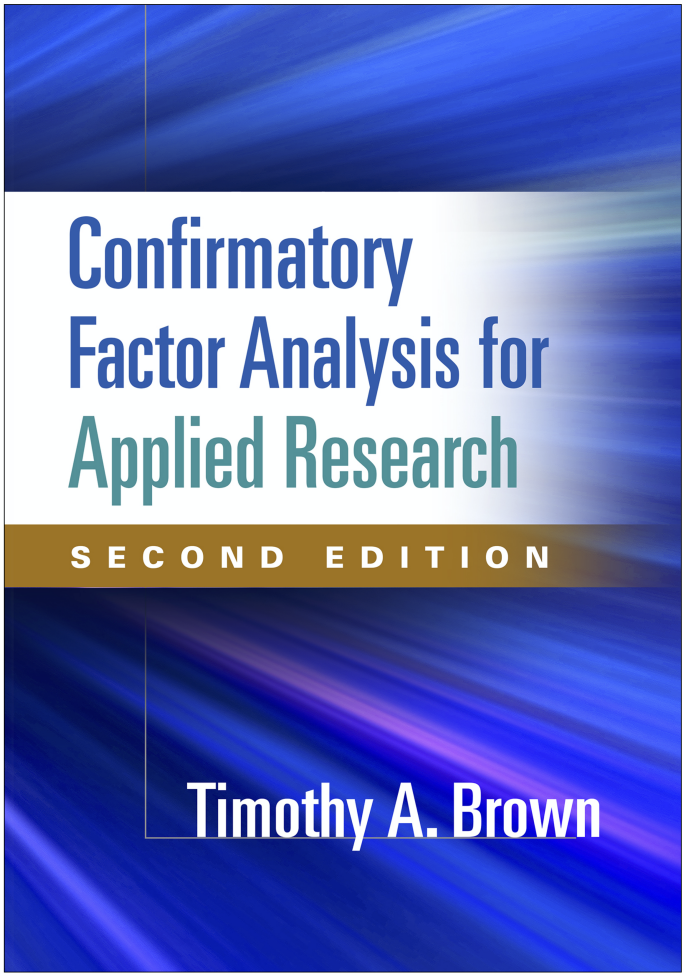
Welcome! Survey/test construction using Mplus

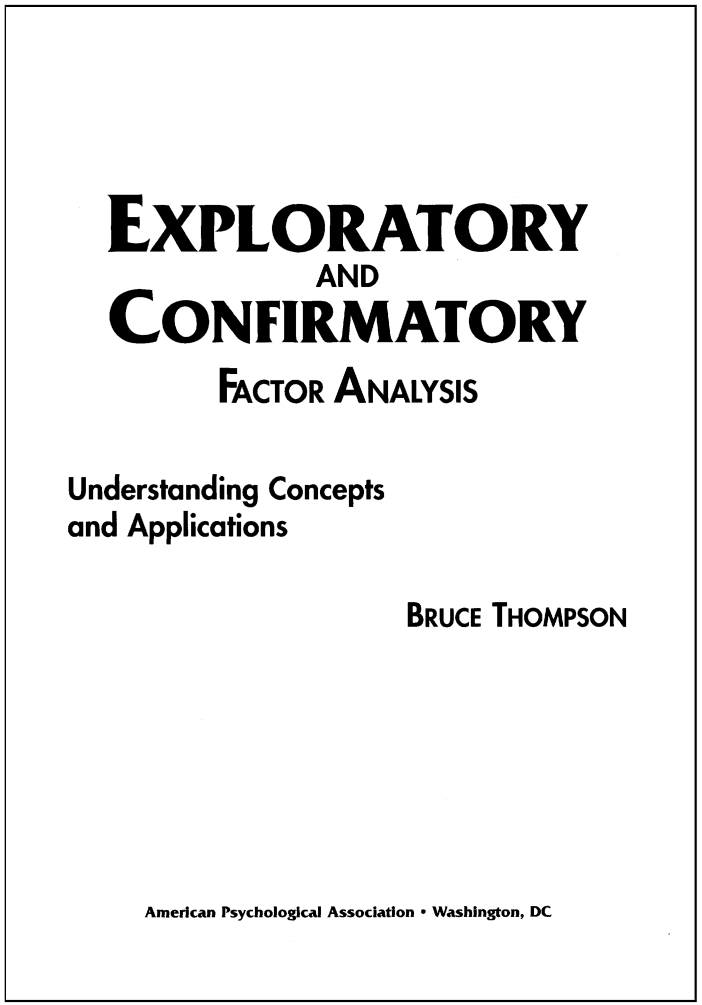
All models are wrong,
but some are useful.

George Box, British statistician (1919 – 2013)

Outline

Resources


Confirmatory
Factor Analysis for
Applied Research
SECOND EDITION
Timothy A. Brown


EXPLORATORY
AND
CONFIRMATORY
FACTOR ANALYSIS
Understanding Concepts
and Applications
BRUCE THOMPSON
American Psychological Association • Washington, DC

Resources

University of Miami virtual lab for Mplus:
https://vlabs.it.miami.edu/

Text editor
https://www.sublimetext.com/

R: `Mplusautomation`
https://www.statmodel.com/usingmplusviar.shtml

Mplus website:
https://www.statmodel.com/

Introduction: EFA

Exploratory factor analysis (EFA): Determine number of latent variables that explains covariance among observed, manifest, variables.

Continuous latent variabales also referred to as factors and the maifest variables are referred to as indicators (can be non-continuous level of measurement).

EFA: informs about factors for tentative items on underlying scale, and which items have "loadings" on the scale.

The main goals are to discover the number factors on a scale and salience of factor loadings (small loadings and large error variances candidates for removal).

Fit Statistics

Joint criteria is recommended
e.g., CFI $\geq$ .96 & SRMR $\leq$ .10

1. Loglikelihood: compares nested models (deviance)
   H_0 Value: null model $p < .05$
2. Information Criteria: non-nested models
   AIC, BIC, SBIC: *smallest value*
3. RMSEA and SRMR
   value $< .05$
4. CFI/TLI: (comparative fit index & Tucker-Lewis Index[1])
   value $> .95$
   literature can support values above .90 as acceptable
5. χ^2 test for baseline
   Baseline is a null model $p < .05$

[1] NNFI is another name for TLI

Research Question

A random sample of respondents (n = 600) are collected for a scale with 11 subtests.

How many latent traits are measured by these subtests?

First, divide the sample in half—unless you have a second dataset as your calibration sample. One way is to randomly sample observations from approximately 50% of the dataset.

1. Run Mplus EFA on calibration sample
2. Revise if needed (example: cross-loadings and/or remove factor with only two-items)
3. Rerun Mplus EFA

Data file: correlation matrix

```
1
.67 1
.65 .64 1
.63 .61 .63 1
.65 .62 .62 .67 1
.62 .67 .63 .65 .65 1
.18 .12 .15 .13 .12 .15 1
.14 .17 .15 .17 .16 .18 .55 1
.17 .13 .12 .14 .15 .12 .55 .52 1
.13 .15 .15 .16 .17 .15 .51 .57 .60 1
.11 .15 .16 .15 .14 .12 .60 .53 .57 .55 1
```

R: `corr_matrix <- cor(<matrix/dataframe object>)`

MPlus input file

```
TITLE: EFA example 1

DATA: FILE = "EFAex1_cor2.dat";
      TYPE = CORRELATION;
      NOBSERVATIONS = 300;

VARIABLE: NAMES = V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11;

ANALYSIS: TYPE = EFA 1 3;
```

`EFA 1 3;`
Declares 1 as minimum number of factors and 3 as highest factor interested in testing.

MPlus output

```
INPUT READING TERMINATED NORMALLY


EFA example 1

SUMMARY OF ANALYSIS

Number of groups                                                 1
Number of observations                                         300

Number of dependent variables                                   11
Number of independent variables                                  0
Number of continuous latent variables                            2

Observed dependent variables

  Continuous
   V1          V2          V3          V4          V5          V6
   V7          V8          V9          V10         V11
```

MPlus output

```
Estimator                                                      ML
Rotation                                                   GEOMIN
Row standardization                                  CORRELATION
Type of rotation                                         OBLIQUE
Epsilon value                                             Varies
Information matrix                                      EXPECTED
Maximum number of iterations                                1000
Convergence criterion                                  0.500D-04
Maximum number of steepest descent iterations                 20
Optimization Specifications for the Exploratory Factor Analysis
Rotation Algorithm
  Number of random starts                                     30
  Maximum number of iterations                             10000
  Derivative convergence criterion                       0.100D-04
```

MPlus output

```
NO CONVERGENCE.  NUMBER OF ITERATIONS EXCEEDED.
PROBLEM OCCURRED IN EXPLORATORY FACTOR ANALYSIS WITH 3 FACTOR(S).
```

```
SUMMARY OF MODEL FIT INFORMATION

                    Number of                      Degrees of
Model             Parameters      Chi-Square       Freedom      P-Value

1-factor                22            618.031         44         0.0000
2-factor                32             39.848         34         0.2260
3-factor               N/A

                                                   Degrees of
Models Compared                   Chi-Square       Freedom      P-Value

1-factor against 2-factor           578.183         10         0.0000
```

Issue of over-specification or overfactoring where solutions attempt to extract too many factors (see on the output there are only 2 factors but the input requested 3). The numerical solution was found.

MPlus output

```
RESULTS FOR EXPLORATORY FACTOR ANALYSIS

EIGENVALUES FOR SAMPLE CORRELATION MATRIX
      1             2             3             4             5
  ________      ________      ________      ________      ________
   4.652         2.772         0.533         0.502         0.442

EIGENVALUES FOR SAMPLE CORRELATION MATRIX
      6             7             8             9            10
  ________      ________      ________      ________      ________
   0.415         0.382         0.370         0.343         0.321

EIGENVALUES FOR SAMPLE CORRELATION MATRIX
     11
  ________
   0.268
```

Note: do not use rules of thumb ("eigenvalues greater than 1 are factors") for judging factors. A modified parallel analysis should be conducted by iteratively updating hypothesis for number of factors.

Zopluoglu, C., & Davenport Jr, E. C. (2017). A note on using eigenvalues in dimensionality assessment. Practical Assessment, Research, and Evaluation, 22(1), 7.

MPlus output

EFA 1 factor:

```
MODEL FIT INFORMATION

Number of Free Parameters                       22

Loglikelihood

          H0 Value                        -4095.126
          H1 Value                        -3786.111

Information Criteria

          Akaike (AIC)                     8234.252
          Bayesian (BIC)                   8315.736
          Sample-Size Adjusted BIC         8245.965
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                             618.031
          Degrees of Freedom                     44
          P-Value                            0.0000

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.209
          90 Percent C.I.                    0.194  0.223
          Probability RMSEA <= .05           0.000

CFI/TLI

          CFI                                0.668
          TLI                                0.584

Chi-Square Test of Model Fit for the Baseline Model

          Value                           1781.755
          Degrees of Freedom                    55
          P-Value                          0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                              0.216
```

MPlus output

EFA 1 factor:

```
MODEL FIT INFORMATION

Number of Free Parameters                       22

Loglikelihood

          H0 Value                        -4095.126
          H1 Value                        -3786.111

Information Criteria

          Akaike (AIC)                     8234.252
          Bayesian (BIC)                   8315.736
          Sample-Size Adjusted BIC         8245.965
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                             618.031
          Degrees of Freedom                     44
          P-Value                            0.0000

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.209
          90 Percent C.I.                    0.194  0.223
          Probability RMSEA <= .05           0.000

CFI/TLI

          CFI                                0.668
          TLI                                0.584

Chi-Square Test of Model Fit for the Baseline Model

          Value                           1781.755
          Degrees of Freedom                    55
          P-Value                           0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                              0.216
```

Notice the χ^2 test as evidence against the null hypothesis:
H_0: one factor accounts for all 11 variables' correlations.
Examine other fit indices as well, especially with large samples.

MPlus output

EFA 2 factors:

```
MODEL FIT INFORMATION

Number of Free Parameters                       32

Loglikelihood

          H0 Value                        -3806.035
          H1 Value                        -3786.111

Information Criteria

          Akaike (AIC)                     7676.070
          Bayesian (BIC)                   7794.591
          Sample-Size Adjusted BIC         7693.106
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                              39.848
          Degrees of Freedom                     34
          P-Value                            0.2260

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.024
          90 Percent C.I.                    0.000  0.050
          Probability RMSEA <= .05           0.948

CFI/TLI

          CFI                                0.997
          TLI                                0.995

Chi-Square Test of Model Fit for the Baseline Model

          Value                           1781.755
          Degrees of Freedom                    55
          P-Value                           0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                              0.019
```

MPlus output

EFA 2 factors:

```
GEOMIN ROTATED LOADINGS (* significant at 5% level)
                1              2
              ________       ________
V1              0.805*        -0.001
V2              0.804*        -0.004
V3              0.789*         0.003
V4              0.795*         0.007
V5              0.802*         0.003
V6              0.807*        -0.006
V7             -0.001          0.741*
V8              0.031          0.714*
V9             -0.009          0.758*
V10             0.010          0.746*
V11            -0.011          0.762*

GEOMIN FACTOR CORRELATIONS (* significant at 5% level)
                1              2
              ________       ________
1             1.000
2             0.240*         1.000


ESTIMATED RESIDUAL VARIANCES
   V1             V2             V3             V4             V5
   ________       ________       ________       ________       ________
    0.352          0.355          0.376          0.366          0.356


ESTIMATED RESIDUAL VARIANCES
   V6             V7             V8             V9             V10
   ________       ________       ________       ________       ________
    0.351          0.452          0.478          0.429          0.439


ESTIMATED RESIDUAL VARIANCES
   V11
   ________
    0.424
```

MPlus output: correlations between indicators & factors

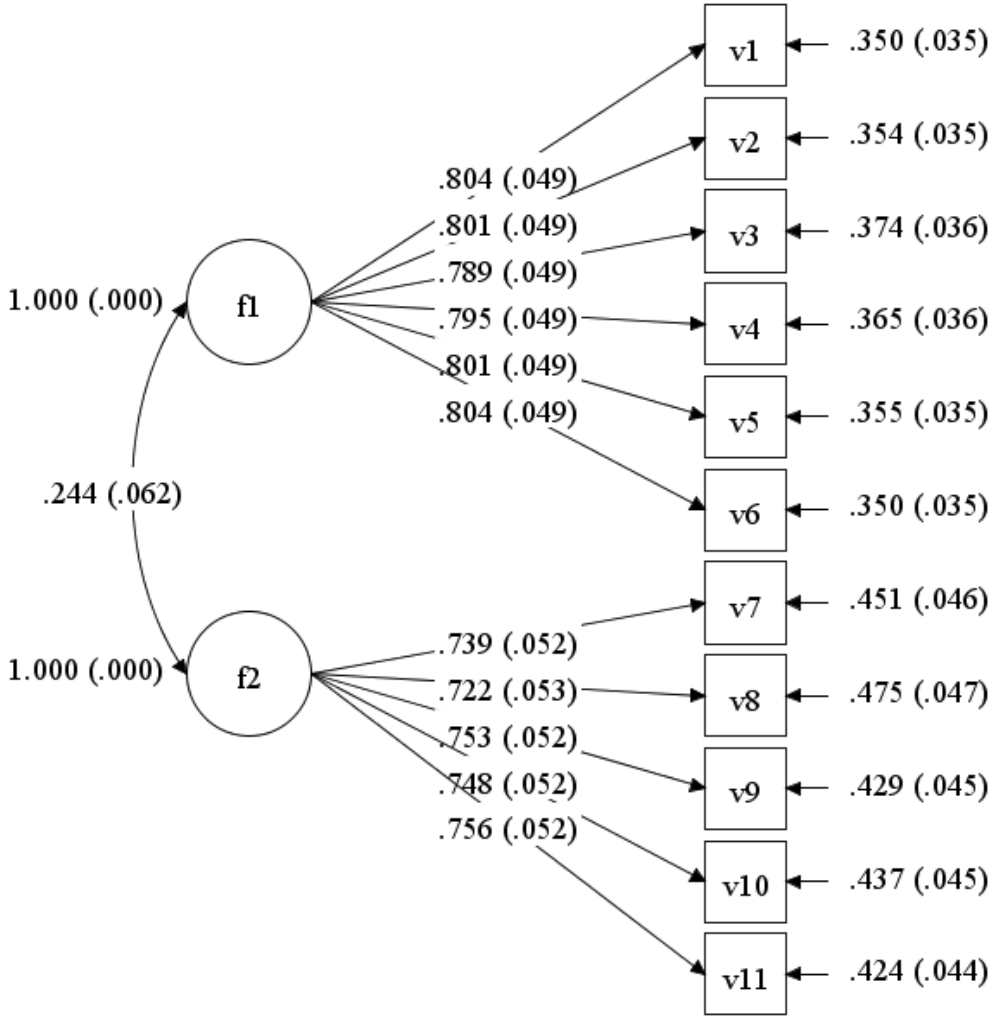
FACTOR STRUCTURE

	1	2
V1	0.805	0.192
V2	0.803	0.189
V3	0.790	0.192
V4	0.796	0.197
V5	0.802	0.195
V6	0.806	0.188
V7	0.177	0.740
V8	0.202	0.722
V9	0.173	0.755
V10	0.189	0.749
V11	0.172	0.759

Diagrammer can be used if rerun model as a CFA with `MODEL` with covariance matrix, not correlation.

```
DATA: FILE = "EFAex1_cor2.dat";
      TYPE = CORRELATION;
      NOBSERVATIONS = 300;

VARIABLE: NAMES = V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11;

MODEL: F1 BY V1* V2-V6;
       F2 BY V7* V8-V11;
       F1-F2@1;
```

EFA model: values = coefficient (SE)



Data file: response data

```
5,5,5,5,4,4,4,4,3,2,4,3
5,5,5,5,5,4,5,4,3,2,4,3
5,5,5,5,4,4,4,4,4,3,4,3
4,4,5,4,4,4,4,3,4,4,4,3
5,5,5,5,5,5,5,5,5,5,4,4
4,4,4,4,4,4,4,3,4,3,4,-9999
4,5,5,-9999,4,4,4,3,4,3,3,4
5,5,5,5,5,5,5,3,3,4,3,3
```

Missing data: -9999

input file

```
Data: File = respdat.dat;

Variable: Names = item13-item24;
          Missing are all (-9999) ;

Analysis: type = EFA 3 3;
```

MPlus output

```
SUMMARY OF ANALYSIS

Number of groups                                                 1
Number of observations                                        1427

Number of dependent variables                                   12
Number of independent variables                                  0
Number of continuous latent variables                            0

Observed dependent variables

  Continuous
   ITEM13      ITEM14      ITEM15      ITEM16      ITEM17      ITEM18
   ITEM19      ITEM20      ITEM21      ITEM22      ITEM23      ITEM24
```

MPlus output

Covariance Coverage

	ITEM13	ITEM14	ITEM15	ITEM16	ITEM17
ITEM13	0.994				
ITEM14	0.994	0.998			
ITEM15	0.993	0.996	0.998		
ITEM16	0.991	0.994	0.994	0.995	
ITEM17	0.992	0.995	0.995	0.994	0.997
ITEM18	0.992	0.996	0.996	0.994	0.996
ITEM19	0.989	0.993	0.994	0.992	0.994
ITEM20	0.974	0.977	0.977	0.975	0.976
ITEM21	0.992	0.994	0.994	0.992	0.994
ITEM22	0.986	0.989	0.989	0.987	0.989
ITEM23	0.992	0.995	0.995	0.992	0.994
ITEM24	0.989	0.992	0.992	0.989	0.991

Covariance Coverage

	ITEM18	ITEM19	ITEM20	ITEM21	ITEM22
ITEM18	0.998				
ITEM19	0.995	0.995			
ITEM20	0.977	0.975	0.978		
ITEM21	0.994	0.992	0.977	0.996	
ITEM22	0.989	0.987	0.973	0.989	0.991
ITEM23	0.995	0.992	0.976	0.995	0.989
ITEM24	0.991	0.988	0.973	0.991	0.986

Covariance Coverage

	ITEM23	ITEM24
ITEM23	0.997	
ITEM24	0.992	0.993

MPlus output

UNIVARIATE HIGHER-ORDER MOMENT DESCRIPTIVE STATISTICS

Variable/ Sample Size	Mean/ Variance	Skewness/ Kurtosis	Minimum/ Maximum	% with Min/Max	Percentiles 20%/60%	40%/80%	Median
ITEM13	4.450	-1.570	1.000	0.63%	4.000	4.000	5.000
1419.000	0.544	3.217	5.000	56.24%	5.000	5.000	
ITEM14	4.517	-1.633	1.000	0.35%	4.000	5.000	5.000
1424.000	0.503	3.106	5.000	61.66%	5.000	5.000	
ITEM15	4.435	-1.475	1.000	0.56%	4.000	4.000	5.000
1424.000	0.559	2.652	5.000	55.69%	5.000	5.000	
ITEM16	4.270	-1.317	1.000	0.92%	4.000	4.000	4.000
1420.000	0.703	1.926	5.000	45.63%	5.000	5.000	
ITEM17	4.159	-1.059	1.000	1.05%	4.000	4.000	4.000
1423.000	0.804	0.892	5.000	41.53%	5.000	5.000	
ITEM18	3.924	-0.838	1.000	2.81%	3.000	4.000	4.000
1424.000	1.064	0.179	5.000	34.41%	4.000	5.000	
ITEM19	4.073	-1.063	1.000	2.11%	3.000	4.000	4.000
1420.000	0.933	0.846	5.000	38.87%	4.000	5.000	
ITEM20	3.771	-0.282	1.000	1.29%	3.000	3.000	4.000
1396.000	0.834	-0.304	5.000	24.64%	4.000	5.000	
ITEM21	3.769	-0.679	1.000	2.32%	3.000	4.000	4.000
1422.000	0.972	0.047	5.000	23.63%	4.000	5.000	
ITEM22	3.593	-0.646	1.000	5.59%	3.000	4.000	4.000
1414.000	1.260	-0.339	5.000	21.29%	4.000	5.000	
ITEM23	3.800	-0.611	1.000	2.32%	3.000	4.000	4.000
1423.000	0.929	0.116	5.000	25.58%	4.000	5.000	
ITEM24	3.653	-0.433	1.000	2.47%	3.000	3.000	4.000
1417.000	0.866	0.128	5.000	18.70%	4.000	4.000	

MPlus output

GEOMIN ROTATED LOADINGS (* significant at 5% level)

	1	2	3
ITEM13	0.858*	-0.087	0.010
ITEM14	0.832*	-0.022	-0.021
ITEM15	0.724*	0.085	0.007
ITEM16	0.645*	0.129*	-0.069
ITEM17	0.515*	0.276*	0.084
ITEM18	0.091*	0.755*	0.012
ITEM19	-0.015	0.842*	-0.095
ITEM20	0.099*	0.559*	-0.011
ITEM21	0.221*	0.408*	0.199*
ITEM22	0.000	0.508*	0.205*
ITEM23	0.123	0.011	0.795*
ITEM24	-0.009	-0.008	0.804*

t-value's indicate cross-loadings: SE and theory most important when deciding how to handle large, significant cross-loadings. Cudeck has published helpful articles on this topic.

Data file: categorical data

3	1	1	1	1
3	3	2	1	4
4	4	4	4	4
1	1	2	1	1
2	3	2	3	2
1	3	2	2	2
2	2	2	3	1
2	2	2	1	1

If unranked then be sure that indicators are converted into a set of binary variables. Also check out mixture modeling on the Mplus website in references section.

input file: categorical data

```
TITLE: EFA example categorical data

DATA: FILE = "EFA_categ.dat";

VARIABLE: NAMES = i1-i5;
          categorical = i1-i5;

ANALYSIS: TYPE = EFA 1 2;
```

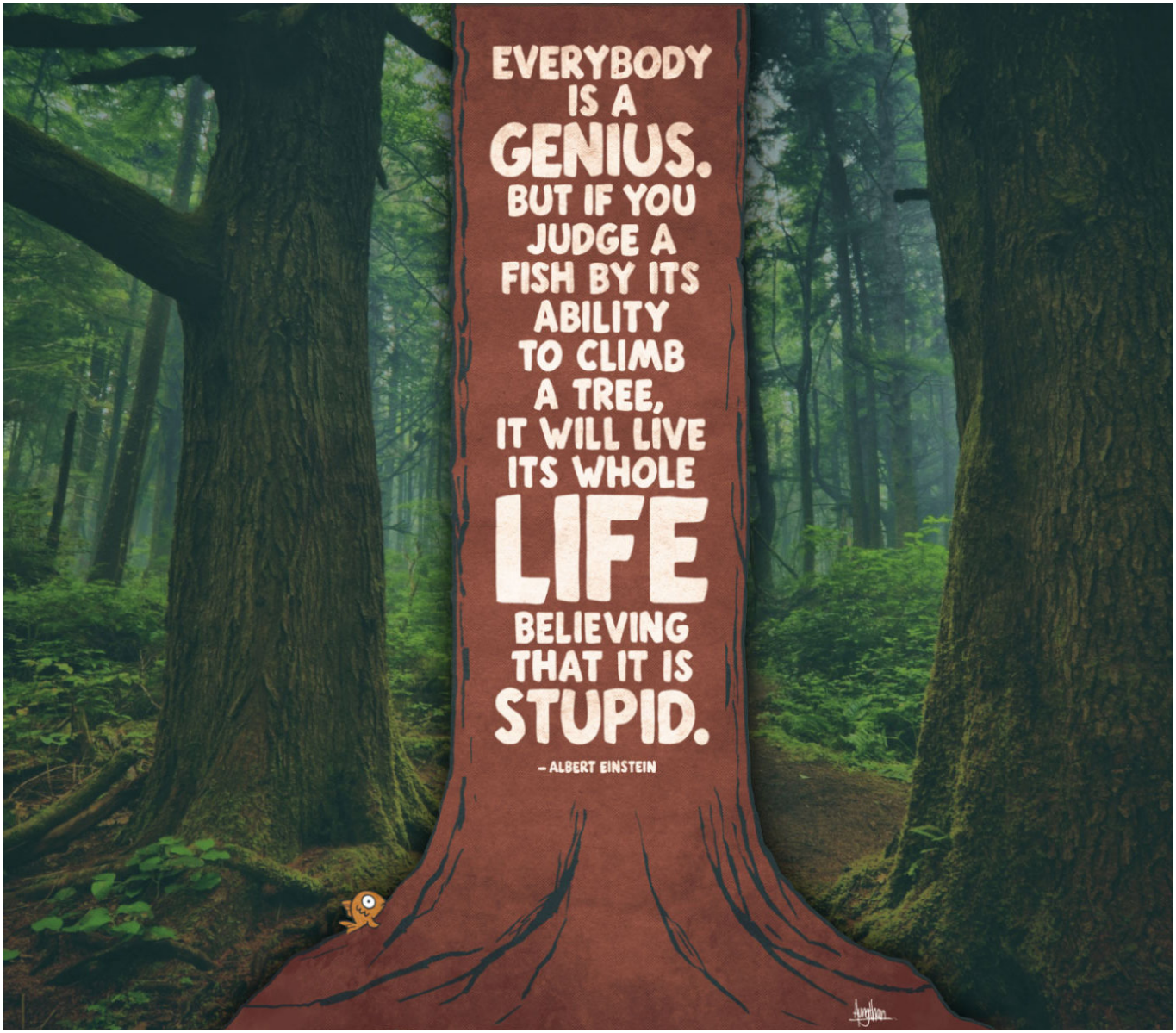
MPlus output

```
EFA example 2 categorical data

SUMMARY OF ANALYSIS

Number of groups                                                 1
Number of observations                                         400

Number of dependent variables                                    5
Number of independent variables                                  0
Number of continuous latent variables                            0

Observed dependent variables

  Binary and ordered categorical (ordinal)
   I1          I2          I3          I4          I5
```

MPlus output

```
UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES

    I1
      Category 1    0.177          71.000
      Category 2    0.295         118.000
      Category 3    0.292         117.000
      Category 4    0.235          94.000
    I2
      Category 1    0.172          69.000
      Category 2    0.310         124.000
      Category 3    0.295         118.000
      Category 4    0.223          89.000
    I3
      Category 1    0.225          90.000
      Category 2    0.247          99.000
      Category 3    0.270         108.000
      Category 4    0.258         103.000
    I4
      Category 1    0.258         103.000
      Category 2    0.212          85.000
      Category 3    0.237          95.000
      Category 4    0.292         117.000
    I5
      Category 1    0.250         100.000
      Category 2    0.223          89.000
      Category 3    0.230          92.000
      Category 4    0.297         119.000
```

Thank you!


EVERYBODY
IS A
GENIUS.
BUT IF YOU
JUDGE A
FISH BY ITS
ABILITY
TO CLIMB
A TREE,
IT WILL LIVE
ITS WHOLE
LIFE
BELIEVING
THAT IT IS
STUPID.
– ALBERT EINSTEIN