE)

```

MIMIC model 2 - probe for covariate -> indicator DIF (Differential Item Functioning)



```
mimic_m2 <- mplusObject(

  TITLE = "MIMIC model2",

  VARIABLE =
    "usevar = frelnch stolen-rac_fght;",

  ANALYSIS =
    "estimator = mlr;",

  MODEL =
    "FACTOR_1 by stolen t_hurt p_fight hit damaged bullied;

    FACTOR_2 by safe disrupt gangs rac_fght;

    FACTOR_1 on frelnch;

    FACTOR_2 on frelnch;

    stolen-rac_fght on frelnch@0; ! to check DIF see modification indices ",

  PLOT = "type = plot3;",
  OUTPUT = "sampstat standardized residual modindices (.1);",
```



```

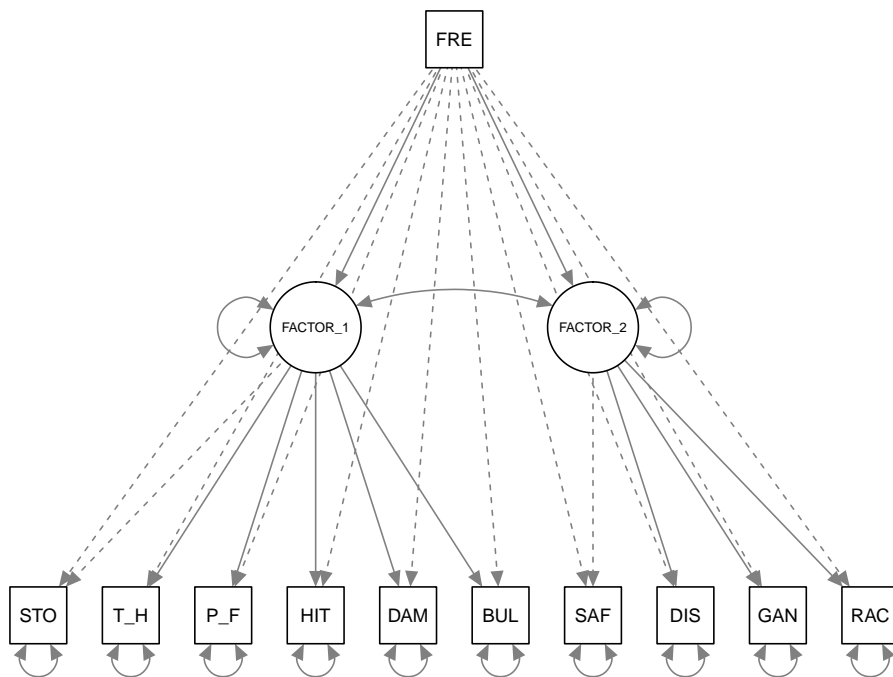
usevariables = colnames(mimic_data),
rdata = mimic_data)

mimic_m2_fit <- mplusModeler(mimic_m2,
  dataout=here("06-MIMIC", "mimic_mplus", "mimic_data.dat"),
  modelout=here("06-MIMIC", "mimic_mplus", "mimic_model2.inp"),
  check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_output2 <- readModels(here("06-MIMIC", "mimic_mplus", "mimic_model2.out"), quiet = TRUE)

# Plot model:
semPaths(mimic_output2,
  intercepts=FALSE,
)

```



MIMIC model 3 - specify covariate -> indicator DIF

Number of parameters for MIMIC model 3 = 34

- 8 indicator loadings (10 items - 2 fixed loadings)
- 10 intercepts
- 10 residual variances
- 2 factor variances
- 1 factor co-variance
- 1 covariate mean
- 1 covariate variance
- 1 DIF (covariate -> indicator)

```

grViz(" digraph mimic_mode_3 {

graph [overlap = true, fontsize = 12, fontname = Times]

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

  node [shape = box, label = 'Percent Free Lunch']
  X;

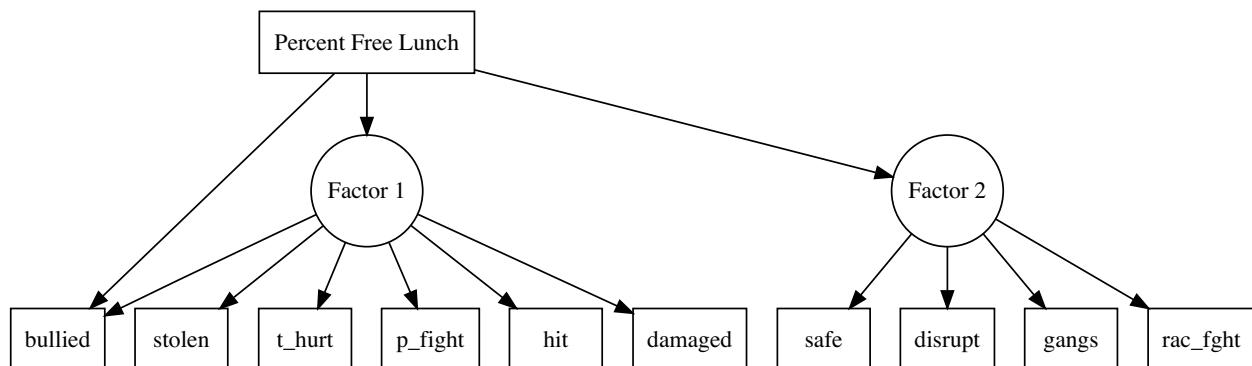
  node [shape = circle, fixedsize = true, width = 0.9, label = 'Factor 1']
  F1;

  node [shape = circle, fixedsize = true, width = 0.9, label = 'Factor 2']
  F2;

  edge [color = black]

  F1->{stolen t_hurt p_fight hit damaged bullied}
  F2->{safe disrupt gangs rac_fght}
  X->F1 X->F2 X->bullied
}")

```



```

mimic_m3 <- mplusObject(

  TITLE = "MIMIC model3",

  VARIABLE =
    "usevar = freelnch stolen-rac_fght;",

  ANALYSIS =
    "estimator = mlr;",

```

```

MODEL =
  "FACTOR_1 by stolen t_hurt p_fight hit damaged bullied;

  FACTOR_2 by safe disrupt gangs rac_fight;

  FACTOR_1 FACTOR_2 on freelnch;

  bullied on freelnch; ",

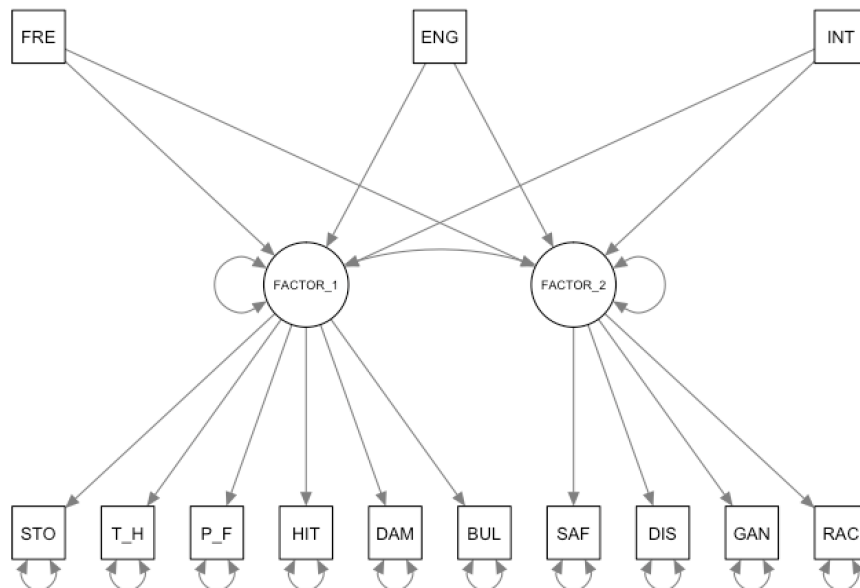
PLOT = "type = plot3;",
OUTPUT = "sampstat standardized residual modindices (3.84);",

usevariables = colnames(mimic_data),
rdata = mimic_data)

mimic_m3_fit <- mplusModeler(mimic_m3,
  dataout=here("06-MIMIC", "mimic_mplus", "mimic_data.dat"),
  modelout=here("06-MIMIC", "mimic_mplus", "mimic_model3.inp"),
  check=TRUE, run = TRUE, hashfilename = FALSE)

```

MIMIC model 4 - two covariates & an interaction term



```

mimic_m4 <- mplusObject(
  TITLE = "MIMIC model4",

```

```

VARIABLE =
  "usevar = freelnch stolen-rac_fght eng_2nd int;",

ANALYSIS =
  "estimator = mlr;",

DEFINE =
  "if bystlang == 1 THEN eng_2nd=0;
  if bystlang == 0 THEN eng_2nd=1;
  int = eng_2nd*freelnch;",

MODEL =
  "FACTOR_1 by stolen t_hurt p_fight hit damaged bullied;

  FACTOR_2 by safe disrupt gangs rac_fght;

  FACTOR_1 FACTOR_2 on freelnch eng_2nd int; ",

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

usevariables = colnames(mimic_data),
rdata = mimic_data)

mimic_m4_fit <- mplusModeler(mimic_m4,
  dataout=here("06-MIMIC", "mimic_mplus", "mimic_data.dat"),
  modelout=here("06-MIMIC", "mimic_mplus", "mimic_model4.inp"),
  check=TRUE, run = TRUE, hashfilename = FALSE)

```

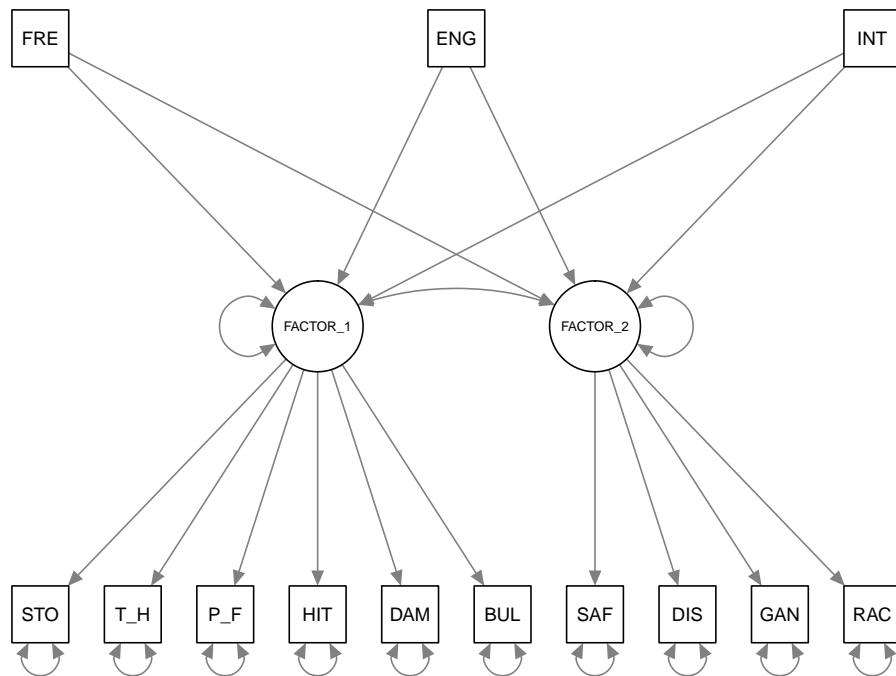
create a path diagram of MIMIC model 4

```

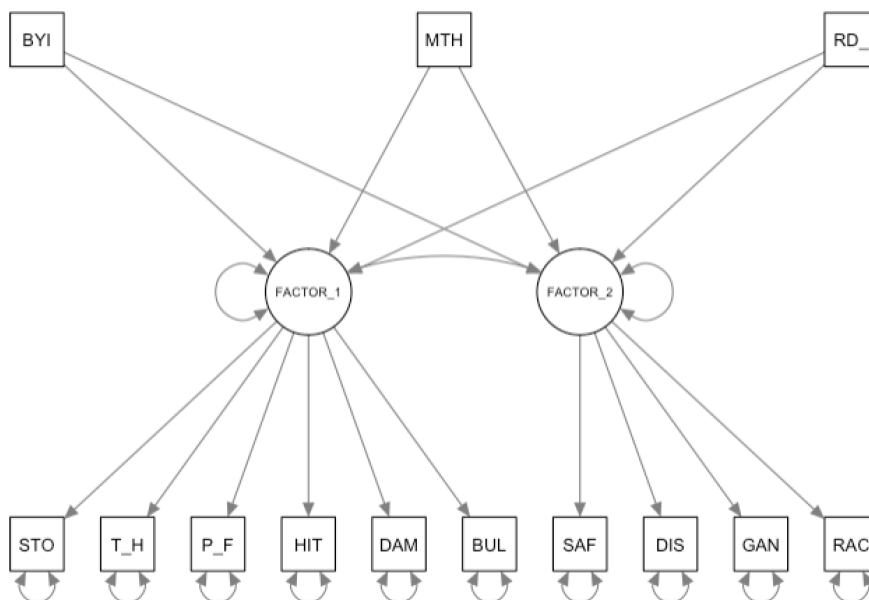
# Read in the model to R within the "cfa_mplus" folder
mimic_output4 <- readModels(here("06-MIMIC", "mimic_mplus", "mimic_model4.out"), quiet = TRUE)

# Plot model:
semPaths(mimic_output4,
  intercepts=FALSE,
  fixedStyle = c(1))

```



MIMIC model 5 - three continuous covariates



```

mimic_m5 <- mplusObject(

  TITLE = "MIMIC model5",

  VARIABLE =
    "usevar = byincome mth_test rd_test stolen-rac_fght;",

  ANALYSIS =
    "estimator = mlr;",

  MODEL =
    "FACTOR_1 by stolen t_hurt p_fight hit damaged bullied;

    FACTOR_2 by safe disrupt gangs rac_fght;

    FACTOR_1 FACTOR_2 on byincome mth_test rd_test; ",

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

  usevariables = colnames(mimic_data),
  rdata = mimic_data)

mimic_m5_fit <- mplusModeler(mimic_m5,
  dataout=here("06-MIMIC", "mimic_mplus", "mimic_data.dat"),
  modelout=here("06-MIMIC", "mimic_mplus", "mimic_model5.inp"),
  check=TRUE, run = TRUE, hashfilename = FALSE)

```

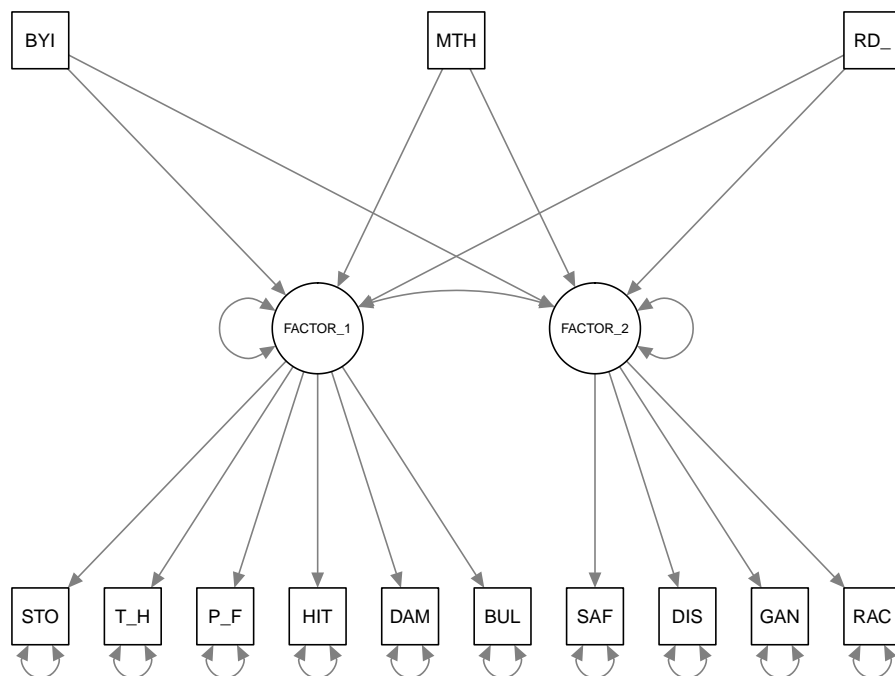
create a path diagram of MIMIC model 5

```

# Read in the model to R
mimic_output5 <- readModels(here("06-MIMIC", "mimic_mplus", "mimic_model5.out"), quiet = TRUE)

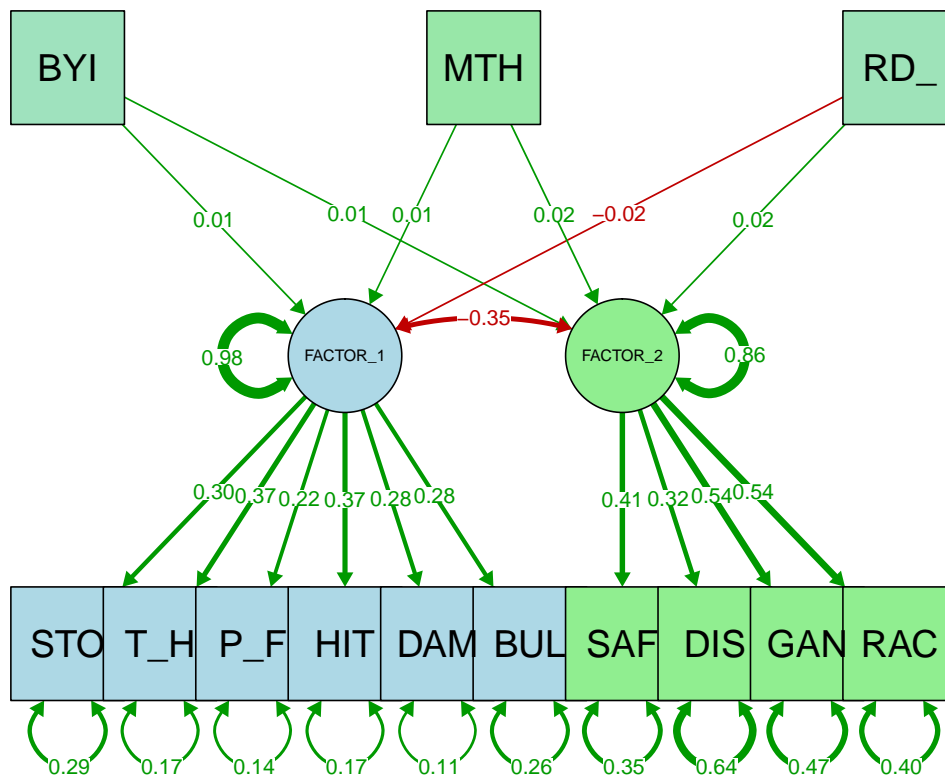
# Plot model:
semPaths(mimic_output5,
  intercepts=FALSE,
  fixedStyle = c(1)
)

```



practice some formatting with `semPlot::semPaths()`

```
semPaths(mimic_output5,
  "stdyx", # plot the standardized parameter estimates (see output section: STDYX)
  intercepts=FALSE,
  fixedStyle = c(1),
  color= list(lat = c("light blue"," light green")),
  sizeMan = 10, sizeInt = 10, sizeLat = 10,
  edge.label.cex=.8,
  fade=FALSE
)
```



read all models and create table

```
all_models <- readModels(here("06-MIMIC", "mimic_mplus"), quiet = TRUE)

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

table %>%
  mutate(Filename = str_remove(Filename, ".out")) %>%
  gt() %>%
  cols_label(
    Filename = "Model",
    Parameters = "Par",
    ChiSqM_Value = "ChiSq",
    CFI = "CFI", TLI = "TLI", SRMR = "SRMR",
    RMSEA_Estimate = "RMSEA",
    RMSEA_90CI_LB = "Lower CI",
    RMSEA_90CI_UB = "Upper CI")
```


Model	Par	ChiSq	CFI	TLI	SRMR	RMSEA	Lower CI	Upper CI
cfa_model0	31	121.460	0.898	0.865	0.043	0.060	0.048	0.071
mimic_model1	33	133.455	0.894	0.862	0.044	0.057	0.047	0.069
mimic_model2	33	133.455	0.894	0.862	0.044	0.057	0.047	0.069
mimic_model3	34	130.058	0.897	0.862	0.043	0.057	0.046	0.069
mimic_model4	37	153.527	0.899	0.869	0.041	0.050	0.040	0.060
mimic_model5	37	169.116	0.894	0.863	0.042	0.052	0.043	0.061

End

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/](https://allisonhorst.github.io/)
- 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/>
