“What is the purpose and process of Exploratory Factor Analysis, and what information can be extracted from the Mplus Annotated Output?”

Authored by
stats writer

June 30, 2024

Exploratory Factor Analysis (EFA) is a statistical technique used to identify underlying factors or dimensions within a set of observed variables. The purpose of EFA is to reduce the complexity of a large data set and to gain a better understanding of the underlying structure of the data. This is achieved by grouping related variables into factors, allowing for a more concise representation of the data.

The process of EFA involves analyzing the intercorrelations between a set of variables and extracting factors based on the strength of these correlations. This is typically done through a series of mathematical calculations and statistical tests, such as the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and the Bartlett's test of sphericity. These tests help determine the suitability of the data for factor analysis and the number of factors to be extracted.

The Mplus Annotated Output is a tool used to interpret the results of EFA. It provides a detailed summary of the factor analysis, including the factor loadings (i.e. the correlation between each variable and the factor), the communalities (i.e. the amount of variance in each variable explained by the factors), and the factor correlations (i.e. the relationship between the extracted factors). This information can be used to identify the underlying factors and their corresponding variables, as well as to evaluate the overall fit of the model. The Mplus Annotated Output also provides additional information, such as the eigenvalues and scree plot, to assist in determining the optimal number of factors to extract. Overall, the output from Mplus can help researchers gain a better understanding of the underlying structure of their data and make informed decisions about the variables and factors in their analysis.

Exploratory
Factor Analysis | Mplus Annotated Output

This page was created using Mplus version 5.2, the output and/or syntax may be different for other versions of Mplus.

This page shows an example exploratory factor analysis with footnotes explaining the output. The data used in this example

were collected on
1428 college students (complete data on 1365 observations) and are responses to
items on a survey. You can obtain
the data set by clicking here. The analysis includes 12 variables, item13 to item24.

Data:
File is m255.dat ;
Variable:
Names are
item13 item14 item15 item16 item17 item18 item19
item20 item21 item22
item23 item24;
Missing are all (-9999) ;
Analysis:
type = efa 3 3;

Comments on syntax

Notes and summary information

The information below printed near the top of the output, it is useful because it lets you know what Mplus did.

You want to check this part of the output to make sure Mplus ran the analysis that you intended.

***** WARNINGa**
Data set contains cases with missing on all variables.
These cases were not included in the analysis.
Number of cases with missing on all variables: 1
1 WARNING(S) FOUND IN THE INPUT INSTRUCTIONS

SUMMARY OF ANALYSIS

Number of groups 1
Number of observationsb 1427

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

Observed dependent variablesd

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

Estimatore ML

Rotationf GEOMIN

Row standardization CORRELATION

Type of rotationg OBLIQUE

Epsilon value Varies

Information matrix OBSERVED

Maximum number of iterations 1000

Convergence criterion 0.500D-04

Maximum number of steepest descent iterations 20

Maximum number of iterations for H1 2000

Convergence criterion for H1 0.100D-03

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

Input data file(s)

m255.dat

Input data format FREE

a. Warning. In this case, the output includes a warning that 1 case

had missing values on all of the variables in the analysis, and hence was excluded from the model.

b. Number of observations. The number of observations used in the analysis. As mentioned above, by default, Mplus will include all cases that have at least partial data on the variables in the analysis.

c. Number of dependent variables. Gives the number of dependent (outcome) variables in the model. Note that Mplus classifies the factor indicators as dependent variables.

d. Observed dependent variables. The list of variables included in this analysis. All of the variables in our model are listed under Continuous. If the model included categorical variables, they would be listed here under the heading Categorical. If this section includes variables you did not intend to include in your analysis, you may need to use the

**usevariables option of
the data command.**

e. Estimator. The method used to estimate the model, in this case, maximum likelihood (ML).

f. Rotation. The specific rotation method used in the model.

**g. Type of rotation. Rotations that allow the factors to be correlated
are oblique, while rotations that force the factors to be uncorrelated are known
as orthogonal. The default geomin rotation is oblique.**

SUMMARY OF DATA

Number of missing data patternsh 24

COVARIANCE COVERAGE OF DATA

Minimum covariance coverage value 0.100

PROPORTION OF DATA PRESENT

Covariance Coveragei
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

h. Number of missing data patterns. This gives the number of different patterns of missingness present in the variables included in the model. Large numbers of missing data patterns can result in difficulty estimating the model.

i. Covariance Coverage. If any of the variables in the model have missing values, Mplus provides information on the number and distribution of missing values. The covariance coverage matrix gives the proportion of values present for each variable individually (on the diagonal) and pairwise combinations of variables (below the diagonal). For example, 99.4% of cases have non-missing values for item13 and 99.3% of cases have valid values for item13 and item15.

RESULTS FOR EXPLORATORY FACTOR ANALYSIS

EIGENVALUES FOR SAMPLE CORRELATION MATRIXj

1 2 3 4 5

________ ________ ________ ________ ________

1 6.289 1.228 0.709 0.606 0.561

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

6 7 8 9 10

________ ________ ________ ________ ________

1 0.499 0.470 0.384 0.366 0.329

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

11 12

________ ________

1 0.309 0.251

EXPLORATORY FACTOR ANALYSIS WITH 3 FACTOR(S):

TESTS OF MODEL FIT

Chi-Square Test of Model Fitk

Value 137.865

Degrees of Freedom 33

P-Value 0.0000

Chi-Square Test of Model Fit for the Baseline Model

Value 9132.568
Degrees of Freedom 66
P-Value 0.0000

CFI/TLII

CFI 0.988
TLI 0.977

Loglikelihood

H0 Value -17776.793
H1 Value -17707.861

Information Criteriam

Number of Free Parameters 57
Akaike (AIC) 35667.586
Bayesian (BIC) 35967.596
Sample-Size Adjusted BIC 35786.527
(n* = (n + 2) / 24)

RMSEA (Root Mean Square Error Of Approximation)n

Estimate 0.047
90 Percent C.I. 0.039 0.055
Probability RMSEA

j. Eigenvalues for sample correlation matrix. An eigenvalue is the variance of the factor. In the initial factor solution,
the first factor will account for the most variance, the second will account for
the next highest amount of variance, and so on.

k. Chi-square test of model fit. Compares the fit of the model to a
model with no restrictions (i.e. all variables correlated freely). Chi-square
values can be used to test the difference in fit between nested models.

l. Fit indices. The Comparative Fit Index (CFI) and the Tucker Lewis
Index (TLI) are measures of model fit. They have a range from 0 to 1 with higher
values indicating better fit.

m. Information Criteria. The Akaike information criterion

(AIC) and
the Bayesian information criterion (BIC, sometimes also called the Schwarz
criterion), can also be used to compare models, including non-nested models.

n. RMSEA. The root mean square error of approximation is another
measure of model fit. Smaller values indicate better model fit.

GEOMIN ROTATED LOADINGSo

	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

GEOMIN FACTOR CORRELATIONSp

1 2 3

________ ________ ________

1 1.000

2 0.591 1.000

3 0.743 0.704 1.000

o. Geomin rotated loadings. The rotated loadings are the linear
combination of variables that make up the factor. In addition to the factor
loadings, to completely interpret an oblique rotation one needs to take into
account both the factor pattern and the factor structure matrices (shown bellow)
and the correlations among the factors. Note that orthogonal rotations produce
only a single matrix, which gives the correlations between the variable and the
factor.

p. Geomin factor correlations. The factor correlations matrix gives the correlations between the factors. For example, the correlation between factor 1 and factor 2 is 0.591.

ESTIMATED RESIDUAL VARIANCESq

	ITEM13	ITEM14	ITEM15	ITEM16	ITEM17
1	0.332	0.354	0.388	0.543	0.388

ESTIMATED RESIDUAL VARIANCES

	ITEM18	ITEM19	ITEM20	ITEM21	ITEM22
1	0.325	0.408	0.623	0.459	0.555

ESTIMATED RESIDUAL VARIANCES

	ITEM23	ITEM24
1	0.194	0.373

q. Estimated residual variances. These are the variances of the observed variables after accounting for all of the variance in the

efa model.

Below are the standard errors for the geomin rotated loadings, factor correlations, and estimated residual variances. These values can be used to perform hypothesis tests and estimate confidence intervals.

S.E. GEOMIN ROTATED LOADINGS

	1	2	3
ITEM13	0.043	0.051	0.031
ITEM14	0.030	0.035	0.040
ITEM15	0.040	0.053	0.039
ITEM16	0.073	0.060	0.099
ITEM17	0.044	0.052	0.061
ITEM18	0.043	0.033	0.029
ITEM19	0.027	0.044	0.059
ITEM20	0.045	0.039	0.040
ITEM21	0.045	0.044	0.051
ITEM22	0.022	0.048	0.057
ITEM23	0.152	0.034	0.180
ITEM24	0.024	0.090	0.104

S.E. GEOMIN FACTOR CORRELATIONS

1 2 3

________ ________ ________

1 0.000

2 0.042 0.000

3 0.054 0.030 0.000

S.E. ESTIMATED RESIDUAL VARIANCES

ITEM13 ITEM14 ITEM15 ITEM16 ITEM17

________ ________ ________ ________ ________

1 0.021 0.021 0.020 0.025 0.018

S.E. ESTIMATED RESIDUAL VARIANCES

ITEM18 ITEM19 ITEM20 ITEM21 ITEM22

________ ________ ________ ________ ________

1 0.021 0.026 0.024 0.020 0.023

S.E. ESTIMATED RESIDUAL VARIANCES

ITEM23 ITEM24

________ ________

1 0.074 0.076

Below are the z-statistics (i.e. estimate/standard error) for the geomin rotated loadings, factor correlations, and estimated residual variances. These values can be compared to a normal

distribution to perform hypothesis tests.

Est./S.E. GEOMIN ROTATED LOADINGS

	1	2	3
ITEM13	20.050	-1.681	0.317
ITEM14	27.757	-0.623	-0.516
ITEM15	18.307	1.618	0.165
ITEM16	8.793	2.157	-0.700
ITEM17	11.680	5.280	1.377
ITEM18	2.136	23.045	0.416
ITEM19	-0.564	19.088	-1.602
ITEM20	2.190	14.343	-0.280
ITEM21	4.966	9.225	3.884
ITEM22	-0.017	10.650	3.567
ITEM23	0.812	0.314	4.426
ITEM24	-0.357	-0.088	7.694

Est./S.E. GEOMIN FACTOR CORRELATIONS

	1	2	3
1	0.000		
2	14.212	0.000	
3	13.724	23.775	0.000

Est./S.E. ESTIMATED RESIDUAL VARIANCES

ITEM13 ITEM14 ITEM15 ITEM16 ITEM17

________ ________ ________ ________ ________

1 16.029 16.792 19.200 21.564 21.324

Est./S.E. ESTIMATED RESIDUAL VARIANCES

ITEM18 ITEM19 ITEM20 ITEM21 ITEM22

________ ________ ________ ________ ________

1 15.574 15.812 26.444 23.392 24.618

Est./S.E. ESTIMATED RESIDUAL VARIANCES

ITEM23 ITEM24

________ ________

1 2.610 4.894

FACTOR STRUCTUREr

1 2 3

________ ________ ________

ITEM13 0.815 0.428 0.587

ITEM14 0.803 0.456 0.582

ITEM15 0.779 0.518 0.605

ITEM16 0.670 0.461 0.501

ITEM17 0.740 0.639 0.660

ITEM18 0.546 0.818 0.611

ITEM19 0.412 0.766 0.486

ITEM20 0.421 0.610 0.456

ITEM21 0.610 0.678 0.650

ITEM22 0.452 0.651 0.562

ITEM23 0.720 0.643 0.894

ITEM24 0.584 0.553 0.792

FACTOR DETERMINACIESs

1 2 3

___________ ___________ ___________

1 0.945 0.926 0.935

r. Factor Structure. With an oblique rotation, the factor structure matrix presents the correlations between the variables and the factors. For example, the correlation between item13 and factor 1 is 0.815. As noted above, the factor structure matrix is used along with the factor loadings and factor correlations to interpret the model.

s. Factor Determinacies are the correlations between the estimated

factor score and the factor.