Setting up a CFA

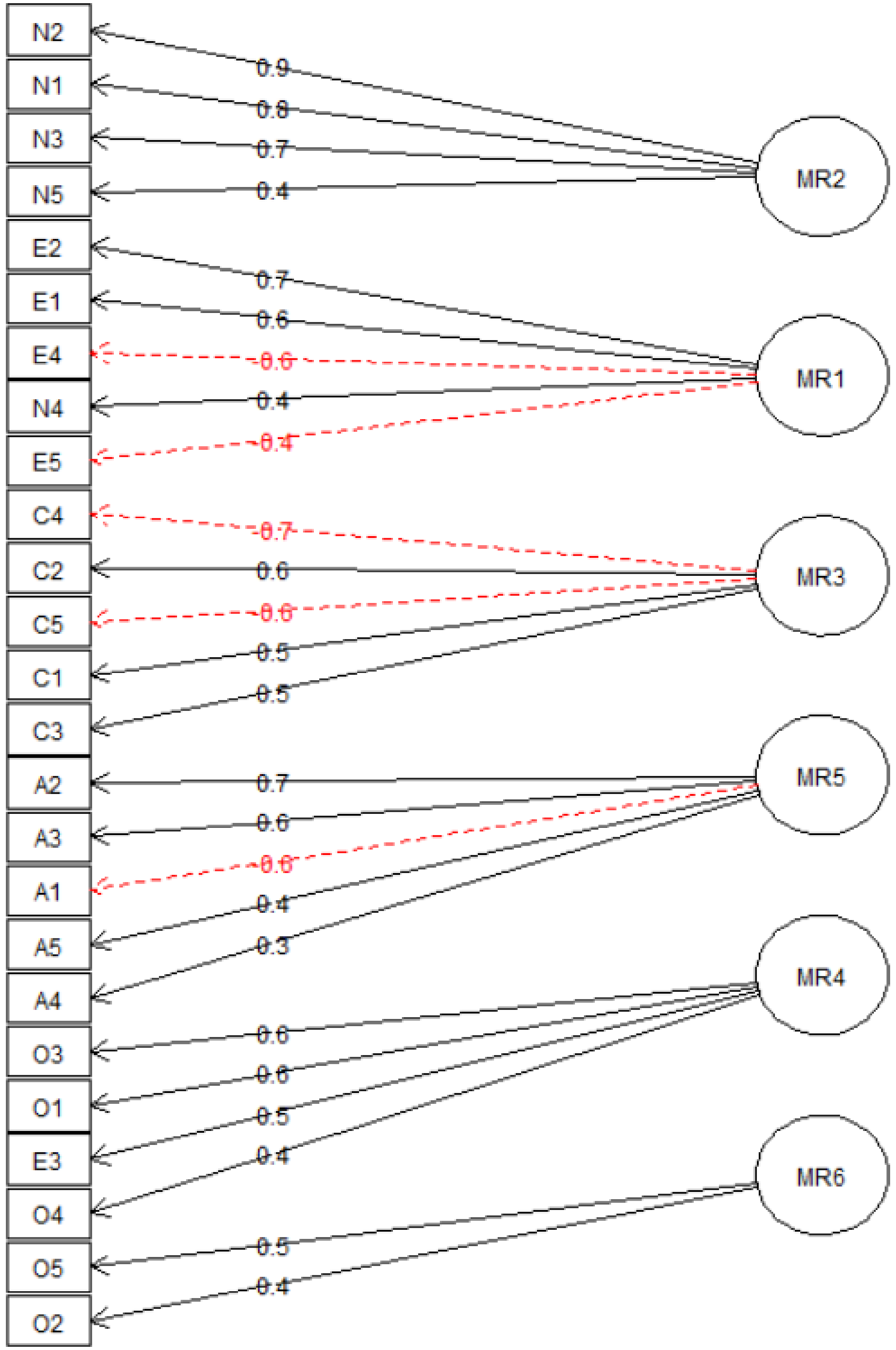
FACTOR ANALYSIS IN R

Jennifer Brussow
Psychometrician

Why a confirmatory analysis?

Benefits of a confirmatory analysis:

- Explicitly specified variable/factor relationships
- Testing a theory that you know in advance
- This is the right thing to publish when you are developing a new measure!


N2
N1
N3
N5
E2
E1
E4
N4
E5
C4
C2
C5
C1
C3
A2
A3
A1
A5
A4
O3
O1
E3
O4
O5
O2
0.9
0.8
0.7
0.4
0.7
0.6
-0.6
0.4
-0.4
-0.7
0.6
-0.6
0.5
0.5
0.7
0.6
-0.6
0.4
0.3
0.6
0.6
0.5
0.4
0.5
0.4
MR2
MR1
MR3
MR5
MR4
MR6

Using the wrapper function to set up a CFA

```
EFA_syn <- structure.sem(EFA_model)
EFA_syn
```

```
      Path         Parameter Value
 [1,] "MR5->A1"    "F4A1"    NA
 [2,] "MR5->A2"    "F4A2"    NA
 [3,] "MR5->A3"    "F4A3"    NA
 [4,] "MR5->A4"    "F4A4"    NA
 [5,] "MR5->A5"    "F4A5"    NA
 [6,] "MR3->C1"    "F3C1"    NA
 [7,] "MR3->C2"    "F3C2"    NA
 [8,] "MR3->C3"    "F3C3"    NA
 [9,] "MR3->C4"    "F3C4"    NA
[10,] "MR3->C5"    "F3C5"    NA
[11,] "MR1->E1"    "F2E1"    NA
...
```

Syntax created from the wrapper function

```
EFA_syn
```

```
     Path         Parameter Value
[1,] "MR5->A1"    "F4A1"    NA
```

- Factor 4 (F4) = Factor MR5 from the EFA
- Examinees' level of a factor predicts item responses
- Wrapper function automatically names parameters
- NA Value = starting value chosen at random

Creating CFA syntax from your theory

```
# Set up syntax specifying which items load onto each factor
theory_syn_eq <- "
AGE: A1, A2, A3, A4, A5     #Agreeableness
CON: C1, C2, C3, C4, C5     #Conscientiousness
EXT: E1, E2, E3, E4, E5     #Extraversion
NEU: N1, N2, N3, N4, N5     #Neuroticism
OPE: O1, O2, O3, O4, O5     #Openness
"
```

- Short, memorable factor names
- Factor name followed by colon
- Items in a comma-separated list

```
theory_syn <- cfa(text = theory_syn_eq, reference.indicators = FALSE)
```

Let's create some syntax!

FACTOR ANALYSIS IN R

Understanding the sem() syntax

FACTOR ANALYSIS IN R

Jennifer Brussow
Psychometrician

Relationships between variables and factors

```
theory_syn
```

```
   Path       Parameter   StartValue
1  AGE-> A1  lam[A1:AGE]
2  AGE-> A2  lam[A2:AGE]
3  AGE-> A3  lam[A3:AGE]
4  AGE-> A4  lam[A4:AGE]
5  AGE-> A5  lam[A5:AGE]
6  CON-> C1  lam[C1:CON]
7  CON-> C2  lam[C2:CON]
8  CON-> C3  lam[C3:CON]
9  CON-> C4  lam[C4:CON]
10 CON-> C5  lam[C5:CON]
11 EXT-> E1  lam[E1:EXT]
...
```

1. Path: Relationships between factors and items
2. Parameter: Automatically assigned names for each parameter
3. Starting value: Blank means they will be randomly generated

Factor variances

```
theory_syn
```

```
   Path         Parameter   StartValue
26 AGE <-> AGE <fixed>      1
27 CON <-> CON <fixed>      1
28 EXT <-> EXT <fixed>      1
29 NEU <-> NEU <fixed>      1
30 OPE <-> OPE <fixed>      1
```

Factor covariances

```
theory_syn
```

```
   Path          Parameter   StartValue
31 AGE <-> CON C[AGE,CON]
32 AGE <-> EXT C[AGE,EXT]
33 AGE <-> NEU C[AGE,NEU]
34 AGE <-> OPE C[AGE,OPE]
35 CON <-> EXT C[CON,EXT]
36 CON <-> NEU C[CON,NEU]
37 CON <-> OPE C[CON,OPE]
38 EXT <-> NEU C[EXT,NEU]
39 EXT <-> OPE C[EXT,OPE]
40 NEU <-> OPE C[NEU,OPE]
```

Item variances

```r
theory_syn
```

```
   Path         Parameter   StartValue
41 A1 <-> A1    V[A1]
42 A2 <-> A2    V[A2]
43 A3 <-> A3    V[A3]
44 A4 <-> A4    V[A4]
45 A5 <-> A5    V[A5]
46 C1 <-> C1    V[C1]
47 C2 <-> C2    V[C2]
48 C3 <-> C3    V[C3]
49 C4 <-> C4    V[C4]
50 C5 <-> C5    V[C5]
51 E1 <-> E1    V[E1]
52 E2 <-> E2    V[E2]
...
```

Running the CFA

Actually running the CFA is *much* easier than setting up the syntax!

```
#Use the sem() function to run a CFA
theory_CFA <- sem(theory_syn, data = bfi_CFA)
```

```r
summary(theory_CFA)
```

```
 Model Chisquare =  2212.032   Df =  265 Pr(>Chisq) = 9.662018e-304
 AIC =  2332.032
 BIC =  326.618
 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-5.5800 -0.3732  1.0350  1.1220  2.4710  8.9000
 R-square for Endogenous Variables
    A1     A2     A3     A4     A5     C1     C2     C3     C4
0.1178 0.4475 0.5731 0.2994 0.4713 0.3006 0.3667 0.2947 0.4886
...
 Parameter Estimates
           Estimate   Std Error  z value    Pr(>|z|)
lam[A1:AGE] -0.5011716 0.04487184 -11.168956  5.785714e-29 A1 <--- AGE
lam[A2:AGE]  0.8230960 0.03447831  23.872862 5.863008e-126 A2 <--- AGE
...
```

Let's practice!

FACTOR ANALYSIS IN R

Investigating model fit

FACTOR ANALYSIS IN R

Jennifer Brussow
Psychometrician

Default fit statistics

Chi-square test (aka the log likelihood test) is only default

```
summary(theory_CFA)
```

```
Model Chisquare =  2231.647   Df =  265 Pr(>Chisq) = 1.695873e-307
```

- Often significant due to sample size
- Desired outcome is lack of significance

Changing the options

```
options(fit.indices = c("CFI", "GFI", "RMSEA", "BIC"))
```

- RMSEA < 0.05
- GFI (Goodness of Fit Index) > 0.90
- CFI (Comparative Fit Index) > 0.90

Absolute model fit

```
summary(theory_CFA)
```

```
Model Chisquare =  2305.159   Df =  271 Pr(>Chisq) = 0
Goodness-of-fit index =  0.8527977
RMSEA index =  0.07815051   90% CI: (NA, NA)
Bentler CFI =  0.7754574
```

Relative fit

```
summary(theory_CFA)
```

```
Model Chisquare =  2305.159   Df =  271 Pr(>Chisq) = 8.422189e-319
Goodness-of-fit index =  0.8527977
RMSEA index =  0.07815051   90% CI: (NA, NA)
Bentler CFI =  0.7754574
BIC =  377.0563
```

```
summary(theory_CFA)$BIC
```

```
326.618
```

Relative fit: comparing models

```
summary(theory_CFA)$BIC
```

```
326.618
```

```
# Run a CFA using the EFA syntax you created earlier
EFA_CFA <- sem(EFA_syn, data = bfi_CFA)
summary(EFA_CFA)$BIC
```

```
377.0563
```

- Useful for nested models that are fit to the same dataset
- Don't use if these conditions are not met!

Let's practice!

FACTOR ANALYSIS IN R