COS-D419 Factor Analysis and Structural Equation Models 2023, Assignment 3

Rong Guang

2023-02-02

CFA & teacher burnout

Exercise 3.1

Specify and test the hypothesis given on the pages 2 and 3 of the lecture material.

Use 1) ML estimator, 2) MLM estimator.

Compare the fit indices and draw conclusions concerning the model fit.

Visualize the model.

Read in the data set:

Start by downloading the data file from Moodle to your Project folder!

1.1 Read in the data set

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

```
library(tidyverse);library(readr)

mbi <- read_csv("ELEMM1.CSV", show_col_types = FALSE)
```

1.2 Write functions

Write some functions to improve the fluency of reporting by minimizing paragraphs frequently interrupted by long codes.

1.2.1 to check unique values

1.2.2 to generate CFA results with improved readability

1.2.2 to generate CFA results with improved readability

1.2.3 to generate a function for correlation matrix with numbers

1.2.4 to generate a function for histogram overlapping with density plot

1.2.4 to generate a function for violin overlapping with box plot

1.2.5 To generate a function describing continuous data set

1.3 Inspect the data

1.3.1 Data structure

Have a quick overview of the data structure

```
dim(mbi);mbi %>% apply(2, function(x)class(x));unique.levels(mbi)
```

```
## [1] 372  22
```

```
##     ITEM1     ITEM2     ITEM3     ITEM4     ITEM5     ITEM6     ITEM7     ITEM8
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##     ITEM9    ITEM10    ITEM11    ITEM12    ITEM13    ITEM14    ITEM15    ITEM16
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##    ITEM17    ITEM18    ITEM19    ITEM20    ITEM21    ITEM22
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
```

```
## [1] "Variable ITEM1 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM2 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM3 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM4 has values of 2,3,4,5,6,7"
## [1] "Variable ITEM5 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM6 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM7 has values of 2,3,4,5,6,7"
## [1] "Variable ITEM8 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM9 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM10 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM11 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM12 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM13 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM14 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM15 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM16 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM17 has values of 2,3,4,5,6,7"
## [1] "Variable ITEM18 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM19 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM20 has values of 1,2,3,4,5,6,7"
## [1] "Variable ITEM21 has values of 2,3,4,5,6,7"
## [1] "Variable ITEM22 has values of 1,2,3,4,5,6,7"
```

The data set contains 22 variables of 372 observations. All of the variable are numeric. Their values appear to follow a consistent patter covering the integer from 1 to 7, except for Item 4, 7, 17 and 21, all of which did not include a value of 1.

1.3.2 Descriptive statistics of measured variables

```
library(finalfit);library(kableExtra)
descriptive(mbi)
```

Table 1: Descriptive statistics for measurements

			Central tendency		Dispersion tendency			
	n	n of NA	Mean	Median	SD	Min	Max	Q1~Q3
ITEM1	372	0	4.4	4.0	1.7	1.0	7.0	3.0 ~ 6.0
ITEM2	372	0	4.9	5.0	1.5	1.0	7.0	4.0 ~ 6.0
ITEM3	372	0	3.5	3.0	1.7	1.0	7.0	2.0 ~ 5.0
ITEM4	372	0	6.3	7.0	1.0	2.0	7.0	6.0 ~ 7.0
ITEM5	372	0	2.2	2.0	1.5	1.0	7.0	1.0 ~ 3.0
ITEM6	372	0	2.7	2.0	1.6	1.0	7.0	2.0 ~ 4.0
ITEM7	372	0	6.3	6.0	0.8	2.0	7.0	6.0 ~ 7.0
ITEM8	372	0	3.0	2.0	1.7	1.0	7.0	2.0 ~ 4.0
ITEM9	372	0	6.0	7.0	1.3	1.0	7.0	6.0 ~ 7.0
ITEM10	372	0	2.2	2.0	1.4	1.0	7.0	1.0 ~ 3.0
ITEM11	372	0	2.2	2.0	1.5	1.0	7.0	1.0 ~ 3.0
ITEM12	372	0	5.7	6.0	1.2	1.0	7.0	5.0 ~ 6.0
ITEM13	372	0	3.6	3.5	1.7	1.0	7.0	2.0 ~ 5.0
ITEM14	372	0	4.0	4.0	1.7	1.0	7.0	3.0 ~ 5.0
ITEM15	372	0	1.8	1.0	1.3	1.0	7.0	1.0 ~ 2.0
ITEM16	372	0	2.5	2.0	1.4	1.0	7.0	1.0 ~ 3.0
ITEM17	372	0	6.4	7.0	0.9	2.0	7.0	6.0 ~ 7.0
ITEM18	372	0	5.7	6.0	1.3	1.0	7.0	5.0 ~ 7.0
ITEM19	372	0	5.9	6.0	1.2	1.0	7.0	6.0 ~ 7.0
ITEM20	372	0	2.2	2.0	1.4	1.0	7.0	1.0 ~ 3.0
ITEM21	372	0	5.9	6.0	1.3	2.0	7.0	5.0 ~ 7.0
ITEM22	372	0	2.6	2.0	1.6	1.0	7.0	1.0 ~ 3.0

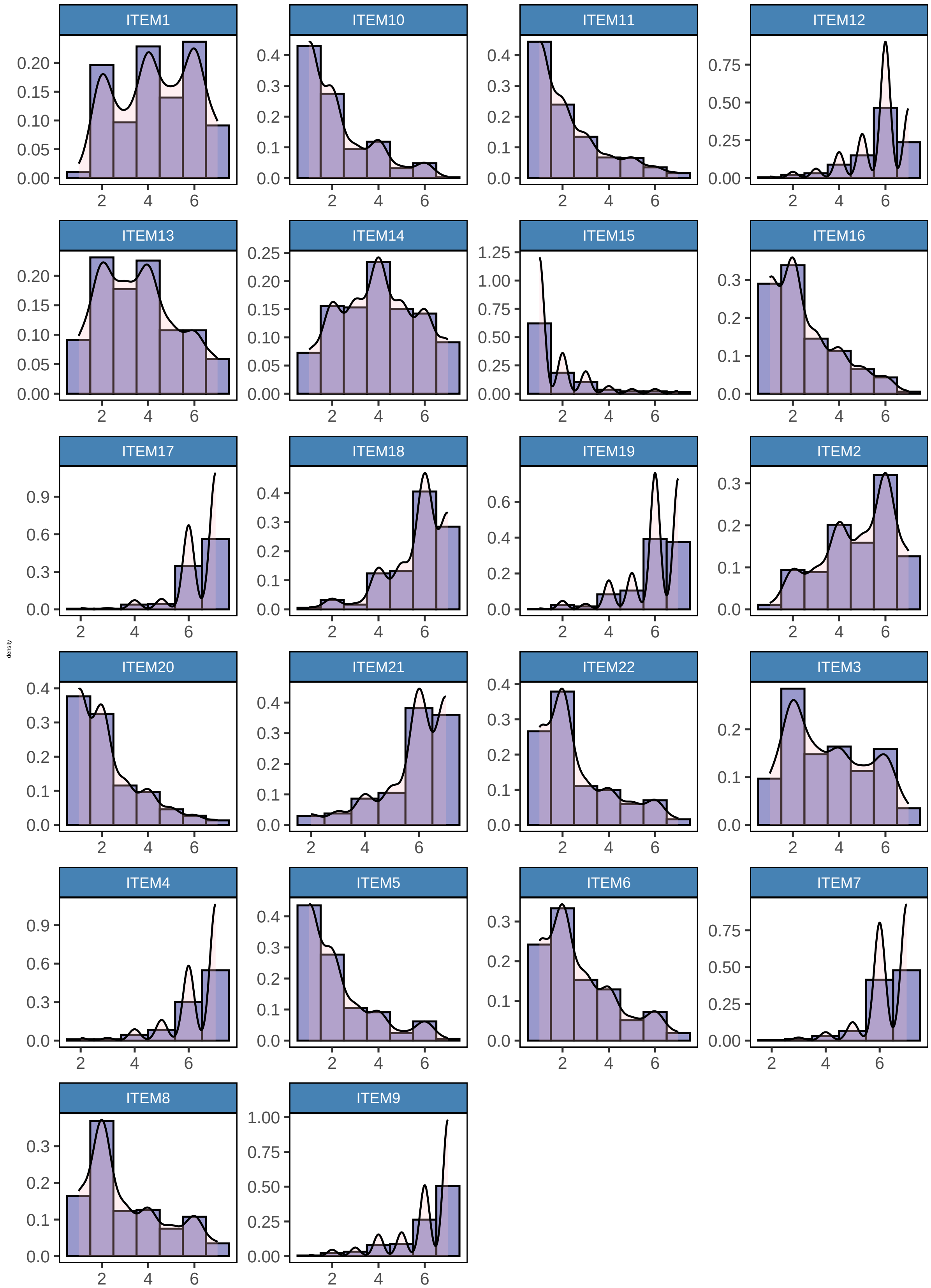
1.3.3 Visualization

(1) Histogram

Distribution of the data was examined via Histogram

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

Figure 1 Distribution of selected items

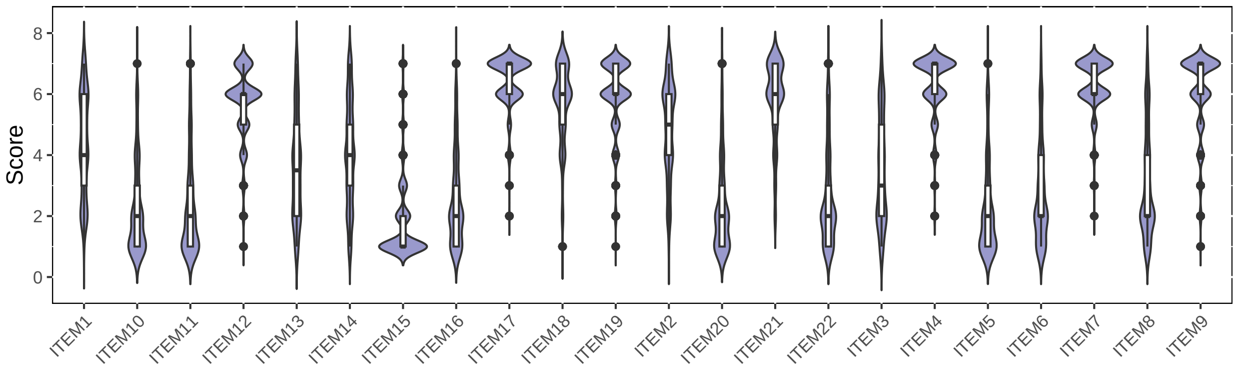


(2) Violin plot

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

```
violin.box(mbi, fig.num = 2)
```

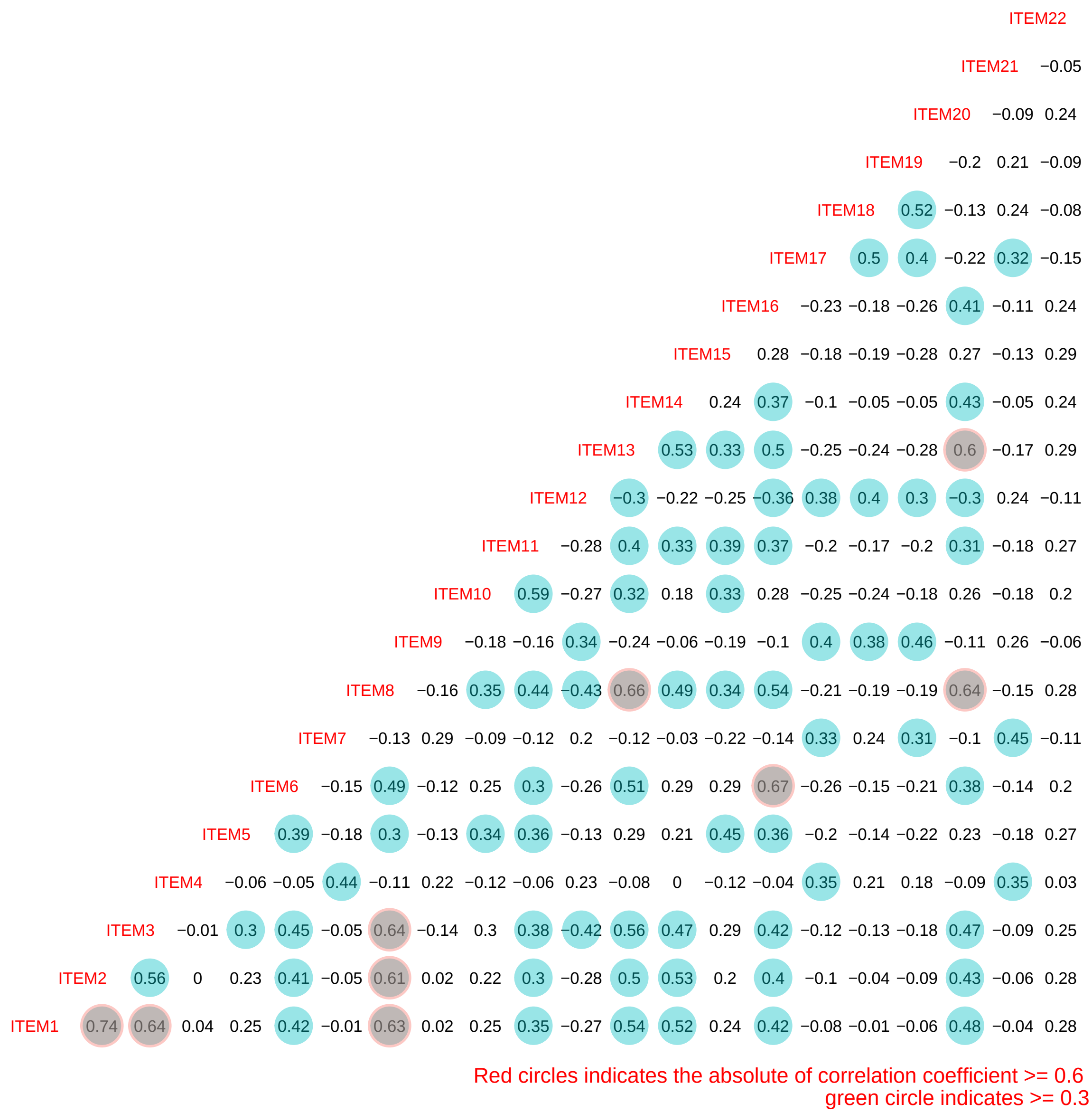
Figure 2 2 Violin plot of the selected items



(3) Correlation among items

```
mymatrix(mbi, fig.num = 3)
```

Figure 3 Pearson correlation matrix of the selected items



All variables had a pearson correlation coefficient >0.3 with at least one other variable, except for ITEM22.

2 Factorial validity

This is to test for the factorial validity of the MBI for elementary teachers using a confirmatory factor analytic approach.

2.1 Define and estimate a CFA model

This report started by estimating the factorial validity of initially postulated model. As was stated by Byrne in the 1991 study:

"The CFA model in the present study hypothesized a priori that: (a) responses to the MBI could be explained by three factors, (b) each item would have a non-zero loading on the burnout factor it was designed to measure, and zero loadings on all other factors, (c) the three factors would be correlated and, (d) the error-uniqueness terms for the item variables would be uncorrelated."

2.1.1 Hypothesis testing via different estimation

(1) ML estimation

```
library(lavaan)
library(kableExtra)
#define model
model1 <- '
# CFA model for the burnout:
# EE: EmotionalExhaustion
# DP: Depersonalization
# PA: PersonalAccomplishment
  EE =~ ITEM1 + ITEM2 + ITEM3 + ITEM6 + ITEM8 +
        ITEM13 + ITEM14 + ITEM16 + ITEM20
  DP =~ ITEM5 + ITEM10 + ITEM11 + ITEM15 +ITEM22
  PA =~ ITEM4 + ITEM7 + ITEM9 + ITEM12 +
        ITEM17 + ITEM18 + ITEM19 + ITEM21
'

# Estimate the model with the default (ML) estimator:
cfa1 <- cfa(model1, data = mbi)

# Numerical summary of the model:
cfa.summary.ml.a(cfa1, 3, 22, "ML");cfa.summary.b(cfa1, 3, 22, "ML");
```

Table 2: Goodness-of-fit and subjective indices of fit for 3 factor model ML

Measure	Value
chi square	695.719
df	206.000
p value	0.000
CFI	0.848
TLI	0.830
RMSEA	0.080
RMSEA p value	0.000
SRMR	0.073

Table 3: Factor Loadings for 3 factor CFA model estimated by ML

Latent Factor	Indicator	B	SE	Z	p-value	Beta
EE	ITEM1	1.000	0.000	NA	NA	0.768
EE	ITEM2	0.887	0.061	14.621	0	0.732
EE	ITEM3	1.021	0.068	15.085	0	0.752
EE	ITEM6	0.764	0.064	12.013	0	0.616

Latent Factor	Indicator	B	SE	Z	p-value	Beta
EE	ITEM8	1.143	0.066	17.299	0	0.845
EE	ITEM13	1.017	0.065	15.544	0	0.772
EE	ITEM14	0.848	0.069	12.251	0	0.627
EE	ITEM16	0.715	0.058	12.410	0	0.634
EE	ITEM20	0.753	0.056	13.410	0	0.679
DP	ITEM5	1.000	0.000	NA	NA	0.565
DP	ITEM10	1.142	0.127	8.986	0	0.663
DP	ITEM11	1.353	0.142	9.511	0	0.743
DP	ITEM15	0.905	0.109	8.318	0	0.586
DP	ITEM22	0.768	0.121	6.361	0	0.408
PA	ITEM4	1.000	0.000	NA	NA	0.440
PA	ITEM7	0.970	0.150	6.482	0	0.507
PA	ITEM9	1.780	0.254	7.007	0	0.594
PA	ITEM12	1.499	0.221	6.769	0	0.552
PA	ITEM17	1.348	0.181	7.463	0	0.695
PA	ITEM18	1.918	0.262	7.329	0	0.662
PA	ITEM19	1.716	0.238	7.205	0	0.634
PA	ITEM21	1.356	0.218	6.219	0	0.471

```
cfa.summary.c(cfa1, 3, 22, "ML");cfa.summary.d(cfa1, 3, 22, "ML")
```

Table 4: Variances for 3 factor model estimated by ML

Type	Indicator	B	SE	Z	p-value	Beta
Residual variance	ITEM1	1.128	0.095	11.861	0	0.410
Residual variance	ITEM2	1.105	0.090	12.214	0	0.464
Residual variance	ITEM3	1.301	0.108	12.031	0	0.434
Residual variance	ITEM6	1.553	0.121	12.888	0	0.621
Residual variance	ITEM8	0.852	0.081	10.553	0	0.286
Residual variance	ITEM13	1.142	0.097	11.821	0	0.404
Residual variance	ITEM14	1.804	0.140	12.844	0	0.607
Residual variance	ITEM16	1.235	0.096	12.812	0	0.598
Residual variance	ITEM20	1.075	0.085	12.585	0	0.539
Residual variance	ITEM5	1.503	0.125	12.026	0	0.681
Residual variance	ITEM10	1.169	0.107	10.901	0	0.560
Residual variance	ITEM11	1.044	0.112	9.330	0	0.447
Residual variance	ITEM15	1.106	0.093	11.838	0	0.657
Residual variance	ITEM22	2.076	0.160	12.958	0	0.833
Residual variance	ITEM4	0.802	0.062	12.901	0	0.806
Residual variance	ITEM7	0.523	0.042	12.572	0	0.743
Residual variance	ITEM9	1.117	0.093	11.952	0	0.647
Residual variance	ITEM12	0.987	0.080	12.287	0	0.695
Residual variance	ITEM17	0.375	0.035	10.739	0	0.517
Residual variance	ITEM18	0.909	0.081	11.224	0	0.562
Residual variance	ITEM19	0.844	0.073	11.557	0	0.598
Residual variance	ITEM21	1.245	0.098	12.764	0	0.778
Total variance	EE	1.625	0.190	8.551	0	1.000
Total variance	DP	0.705	0.132	5.321	0	1.000
Total variance	PA	0.193	0.048	4.047	0	1.000

Table 5: Covariances for 3 factor model estimated by ML

Type	B	SE	Z	p-value	Beta
EE with DP	0.701	0.099	7.061	0	0.655
EE with PA	-0.192	0.042	-4.537	0	-0.343

(2) MLM estimation

```
# Use a robust (MLM) estimator:
cfa2 <- cfa(model1, data = mbi, estimator = "MLM")

# Numerical summary of the model:
cfa.summary.mlm.a(cfa2, 3, 22, "MLM");cfa.summary.b(cfa2, 3, 22, "MLM");
```

Table 6: Goodness-of-fit and subjective indices of fit for 3 factor model MLM

Measure	Value
chi square	567.753
df	206.000
p value	0.000
CFI	0.865
TLI	0.849
RMSEA	0.069
RMSEA p value	0.000
SRMR	0.073
NA	1.225

Table 7: Factor Loadings for 3 factor CFA model estimated by MLM

Latent Factor	Indicator	B	SE	Z	p-value	Beta
EE	ITEM1	1.000	0.000	NA	NA	0.768
EE	ITEM2	0.887	0.040	22.391	0	0.732
EE	ITEM3	1.021	0.053	19.310	0	0.752
EE	ITEM6	0.764	0.070	10.974	0	0.616
EE	ITEM8	1.143	0.059	19.366	0	0.845
EE	ITEM13	1.017	0.062	16.340	0	0.772
EE	ITEM14	0.848	0.058	14.584	0	0.627
EE	ITEM16	0.715	0.066	10.826	0	0.634
EE	ITEM20	0.753	0.061	12.303	0	0.679
DP	ITEM5	1.000	0.000	NA	NA	0.565
DP	ITEM10	1.142	0.152	7.509	0	0.663
DP	ITEM11	1.353	0.162	8.368	0	0.743
DP	ITEM15	0.905	0.123	7.366	0	0.586
DP	ITEM22	0.768	0.122	6.284	0	0.408
PA	ITEM4	1.000	0.000	NA	NA	0.440
PA	ITEM7	0.970	0.128	7.563	0	0.507
PA	ITEM9	1.780	0.322	5.529	0	0.594
PA	ITEM12	1.499	0.241	6.232	0	0.552

Latent Factor	Indicator	B	SE	Z	p-value	Beta
PA	ITEM17	1.348	0.200	6.757	0	0.695
PA	ITEM18	1.918	0.298	6.435	0	0.662
PA	ITEM19	1.716	0.287	5.978	0	0.634
PA	ITEM21	1.356	0.227	5.984	0	0.471

```
cfa.summary.c(cfa2, 3, 22, "MLM");cfa.summary.d(cfa2, 3, 22, "MLM")
```

Table 8: Variances for 3 factor model estimated by MLM

Type	Indicator	B	SE	Z	p-value	Beta
Residual variance	ITEM1	1.128	0.093	12.177	0	0.410
Residual variance	ITEM2	1.105	0.088	12.506	0	0.464
Residual variance	ITEM3	1.301	0.106	12.317	0	0.434
Residual variance	ITEM6	1.553	0.134	11.550	0	0.621
Residual variance	ITEM8	0.852	0.082	10.450	0	0.286
Residual variance	ITEM13	1.142	0.124	9.173	0	0.404
Residual variance	ITEM14	1.804	0.142	12.730	0	0.607
Residual variance	ITEM16	1.235	0.110	11.278	0	0.598
Residual variance	ITEM20	1.075	0.137	7.860	0	0.539
Residual variance	ITEM5	1.503	0.179	8.381	0	0.681
Residual variance	ITEM10	1.169	0.147	7.959	0	0.560
Residual variance	ITEM11	1.044	0.141	7.398	0	0.447
Residual variance	ITEM15	1.106	0.153	7.220	0	0.657
Residual variance	ITEM22	2.076	0.184	11.266	0	0.833
Residual variance	ITEM4	0.802	0.113	7.124	0	0.806
Residual variance	ITEM7	0.523	0.075	7.010	0	0.743
Residual variance	ITEM9	1.117	0.149	7.487	0	0.647
Residual variance	ITEM12	0.987	0.126	7.852	0	0.695
Residual variance	ITEM17	0.375	0.056	6.635	0	0.517
Residual variance	ITEM18	0.909	0.143	6.376	0	0.562
Residual variance	ITEM19	0.844	0.111	7.622	0	0.598
Residual variance	ITEM21	1.245	0.133	9.338	0	0.778
Total variance	EE	1.625	0.148	11.004	0	1.000
Total variance	DP	0.705	0.158	4.452	0	1.000
Total variance	PA	0.193	0.050	3.839	0	1.000

Table 9: Covariances for 3 factor model estimated by MLM

Type	B	SE	Z	p-value	Beta
EE with DP	0.701	0.106	6.608	0	0.655
EE with PA	-0.192	0.040	-4.796	0	-0.343

```
#summary(cfa2, fit.measures = TRUE, standardized = TRUE)
```

2.1.2 Results comparison (ML vs MLM)

```r
options(scipen = 999)
#obtain measures from ML estimation
  cfa.measure.ml <- fitMeasures(cfa1,    #obtain specified measured.
                            c("chisq",
                              "df",
                              "pvalue",
                              "cfi",
                              "tli",
                              "rmsea",
                              "rmsea.pvalue",
                              "srmr"))
  names(cfa.measure.ml) <- c("chi square",
                             "df",
                             "p value",
                             "CFI",
                             "TLI",
                             "RMSEA",
                             "RMSEA p value",
                             "SRMR")

cfa.measure.ml <- cfa.measure.ml %>%
  tibble(Indicator = names(cfa.measure.ml), "By ML" = round(cfa.measure.ml, 3)) %>%
  select(Indicator, "By ML")


#obtain measures from MLM estimation
    cfa.measure.mlm <- fitMeasures(cfa2,    #obtain specified measured.
                            c("chisq.scaled",
                              "df.scaled",
                              "pvalue.scaled",
                              "cfi.scaled",
                              "tli.scaled",
                              "rmsea.scaled",
                              "rmsea.pvalue.scaled",
                              "srmr_bentler",
                              "chisq.scaling.factor"))

  names(cfa.measure.mlm) <- c("chi square",
                              "df",
                              "p value",
                              "CFI",
                              "TLI",
                              "RMSEA",
                              "RMSEA p value",
                              "SRMR",
                              "Scaling correct factor")

#combine the indicators from both estimations
cfa.measure.mlm <- cfa.measure.mlm %>%
  tibble(Indicator = names(cfa.measure.mlm), "By MLM" = round(cfa.measure.mlm, 3)) %>%
  select(Indicator, "By MLM")

#ordered row
```

```r
name.order <- c("chi square",
                "df",
                "p value",
                "CFI",
                "TLI",
                "RMSEA",
                "RMSEA p value",
                "SRMR",
                "Scaling correct factor")

#generate the table
estimation.comparison <- merge(cfa.measure.ml,
                               cfa.measure.mlm,
                               by = "Indicator",
                               all.y = T) %>%
  arrange(factor(Indicator, levels = name.order))


estimation.comparison$'Contrast: MLM-ML' <- estimation.comparison$`By MLM`-
  estimation.comparison$`By ML`



#display the table
estimation.comparison %>%
  kable(booktabs = T,
        linesep = "",
        caption = "Comparing indicators generated by ML vs MLM",
        digits = 3) %>%
  kable_styling() %>%
  row_spec(c(1,4,5,6,8,9), background = "#D3D3D3") %>%
  footnote(symbol = c("Larger value indicates better fit",
                      "Smaller value indicates better fit",
                      ">1 indicates the data violates normality assumption and MLM is a better estimato
  column_spec(3, width = "2.5cm") %>%
  column_spec(0, width = "3cm")
```

According to the table, comparing to the statistics estimated by ML, MLM estimation provided smaller chi square (by 127), higher CFI (by 0.017), higher TLI by 0.19 and smaller RMSEA (by 0.011). The scaling correct factor is 1.225 (>1), indicating some violence of the normality assumption and MLM is more optimal a estimator for the data.

2.1.3

Table 10: Comparing indicators generated by ML vs MLM

Indicator	By ML	By MLM	Contrast: MLM-ML
chi square	695.719	567.753	-127.966
df	206.000	206.000	0.000
p value	0.000	0.000	0.000
CFI	0.848	0.865	0.017
TLI	0.830	0.849	0.019
RMSEA	0.080	0.069	-0.011
RMSEA p value	0.000	0.000	0.000
SRMR	0.073	0.073	0.000
Scaling correct factor	NA	1.225	NA

[*] Larger value indicates better fit

[†] Smaller value indicates better fit

[‡] >1 indicates the data violates normality assumption and MLM is a better estimator