Weighted Least Squares Estimator - Factor Analysis with Categorical Indicators

Adam Garber

Norwegian University of Science and Technology - A Course in `MplusAutomation`

June 01, 2021

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

FURTHER READING: A description of the models run in this lab exercise can be found in chapter 16 (page 411) in Kline, 2016. All specifications for invariance with the WLS estimator follow the recomendations found in this book. Example Mplus Syntax

---

When to use the Weighted Least Squared (`WSL`) method?

1. you have ordinal indicators (e.g., Likert scales)
2. you have categorical indicators (e.g., binary, or nominal)
3. you have a combination of ordinal, categorical, and/or continuous indicators
4. you have skewed continuous indicators

How to specify the Robust WSL estimator (`WLSMV`) in Mplus

Include the following lines of syntax:

- `categorical = X1 X2 X3... ;` *identify any categorical indicators*
- `estimator = wlsmv;`
- `parameterization = theta;` *to allow for error variances to be freely estimated across groups*

Interpretation of χ^2 with WLS estimator

- Cannot do normal chi-square difference tests (use the `DIFFTEST` option)
- The scaled Satorra-Bentler correction is necessary

```r
# load packages
library(MplusAutomation)
library(tidyverse)
library(here)
library(corrplot)
library(semPlot)
library(gt)
```

Prepare data

Read in data

```r
data_raw <- read_csv("https://garberadamc.github.io/project-site/data/hsls_fa_data_subset.csv")
```

Reverse code for factor interpretation

```r
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]
```

Check correlations of science indicators

```r
sci_cor <- cor(hsls_data[27:38], use = "pairwise.complete.obs")
#sci_cor

corrplot(sci_cor,
         method = "circle",
         type = "upper")
```
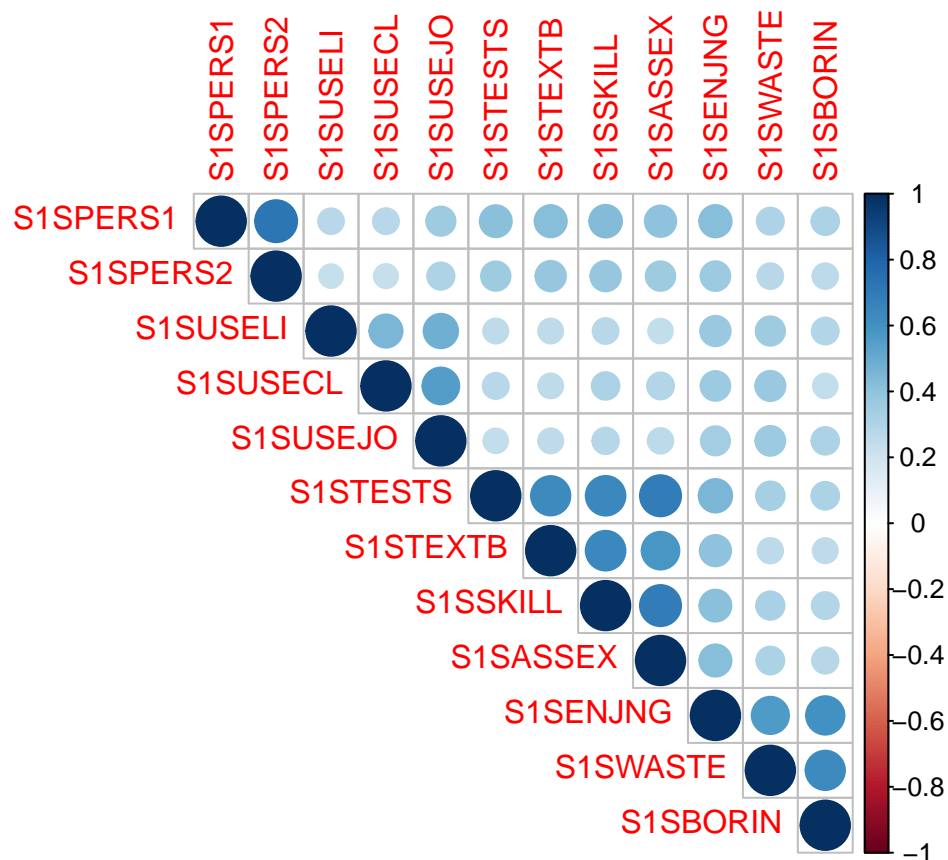
```
# NOTE: correlations should be reported from Mplus output as
# the matrix estimated using `cor()` are not polychoric correlations.
```

Model 0 - Exploratory Factor Analysis (EFA) with WLS Estimator

```r
m.step0  <- mplusObject(
  TITLE = " WSL FACTOR ANALYSIS EFA - HSLS SCIENCE DEMO",
  VARIABLE =
"usevar =
S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
S1SUSEJO S1STESTS S1STEXTB S1SSKILL
S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

categorical =
S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
S1SUSEJO S1STESTS S1STEXTB S1SSKILL
S1SASSEX S1SENJNG S1SWASTE S1SBORIN;",

  ANALYSIS =
"type = efa 1 7;
```

```
    estimator=wlsmv;
    parameterization=theta;",

       MODEL = "" ,

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

  usevariables = colnames(hsls_data),
  rdata = hsls_data)

m.step0.fit <- mplusModeler(m.step0,
               dataout=here("10-wls", "wls_efa", "efa_sci_HSLS_wls.dat"),
               modelout=here("10-wls", "wls_efa", "efa_sci_HSLS_wls.inp"),
               check=TRUE, run = TRUE, hashfilename = FALSE)
```

BEGIN: SINGLE GROUP INVARIANCE MODELS

Model 1 - Confirmatory Factor Analysis (CFA) - Full Sample

```
m.step1  <- mplusObject(
  TITLE = " WSL FACTOR ANALYSIS CFA - HSLS SCIENCE DEMO",
  VARIABLE =
 "usevar =
 S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
 S1SUSEJO S1STESTS S1STEXTB S1SSKILL
 S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

 categorical =
 S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
 S1SUSEJO S1STESTS S1STEXTB S1SSKILL
 S1SASSEX S1SENJNG S1SWASTE S1SBORIN;",

  ANALYSIS =
 "estimator=wlsmv;
  parameterization=theta;",

       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.step1.fit <- mplusModeler(m.step1,
              dataout=here("10-wls", "wls_invar", "m1_cfa_wls.dat"),
              modelout=here("10-wls", "wls_invar", "m1_cfa_wls.inp"),
              check=TRUE, run = TRUE, hashfilename = FALSE)

## END: CONFIRMATORY FACTOR ANALYSIS OF SCIENCE INDICATORS
```

Model 2 - CFA for group SEX = 0 (Male)

```
m.step2  <- mplusObject(
  TITLE = " WSL FACTOR ANALYSIS CFA MALE GROUP SCIENCE",
  VARIABLE =
 "usevar =
 S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
 S1SUSEJO S1STESTS S1STEXTB S1SSKILL
 S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

  categorical =
 S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
 S1SUSEJO S1STESTS S1STEXTB S1SSKILL
 S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

 !X1SEX (1 = MALE, 2 = FEMALE)
 USEOBS = X1SEX == 1;",

  ANALYSIS =
   "estimator=wlsmv;  parameterization=theta;",

     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.step2.fit <- mplusModeler(m.step2,
              dataout=here("10-wls", "wls_invar", "m2_cfa_males_wls.dat"),
              modelout=here("10-wls", "wls_invar", "m2_cfa_males_wls.inp"),
              check=TRUE, run = TRUE, hashfilename = FALSE)
```

Model 3 - CFA for group SEX = 1 (Female)

```
m.step3  <- mplusObject(
  TITLE = " WSL FACTOR ANALYSIS CFA - FEMALE GROUP SCIENCE",
  VARIABLE =
 "usevar =
 S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
 S1SUSEJO S1STESTS S1STEXTB S1SSKILL
 S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

  categorical =
 S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
 S1SUSEJO S1STESTS S1STEXTB S1SSKILL
 S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

 !X1SEX (1 = MALE, 2 = FEMALE)
 USEOBS = X1SEX==2;",

  ANALYSIS =  "estimator=wlsmv; parameterization=theta;",

     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.step3.fit <- mplusModeler(m.step3,
              dataout=here("10-wls", "wls_invar", "m3_cfa_females_wls.dat"),
              modelout=here("10-wls", "wls_invar", "m3_cfa_females_wls.inp"),
              check=TRUE, run = TRUE, hashfilename = FALSE)
```

BEGIN: MULTI-GROUP INVARIANCE MODELS

How does `update()` work with MplusAutomation?

- Instead of copying the whole model above and altering one line of syntax we can use the `update()` model function.
- A key advantage of this is that it highlights changes in code which otherwise might be easy to overlook.
- **tilde** (~) will replace everything in that section of the input.
- **tilde-dot-plus** (~.+) will update the section by adding the specified code into that section.

This coding strategy although space-efficient was not utilized for following models. Rather, code is made explicit for pedagogy.

Model 4 - Configural Invariance:

- free item loadings
- free intercepts
- free residuals (group 2)

Number of parameters for configural model = 108

- 16 item loadings (`8 freeloadings * 2 groups = 16`)
- 56 intercepts (`36 * 2groups - 16 fixed  = 56`)
- 12 residual variances
- 12 factor co-variances
- 04 factor means
- 08 factor variances

```
m.step4  <- mplusObject(
 TITLE = " WSL FACTOR ANALYSIS - CONFIGURAL",
 VARIABLE =
"usevar =
 S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
 S1SUSEJO S1STESTS S1STEXTB S1SSKILL
 S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

  categorical =
 S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
 S1SUSEJO S1STESTS S1STEXTB S1SSKILL
 S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

 grouping = X1SEX (1=MALE 2=FEMALE);",

 ANALYSIS =  "estimator=wlsmv; parameterization=theta;",

 MODEL =
"SCI_ID  BY S1SPERS1@1 S1SPERS2;
 SCI_UT  BY S1SUSELI@1 S1SUSECL S1SUSEJO;
 SCI_EFF BY S1STESTS@1 S1STEXTB S1SSKILL S1SASSEX ;
 SCI_INT BY S1SENJNG@1 S1SWASTE S1SBORIN;

! Threshold 1 for each item is constrained across groups1
! One additional threshold constrained across groups for
! the identification variable (ULI).

[S1SPERS1$1](f1_11);   [S1SPERS1$2](f1_12);   [S1SPERS1$3] ;
[S1SPERS2$1](f1_21);   [S1SPERS2$2]        ;   [S1SPERS2$3] ;
[S1SUSELI$1](f2_11);   [S1SUSELI$2](f2_12);   [S1SUSELI$3] ;
[S1SUSECL$1](f2_21);   [S1SUSECL$2]        ;   [S1SUSECL$3] ;
[S1SUSEJO$1](f2_31);   [S1SUSEJO$2]        ;   [S1SUSEJO$3] ;
[S1STESTS$1](f3_11);   [S1STESTS$2](f3_12);   [S1STESTS$3] ;
[S1STEXTB$1](f3_21);   [S1STEXTB$2]        ;   [S1STEXTB$3] ;
[S1SSKILL$1](f3_31);   [S1SSKILL$2]        ;   [S1SSKILL$3] ;
[S1SASSEX$1](f3_41);   [S1SASSEX$2]        ;   [S1SASSEX$3] ;
[S1SENJNG$1](f4_11);   [S1SENJNG$2](f4_12);   [S1SENJNG$3] ;
[S1SWASTE$1](f4_21);   [S1SWASTE$2]        ;   [S1SWASTE$3] ;
[S1SBORIN$1](f4_31);   [S1SBORIN$2]        ;   [S1SBORIN$3] ;

[SCI_ID-SCI_INT@0];   ! factor means are fixed to 0 in group 1
SCI_ID-SCI_INT;       ! factor variances are free in group 1
S1SPERS1-S1SBORIN@1;  ! residual variances are fixed to 1 in group 1

MODEL FEMALE:

SCI_ID  BY S1SPERS1@1 S1SPERS2;
SCI_UT  BY S1SUSELI@1 S1SUSECL S1SUSEJO;
SCI_EFF BY S1STESTS@1 S1STEXTB S1SSKILL S1SASSEX ;
SCI_INT BY S1SENJNG@1 S1SWASTE S1SBORIN;

[S1SPERS1$1](f1_11);   [S1SPERS1$2](f1_12);   [S1SPERS1$3] ;
[S1SPERS2$1](f1_21);   [S1SPERS2$2]        ;   [S1SPERS2$3] ;
```

```
 [S1SUSELI$1](f2_11);  [S1SUSELI$2](f2_12);  [S1SUSELI$3] ;
 [S1SUSECL$1](f2_21);  [S1SUSECL$2]       ;  [S1SUSECL$3] ;
 [S1SUSEJO$1](f2_31);  [S1SUSEJO$2]       ;  [S1SUSEJO$3] ;
 [S1STESTS$1](f3_11);  [S1STESTS$2](f3_12);  [S1STESTS$3] ;
 [S1STEXTB$1](f3_21);  [S1STEXTB$2]       ;  [S1STEXTB$3] ;
 [S1SSKILL$1](f3_31);  [S1SSKILL$2]       ;  [S1SSKILL$3] ;
 [S1SASSEX$1](f3_41);  [S1SASSEX$2]       ;  [S1SASSEX$3] ;
 [S1SENJNG$1](f4_11);  [S1SENJNG$2](f4_12);  [S1SENJNG$3] ;
 [S1SWASTE$1](f4_21);  [S1SWASTE$2]       ;  [S1SWASTE$3] ;
 [S1SBORIN$1](f4_31);  [S1SBORIN$2]       ;  [S1SBORIN$3] ;

[SCI_ID-SCI_INT];        ! factor means are free in group 2
SCI_ID-SCI_INT;          ! factor variances are free in group 2
S1SPERS1-S1SBORIN;       ! residual variances are free in group 2

savedata: DIFFTEST=diff_1_2.dat;",

  usevariables = colnames(hsls_data),
  rdata = hsls_data)

m.step4.fit <- mplusModeler(m.step4,
              dataout=here("10-wls", "wls_invar", "m4_configural_wls.dat"),
              modelout=here("10-wls", "wls_invar", "m4_configural_wls.inp"),
              check=TRUE, run = TRUE, hashfilename = FALSE)
```

Model 5 - Metric Invariance:

- item loadings (set to equal)
- free intercepts
- free residuals (group 2)

Number of parameters for metric model = 100

- 08 item loadings (8 free loadings)
- 56 intercepts (36*2groups - 16 fixed = 56)
- 12 residual variances
- 12 factor co-variances
- 04 factor means
- 08 factor variances

```
m.step5 <- mplusObject(
  TITLE = "WSL FACTOR ANALYSIS - METRIC",
  VARIABLE =
 "usevar =
  S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
```

```
   S1SUSEJO S1STESTS S1STEXTB S1SSKILL
   S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

    categorical =
   S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
   S1SUSEJO S1STESTS S1STEXTB S1SSKILL
   S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

   grouping = X1SEX (1=MALE 2=FEMALE);",

   ANALYSIS =  "estimator=wlsmv; parameterization=theta;",

   MODEL =
  "SCI_ID  BY S1SPERS1@1 S1SPERS2;
   SCI_UT  BY S1SUSELI@1 S1SUSECL S1SUSEJO;
   SCI_EFF BY S1STESTS@1 S1STEXTB S1SSKILL S1SASSEX ;
   SCI_INT BY S1SENJNG@1 S1SWASTE S1SBORIN;

  ! threshold 1 for each item is constrained across groups
  ! one additional threshold constrained across groups for the identification variable (ULI)
  [S1SPERS1$1](f1_11);  [S1SPERS1$2](f1_12);  [S1SPERS1$3] ;
  [S1SPERS2$1](f1_21);  [S1SPERS2$2]       ;  [S1SPERS2$3] ;
  [S1SUSELI$1](f2_11);  [S1SUSELI$2](f2_12);  [S1SUSELI$3] ;
  [S1SUSECL$1](f2_21);  [S1SUSECL$2]       ;  [S1SUSECL$3] ;
  [S1SUSEJO$1](f2_31);  [S1SUSEJO$2]       ;  [S1SUSEJO$3] ;
  [S1STESTS$1](f3_11);  [S1STESTS$2](f3_12);  [S1STESTS$3] ;
  [S1STEXTB$1](f3_21);  [S1STEXTB$2]       ;  [S1STEXTB$3] ;
  [S1SSKILL$1](f3_31);  [S1SSKILL$2]       ;  [S1SSKILL$3] ;
  [S1SASSEX$1](f3_41);  [S1SASSEX$2]       ;  [S1SASSEX$3] ;
  [S1SENJNG$1](f4_11);  [S1SENJNG$2](f4_12);  [S1SENJNG$3] ;
  [S1SWASTE$1](f4_21);  [S1SWASTE$2]       ;  [S1SWASTE$3] ;
  [S1SBORIN$1](f4_31);  [S1SBORIN$2]       ;  [S1SBORIN$3] ;

  [SCI_ID-SCI_INT@0];   ! factor means are fixed to 0 in group 1
  SCI_ID-SCI_INT;       ! factor variances are free in group 1
  S1SPERS1-S1SBORIN@1;  ! residual variances are fixed to 1 in group 1

  MODEL FEMALE:

  ! ***NEW TO METRIC INV*** (comment out to hold loadings equal)
  !SCI_ID  BY S1SPERS1@1 S1SPERS2;
  !SCI_UT  BY S1SUSELI@1 S1SUSECL S1SUSEJO;
  !SCI_EFF BY S1STESTS@1 S1STEXTB S1SSKILL S1SASSEX ;
  !SCI_INT BY S1SENJNG@1 S1SWASTE S1SBORIN;

  [S1SPERS1$1](f1_11);  [S1SPERS1$2](f1_12);  [S1SPERS1$3] ;
  [S1SPERS2$1](f1_21);  [S1SPERS2$2]       ;  [S1SPERS2$3] ;
  [S1SUSELI$1](f2_11);  [S1SUSELI$2](f2_12);  [S1SUSELI$3] ;
  [S1SUSECL$1](f2_21);  [S1SUSECL$2]       ;  [S1SUSECL$3] ;
  [S1SUSEJO$1](f2_31);  [S1SUSEJO$2]       ;  [S1SUSEJO$3] ;
  [S1STESTS$1](f3_11);  [S1STESTS$2](f3_12);  [S1STESTS$3] ;
  [S1STEXTB$1](f3_21);  [S1STEXTB$2]       ;  [S1STEXTB$3] ;
  [S1SSKILL$1](f3_31);  [S1SSKILL$2]       ;  [S1SSKILL$3] ;
```

```
 [S1SASSEX$1](f3_41);  [S1SASSEX$2]          ;  [S1SASSEX$3] ;
 [S1SENJNG$1](f4_11);  [S1SENJNG$2](f4_12);  [S1SENJNG$3] ;
 [S1SWASTE$1](f4_21);  [S1SWASTE$2]          ;  [S1SWASTE$3] ;
 [S1SBORIN$1](f4_31);  [S1SBORIN$2]          ;  [S1SBORIN$3] ;

[SCI_ID-SCI_INT];        ! factor means are free in group 2
SCI_ID-SCI_INT;          ! factor variances are free in group 2
S1SPERS1-S1SBORIN;       ! residual variances are free in group 2

savedata: DIFFTEST=diff_2_3.dat;",

  usevariables = colnames(hsls_data),
  rdata = hsls_data)

m.step5.fit <- mplusModeler(m.step5,
              dataout=here("10-wls", "wls_invar", "m5_metric_wls.dat"),
              modelout=here("10-wls", "wls_invar", "m5_metric_wls.inp"),
              check=TRUE, run = TRUE, hashfilename = FALSE)
```

Model 6 - Scalar Invariance:

- item loadings (set to equal)
- intercepts (set to equal)
- free residuals (group 2)

Number of parameters for metric model = 80

- 08 item loadings (8 free loadings)
- 36 intercepts

- 12 residual variances
- 12 factor co-variances
- 04 factor means
- 08 factor variances

```
m.step6  <- mplusObject(
  TITLE = "WSL FACTOR ANALYSIS - SCALAR",
  VARIABLE =
 "usevar =
 S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
 S1SUSEJO S1STESTS S1STEXTB S1SSKILL
 S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

   categorical =
 S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
```

```
     S1SUSEJO S1STESTS S1STEXTB S1SSKILL
     S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

     grouping = X1SEX (1=MALE 2=FEMALE);",

     ANALYSIS =  "estimator=wlsmv; parameterization=theta;",

     MODEL =
    "SCI_ID  BY S1SPERS1@1 S1SPERS2;
     SCI_UT  BY S1SUSELI@1 S1SUSECL S1SUSEJO;
     SCI_EFF BY S1STESTS@1 S1STEXTB S1SSKILL S1SASSEX ;
     SCI_INT BY S1SENJNG@1 S1SWASTE S1SBORIN;

    ! ***NEW TO SCALAR INV*** (all thresholds set to equality)
    [S1SPERS1$1](f1_11);   [S1SPERS1$2](f1_12);   [S1SPERS1$3](f1_13);
    [S1SPERS2$1](f1_21);   [S1SPERS2$2](f1_22);   [S1SPERS2$3](f1_23);
    [S1SUSELI$1](f2_11);   [S1SUSELI$2](f2_12);   [S1SUSELI$3](f2_13);
    [S1SUSECL$1](f2_21);   [S1SUSECL$2](f2_22);   [S1SUSECL$3](f2_23);
    [S1SUSEJO$1](f2_31);   [S1SUSEJO$2](f2_32);   [S1SUSEJO$3](f2_33);
    [S1STESTS$1](f3_11);   [S1STESTS$2](f3_12);   [S1STESTS$3](f3_13);
    [S1STEXTB$1](f3_21);   [S1STEXTB$2](f3_22);   [S1STEXTB$3](f3_23);
    [S1SSKILL$1](f3_31);   [S1SSKILL$2](f3_32);   [S1SSKILL$3](f3_33);
    [S1SASSEX$1](f3_41);   [S1SASSEX$2](f3_42);   [S1SASSEX$3](f3_43);
    [S1SENJNG$1](f4_11);   [S1SENJNG$2](f4_12);   [S1SENJNG$3](f4_13);
    [S1SWASTE$1](f4_21);   [S1SWASTE$2](f4_22);   [S1SWASTE$3](f4_23);
    [S1SBORIN$1](f4_31);   [S1SBORIN$2](f4_32);   [S1SBORIN$3](f4_33);

    [SCI_ID-SCI_INT@0];   ! factor means are fixed to 0 in group 1
    SCI_ID-SCI_INT;       ! factor variances are free in group 1
    S1SPERS1-S1SBORIN@1;  ! residual variances are fixed to 1 in group 1

    MODEL FEMALE:

    !SCI_ID  BY S1SPERS1@1 S1SPERS2;
    !SCI_UT  BY S1SUSELI@1 S1SUSECL S1SUSEJO;
    !SCI_EFF BY S1STESTS@1 S1STEXTB S1SSKILL S1SASSEX ;
    !SCI_INT BY S1SENJNG@1 S1SWASTE S1SBORIN;

    ! ***NEW TO SCALAR INV*** (all thresholds set to equality)
    [S1SPERS1$1](f1_11);   [S1SPERS1$2](f1_12);   [S1SPERS1$3](f1_13);
    [S1SPERS2$1](f1_21);   [S1SPERS2$2](f1_22);   [S1SPERS2$3](f1_23);
    [S1SUSELI$1](f2_11);   [S1SUSELI$2](f2_12);   [S1SUSELI$3](f2_13);
    [S1SUSECL$1](f2_21);   [S1SUSECL$2](f2_22);   [S1SUSECL$3](f2_23);
    [S1SUSEJO$1](f2_31);   [S1SUSEJO$2](f2_32);   [S1SUSEJO$3](f2_33);
    [S1STESTS$1](f3_11);   [S1STESTS$2](f3_12);   [S1STESTS$3](f3_13);
    [S1STEXTB$1](f3_21);   [S1STEXTB$2](f3_22);   [S1STEXTB$3](f3_23);
    [S1SSKILL$1](f3_31);   [S1SSKILL$2](f3_32);   [S1SSKILL$3](f3_33);
    [S1SASSEX$1](f3_41);   [S1SASSEX$2](f3_42);   [S1SASSEX$3](f3_43);
    [S1SENJNG$1](f4_11);   [S1SENJNG$2](f4_12);   [S1SENJNG$3](f4_13);
    [S1SWASTE$1](f4_21);   [S1SWASTE$2](f4_22);   [S1SWASTE$3](f4_23);
    [S1SBORIN$1](f4_31);   [S1SBORIN$2](f4_32);   [S1SBORIN$3](f4_33);

    [SCI_ID-SCI_INT];       ! factor means are free in group 2
```

```
SCI_ID-SCI_INT;          ! factor variances are free in group 2
S1SPERS1-S1SBORIN;       ! residual variances are free in group 2

savedata: DIFFTEST=diff_3_4.dat;",

  usevariables = colnames(hsls_data),
  rdata = hsls_data)

m.step6.fit <- mplusModeler(m.step6,
              dataout=here("10-wls", "wls_invar", "m6_scalar_wls.dat"),
              modelout=here("10-wls", "wls_invar", "m6_scalar_wls.inp"),
              check=TRUE, run = TRUE, hashfilename = FALSE)
```

Model 7 - Strict Invariance:

- item loadings (set to equal)
- intercepts (set to equal)
- residuals (fixed to 1)

Number of parameters for metric model = 68

- 08 item loadings (8 free loadings)
- 36 intercepts

- 00 residual variances
- 12 factor co-variances
- 04 factor means
- 08 factor variances

```
m.step7  <- mplusObject(
  TITLE = "WSL FACTOR ANALYSIS - STRICT",
  VARIABLE =
 "usevar =
  S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
  S1SUSEJO S1STESTS S1STEXTB S1SSKILL
  S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

   categorical =
  S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
  S1SUSEJO S1STESTS S1STEXTB S1SSKILL
  S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

  grouping = X1SEX (1=MALE 2=FEMALE);",

  ANALYSIS =  "estimator=wlsmv; parameterization=theta;",
```

```
 MODEL =
"SCI_ID  BY S1SPERS1@1 S1SPERS2;
 SCI_UT  BY S1SUSELI@1 S1SUSECL S1SUSEJO;
 SCI_EFF BY S1STESTS@1 S1STEXTB S1SSKILL S1SASSEX ;
 SCI_INT BY S1SENJNG@1 S1SWASTE S1SBORIN;

 [S1SPERS1$1](f1_11);   [S1SPERS1$2](f1_12);   [S1SPERS1$3](f1_13);
 [S1SPERS2$1](f1_21);   [S1SPERS2$2](f1_22);   [S1SPERS2$3](f1_23);
 [S1SUSELI$1](f2_11);   [S1SUSELI$2](f2_12);   [S1SUSELI$3](f2_13);
 [S1SUSECL$1](f2_21);   [S1SUSECL$2](f2_22);   [S1SUSECL$3](f2_23);
 [S1SUSEJO$1](f2_31);   [S1SUSEJO$2](f2_32);   [S1SUSEJO$3](f2_33);
 [S1STESTS$1](f3_11);   [S1STESTS$2](f3_12);   [S1STESTS$3](f3_13);
 [S1STEXTB$1](f3_21);   [S1STEXTB$2](f3_22);   [S1STEXTB$3](f3_23);
 [S1SSKILL$1](f3_31);   [S1SSKILL$2](f3_32);   [S1SSKILL$3](f3_33);
 [S1SASSEX$1](f3_41);   [S1SASSEX$2](f3_42);   [S1SASSEX$3](f3_43);
 [S1SENJNG$1](f4_11);   [S1SENJNG$2](f4_12);   [S1SENJNG$3](f4_13);
 [S1SWASTE$1](f4_21);   [S1SWASTE$2](f4_22);   [S1SWASTE$3](f4_23);
 [S1SBORIN$1](f4_31);   [S1SBORIN$2](f4_32);   [S1SBORIN$3](f4_33);


 [SCI_ID-SCI_INT@0];   ! factor means are fixed to 0 in group 1
 SCI_ID-SCI_INT;       ! factor variances are free in group 1
 S1SPERS1-S1SBORIN@1;  ! residual variances are fixed to 1 in group 1

 MODEL FEMALE:

 !SCI_ID  BY S1SPERS1@1 S1SPERS2;
 !SCI_UT  BY S1SUSELI@1 S1SUSECL S1SUSEJO;
 !SCI_EFF BY S1STESTS@1 S1STEXTB S1SSKILL S1SASSEX ;
 !SCI_INT BY S1SENJNG@1 S1SWASTE S1SBORIN;

 [S1SPERS1$1](f1_11);   [S1SPERS1$2](f1_12);   [S1SPERS1$3](f1_13);
 [S1SPERS2$1](f1_21);   [S1SPERS2$2](f1_22);   [S1SPERS2$3](f1_23);
 [S1SUSELI$1](f2_11);   [S1SUSELI$2](f2_12);   [S1SUSELI$3](f2_13);
 [S1SUSECL$1](f2_21);   [S1SUSECL$2](f2_22);   [S1SUSECL$3](f2_23);
 [S1SUSEJO$1](f2_31);   [S1SUSEJO$2](f2_32);   [S1SUSEJO$3](f2_33);
 [S1STESTS$1](f3_11);   [S1STESTS$2](f3_12);   [S1STESTS$3](f3_13);
 [S1STEXTB$1](f3_21);   [S1STEXTB$2](f3_22);   [S1STEXTB$3](f3_23);
 [S1SSKILL$1](f3_31);   [S1SSKILL$2](f3_32);   [S1SSKILL$3](f3_33);
 [S1SASSEX$1](f3_41);   [S1SASSEX$2](f3_42);   [S1SASSEX$3](f3_43);
 [S1SENJNG$1](f4_11);   [S1SENJNG$2](f4_12);   [S1SENJNG$3](f4_13);
 [S1SWASTE$1](f4_21);   [S1SWASTE$2](f4_22);   [S1SWASTE$3](f4_23);
 [S1SBORIN$1](f4_31);   [S1SBORIN$2](f4_32);   [S1SBORIN$3](f4_33);

[SCI_ID-SCI_INT];       ! factor means are free in group 2
SCI_ID-SCI_INT;         ! factor variances are free in group 2
! ***NEW TO STRICT INV*** (all residual variances fixed to 1 in group 2)
S1SPERS1-S1SBORIN@1; ! residual variances are fixed in group 2

savedata: DIFFTEST=diff_4_5.dat;",

 usevariables = colnames(hsls_data),
 rdata = hsls_data)
```

14

```r
m.step7.fit <- mplusModeler(m.step7,
                dataout=here("10-wls", "wls_invar", "m7_strict_wls.dat"),
                modelout=here("10-wls", "wls_invar", "m7_strict_wls.inp"),
                check=TRUE, run = TRUE, hashfilename = FALSE)
```

Model 8 - Structural Invariance (using scalar model):

- item loadings (set to equal)
- intercepts (set to equal)
- free residuals (in group 2)
- **factor variances (set to equal)**
- **factor covariances (set to equal)**

Number of parameters for metric model = 70

- 08 item loadings (8 free loadings)
- 36 intercepts

- 12 residual variances
- 06 factor co-variances (set to equal)
- 04 factor means
- 04 factor variances (set to equal)

```r
m.step8  <- mplusObject(
  TITLE = "WSL FACTOR ANALYSIS - STRICT",
  VARIABLE =
 "usevar =
  S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
  S1SUSEJO S1STESTS S1STEXTB S1SSKILL
  S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

   categorical =
  S1SPERS1 S1SPERS2 S1SUSELI S1SUSECL
  S1SUSEJO S1STESTS S1STEXTB S1SSKILL
  S1SASSEX S1SENJNG S1SWASTE S1SBORIN;

  grouping = X1SEX (1=MALE 2=FEMALE);",

  ANALYSIS =  "estimator=wlsmv; parameterization=theta;",

  MODEL =
 "SCI_ID  BY S1SPERS1@1 S1SPERS2;
  SCI_UT  BY S1SUSELI@1 S1SUSECL S1SUSEJO;
  SCI_EFF BY S1STESTS@1 S1STEXTB S1SSKILL S1SASSEX ;
  SCI_INT BY S1SENJNG@1 S1SWASTE S1SBORIN;
```

```
  [S1SPERS1$1](f1_11);   [S1SPERS1$2](f1_12);   [S1SPERS1$3](f1_13);
  [S1SPERS2$1](f1_21);   [S1SPERS2$2](f1_22);   [S1SPERS2$3](f1_23);
  [S1SUSELI$1](f2_11);   [S1SUSELI$2](f2_12);   [S1SUSELI$3](f2_13);
  [S1SUSECL$1](f2_21);   [S1SUSECL$2](f2_22);   [S1SUSECL$3](f2_23);
  [S1SUSEJO$1](f2_31);   [S1SUSEJO$2](f2_32);   [S1SUSEJO$3](f2_33);
  [S1STESTS$1](f3_11);   [S1STESTS$2](f3_12);   [S1STESTS$3](f3_13);
  [S1STEXTB$1](f3_21);   [S1STEXTB$2](f3_22);   [S1STEXTB$3](f3_23);
  [S1SSKILL$1](f3_31);   [S1SSKILL$2](f3_32);   [S1SSKILL$3](f3_33);
  [S1SASSEX$1](f3_41);   [S1SASSEX$2](f3_42);   [S1SASSEX$3](f3_43);
  [S1SENJNG$1](f4_11);   [S1SENJNG$2](f4_12);   [S1SENJNG$3](f4_13);
  [S1SWASTE$1](f4_21);   [S1SWASTE$2](f4_22);   [S1SWASTE$3](f4_23);
  [S1SBORIN$1](f4_31);   [S1SBORIN$2](f4_32);   [S1SBORIN$3](f4_33);

  [SCI_ID-SCI_INT@0];      ! factor means are fixed to 0 in group 1
  SCI_ID-SCI_INT (V1-V4); ! factor variances (set to equality)
  S1SPERS1-S1SBORIN@1;     ! residual variances are fixed to 1 in group 1

  ! factor covariances set to equality
  SCI_ID with SCI_UT-SCI_INT(C1-C3);
  SCI_EFF with SCI_ID(C4);
  SCI_EFF with SCI_UT(C5);
  SCI_UT with SCI_ID (C6);

  MODEL FEMALE:

  [S1SPERS1$1](f1_11);   [S1SPERS1$2](f1_12);   [S1SPERS1$3](f1_13);
  [S1SPERS2$1](f1_21);   [S1SPERS2$2](f1_22);   [S1SPERS2$3](f1_23);
  [S1SUSELI$1](f2_11);   [S1SUSELI$2](f2_12);   [S1SUSELI$3](f2_13);
  [S1SUSECL$1](f2_21);   [S1SUSECL$2](f2_22);   [S1SUSECL$3](f2_23);
  [S1SUSEJO$1](f2_31);   [S1SUSEJO$2](f2_32);   [S1SUSEJO$3](f2_33);
  [S1STESTS$1](f3_11);   [S1STESTS$2](f3_12);   [S1STESTS$3](f3_13);
  [S1STEXTB$1](f3_21);   [S1STEXTB$2](f3_22);   [S1STEXTB$3](f3_23);
  [S1SSKILL$1](f3_31);   [S1SSKILL$2](f3_32);   [S1SSKILL$3](f3_33);
  [S1SASSEX$1](f3_41);   [S1SASSEX$2](f3_42);   [S1SASSEX$3](f3_43);
  [S1SENJNG$1](f4_11);   [S1SENJNG$2](f4_12);   [S1SENJNG$3](f4_13);
  [S1SWASTE$1](f4_21);   [S1SWASTE$2](f4_22);   [S1SWASTE$3](f4_23);
  [S1SBORIN$1](f4_31);   [S1SBORIN$2](f4_32);   [S1SBORIN$3](f4_33);

  [SCI_ID-SCI_INT];        ! factor means are free in group 2
  SCI_ID-SCI_INT (V1-V4); ! factor variances (set to equality)
  S1SPERS1-S1SBORIN;       ! residual variances are free group 2

  ! factor covariances set to equality
  SCI_ID with SCI_UT-SCI_INT(C1-C3);
  SCI_EFF with SCI_ID(C4);
  SCI_EFF with SCI_UT(C5);
  SCI_UT with SCI_ID (C6); ",

  usevariables = colnames(hsls_data),
  rdata = hsls_data)

m.step8.fit <- mplusModeler(m.step8,
              dataout=here("10-wls", "wls_invar", "m8_structural_wls.dat"),
```

```
              modelout=here("10-wls", "wls_invar", "m8_structural_wls.inp"),
              check=TRUE, run = TRUE, hashfilename = FALSE)
```

Mean differences: Females are...

```
Means  SCI_ID            -0.866      0.452     -1.914      0.056  SCI_UT            0.010
0.054       0.187        0.852      SCI_EFF            -0.577      0.094     -6.149      0.000
SCI_INT           -0.062      0.094     -0.656      0.512
```

Create and customize tables using the package {gt}

Read into R summary of all models

```
all_models <- readModels(here("10-wls", "wls_invar"), quiet = TRUE)
```

Extract relevant data and generate table

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

model_table <- invar_summary %>%
  gt() %>%
  tab_header(
    title = "Fit Indices",    # Add a title
    subtitle = md(" ")
  ) %>%
  tab_options(
    table.width = pct(80)
  ) %>%
  tab_footnote(
    footnote = "Data from HSLS 2009",
    location = cells_title()
  ) %>%
  data_color(                  # Update cell colors...
    columns = vars(CFI),      # ...for column
    alpha = .5,
    autocolor_text = FALSE,
    colors = scales::col_numeric(
```

17

```
    palette = c(
      "yellow", "blue"),   # choose color
    domain = c(.89,.91))   # Scale endpoints (outside will be gray)
  ) %>%
cols_label(
  Filename = "Model",
  Parameters =  "Par",
  ChiSqM_Value = "ChiSq",
  RMSEA_Estimate = "RMSEA",
  RMSEA_90CI_LB = "Lower CI",
  RMSEA_90CI_UB = "Upper CI")
```

```
model_table
```

Fit Indices[1]

Model	Par	ChiSq	CFI	TLI	SRMR	RMSEA	Lower CI	Upper CI
m1_cfa_wls.out	54	657.142	0.983	0.977	0.028	0.068	0.063	0.073
m2_cfa_males_wls.out	54	345.034	0.986	0.980	0.027	0.067	0.060	0.074
m3_cfa_females_wls.out	54	354.316	0.981	0.974	0.032	0.068	0.062	0.075
m4_configural_wls.out	108	699.033	0.984	0.978	0.030	0.068	0.063	0.072
m5_metric_wls.out	100	685.895	0.984	0.980	0.030	0.064	0.059	0.068
m6_scalar_wls.out	80	749.934	0.983	0.982	0.031	0.061	0.057	0.065
m7_strict_wls.out	68	764.189	0.983	0.984	0.032	0.058	0.054	0.062
m8_structural_wls.out	72	565.901	0.988	0.988	0.034	0.049	0.045	0.053

[1]Data from HSLS 2009

Save table

```
model_table %>% gtsave("model_fit_table_wls.png", path = here("10-wls", "figures"))
```

How to use `DIFFTEST`?

https://stats.idre.ucla.edu/mplus/faq/how-can-i-compute-a-chi-squared-test-for-nested-models-with-the-mlmv-or-wlsmv-estimators-difftest/

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/

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