Course: Multivariate statistics (AUT23)

Chapter 9: Confirmatory factor analysis

10.18.1 Exercise 1: Big-5

To illustrate CFA, let us use the same International Personality Item Pool data available in the psych package to confirm the big 5 personality structure.

Start by defining the model using the lavaan syntax and extract the parameters.

- Show the code

Latent Factor	Indicator		B	SE	Z	p-value	Beta
Neuroticism	N1	1.000	0.000	NA	NA	0.825	
Neuroticism	N2	0.947	0.024	39.899	0	0.803	
Neuroticism	N3	0.884	0.025	35.919	0	0.721	
Neuroticism	N4	0.692	0.025	27.753	0	0.573	
Neuroticism	N5	0.628	0.026	24.027	0	0.503	
Conscientiousness	C1	1.000	0.000	NA	NA	0.551	
Conscientiousness	C2	1.148	0.057	20.152	0	0.592	
Conscientiousness	C3	1.036	0.054	19.172	0	0.546	
Conscientiousness	C4	-1.421	0.065	-21.924	0	-0.702	
Conscientiousness	C5	-1.489	0.072	-20.694	0	-0.620	

Also get the fit statistics of the model. Is it a good model?

- Show the code

Lastly, make a visualization of the output using the semPaths() function.

- Show the code

10.18.2 Exercise 2: Calculate the degree of freedom for one-factor CFA with more than 3 items

The benefit of performing a one-factor CFA with more than three items is that:

- your model is automatically identified (there will be more than 6 free parameters)
- your model will not be saturated (there will be degrees of freedom left over to assess model fit).

Imagine that we have specified the following model in lavaan with 8 items.

- Show the code

From this model, explain how to obtain 20 degrees of freedom from the 8-item one factor CFA by first calculating the number of free parameters and comparing that to the number of known values.

- Solution

10.18.3 Exercise 3: Calculate the degree of freedom for two-factor CFA with more than 3 items

Imagine that we have specified the following model in lavaan with two factors:

- Show the code

From this model, explain how to obtain 15 degrees of freedom from the 7-item two-factor CFA by first calculating the number of free parameters and comparing that to the number of known values. We also make the assumption of uncorrelated (orthogonal) factors.

- Solution

Now, we make the assumption of correlated (oblique) factors, which gives us the following model:

- Show the code

Explain how to obtain 13 degrees of freedom from the 7-item two-factor CFA by first calculating the number of free parameters and comparing that to the number of known values.

- Solution

Chapter 9: Confirmatory factor analysis (answers)

10.18.1 Exercise 1: Big-5

To illustrate CFA, let us use the same International Personality Item Pool data available in the psych package to confirm the big 5 personality structure.

Start by defining the model using the lavaan syntax and extract the parameters.

➢ Show the code

```
model_measurement <- "
  Neuroticism =~ N1 + N2 + N3 + N4 + N5
  Conscientiousness =~ C1 + C2 + C3 + C4 + C5
  Extraversion =~ E1 + E2 + E3 + E4 + E5
  Agreeableness =~ A1 + A2 + A3 + A4 + A5
  Opennness =~ O1 + O2 + O3 + O4 + O5
"
fit_measurement <- lavaan::sem(model_measurement, data = data)
# summary(fit_measurement, fit.measures = TRUE, standardized = TRUE)
coef <- lavaan::parameterestimates(fit_measurement)
# output: only 10 first rows
lavaan::parameterEstimates(fit_measurement, standardized=TRUE) |>
  dplyr::select('Latent Factor'=lhs,
        Indicator=rhs,
        B=est,
        SE=se,
        Z=z,
        'p-value'=pvalue,
        Beta=std.all) |>
  dplyr::slice_head(n = 10) |> # select only the 10 first rows
  knitr::kable(digits = 3, booktabs=TRUE) |>
  kableExtra::kable_styling(font_size = 11)
```

Latent Factor	Indicator		B	SE	Z	p-value	Beta	
Neuroticism	N1	1.000	0.000	NA	NA	0.825		
Neuroticism	N2	0.947	0.024	39.899	0	0.803		
Neuroticism	N3	0.884	0.025	35.919	0	0.721		
Neuroticism	N4	0.692	0.025	27.753	0	0.573		
Neuroticism	N5	0.628	0.026	24.027	0	0.503		
Conscientiousness	C1	1.000	0.000	NA	NA	0.551		
Conscientiousness	C2	1.148	0.057	20.152	0	0.592		
Conscientiousness	C3	1.036	0.054	19.172	0	0.546		
Conscientiousness	C4	-1.421	0.065	-21.924		0	-0.702	
Conscientiousness	C5	-1.489	0.072	-20.694		0	-0.620	

Also get the fit statistics of the model. Is it a good model?

- Show the code

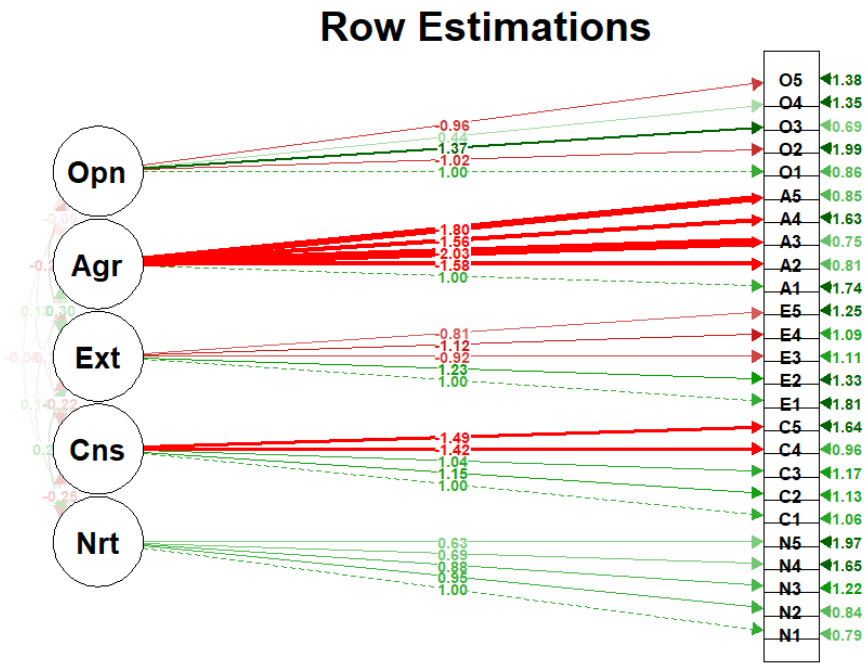
```
fits <- lavaan::fitMeasures(fit_measurement)
data.frame(fits = round(fits[c("ntotal","df",
                               "chisq","pvalue",
                               "rmsea","rmsea.pvalue","srmr")], 2))
##              fits
## ntotal     2436.00
## df          265.00
## chisq      4165.47
## pvalue        0.00
## rmsea         0.08
## rmsea.pvalue  0.00
## srmr          0.08
```

Lastly, make a visualization of the output using the semPaths() function.

- Show the code

```
semPlot::semPaths(fit_measurement, what = "est", rotation = 2,
```

```
        style = "lisrel", font = 2)
title("Row Estimations")
```



10.18.2 Exercise 2: Calculate the degree of freedom for one-factor CFA with more than 3 items

The benefit of performing a one-factor CFA with more than three items is that:

- your model is automatically identified (there will be more than 6 free parameters)
- your model will not be saturated (there will be degrees of freedom left over to assess model fit).

Imagine that we have specified the following model in lavaan with 8 items.

➢ Show the code

```
#one factor eight items, variance std
mod <- 'f =~ q01 + q02 + q03 + q04 + q05 + q06 + q07 + q08'
onefac8items <- cfa(mod, data=dat,std.lv=TRUE)
summary(onefac8items, fit.measures=TRUE, standardized=TRUE)
```

From this model, explain how to obtain 20 degrees of freedom from the 8-item one factor CFA by first calculating the number of free parameters and comparing that to the number of known values.

➢ Solution

The number of elements in the variance-covariance matrix is:

$$\frac{p*(p+1)}{2} = \frac{8*(8+1)}{2} = 36$$

We also have 8 loadings (λ_i), 8 residual variances (θ_i) and 1 variance of the factor (ψ_i). Thus, in total we have 17 unique parameters.

This gives us 17-1=16 free parameters, where we have fixed 1 parameter (using the variance standardization method).

Therefore, the degrees of freedom is 36-16=20. This suggests that we have an over-identified model (degrees of freedom above 0).

10.18.3 Exercise 3: Calculate the degree of freedom for two-factor CFA with more than 3 items

Imagine that we have specified the following model in lavaan with two factors:

- Show the code

```
#one factor eight items, variance std
mod <- '
 f1 =~ q01 + q03 + q04 + q05 + q08
 f2 =~ a*q06 + a*q07
 ## equate the 2 items on the same factors
 ## while setting the factor variance at 1
 f1 ~~ 0*f2' ## orthogonal factors
twofac7items <- cfa(mod, data=dat,std.lv=TRUE)
summary(twofac7items, fit.measures=TRUE, standardized=TRUE)
```

From this model, explain how to obtain 15 degrees of freedom from the 7-item two-factor CFA by first calculating the number of free parameters and comparing that to the number of known values. We also make the assumption of uncorrelated (orthogonal) factors.

- Solution

The number of elements in the variance-covariance matrix is:

$$\frac{p*(p+1)}{2} = \frac{7*(7+1)}{2} = 28$$

We also have 7 loadings (λ_i), 7 residual variances (θ_i) and 2 variance of the factors (ψ_i). Thus, in total we have 16 unique parameters. But we make the assumption of uncorrelated factors, thus we have 14 free parameters.

Therefore, the degrees of freedom is 28-14=14. However, as we constrained the loadings of q06 and q07 to be equal, it frees up a parameter and thus we have 14+1=15 degrees of freedom.

Now, we make the assumption of correlated (oblique) factors, which gives us the following model:

- Show the code

```
#one factor eight items, variance std
mod <- '
 f1 =~ q01 + q03 + q04 + q05 + q08
 f2 =~ q06 + q07'
twofac7items <- cfa(mod, data=dat,std.lv=TRUE)
summary(twofac7items, fit.measures=TRUE, standardized=TRUE)
```

Explain how to obtain 13 degrees of freedom from the 7-item two-factor CFA by first calculating the number of free parameters and comparing that to the number of known values.

- Solution

The number of elements in the variance-covariance matrix is:

$$\frac{p*(p+1)}{2} = \frac{7*(7+1)}{2} = 28$$

We also have 7 loadings (λ_i), 7 residual variances (θ_i) and 1 covariance of the factors (ψ_{21}). Thus, in total we have 15 unique parameters. Therefore, the degrees of freedom is 28-15=13.