

# Confirmatory factor analysis and Raykov's rho

Questionnaire Validation Workshop

Wan Nor Arifin ([wnarifin@usm.my](mailto:wnarifin@usm.my)), *Universiti Sains Malaysia*  
Website: [wnarifin.github.io](http://wnarifin.github.io)



©Wan Nor Arifin under the Creative Commons Attribution-ShareAlike 4.0 International License.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Preliminaries</b>	<b>2</b>
2.1	Load libraries . . . . .	2
2.2	Load data set . . . . .	2
<b>3</b>	<b>Confirmatory factor analysis</b>	<b>2</b>
3.1	Preliminary steps . . . . .	2
3.2	Step 1 . . . . .	4
3.3	Step 2 . . . . .	4
3.4	Step 3 . . . . .	9
<b>4</b>	<b>Construct reliability</b>	<b>12</b>
<b>5</b>	<b>Path diagram</b>	<b>12</b>
<b>6</b>	<b>Results presentation</b>	<b>13</b>
	<b>References</b>	<b>14</b>

## 1 Introduction

In this hands-on, we are going to further validate our model based on the EFA findings. The same data set, “Attitude\_Statistics v3.sav” will be used.

The evidence of internal structure will be provided by

1. Confirmatory factor analysis
  - Model fit
  - Factor loadings
  - Factor correlations (no multicollinearity)
2. Construct reliability
  - Raykov's rho

## 2 Preliminaries

### 2.1 Load libraries

In addition to `psych` (Revelle, 2017) and `MVN` (Korkmaz, Goksuluk, & Zararsiz, 2016), we are going to use `lavaan` (Rosseel, 2017), `semTools` (Jorgensen, Pornprasertmanit, Miller, Schoemann, & Rosseel, 2016) and `semPlot` (Epskamp & Simon Stuber, 2017) in our analysis. Again, make sure you already installed all of them.

```
library(foreign)
library(psych)
library(MVN)
library(lavaan) # for CFA
library(semTools) # for reliability
library(semPlot) # for path diagram
```

### 2.2 Load data set

We include only good items from **PA1** and **PA2** in `data.cfa` data frame.

```
data = read.spss("Attitude_Statistics v3.sav", F, T) # Shortform
# Include selected items from PA1 & PA2 in 'data.cfa'
data.cfa = data[c("Q4", "Q5", "Q6", "Q7", "Q8", "Q9", "Q10", "Q11")]
dim(data.cfa)
```

```
## [1] 150 8
```

```
names(data.cfa)
```

```
## [1] "Q4" "Q5" "Q6" "Q7" "Q8" "Q9" "Q10" "Q11"
```

```
head(data.cfa)
```

```
##  Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11
## 1  3  4  4  3  3  3  3  4
## 2  3  4  4  4  3  3  3  4
## 3  1  1  1  1  4  4  5  1
## 4  4  3  2  2  2  2  2  3
## 5  2  5  1  4  5  5  3  4
## 6  3  4  4  4  3  4  4  4
```

## 3 Confirmatory factor analysis

### 3.1 Preliminary steps

#### Descriptive statistics

Check minimum/maximum values per item, and screen for any missing values,

```
describe(data.cfa)
```

```
##      vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
## Q4      1 150 2.81 1.17      3   2.77 1.48   1  5   4  0.19  -0.81 0.10
## Q5      2 150 3.31 1.01      3   3.32 1.48   1  5   4 -0.22  -0.48 0.08
## Q6      3 150 3.05 1.09      3   3.05 1.48   1  5   4 -0.04  -0.71 0.09
```

```
## Q7      4 150 2.92 1.19      3    2.92 1.48  1  5    4 -0.04  -1.06 0.10
## Q8      5 150 3.33 1.00      3    3.34 1.48  1  5    4 -0.08  -0.12 0.08
## Q9      6 150 3.44 1.05      3    3.48 1.48  1  5    4 -0.21  -0.32 0.09
## Q10     7 150 3.31 1.10      3    3.36 1.48  1  5    4 -0.22  -0.39 0.09
## Q11     8 150 3.35 0.94      3    3.37 1.48  1  5    4 -0.31  -0.33 0.08
```

Note that all  $n = 150$ , no missing values. `min-max` cover the whole range of response options.

% of response to options per item,

```
response.frequencies(data.cfa)
```

```
##          1      2      3      4      5 miss
## Q4  0.140 0.280 0.30 0.19 0.093  0
## Q5  0.040 0.167 0.35 0.33 0.113  0
## Q6  0.080 0.233 0.33 0.26 0.093  0
## Q7  0.133 0.267 0.23 0.29 0.080  0
## Q8  0.047 0.100 0.48 0.23 0.147  0
## Q9  0.047 0.093 0.42 0.25 0.187  0
## Q10 0.073 0.107 0.42 0.23 0.167  0
## Q11 0.027 0.153 0.35 0.39 0.087  0
```

All response options are used, and there are no missing values.

### Multivariate normality

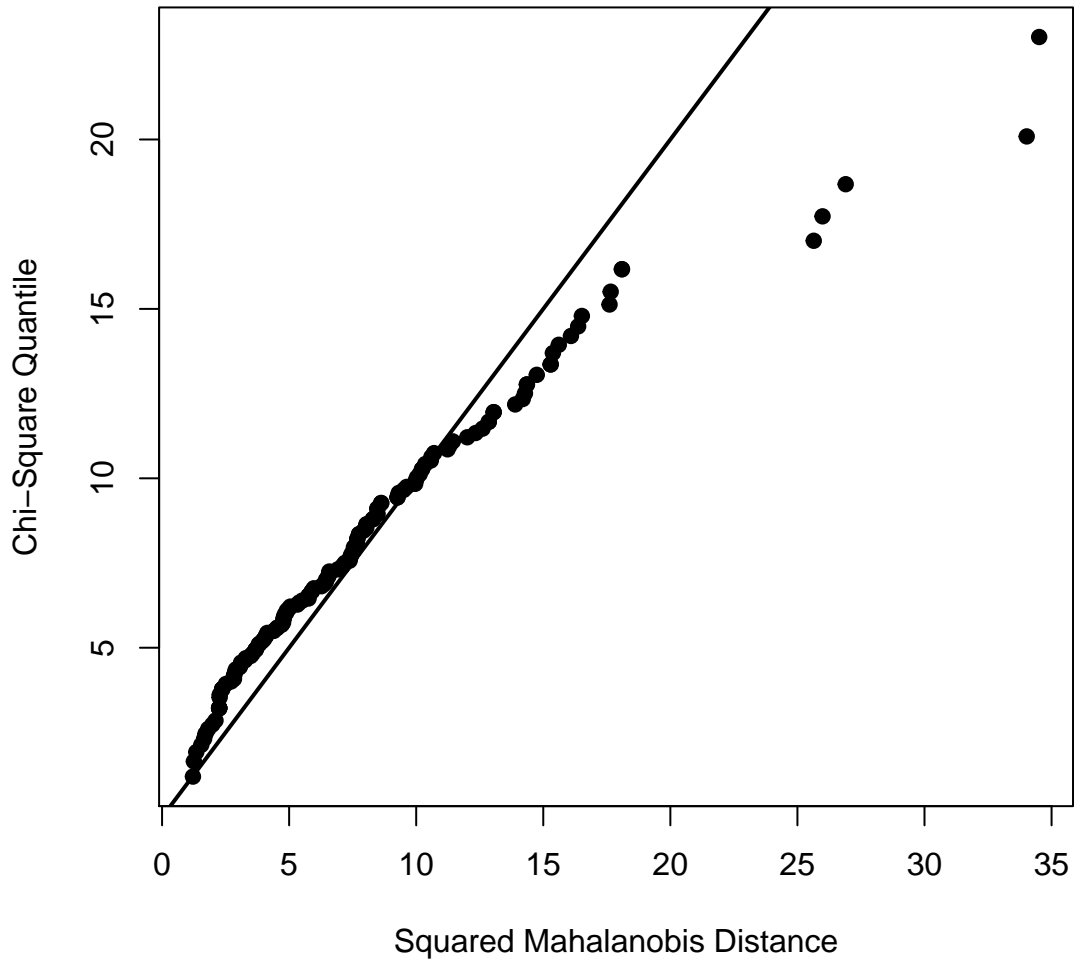
This is done to check the multivariate normality of the data. If the data are normally distributed, we may use maximum likelihood (ML) estimation method for the CFA. In `lavaan`, we have a number of alternative estimation methods (the full list is available at <http://lavaan.ugent.be/tutorial/est.html> or by typing `?lavOptions`). Two common alternatives are:

1. **MLR** (robust ML), suitable for complete and incomplete, non-normal data (Rosseel, 2017).
2. **WLSMV** (robust weighted least squares), suitable for categorical response options (e.g. dichotomous, polynomous, ordinal (Brown, 2015))

```
mardiaTest(data.cfa, qqplot = TRUE)
```

```
##      Mardia's Multivariate Normality Test
## -----
##      data : data.cfa
##
##      g1p          : 11.60132
##      chi.skew     : 290.033
##      p.value.skew : 4.219453e-16
##
##      g2p          : 98.20544
##      z.kurtosis   : 8.813671
##      p.value.kurt : 0
##
##      chi.small.skew : 297.1455
##      p.value.small  : 4.95166e-17
##
##      Result          : Data are not multivariate normal.
## -----
```

## Chi-Square Q-Q Plot



the data are not multivariate normal ( $z\text{-kurtosis} > 5$ ,  $P < 0.05$ ). We will use **MLR** in our analysis.

### 3.2 Step 1

#### Specify the measurement model

Specify the measurement model according to lavaan syntax.

```
model = "  
PA1 =~ Q4 + Q5 + Q6 + Q7 + Q11  
PA2 =~ Q8 + Q9 + Q10  
"
```

`=~`  indicates “measured by”, thus the items represent the factor.

By default, lavaan will correlate PA1 and PA2 (i.e. `PA1 ~~ PA2`), somewhat similar to oblique rotation in EFA.  `~~`  means “correlation”. We will use  `~~`  when we add correlated errors later.

### 3.3 Step 2

#### Fit the model

Here, we fit the specified model. By default, marker indicator variable approach<sup>1</sup> is used in lavaan to scale a factor<sup>2</sup>. We use MLR as the estimation method.

```
cfa.model = cfa(model, data = data.cfa, estimator = "MLR")
# cfa.model = cfa(model, data = data.cfa, std.lv = 1) # factor variance = 1
summary(cfa.model, fit.measures = T, standardized = T)
```

```
## lavaan (0.5-23.1097) converged normally after 21 iterations
##
##   Number of observations                150
##
##   Estimator                            ML      Robust
##   Minimum Function Test Statistic      37.063  27.373
##   Degrees of freedom                    19      19
##   P-value (Chi-square)                  0.008   0.096
##   Scaling correction factor
##     for the Yuan-Bentler correction
##
## Model test baseline model:
##
##   Minimum Function Test Statistic      453.795  325.195
##   Degrees of freedom                    28      28
##   P-value                                0.000   0.000
##
## User model versus baseline model:
##
##   Comparative Fit Index (CFI)           0.958   0.972
##   Tucker-Lewis Index (TLI)            0.937   0.958
##
##   Robust Comparative Fit Index (CFI)    0.973
##   Robust Tucker-Lewis Index (TLI)      0.960
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)         -1566.019 -1566.019
##   Scaling correction factor
##     for the MLR correction
##   Loglikelihood unrestricted model (H1) -1547.487 -1547.487
##   Scaling correction factor
##     for the MLR correction
##
##   Number of free parameters             25      25
##   Akaike (AIC)                          3182.037 3182.037
##   Bayesian (BIC)                        3257.303 3257.303
##   Sample-size adjusted Bayesian (BIC)   3178.183 3178.183
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                0.080   0.054
##   90 Percent Confidence Interval         0.040  0.118   0.000  0.091
##   P-value RMSEA <= 0.05                 0.098   0.397
```

<sup>1</sup>The regression weight of an item from a factor is fixed to 1. Another approach in CFA is to fix the factor variance to 1 (Brown, 2015).

<sup>2</sup>The latent variable (factor) is an unobserved variable, thus it has to be scaled by a method to define its metric/unit of measurement. This is done by fixing either the item regression weight or the factor variance to 1.

```

##
## Robust RMSEA 0.063
## 90 Percent Confidence Interval 0.000 0.112
##
## Standardized Root Mean Square Residual:
##
## SRMR 0.072 0.072
##
## Parameter Estimates:
##
## Information Observed
## Standard Errors Robust.huber.white
##
## Latent Variables:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## PA1 =~
## Q4 1.000 0.952 0.814
## Q5 0.660 0.092 7.218 0.000 0.629 0.624
## Q6 0.810 0.090 9.010 0.000 0.771 0.708
## Q7 0.916 0.086 10.641 0.000 0.872 0.735
## Q11 0.533 0.093 5.719 0.000 0.507 0.544
## PA2 =~
## Q8 1.000 0.653 0.655
## Q9 1.347 0.156 8.654 0.000 0.880 0.844
## Q10 1.436 0.199 7.206 0.000 0.938 0.856
##
## Covariances:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## PA1 ~~
## PA2 0.077 0.075 1.035 0.301 0.124 0.124
##
## Intercepts:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## .Q4 2.813 0.095 29.490 0.000 2.813 2.408
## .Q5 3.313 0.082 40.276 0.000 3.313 3.289
## .Q6 3.053 0.089 34.370 0.000 3.053 2.806
## .Q7 2.920 0.097 30.150 0.000 2.920 2.462
## .Q11 3.353 0.076 44.070 0.000 3.353 3.598
## .Q8 3.327 0.081 40.881 0.000 3.327 3.338
## .Q9 3.440 0.085 40.421 0.000 3.440 3.300
## .Q10 3.313 0.090 37.016 0.000 3.313 3.022
## PA1 0.000 0.000 0.000
## PA2 0.000 0.000 0.000
##
## Variances:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## .Q4 0.460 0.089 5.182 0.000 0.460 0.337
## .Q5 0.620 0.086 7.178 0.000 0.620 0.611
## .Q6 0.590 0.101 5.836 0.000 0.590 0.498
## .Q7 0.647 0.125 5.181 0.000 0.647 0.460
## .Q11 0.611 0.077 7.934 0.000 0.611 0.704
## .Q8 0.567 0.101 5.628 0.000 0.567 0.570
## .Q9 0.312 0.094 3.325 0.001 0.312 0.287
## .Q10 0.321 0.106 3.046 0.002 0.321 0.267

```

##	PA1	0.906	0.137	6.587	0.000	1.000	1.000
##	PA2	0.427	0.106	4.009	0.000	1.000	1.000

## Results

Read the results marked as **Robust**. These represent the results of **MLR**.

To interpret the results, we must look at

1. Overall model fit - by fit indices.
  2. Localized areas of misfit
    - Residuals.
    - Modification indices.
  3. Parameter estimates
    - Factor loadings (**Std.all** column under **Latent Variables** table).
    - Factor correlations (**Std.all** column under **Covariances** table).
1. Fit indices.

The following are a number of selected fit indices and the recommended cut-off values (Brown, 2015; Schreiber, Nora, Stage, Barlow, & King, 2006),

Category	Fit index	Cut-off
Absolute fit	$\chi^2$	$P > 0.05$
	Standardized root mean square (SRMR)	$\leq 0.08$
Parsimony correction	Root mean square error of approximation (RMSEA)	and its 90% CI $\leq 0.08$ , CFit $P > 0.05$
	Comparative fit index (CFI)	$\geq 0.95$
Comparative fit	Tucker-Lewis index (TLI)	

2. Localized areas of misfit (Brown, 2015)

- Residuals

Residuals are the difference between the values in the sample and model-implied variance-covariance matrices.

Standardized residuals (SRs)  $> |2.58|$  indicate the standardized discrepancy between the matrices.

- Modification indices (MIs)

A modification index indicates the expected parameter change if we include a particular specification in the model (i.e. a constrained/fixed parameter is freely estimated, e.g. by correlating between errors of Q1 and Q2).

Specifications with MIs  $> 3.84$  should be investigated.

3. Parameter estimates

- Factor loadings (FLs) (**Std.all** column under **Latent Variables** table).

The guideline for EFA is applicable also to CFA. For example, FLs  $\geq 0.5$  are practically significant. In addition, the  $P$ -values of the FLs must be significant (at  $\alpha = 0.05$ ).

Also look for out-of-range values. FLs should be in range of 0 to 1 (absolute values), thus values  $> 1$  are called *Heywood cases* or *offending estimates* (Brown, 2015)

- Factor correlations (**Std.all** column under **Covariances** table).

Similar to EFA, a factor correlation must be  $< 0.85$ , which indicates that the factors are distinct. A correlation  $> 0.85$  indicates multicollinearity problem. Also look for out-of-range values. Factor correlations should be in range of 0 to 1 (absolute values).

In addition, when a model has Heywood cases, the solution is not acceptable. The variance-covariance matrix (of our data) could be *non-positive definite* i.e. the matrix is not invertible for the analysis.

*In our output:*

Fit indices,

```
## Estimator ML Robust
## Minimum Function Test Statistic 37.063 27.373
## Degrees of freedom 19 19
## P-value (Chi-square) 0.008 0.096
## Scaling correction factor 1.354
## for the Yuan-Bentler correction
## Robust Comparative Fit Index (CFI) 0.973
## Robust Tucker-Lewis Index (TLI) 0.960
##
## Robust RMSEA 0.063
## 90 Percent Confidence Interval 0.000 0.112
##
## SRMR 0.072 0.072
```

The model has good model fit based on all indices, with the exception of the upper 90% CI of robust RMSEA = 0.112. Please note there is no CFit *P*-value for robust RMSEA.

FLs and factor correlation,

```
## Latent Variables:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## PA1 =~
## Q4 1.000 0.952 0.814
## Q5 0.660 0.092 7.218 0.000 0.629 0.624
## Q6 0.810 0.090 9.010 0.000 0.771 0.708
## Q7 0.916 0.086 10.641 0.000 0.872 0.735
## Q11 0.533 0.093 5.719 0.000 0.507 0.544
## PA2 =~
## Q8 1.000 0.653 0.655
## Q9 1.347 0.156 8.654 0.000 0.880 0.844
## Q10 1.436 0.199 7.206 0.000 0.938 0.856
## Covariances:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## PA1 ~~
## PA2 0.077 0.075 1.035 0.301 0.124 0.124
```

Remember to read the results down the `Std.all` column. All FLs > 0.5 and the factor correlation < 0.85. There is no problem with the item quality and the factors are distinct.

Localized areas of misfit,

```
mi = modificationIndices(cfa.model)
subset(mi, mi.scaled > 3.84) # since we are using MLR, look at 'mi.scaled'
```

```
## lhs op rhs mi mi.scaled epc sepc.lv sepc.all sepc.nox
## 30 PA1 =~ Q8 10.264 7.581 0.244 0.232 0.233 0.233
## 34 PA2 =~ Q5 8.359 6.174 0.332 0.217 0.215 0.215
## 47 Q5 ~~ Q11 6.301 4.654 0.144 0.144 0.154 0.154
## 65 Q9 ~~ Q10 10.264 7.581 2.325 2.325 2.035 2.035
```

```
sr = residuals(cfa.model, type = "standardized")
sr
```



```

## $type
## [1] "standardized"
##
## $cov
##      Q4      Q5      Q6      Q7      Q11      Q8      Q9      Q10
## Q4   0.000
## Q5   0.011      NA
## Q6  -0.240 -0.438      NA
## Q7   0.353 -1.221      NA  0.000
## Q11 -0.098  2.893 -1.112 -1.039  0.000
## Q8   2.174  3.364  1.648  1.621  1.179  0.000
## Q9  -1.581  2.365  0.164 -2.069  1.120      NA  0.000
## Q10 -1.150  1.510 -1.231 -1.720  0.343      NA  0.602      NA
##
## $mean
##  Q4  Q5  Q6  Q7  Q11  Q8  Q9  Q10
##   0   0   0   0   0   0   0   0

```

There are four suggested specifications with MIs > 3.84. We may ignore PA1 =~ Q8 and PA2 =~ Q5 based on content, because it is not justifiable to allow these two items specified under other factors. Q9 ~~ Q10 is justifiable, based on the wording “is important”. But Q5 ~~ Q11 is not justifiable.

Q5 has two SRs with Q11 (SR = 2.893) and Q8 (SR = 3.364). So we may focus on Q5.

### 3.4 Step 3

Whenever the model do not fit well, we must revise the model. To do so, we must look for the causes of the poor fit to the data. The causes in CFA could be:

1. Item – the item has low FL (< 0.3), is specified to load on wrong factor or has cross-loading issue.
2. Factor – the factors have multicollinearity problem (correlation > 0.85), or the presence of redundant factors in a model. This can be detected by residuals and MIs.
3. Correlated error (method effect) – some items are similarly worded (e.g. “I like . . .”, “I believe . . .”) or have almost similar meaning/content. This is usually detected by residuals and MIs.
4. Improper solution – the solution with Heywood cases. It could be because the specified model is not supported by the data and the misspecification could be a combination of all the first three causes listed above. A small sample may also lead to improper solution.

The problems might not surface if a proper EFA is done in the first place and the model is theoretically sound.

Model-to-model comparison following revision is done based on:

1.  $\chi^2$  difference
  - for nested<sup>3</sup> models only.
2. AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion)
  - for nested and unnested models.
  - an improvement in the model is shown as a reduction in AIC and BIC values (Brown, 2015). Better model = Smaller AIC/BIC.

#### Model revision

*Revision 1:* Based on MI, Q9 ~~ Q10?

Both from PA2, reasonable by the wording of the questions.

<sup>3</sup>model with same number of items, but with different model specifications e.g. number of factors

```

modell1 = "
PA1 =~ Q4 + Q5 + Q6 + Q7 + Q11
PA2 =~ Q8 + Q9 + Q10
Q9 ~~ Q10
"
cfa.modell1 = cfa(modell1, data = data.cfa, estimator = "MLR")

## Warning in lav_object_post_check(object): lavaan WARNING: some estimated ov variances are
## negative

```

```
summary(cfa.modell1, fit.measures = T, standardized = T)
```

```

## Estimator ML Robust
## Minimum Function Test Statistic 26.487 19.771
## Degrees of freedom 18 18
## P-value (Chi-square) 0.089 0.346
## Scaling correction factor 1.340
## for the Yuan-Bentler correction
## Robust Comparative Fit Index (CFI) 0.994
## Robust Tucker-Lewis Index (TLI) 0.991
##
## Robust RMSEA 0.030
## 90 Percent Confidence Interval 0.000 0.091
##
## SRMR 0.053 0.053

```

The upper 90% CI of RMSEA is smaller, but

```

## Latent Variables:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## PA1 =~
## Q4 1.000 0.950 0.813
## Q5 0.663 0.089 7.470 0.000 0.630 0.626
## Q6 0.811 0.091 8.876 0.000 0.771 0.708
## Q7 0.923 0.089 10.409 0.000 0.877 0.739
## Q11 0.528 0.092 5.729 0.000 0.502 0.538
## PA2 =~
## Q8 1.000 1.517 1.522
## Q9 0.247 0.314 0.786 0.432 0.374 0.359
## Q10 0.267 0.332 0.803 0.422 0.404 0.369
## Covariances:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## .Q9 ~~
## .Q10 0.678 0.198 3.419 0.001 0.678 0.683
## PA1 ~~
## PA2 0.267 0.096 2.787 0.005 0.185 0.185

```

we have a serious Heywood case here! Q8 FL = 1.522. Thus this solution is not acceptable.

*Revision 2:* Remove Q5? Because Q5 has two SRs with other Q8 and Q11. (You may try removing Q11 and Q8 instead)

```

modell2 = "
PA1 =~ Q4 + Q6 + Q7 + Q11
PA2 =~ Q8 + Q9 + Q10
"
cfa.model2 = cfa(modell2, data = data.cfa, estimator = "MLR")

```

```
summary(cfa.model2, fit.measures = T, standardized = T)
```

```
## Estimator ML Robust
## Minimum Function Test Statistic 20.451 14.467
## Degrees of freedom 13 13
## P-value (Chi-square) 0.085 0.342
## Scaling correction factor 1.414
## for the Yuan-Bentler correction
## Robust Comparative Fit Index (CFI) 0.994
## Robust Tucker-Lewis Index (TLI) 0.990
##
## Robust RMSEA 0.033
## 90 Percent Confidence Interval 0.000 0.104
##
## SRMR 0.057 0.057
```

The upper 90% CI of RMSEA has reduced from 0.112 to 0.104.

```
## Latent Variables:
```

```
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## PA1 =~
## Q4 1.000 0.941 0.806
## Q6 0.830 0.099 8.391 0.000 0.781 0.718
## Q7 0.960 0.100 9.597 0.000 0.904 0.762
## Q11 0.504 0.091 5.547 0.000 0.474 0.509
## PA2 =~
## Q8 1.000 0.651 0.653
## Q9 1.351 0.155 8.692 0.000 0.880 0.844
## Q10 1.444 0.202 7.143 0.000 0.940 0.858
## Covariances:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## PA1 ~~
## PA2 0.048 0.068 0.705 0.481 0.078 0.078
```

The FLs and factor correlation are acceptable. No Heywood's case.

```
mi2 = modificationIndices(cfa.model2)
subset(mi2, mi.scaled > 5)
```

```
## lhs op rhs mi mi.scaled epc sepc.lv sepc.all sepc.nox
## 27 PA1 =~ Q8 9.707 6.867 0.241 0.227 0.228 0.228
## 54 Q9 ~~ Q10 9.707 6.867 3.661 3.661 3.204 3.204
```

```
sr2 = residuals(cfa.model2, type = "standardized")
sr2
```

```
## $type
## [1] "standardized"
##
## $cov
## Q4 Q6 Q7 Q11 Q8 Q9 Q10
## Q4 0.000
## Q6 -0.293 0.000
## Q7 -0.131 2.186 NA
## Q11 0.910 -0.587 -0.665 NA
## Q8 2.374 1.861 1.868 1.359 0.000
## Q9 -0.607 0.552 -1.546 1.371 NA 0.000
```

```
## Q10 -0.486 -0.775 -1.221 0.594 NA 0.405 NA
##
## $mean
## Q4 Q6 Q7 Q11 Q8 Q9 Q10
## 0 0 0 0 0 0 0
```

There are no more SRs > 2.56.

So we may stop at **model2**, although the upper 90% CI of RMSEA is still > 0.08, but there is no more localized areas of misfit by SR.

### Model-to-model comparison

Because **model2** is not nested in **model**, we compare mainly by AIC and BIC, and additionally by  $\chi^2$  difference (in our case scaled  $\chi^2$  difference),

```
anova(cfa.model, cfa.model2, method = "satorra.bentler.2010")
```

```
## Scaled Chi Square Difference Test (method = "satorra.bentler.2001")
##
##           Df  AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## cfa.model2 13 2805 2871.3 20.451
## cfa.model  19 3182 3257.3 37.063      13.562      6    0.03493 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Clearly, the AIC and BIC are reduced (**model2** [without Q5] vs **model** [with Q5]). The  $\chi^2$  difference is significant, which indicates an improvement in model fit.

## 4 Construct reliability

### Raykov's rho

Raykov's rho is one of the reliability indices applicable to CFA. It takes into account the correlated errors.

Construct reliability  $\geq 0.7$  (Hair, Black, Babin, & Anderson, 2010) is acceptable.

Look at the omega row in the output,

```
rel.model2 = reliability(cfa.model2)
print(rel.model2, digits = 3)
```

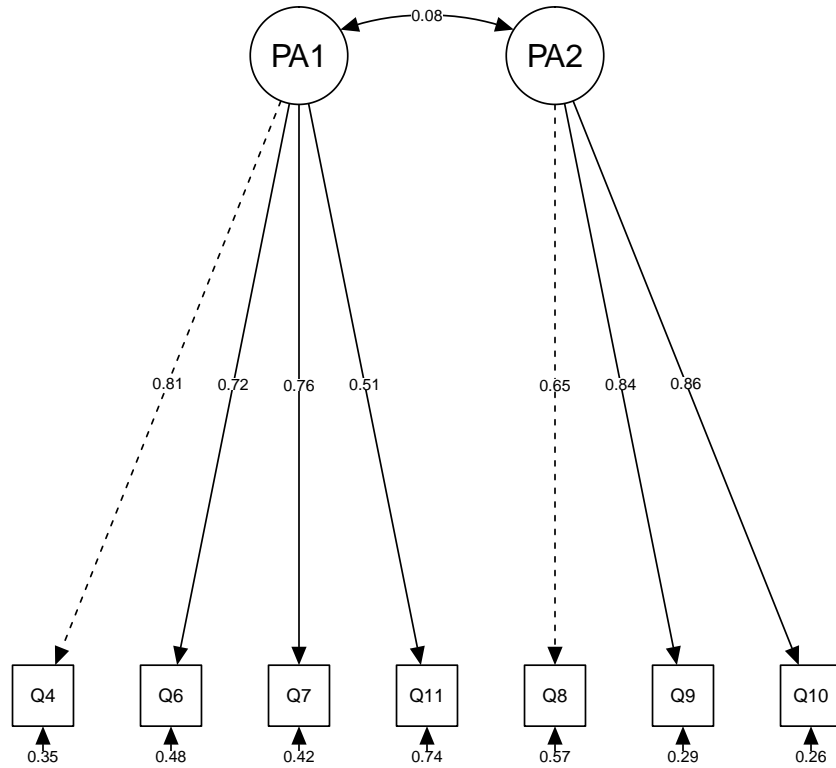
```
##           PA1  PA2 total
## alpha  0.792 0.826 0.723
## omega  0.808 0.836 0.829
## omega2 0.808 0.836 0.829
## omega3 0.809 0.836 0.793
## avevar 0.526 0.634 0.570
```

Raykov's rho (the omega): **PA1** = 0.808, **PA2** = 0.836. Both factors are reliable.

## 5 Path diagram

A CFA model can be nicely presented in the form of path diagram.

```
semPaths(cfa.model2, "path", "std", style = "lisrel", edge.color = "black", intercepts = F)
```



## 6 Results presentation

In the report, you must include a number of important statements and results pertaining to the CFA,

1. The estimation method e.g. ML, MLR, WLSMV etc.
2. The model specification and the theoretical background supporting the model.
3. Details about the selected fit indices, residuals, MIs, FLs and factor correlations and the accepted cut-off values.
4. Detailed comments on the fit and parameters of the tested models. This is usually done in reference to summary tables.
5. Details about the revision process, i.e. item deletion, addition of correlated errors or any other modifications and the effects on the model fit. Also mention the reasons e.g. high SRs, low FLs etc.
6. Summary tables, which outlines the model fit indices, model comparison, FLs, communalities, Raykov's rho, and factor correlations.
7. The path diagram (most of the time, of the final model). This may be requested by some journals.

Fit indices of the models.

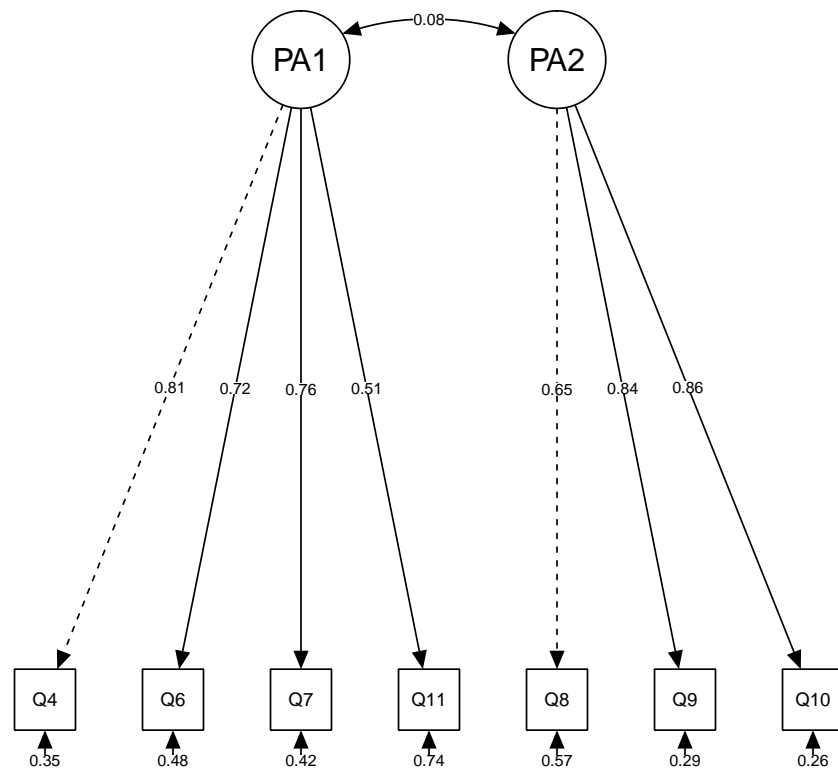
Model	$\chi^2(df)$	$P$	$\chi^2_{diff}(df)$	$P$	SRMR	RMSEA	90% CI	CFI	TLI	AIC	BIC
Model 1	27.4(19)	0.096	-		0.072	0.063	0.000, 0.112	0.973	0.960	3182	3257
Model 2	14.5(13)	0.342	13.6 (6)	0.035	0.057	0.033	0.000, 0.104	0.994	0.990	2805	2871

Factor loadings and reliability of Model 2.

Factor	Item	Factor loading	Raykov's rho
Affinity	Q4	0.806	0.808
	Q6	0.718	
	Q7	0.762	
	Q11	0.509	
Importance	Q8	0.653	0.836
	Q9	0.844	
	Q10	0.858	

Factor correlation: Affinity  $\leftrightarrow$  Importance  $r = 0.078$ .

The path diagram of Model 2.



## References

- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Epskamp, S., & Simon Stuber. (2017). *SemPlot: Path diagrams and visual analysis of various sem packages' output*. Retrieved from <https://CRAN.R-project.org/package=semPlot>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. New Jersey: Prentice Hall.
- Jorgensen, T. D., Pornprasertmanit, S., Miller, P., Schoemann, A., & Rosseel, Y. (2016). *SemTools: Useful tools for structural equation modeling*. Retrieved from <https://CRAN.R-project.org/package=semTools>
- Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2016). *MVN: Multivariate normality tests*. Retrieved from <https://CRAN.R-project.org/package=MVN>
- Revelle, W. (2017). *Psych: Procedures for psychological, psychometric, and personality research*. Retrieved

from <https://CRAN.R-project.org/package=psych>

Rosseel, Y. (2017). *Lavaan: Latent variable analysis*. Retrieved from <https://CRAN.R-project.org/package=lavaan>

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323–338.