EFA vs. CFA revisited

FACTOR ANALYSIS IN R

Jennifer Brussow
Psychometrician

Review of the differences between EFA & CFA

EFA:

- Estimates all possible variable/factor relationships
- Looking for patterns in the data
- Use when you don't have a well-developed theory

CFA:

- Only specified variable/factor relationships
- Testing a theory that you know in advance
- This is the right thing to publish!

```r
# View the first five rows of the EFA loadings
EFA_model$loadings[1:5,]
```

```
           MR2         MR1          MR3        MR5         MR4
A1  0.24282081 -0.15387946 0.0780303740 -0.3897470 -0.08461786
A2 -0.02320759  0.01798410 0.0679900414  0.6584172 -0.01095097
A3 -0.05917275 -0.12693134 0.0238309309  0.6154942  0.05036830
A4 -0.03852599 -0.08709392 0.1936172346  0.4005924 -0.17361760
A5 -0.13355262 -0.23429925 0.0001429341  0.5075002  0.07523716
```

```r
# View the first five loadings from the CFA estimated from the EFA results
summary(EFA_CFA)$coeff[1:5,]
```

```
       Estimate  Std Error   z value      Pr(>|z|)
F4A1 -0.5038817 0.04497739 -11.20300  3.941591e-29 A1 <--- MR5
F4A2  0.8207622 0.03465055  23.68684 4.927422e-124 A2 <--- MR5
F4A3  1.0360812 0.03700471  27.99863 1.688392e-172 A3 <--- MR5
F4A4  0.8264718 0.04471746  18.48208  2.878650e-76 A4 <--- MR5
F4A5  0.9012645 0.03688629  24.43359 7.520155e-132 A5 <--- MR5
```

Comparing factor loadings

```
# View the first five loadings from the CFA estimated from the EFA results
summary(EFA_CFA)$coeff[1:5, ]
```

```
       Estimate  Std Error    z value       Pr(>|z|)
F4A1 -0.5038817 0.04497739 -11.20300  3.941591e-29 A1 <--- MR5
```

```
# View the first five rows of the EFA loadings
EFA_model$loadings[1:5, ]
```

```
           MR2         MR1          MR3         MR5         MR4
A1                                -0.3897470
```

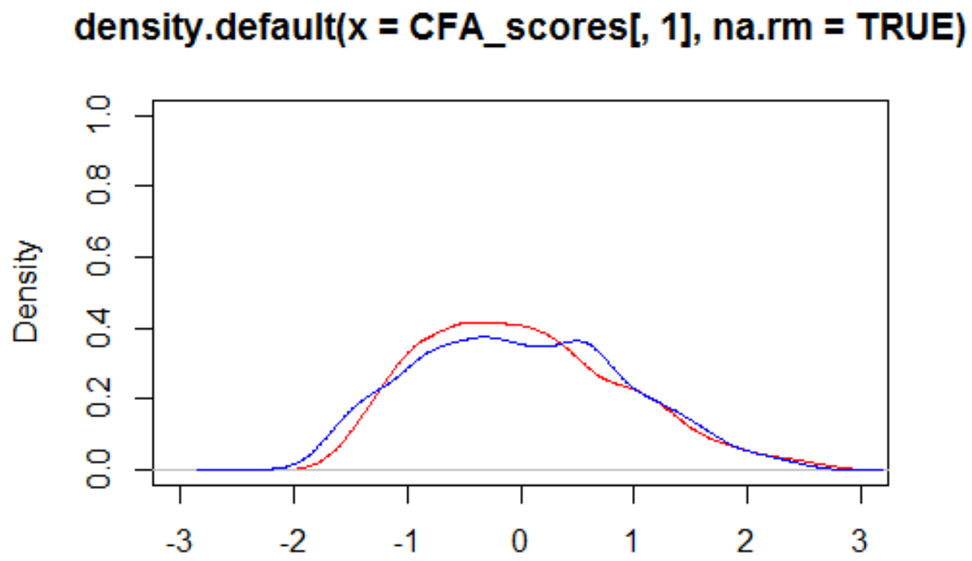
Differences in factor scores

```
# Extracting factor scores from the EFA model
EFA_scores <- EFA_model$scores
```

```
# Calculate factor scores for the EFA dataset
CFA_scores <- fscores(EFA_CFA, data = bfi_EFA)
```

Differences in factor scores, visualized

```
plot(density(EFA_scores[,1], na.rm = TRUE),
     xlim = c(-3, 3), ylim = c(0, 1), col = "blue")
lines(density(CFA_scores[,1], na.rm = TRUE),
     xlim = c(-3, 3), ylim = c(0, 1), col = "red")
```



Let's practice!

FACTOR ANALYSIS IN R

Adding loadings to improve model fit

FACTOR ANALYSIS IN R

Jennifer Brussow
Psychometrician

When to make adjustments

Remember:

- EFAs estimate all item/factor loadings
- CFAs only estimate specified loadings
- Poor model fit could be due to excluded loadings

Adding loadings to the syntax

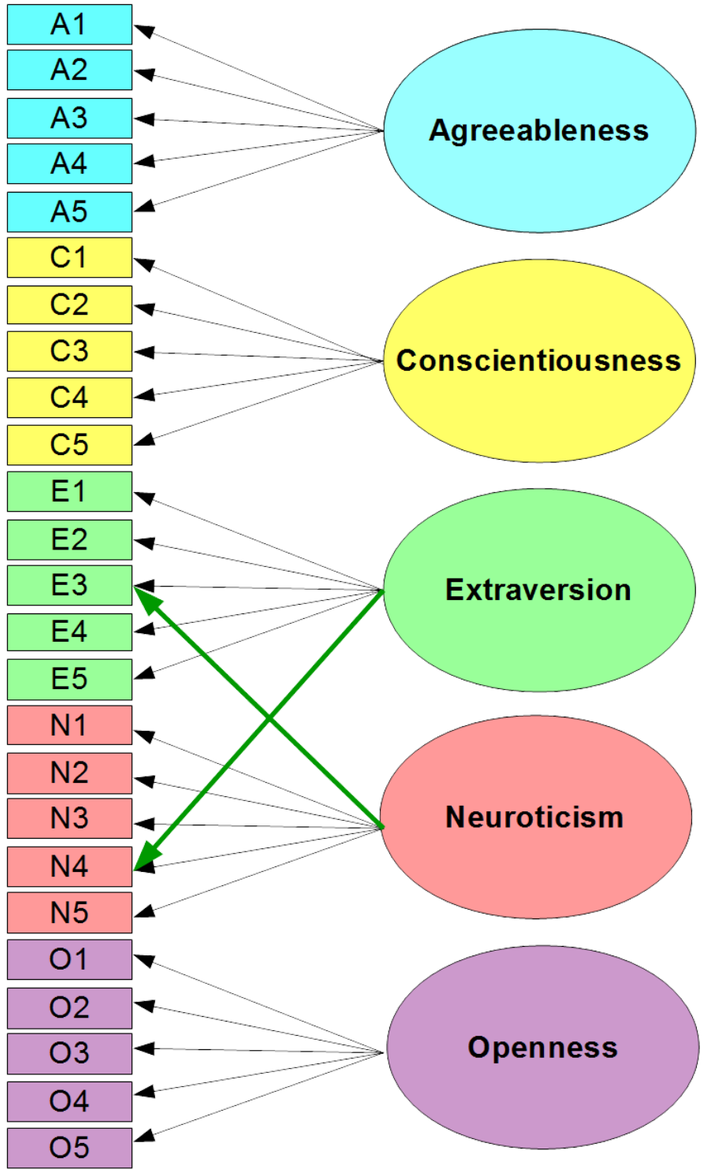
Two promising item/factor relationships to add:

- Extraversion ? Item N4
- Neuroticism ? Item E3

```
summary(theory_CFA)
```

```
...

Parameter Estimates
            Estimate   Std Error  z value    Pr(>|z|)
C[EXT,NEU]   0.2362614 0.03364096   7.023029  2.171093e-12 NEU <--> EXT
...
```


A1
A2
A3
A4
A5
C1
C2
C3
C4
C5
E1
E2
E3
E4
E5
N1
N2
N3
N4
N5
O1
O2
O3
O4
O5
Agreeableness
Conscientiousness
Extraversion
Neuroticism
Openness

Adding new loadings to the syntax

```
# Add some plausible item/factor loadings to the syntax
theory_syn_add <- "
AGE: A1, A2, A3, A4, A5
CON: C1, C2, C3, C4, C5
EXT: E1, E2, E3, E4, E5, N4
NEU: N1, N2, N3, N4, N5, E3
OPE: O1, O2, O3, O4, O5 "
# As before, convert your equations to sem-compatible syntax
theory_syn2 <- cfa(text = theory_syn_add, reference.indicators = FALSE)
# Run a CFA with the revised syntax
theory_CFA_add <- sem(model = theory_syn2, data = bfi_CFA)
```

Comparing the original and revised models

```
# Conduct a likelihood ratio test
anova(theory_CFA, theory_CFA_add)
```

```
LR Test for Difference Between Models

               Model Df Model Chisq Df LR Chisq Pr(>Chisq)
theory_CFA          265      2212.0
theory_CFA_rev      263      2097.8  2   114.28  < 2.2e-16 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparing the original and revised models

```
# Compare the comparative fit indices - higher is better!
summary(theory_CFA)$CFI
```

```
0.785075
```

```
summary(theory_CFA_add)$CFI
```

```
0.7974694
```

Comparing the original and revised models

```
# Compare the RMSEA values - lower is better!
summary(theory_CFA)$RMSEA
```

```
0.07731925          NA          NA 0.90
```

```
summary(theory_CFA_add)$RMSEA
```

```
0.07534156          NA          NA 0.90
```

- More information about fit indices: **http://davidakenny.net/cm/fit.htm**

Let's practice!

FACTOR ANALYSIS IN R

Removing loadings to improve fit

FACTOR ANALYSIS IN R

Jennifer Brussow
Psychometrician

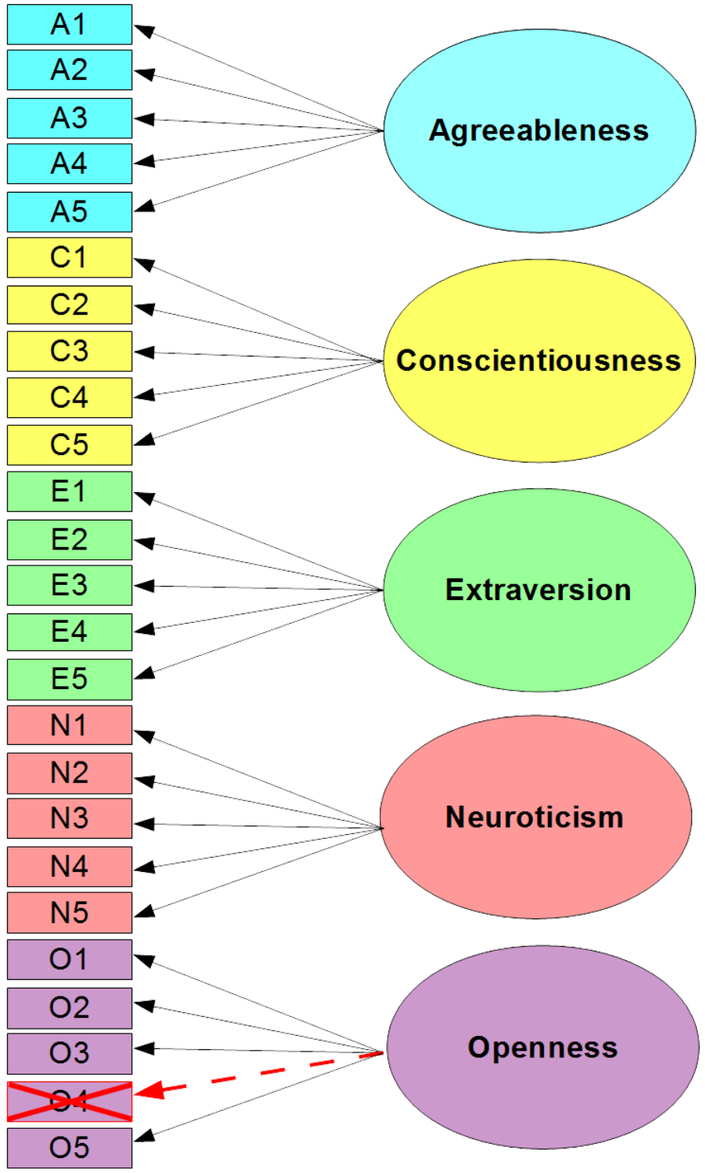
What does it mean to remove a loading?

```
summary(theory_CFA)
```

```
 Parameter Estimates
            Estimate Std Error z value Pr(>|z|)
lam[A1:AGE] -0.501   0.0449    -11.17   5.79e-29 A1 <--- AGE

...

lam[O1:OPE]  0.636   0.0379     16.79   2.88e-63 O1 <--- OPE
lam[O2:OPE] -0.731   0.0532    -13.75   5.39e-43 O2 <--- OPE
lam[O3:OPE]  0.809   0.0399     20.25   3.24e-91 O3 <--- OPE
lam[O4:OPE]  0.287   0.0413      6.95   3.69e-12 O4 <--- OPE
lam[O5:OPE] -0.624   0.0444    -14.06   7.07e-45 O5 <--- OPE
```


A1
A2
A3
A4
A5
C1
C2
C3
C4
C5
E1
E2
E3
E4
E5
N1
N2
N3
N4
N5
O1
O2
O3
O4
O5
Agreeableness
Conscientiousness
Extraversion
Neuroticism
Openness

Removing a loading in the syntax

```
# The original syntax
theory_syn_eq <- "
AGE: A1, A2, A3, A4, A5
CON: C1, C2, C3, C4, C5
EXT: E1, E2, E3, E4, E5
NEU: N1, N2, N3, N4, N5
OPE: O1, O2, O3, O4, O5 "
# Remove the worst item/factor loadings from the syntax
theory_syn_del <- "
AGE: A1, A2, A3, A4, A5
CON: C1, C2, C3, C4, C5
EXT: E1, E2, E3, E4, E5
NEU: N1, N2, N3, N4, N5
OPE: O1, O2, O3, O5 "
# As before, convert your equations to sem-compatible syntax
theory_syn3 <- cfa(text = theory_syn_del, reference.indicators = FALSE)
```

Running the revised CFA

```
# Run a CFA with the revised syntax
theory_CFA_del <- sem(model = theory_syn3, data = bfi_CFA)
```

```
Warning messages:
1: In sem.semmod(model = theory_syn3, data = bfi_CFA) :
  -170 observations removed due to missingness
2: In sem.semmod(model = theory_syn3, data = bfi_CFA) :
  The following observed variables are in the input covariance
  or raw-moment matrix but do not appear in the model:
O4
```

Comparing the original and revised models

```
anova(theory_CFA, theory_CFA_del)
```

```
Error in anova.objectiveML(theory_CFA, theory_CFA_del) :
  the models are fit to different moment matrices
```

```
# Compare the comparative fit indices - higher is better!
summary(theory_CFA)$CFI
```

```
0.785075
```

```
summary(theory_CFA_del)$CFI
```

```
0.7995587
```

Comparing the original and revised models

```
# Compare the RMSEA values - lower is better!
summary(theory_CFA)$RMSEA
```

```
0.07731925          NA          NA 0.90000000
```

```
summary(theory_CFA_del)$RMSEA
```

```
0.07718057          NA          NA 0.90000000
```

More information about fit indices: **http://davidakenny.net/cm/fit.htm**

Let's practice!

FACTOR ANALYSIS IN R

Now you can conduct and interpret EFAs and CFAs!

FACTOR ANALYSIS IN R

Jennifer Brussow
Psychometrician

Things you can do

- Conduct a unidimensional EFA
- Conduct a multidimensional EFA
- Conduct a CFA based on EFA results
- Conduct a CFA based on theory
- Interpret fit statistics
- Compare and refine models

More information

- `psych` and `sem` package documentation
- Books on multivariate analysis
- **Applied Multivariate Statistical Analysis by Johnson & Wichern**
- **Structural Equation Modeling with lavaan in R**

Congrats!

FACTOR ANALYSIS IN R