COS-D419 Factor Analysis and Structural Equation Models 2023, Assignment 4

Rong Guang

2023-02-10

Contents

1 SEM & teacher burnout

1.1 Exercise 4.1

Draw the graph of the initial, full structural equation model. Make sure that you have included all the specified paths.

Estimate the initial model using the robust MLM estimator *(robust variant of the ML estimator, to be precise!)* and present a brief summary of the model fit.

2 Preparation

2.1 Read in the data set

Start by downloading the data file from Moodle to Project folder.

```
library(tidyverse)
library(readr)
library(here)

#This week's file name
latest.name <- "ALLSEC.CSV"

#read in the data
mbi <- read_csv(file.path(here(),
                'data',
                latest.name))
```

2.2 Write functions

To control length of reports, codes already shown in the previous homework were not showing in the current report. Yet they are available in .rmd report.

2.2.1 to check unique values

2.2.2 to generate CFA results with improved readability

2.2.3 to generate functions for improving aethetics of correlation matrix

2.2.4 to generate a function for histogram overlapping with density plot

2.2.5 to generate a function for violin overlapping with box plot

2.2.6 To generate a function describing continuous data set

2.3 Inspect the data

2.3.1 Data structure

Have a quick overview of the data structure

```
library(knitr)
library(broom)
dim(mbi);mbi %>% apply(2, function(x)class(x));
```

```
## [1] 1430   32

##    ROLEA1    ROLEA2    ROLEC1    ROLEC2     WORK1     WORK2    CCLIM1    CCLIM2
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##    CCLIM3    CCLIM4      DEC1      DEC2     SSUP1     SSUP2     PSUP1     PSUP2
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##     SELF1     SELF2     SELF3      ELC1      ELC2      ELC3      ELC4      ELC5
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##       EE1       EE2       EE3       DP1       DP2       PA1       PA2       PA3
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
```

The data set contains 22 numeric variables of 372 obs. Their values appear to follow a consistent pattern covering the integer from 1 to 7, except for Items 4, 7, 17 and 21, which did not include a value of 1.

2.3.2 Descriptive statistics of measured variables

```
library(finalfit);library(kableExtra);

descriptive(mbi) |>
  pack_rows(index =
              c("Factor 1*: Role Ambiguity \n(high score means negative)" = 2,
                "Factor 2*: Role conflict \n(high score means negative)" = 2,
                "Factor 3*: Work overload \n(high score means negative)" = 2,
                "Factor 4‡: classroom climate" = 4,
                "Factor 5*: Decision-making" = 2,
                "Factor 6*: Superior support" =  2,
                "Factor 7*: Peer support" = 2,
                "Factor 8‡: Self-esteem" = 3,
                "Factor 9‡: External locus of control" = 5,
                "Factor 10†: Emotional Exhaustion \n(high score means negative)" = 3,
                "Factor 11†: Depersonalization \n(high score means negative)" = 2,
                "Factor 12†: Personal Accomplishment" = 3)) |>
  footnote(general =
             "Indicators variables were formulated through item parcels.",
    symbol = c("These indicators are parcels from Teacher Stress Scale instrument",
               "These indicators are parcels from BMI instrument",
               "These parcels consist of items from single unidimensional scales")
           )
```

Table 1: Descriptive statistics for measurements

			Central tendency		Dispersion tendency			
	n	n of NA	Mean	Median	SD	Min	Max	Q1~Q3
Factor 1*: Role Ambiguity (high score means negative)								
ROLEA1	1430	0	2.4	2.3	0.9	1.0	6.0	1.7 ~ 3.0
ROLEA2	1430	0	2.1	2.0	1.0	1.0	6.0	1.5 ~ 2.5
Factor 2*: Role conflict (high score means negative)								
ROLEC1	1430	0	3.0	3.0	1.1	1.0	6.0	2.3 ~ 3.7
ROLEC2	1430	0	3.0	3.0	1.2	1.0	6.0	2.0 ~ 4.0
Factor 3*: Work overload (high score means negative)								
WORK1	1430	0	3.2	3.3	1.2	1.0	6.0	2.3 ~ 4.0
WORK2	1430	0	2.2	2.0	1.1	1.0	6.0	1.5 ~ 3.0
Factor 4‡: classroom climate								
CCLIM1	1430	0	3.0	3.0	0.5	1.0	4.0	2.7 ~ 3.3
CCLIM2	1430	0	2.7	2.7	0.6	1.0	4.0	2.3 ~ 3.0
CCLIM3	1430	0	2.9	3.0	0.5	1.0	4.0	2.7 ~ 3.3
CCLIM4	1430	0	3.1	3.0	0.7	1.0	4.0	2.5 ~ 3.5
Factor 5*: Decision-making								
DEC1	1430	0	4.0	4.0	1.0	1.0	6.0	3.3 ~ 4.7
DEC2	1430	0	4.2	4.5	1.3	1.0	6.0	3.5 ~ 5.5
Factor 6*: Superior support								

Table 1: Descriptive statistics for measurements *(continued)*

	n	n of NA	Central tendency		Dispersion tendency			
			Mean	Median	SD	Min	Max	Q1~Q3
SSUP1	1430	0	4.3	4.3	1.2	1.0	6.0	3.7 ~ 5.3
SSUP2	1430	0	4.4	4.5	1.3	1.0	6.0	3.5 ~ 5.5
Factor 7*: Peer support								
PSUP1	1430	0	4.6	4.7	1.0	1.0	6.0	4.0 ~ 5.3
PSUP2	1430	0	4.6	4.5	0.9	1.0	6.0	4.0 ~ 5.0
Factor 8‡: Self-esteem								
SELF1	1430	0	3.6	3.7	0.4	1.0	4.0	3.3 ~ 4.0
SELF2	1430	0	3.6	3.8	0.5	1.0	4.0	3.4 ~ 4.0
SELF3	1430	0	3.5	3.7	0.5	1.0	4.0	3.3 ~ 4.0
Factor 9‡: External locus of control								
ELC1	1430	0	2.9	3.0	0.6	1.0	4.8	2.6 ~ 3.4
ELC2	1430	0	3.0	3.0	0.6	1.0	5.0	2.5 ~ 3.5
ELC3	1430	0	2.8	2.8	0.5	1.0	4.8	2.4 ~ 3.2
ELC4	1430	0	2.2	2.2	0.6	1.0	4.5	1.8 ~ 2.5
ELC5	1430	0	2.5	2.4	0.6	1.0	4.8	2.0 ~ 3.0
Factor 10†: Emotional Exhaustion (high score means negative)								
EE1	1430	0	3.9	4.0	1.3	1.0	7.0	3.0 ~ 4.7
EE2	1430	0	3.5	3.3	1.3	1.0	7.0	2.7 ~ 4.3
EE3	1430	0	3.2	3.0	1.3	1.0	7.0	2.0 ~ 4.0
Factor 11†: Depersonalization (high score means negative)								
DP1	1430	0	2.3	2.0	1.1	1.0	6.7	1.3 ~ 3.0
DP2	1430	0	2.1	1.5	1.2	1.0	7.0	1.0 ~ 2.5
Factor 12†: Personal Accomplishment								
PA1	1430	0	5.7	6.0	0.9	2.0	7.0	5.3 ~ 6.3
PA2	1430	0	5.8	6.0	1.0	2.0	7.0	5.5 ~ 6.5
PA3	1430	0	5.8	6.0	1.0	2.0	7.0	5.3 ~ 6.7

Note:

Indicators variables were formulated through item parcels.

* These indicators are parcels from Teacher Stress Scale instrument

† These indicators are parcels from BMI instrument

‡ These parcels consist of items from single unidimensional scales

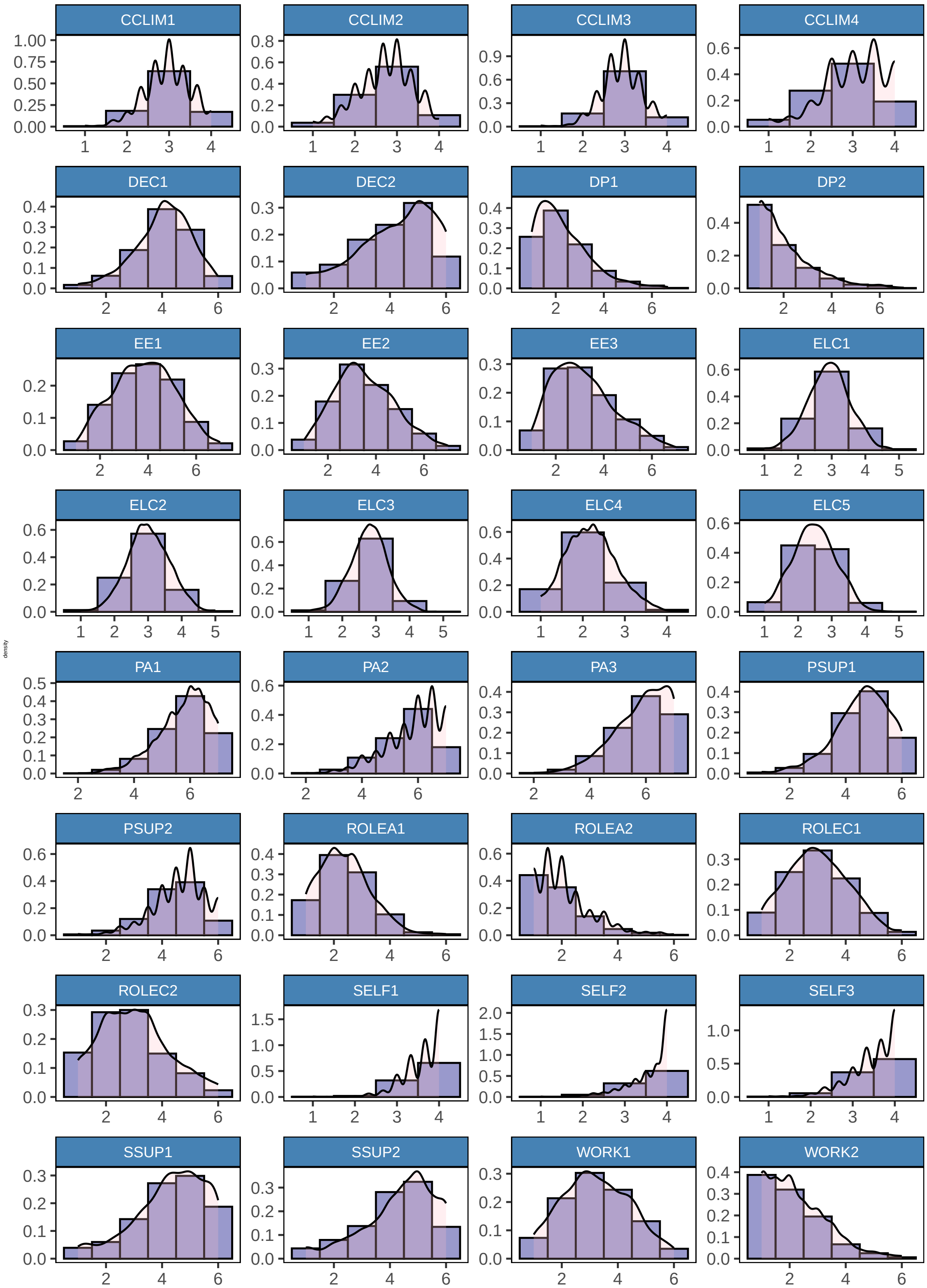
2.3.3 Visualization

(1) Histogram

Distribution of the data was examined via Histogram

```
corr.density(mbi, fig.num = 1)
```

Figure 1 Distribution of selected items



Ridge-line plots were generated for TSS and MBI indicators, respectively. By partially overlaying, it is a demonstration viable for comparing multiple distributions.

```
library(ggridges)
library(viridis)
library(hrbrthemes)
library(patchwork)
a <- mbi |>
  select(
    starts_with("EE")|starts_with("DP")|starts_with("PA")
    ) |>
  pivot_longer(everything(), names_to = "variable", values_to = "value") |>
  ggplot(aes(x = value, y = variable, fill = ..x..)) +
  geom_density_ridges_gradient(scale = 3, rel_min_height = 0.01) +
  scale_fill_viridis(name = "parcelled score", option = "C") +
  labs(title =
         'Fig2 (a). Distribution of indicator scores from BMI instrument') +
  labs(x = "Indicator scores", y = "Indicators") +
    theme(
      legend.position="none",
      panel.spacing = unit(0.1, "lines"),
      plot.title = element_text(size = 12),
      panel.grid.major = element_blank(),
      panel.background = element_rect(color = "black",
                                      fill = "white")
      )

b <- mbi |>
  select(
    starts_with("ROL")|starts_with("WOR")|starts_with("DEC")|contains("SUP")
    ) |>
  pivot_longer(everything(), names_to = "variable", values_to = "value") |>
  ggplot(aes(x = value, y = variable, fill = ..x..)) +
  geom_density_ridges_gradient(scale = 3, rel_min_height = 0.01) +
  scale_fill_viridis(name = "parcelled score", option = "C") +
  labs(title =
         'Fig2 (b). Distribution of indicator scores from TSS instrument') +
  labs(x = "Indicator scores", y = "Indicators")+
    theme(
      legend.position="none",
      panel.spacing = unit(0.1, "lines"),
      plot.title = element_text(size = 12),
      panel.grid.major = element_blank(),
      panel.background = element_rect(color = "black",
                                      fill = "white")
    )

a/b
```

Fig2 (a). Distribution of indicator scores from BMI instrument

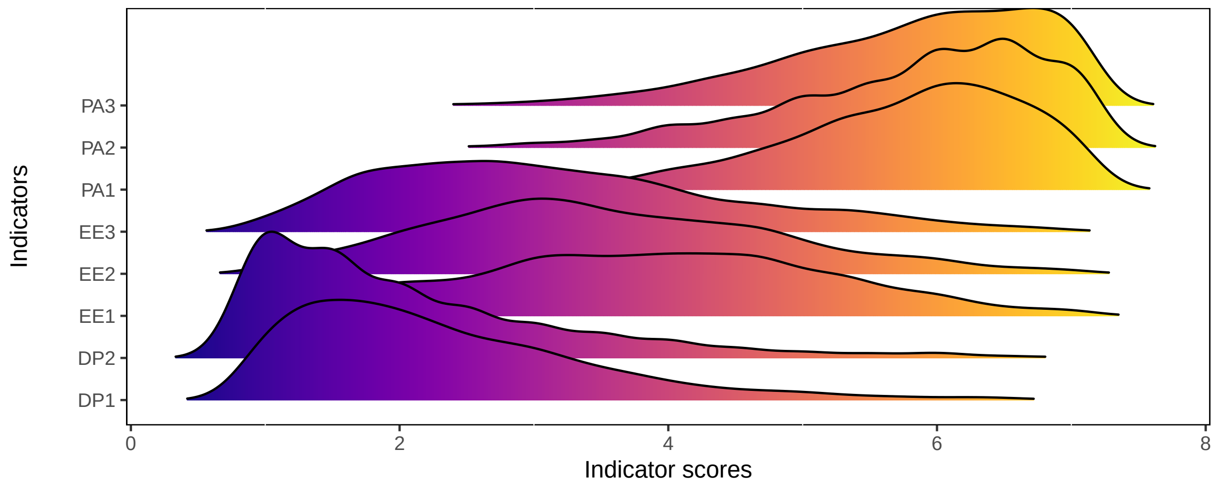
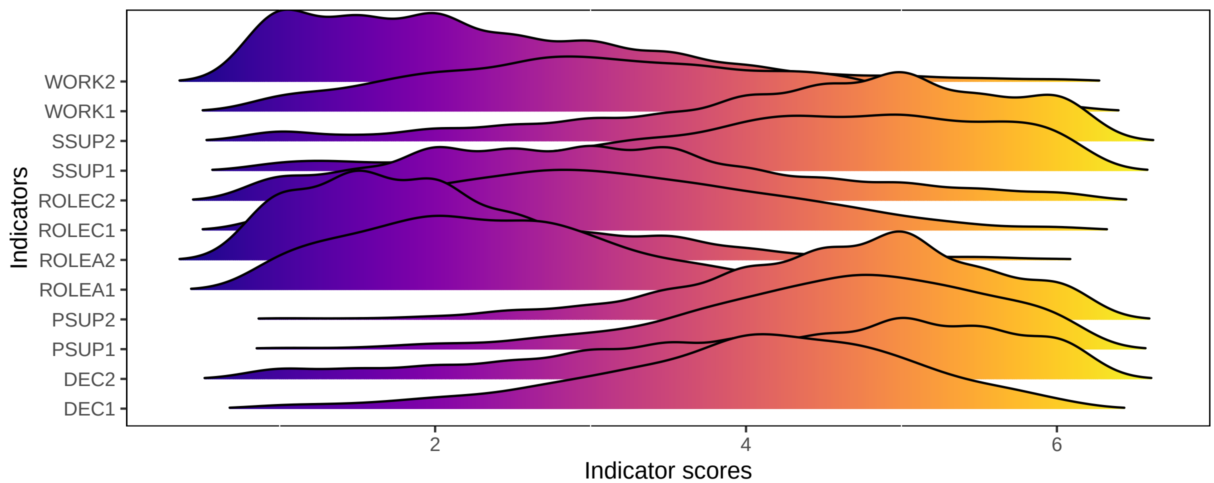


Fig2 (b). Distribution of indicator scores from TSS instrument



Clearly, within each instrument, indicators for same factor tend to show similar distribution features.

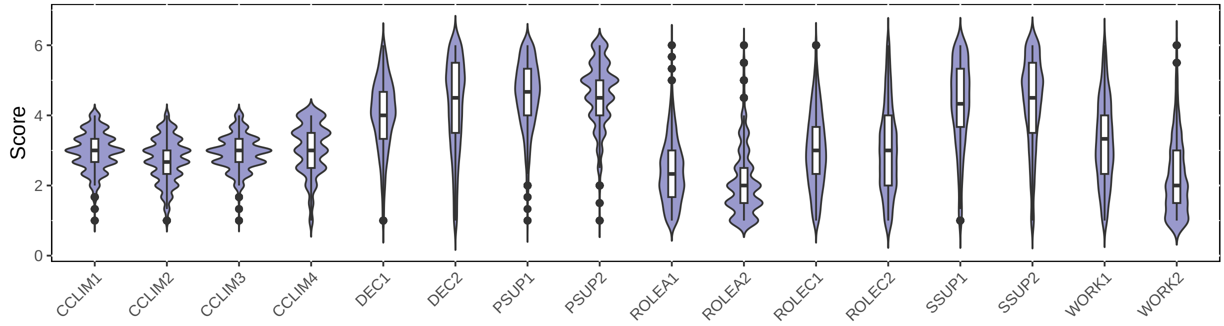
(2) Violin plot

Violin plot also provides information on distribution, plus ideas on out-liers.

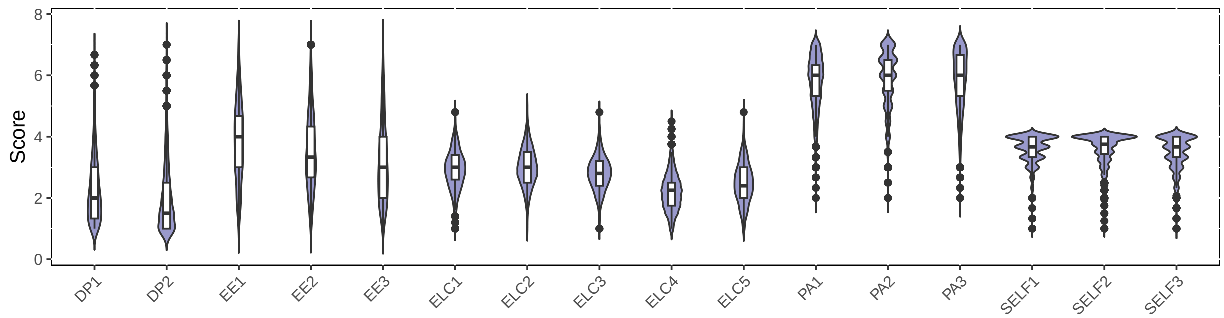
```
a <- mbi |>
  select(1:16) |>
  violin.box(fig.num = "3(a)")

b <- mbi |>
  select(17:32) |>
  violin.box(fig.num = "3(b)")
library(patchwork)
a/b
```

Figureou 3(a) Violin plot of the selected items



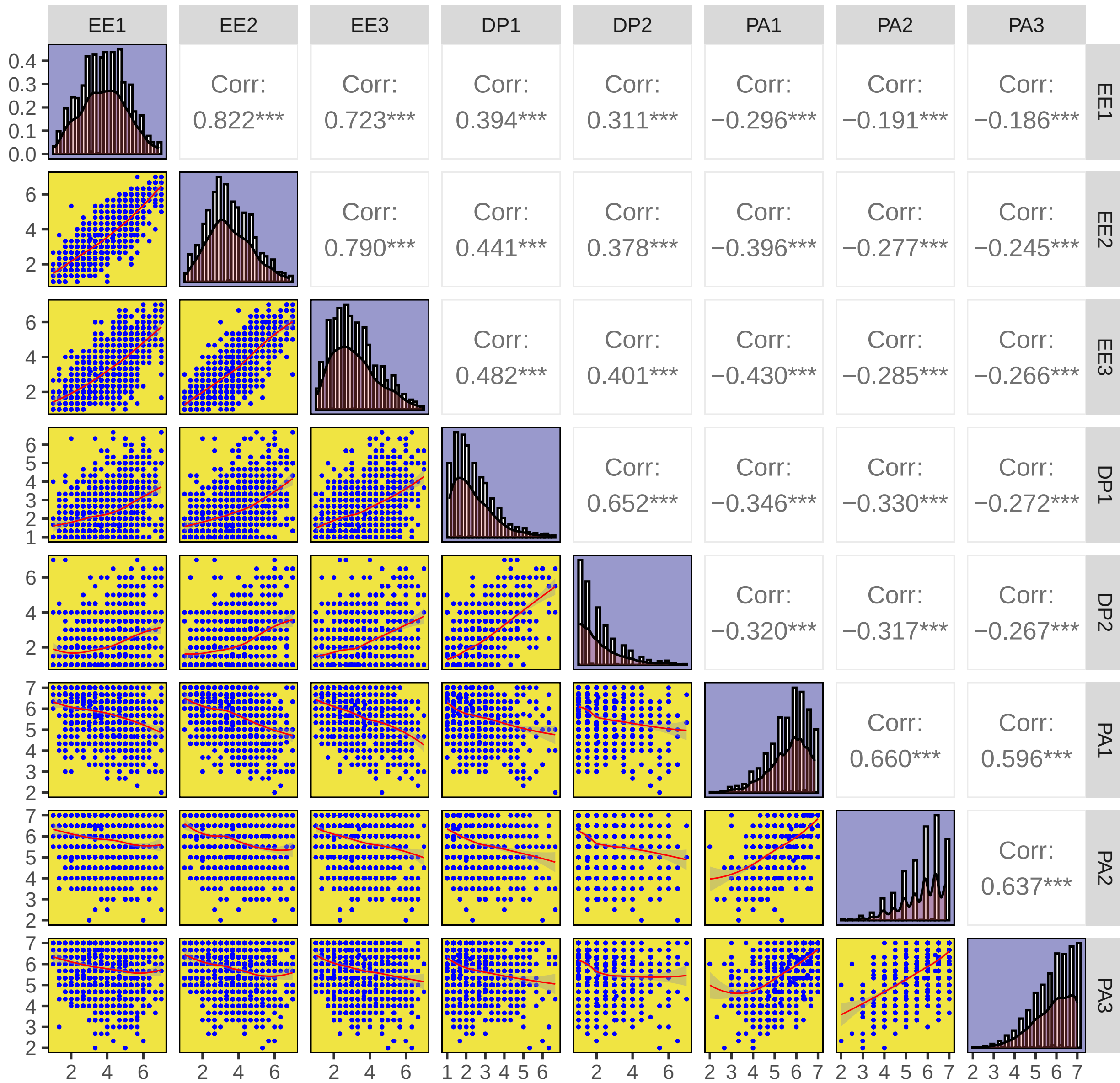
Figureou 3(b) Violin plot of the selected items



(3) Correlation among items

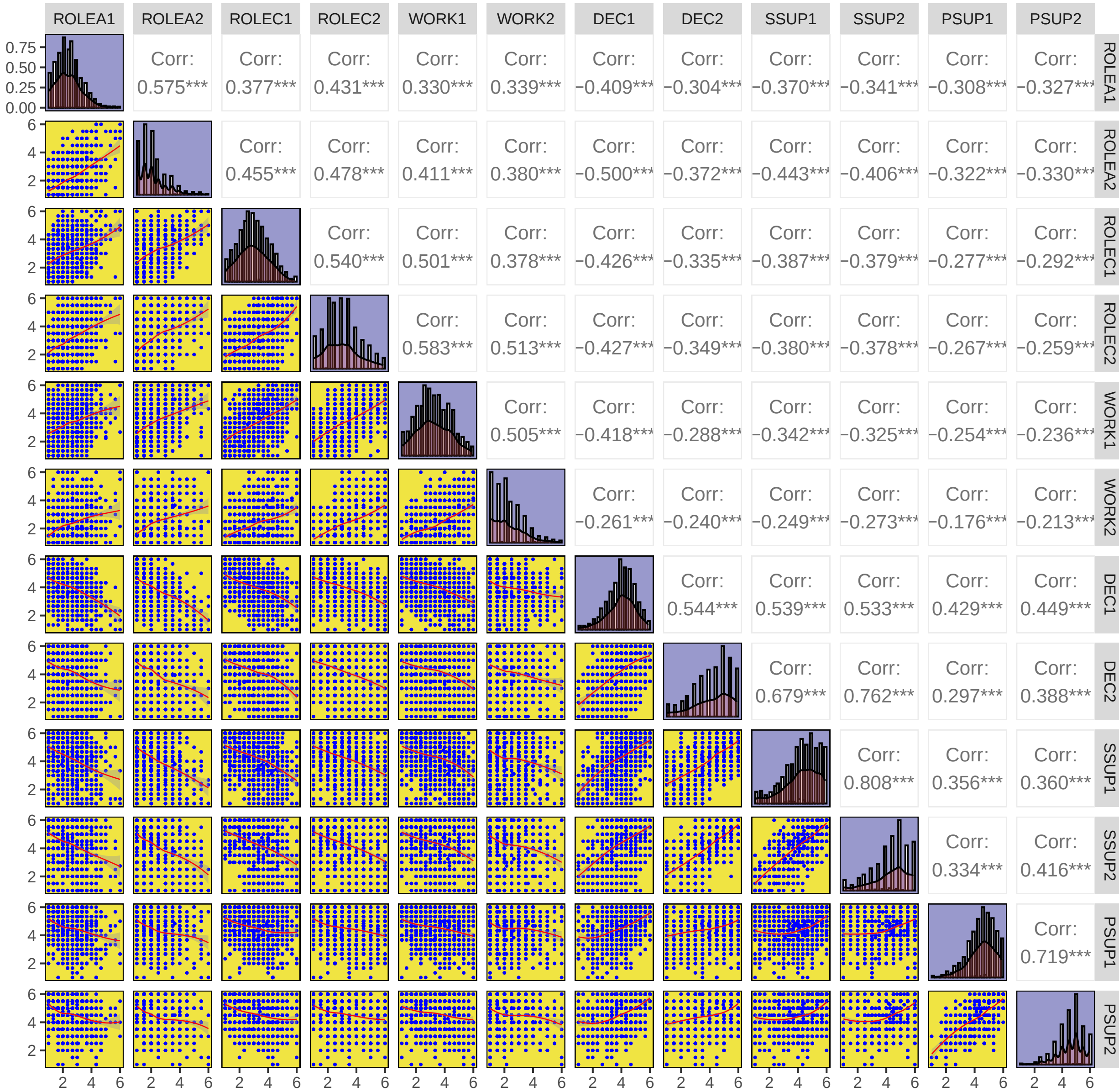
```
#draw it
mbi |>  select(starts_with("EE")|starts_with("DP")|starts_with("PA")) |>
  ggpairs(lower =
            list(continuous = my.fun.smooth), #lower half show points with fitted line
          diag =
            list(continuous = my.fun.density), #diagonal grids show density plots
          title = "Fig4(a). Relationships among parcels of MBI instrument") + #title
  theme (plot.title = element_text(size = 12,  #adjust title visuals
                                   face = "bold"))
```

Fig4(a). Relationships among parcels of MBI instrument




```
mbi |>  select(starts_with("ROL")|starts_with("WOR")|starts_with("DEC")|contains("SUP")) |>
  ggpairs(lower =
            list(continuous = my.fun.smooth), #lower half show points with fitted line
          diag =
            list(continuous = my.fun.density), #diagonal grids show density plots
          title = "Fig4(b). Relationships among parcels of TSS instrument") + #title
  theme (plot.title = element_text(size = 14,  #adjust title visuals
                                   face = "bold"))
```

Fig4(b). Relationships among parcels of TSS instrument



3 Testing the for the validity of causal structure of burnout

3.1 Initial full structural equation model (hypothesized model modified according to CFA)

This full structural equation model was a hypothesized model. I have established causal structure linking several stressor variables considered to contribute to the presence of burnout (Fig. 4). These postulated causal relations linking variables were supported in theory and/or empirical research. I wanted to test the hypothesis that the causal pattern was true. The findings would contribute to the understanding of key determinants of teacher burnout.

Note that *"an important preliminary step in the analysis of full latent variable models is to test first for the validity of the measurement model before making any attempt to evaluate the structural model. Accordingly,*

CFA procedures are used in testing the validity of the indicator variables. Once it is known that the measurement model is operating adequately, one can then have more confidence in findings related to assessment of the hypothesized structural model." The current analysis started at the point where CFA had already been done. As described by Byrne, the analysis produced fit indices showing exceptionally good fit to the data; nonetheless, CFA model for the TSS was re-specified to include two additional parameters, both about allowing for cross-loading terms (DEC2 cross-loads onto Factor 1; DEC2 cross-loads onto Factor 5). These parameters set free were incorporated into the initial hypothesized model (Fig. 4).
Define the initial model

```
library(semPlot)#install.packages("semPlot")

initial_model <- '
# Factors:
 F1ROLA =~ ROLEA1 + ROLEA2 + DEC2
 F2ROLC =~ ROLEC1 + ROLEC2
 F3WORK =~ WORK1 + WORK2
 F4CLIM =~ CCLIM1 + CCLIM2 + CCLIM3 + CCLIM4
 F5DEC =~ DEC1 + DEC2
 F6SSUP =~ SSUP1 + SSUP2 + DEC2
 F7PSUP =~ PSUP1 + PSUP2
 F8SELF =~ SELF1 + SELF2 + SELF3
 F9ELC =~ ELC1 + ELC2 + ELC3 + ELC4 + ELC5
 F10EE =~ EE1 + EE2 + EE3
 F11DP =~ DP1 + DP2
 F12PA =~ PA1 + PA2 + PA3
# Regressions:
 F8SELF ~ F5DEC + F6SSUP + F7PSUP
 F9ELC ~ F5DEC
 F10EE ~ F2ROLC + F3WORK + F4CLIM
 F11DP ~ F2ROLC + F10EE
 F12PA ~ F1ROLA + F8SELF + F9ELC + F10EE + F11DP
'
```

3.1.1 Visualize the initial model

To approximate the visual effect on slides, the coordinates for each nodes were defined on a 60 by 72 matrix.

```
#generate a matrix
m <- matrix(NA, 60, 72)

#define positions of the factors
m[12, 68] <- "F1ROLA"
m[12, 40] <- "F2ROLC"
m[12, 28] <- "F3WORK"
m[12,12] <- "F4CLIM"
m[21,12] <-"F5DEC"
m[40,12] <-"F6SSUP"
m[53,9] <-"F7PSUP"
m[44,24] <-"F8SELF"
m[52,40] <-"F9ELC"
m[37,48] <-"F10EE"
m[26,60] <-"F11DP"
m[48,64] <-"F12PA"
```

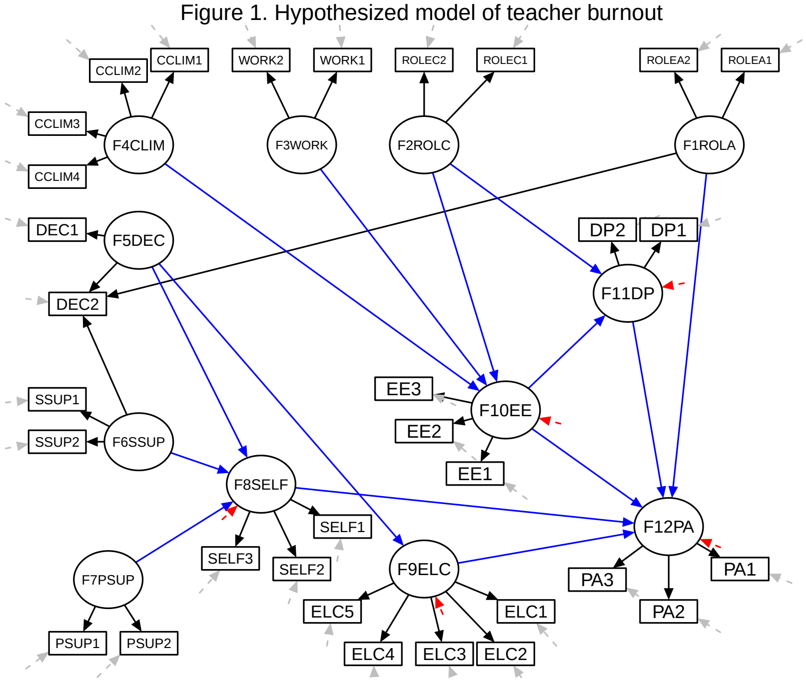
```r
#define the posiitions of the indicators (parcelled items)
m[4, 72] <- "ROLEA1"
m[4, 64] <- "ROLEA2"
m[4, 48] <- "ROLEC1"
m[4, 40] <- "ROLEC2"
m[4, 32] <- "WORK1"
m[4, 24] <- "WORK2"
m[4, 16] <- "CCLIM1"
m[5, 10] <- "CCLIM2"
m[10, 4] <- "CCLIM3"
m[15, 4] <- "CCLIM4"
m[20, 4] <- "DEC1"
m[27, 6] <- "DEC2"
m[36, 4] <- "SSUP1"
m[40, 4] <- "SSUP2"
m[59, 6] <- "PSUP1"
m[59, 13] <- "PSUP2"
m[48, 32] <- "SELF1"
m[52, 28] <- "SELF2"
m[51, 21] <- "SELF3"
m[56, 50] <- "ELC1"
m[60, 48] <- "ELC2"
m[60, 42] <- "ELC3"
m[60, 35] <- "ELC4"
m[56, 31] <- "ELC5"
m[43, 45] <- "EE1"
m[39, 40] <- "EE2"
m[35, 38] <- "EE3"
m[20, 64] <- "DP1"
m[20, 58] <- "DP2"
m[52, 71] <- "PA1"
m[56, 64] <- "PA2"
m[53, 57] <- "PA3"
```

The diagram of the inital model was generated.

```r
semPaths(semPlotModel(initial_model),
             style = "lisrel",
             rotation = 2,
             sizeLat = 6,
             sizeLat2 = 5,
             sizeMan = 5,
             sizeMan2 = 2,
             residScale = 4,
             shapeMan = "rectangle",
             edge.color = c(rep("black",34),
                            rep("blue",14),
                            rep("gray",32),
                            rep("red",5)),
         residuals = T,
         layout = m,
         nCharNodes=0,
         optimizeLatRes = T,
```

```
         exoVar = F)
title(main = list("Figure 1. Hypothesized model of teacher burnout", cex = 1.5, font =1),
     outer = F, line = -2)
title(sub = "Notes: Red arrow indicates factor residuals; gray arrow indicates error residuals;
     blue arrow indicates regression path; black arrow indicates factor loading", line = 0, adj = 0.7)
```

Figure 1. Hypothesized model of teacher burnout



Notes: Red arrow indicates factor residuals; gray arrow indicates error residuals;
blue arrow indicates regression path; black arrow indicates factor loading

3.2 Estimate the SEM model

```
library(lavaan)
model1 <- initial_model # defined above

# Estimate the model with the robust (MLM) estimator:
sem1 <- sem(model1, data = mbi, estimator = "MLM")

# Numerical summary of the model:
sem.summary.mlm.a(sem1, 12, 32, "mlm", "Model fit indices for initial model")
```

Table 2: Model fit indices for initial model

Measure	Value
chi square	1541.844
df	427.000
p value	0.000
CFI	0.940
TLI	0.930
RMSEA	0.043
RMSEA p value	1.000
SRMR	0.053
CSF	1.127

```
#summary(sem1, fit.measures = TRUE, standardized = TRUE)
```

3.3 Comments on the result

Here we see that the rescaled 2 value (i.e., the MLM 2) is 1541.844 with 427 degrees of freedom. The reported chi square scaling factor value for the MLM estimator indicates that if the MLM 2 were multiplied by 1.127, it would approximate the uncorrected ML 2 value (1737.658).

```
sem.summary.c(sem1, 12, 32, "mlm")
```

Table 3: Variances for 12 factor model estimated by mlm

Type	Indicator	B	SE	Z	p-value	Beta
Residual variance	ROLEA1	0.422	0.024	17.386	0.000	0.505
Residual variance	ROLEA2	0.313	0.027	11.556	0.000	0.331
Residual variance	DEC2	0.598	0.033	18.016	0.000	0.339
Residual variance	ROLEC1	0.642	0.029	22.304	0.000	0.530
Residual variance	ROLEC2	0.546	0.037	14.798	0.000	0.359
Residual variance	WORK1	0.646	0.030	21.317	0.000	0.448
Residual variance	WORK2	0.739	0.035	20.903	0.000	0.623
Residual variance	CCLIM1	0.180	0.008	22.588	0.000	0.618
Residual variance	CCLIM2	0.151	0.010	14.874	0.000	0.382
Residual variance	CCLIM3	0.141	0.007	19.356	0.000	0.579
Residual variance	CCLIM4	0.337	0.015	21.843	0.000	0.629
Residual variance	DEC1	0.515	0.027	19.233	0.000	0.505
Residual variance	SSUP1	0.398	0.026	15.205	0.000	0.257
Residual variance	SSUP2	0.200	0.022	8.898	0.000	0.126
Residual variance	PSUP1	0.336	0.027	12.238	0.000	0.361
Residual variance	PSUP2	0.164	0.026	6.287	0.000	0.191
Residual variance	SELF1	0.082	0.005	16.563	0.000	0.415
Residual variance	SELF2	0.065	0.005	13.033	0.000	0.256
Residual variance	SELF3	0.083	0.006	13.021	0.000	0.281
Residual variance	ELC1	0.204	0.010	20.506	0.000	0.524
Residual variance	ELC2	0.259	0.011	23.326	0.000	0.661
Residual variance	ELC3	0.135	0.007	18.174	0.000	0.451
Residual variance	ELC4	0.215	0.010	21.720	0.000	0.587
Residual variance	ELC5	0.187	0.010	18.595	0.000	0.451

Type	Indicator	B	SE	Z	p-value	Beta
Residual variance	EE1	0.413	0.024	17.250	0.000	0.241
Residual variance	EE2	0.225	0.019	11.753	0.000	0.142
Residual variance	EE3	0.449	0.025	17.799	0.000	0.267
Residual variance	DP1	0.278	0.045	6.145	0.000	0.232
Residual variance	DP2	0.622	0.049	12.655	0.000	0.445
Residual variance	PA1	0.270	0.022	12.414	0.000	0.329
Residual variance	PA2	0.319	0.025	12.783	0.000	0.349
Residual variance	PA3	0.407	0.024	17.000	0.000	0.443
Total variance	F1ROLA	0.413	0.033	12.434	0.000	1.000
Total variance	F2ROLC	0.571	0.041	13.844	0.000	1.000
Total variance	F3WORK	0.797	0.047	16.926	0.000	1.000
Total variance	F4CLIM	0.112	0.010	10.732	0.000	1.000
Total variance	F5DEC	0.504	0.038	13.435	0.000	1.000
Total variance	F6SSUP	1.151	0.061	18.983	0.000	1.000
Total variance	F7PSUP	0.595	0.043	13.894	0.000	1.000
Total variance	F8SELF	0.079	0.008	9.693	0.000	0.682
Total variance	F9ELC	0.143	0.012	12.447	0.000	0.774
Total variance	F10EE	-0.432	0.816	-0.530	0.596	-0.332
Total variance	F11DP	0.605	0.053	11.482	0.000	0.658
Total variance	F12PA	0.383	0.025	15.337	0.000	0.695

```r
  options(scipen = 999)
#regression path estimates
sem.parameter <- parameterEstimates(sem1, standardized=TRUE) |>  # obtain estimates
  filter(op == "~") |>   #select "is measured by" rows
  select('DV*'=lhs, #left hand side column
         'IV*'=rhs, #right hand side column
         'B†'=est, #estimates
         'Beta‡'=std.all,#estimates standardized
         SE=se, #standard error
         Z=z, #z statistics
         'p-value'=pvalue #p value
         )

#round the p-value column
sem.parameter$`p-value` <- sem.parameter$`p-value` |>
  round(3)

#add a conditional logic to the p-value column that >0.05 cell shows in red
sem.parameter$`p-value` <- cell_spec(sem.parameter$`p-value`,
                                     color = ifelse(
                                       sem.parameter$`p-value` > 0.05,
                                       "red",
                                       "black")
                                     )
#furhter aesthetics
sem.parameter |>
  kable(digits = 3, #rounded to 3
        #format="latex", #Latex markdown
        booktabs=TRUE, #Latex booktabs
        linesep = "",
```

Table 4: Structural regression path and residual variance estimates.

DV*	IV*	B†	Beta‡	SE	Z	p-value
F8SELF	F5DEC	0.475	0.990	0.054	8.784	0
F8SELF	F6SSUP	-0.155	-0.490	0.026	-5.890	0
F8SELF	F7PSUP	-0.066	-0.150	0.030	-2.223	0.026
F9ELC	F5DEC	-0.288	-0.476	0.023	-12.787	0
F10EE	**F2ROLC**	**-8.707**	**-5.765**	**6.705**	**-1.298**	**0.194**
F10EE	**F3WORK**	**8.082**	**6.325**	**5.647**	**1.431**	**0.152**
F10EE	**F4CLIM**	**-0.930**	**-0.272**	**0.740**	**-1.257**	**0.209**
F11DP	F2ROLC	0.258	0.203	0.054	4.789	0
F11DP	F10EE	0.373	0.444	0.036	10.242	0
F12PA	F1ROLA	-0.071	-0.062	0.048	-1.474	0.14
F12PA	F8SELF	0.472	0.217	0.090	5.245	0
F12PA	F9ELC	-0.208	-0.121	0.052	-3.975	0
F12PA	F10EE	-0.064	-0.098	0.026	-2.416	0.016
F12PA	F11DP	-0.218	-0.281	0.033	-6.556	0

Note:

Rows in bold have insignificant parameters.

* DV: dependent variable; IV: independent variable

† Crude estimates

‡ Standardized estimates

```
        caption= "Structural regression path and residual variance estimates.",
        escape = F) %>% #caption
  kable_styling(latex_options = "striped") %>% #gray every other row
  footnote(general = "Rows in bold have insignificant parameters.",
           symbol = c("DV: dependent variable; IV: independent variable",
                      "Crude estimates",
                      "Standardized estimates")) |>
  row_spec(5, bold = T) |>
  row_spec(6, bold = T) |>
  row_spec(7, bold = T)
```

```
#Variance
type <- "Total variance" #create a new row clarifying types of variance
#write a function for minus calcuation
minus <- function(x,y) {x - y}

variance <- parameterEstimates(sem1, standardized=TRUE)  |>  #obtain estimates
  filter(op == "~~") #select "is correlated with" rows
variance <- variance[minus(sum(32,12), 5):sum(32,12),] #subset needed rows (variance row)
variance <- cbind(type, variance) #add column
sem.tab.variance <- variance %>%select(Type = type, #select and rename variables
                   Factor=rhs, #right hand side column
                   B=est, #estimates
                   SE=se,#standard error
                   Z=z, #z statistics
                   'p-value'=pvalue, #p value
                   Beta=std.all)
```

Table 5:

	Type	Factor	B	SE	Z	p-value	Beta
39	Total variance	F7PSUP	\textcolor{black}{0.595}	0.043	13.894	0.000	1.000
40	Total variance	F8SELF	\textcolor{black}{0.079}	0.008	9.693	0.000	0.682
41	Total variance	F9ELC	\textcolor{black}{0.143}	0.012	12.447	0.000	0.774
42	Total variance	F10EE	\textcolor{black}{-0.432}	0.816	-0.530	0.596	-0.332
43	Total variance	F11DP	\textcolor{black}{0.605}	0.053	11.482	0.000	0.658
44	Total variance	F12PA	\textcolor{black}{0.383}	0.025	15.337	0.000	0.695

```r
#round the p-value column
sem.tab.variance$B <- sem.tab.variance$B |>
  round(3)

#add a conditional logic to the p-value column that >0.05 cell shows in red
sem.tab.variance$B <- cell_spec(sem.tab.variance$B,
                                     color = ifelse(
                                       sem.tab.variance$`p-value` < 0,
                                       "red",
                                       "black")
                                     )

sem.tab.variance |>
  kable(digits = 3, #rounded
        #format="markdown",  #Latex markdown
        booktabs=TRUE, #Latex booktabs
        caption="") |>  #caption
  kable_styling(latex_options = "striped")  |>  # gray every other row
  row_spec(0, background = "#9999CC") # color the variable row
```

3.3.1 Estimate the SEM model:

Note: Here, we will use the sem() function for the estimation, instead of the cfa() function, as we are now working with a full SEM (i.e., CFA + regression paths).

3.4 Exercise 4.2

Proceed **step by step** following the guidelines given in the lecture material, i.e., implement the modifications **one at a time**, testing and studying each step. See (and report) how the fit improves and which parameters are suggested to be modified. Please be careful! There will (always) be a lot of suggestions… Do not list all the MIs (only a few of them are useful!), try to keep your report as concise as possible.

Note: A good way to proceed is to collect the necessary information (i.e., which parameter was modified and how, MI, EPC, chi-square, df, CFI, TLI, scaling correction factor, RMSEA, and SRMR) of each modelling step to a **table** (in a way or another). (Some examples in R code were given in Assignment 3, consult also the reports by other students, if you do not know how to proceed.) **Such tables makes it easy to see how the results of the modelling develop through each step.**

The best practice is to build the tables step by step: In the first table you will have only one row, then two rows, then three rows etc., and in the final version of the table you will have all the steps collected together on k rows, representing the k steps of the modelling process.

3.4.1 Calculating the MLM χ^2 difference tests

Calculate the MLM χ^2 difference tests between the consecutive models of the above steps, as advised in the lecture material (p.14-15). Do those calculations in detail at least once or twice so that you get the idea.

Note: The formulas are simpler than they are in Byrne's book (p.168-169), where both MLM and ML estimations are needed. For more information, see: https://statmodel.com/chidiff.shtml.

For the calculations, you may use R (of course!) or Excel, or some ready-made calculation forms found on the web, such as https://www.thestatisticalmind.com/calculators/SBChiSquareDifferenceTest.htm.

```
# (copy and modify the R codes given earlier)
```

3.5 Exercise 4.3

Draw the graph of the final model and present its fit indices and the essential, standardized parameter estimates. **Pay attention to the factor correlations.**

Compare the initial and final graphs and make sure that you understand the whole modelling process and the final conclusions.

```
# (copy and modify the R codes given earlier)
```