How can I perform an exploratory factor analysis with categorical (or categorical and continuous) variables?

Authored by

stats writer

July 1, 2024

Exploratory factor analysis is a statistical technique used to identify underlying factors or dimensions within a set of variables. This method is commonly used in social sciences, psychology, and market research. It can also be applied to data sets consisting of both categorical and continuous variables.

To perform an exploratory factor analysis with categorical or mixed data, the first step is to choose an appropriate method based on the nature of the data. This can include principal component analysis, principal axis factoring, or maximum likelihood estimation.

Next, the researcher must decide on the number of factors to be extracted, which can be determined using various techniques such as the Kaiser criterion or scree plot. Once the number of factors is determined, the data is then subjected to rotation, which helps to simplify and interpret the factor structure.

The rotation method can also vary, with options such as Varimax, Promax, and Oblimin. After rotation, the researcher should examine the factor loadings to determine which variables are most strongly associated with each factor.

The final step is to interpret and label the factors based on the variables with high loadings. This process can help to identify underlying themes or concepts within the data and provide valuable insights for further analysis.

Overall, performing an exploratory factor analysis with categorical or mixed data requires careful consideration of the data, appropriate statistical methods, and thorough interpretation of the results to gain a deeper understanding of the underlying factors influencing the variables.

How can I perform an exploratory factor analysis with categorical (or categorical and continuous) variables? | Mplus FAQ

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 in Mplus with both
categorical and continuous variables. The dataset for this example includes data on 1428 college students and their instructors. You can download the dataset by clicking on
https://stats.idre.ucla.edu/wp-content/uploads/2016/02/fa_categorical.dat. The factor analysis will
include dichotomous variables, including
faculty sex (facsex) and faculty nationality (US citizen or foreign citizen,
facnat); ordered categorical variables, including faculty rank (facrank), student
rank (studrank) and grade (A, B, C, etc., grade);
and the continuous variables faculty salary (salary), years teaching at
the University of Texas (yrsut), and number of students in the class (nstud) in this analysis. These
variables were selected to represent a range of types of variables (i.e. dichotomous, ordered categorical, and continuous),
and do not necessarily form substantively meaningful factors.

Below is the Mplus input file for our model. The

categorical variables, both dichotomous and ordered categorical, are
listed in the categorical option of the variable command. Note
that the nominal option is used to specify that variables are unordered
categorical (none of the variables in this model are nominal so that option was
not used). We indicate the type of analysis that we would like to do,
exploratory factor analysis (efa), using the type option of the analysis command. The numbers after
efa indicate the minimum and maximum number of factors to be extracted.
By using 3 3, we indicate that we want only a three-factor solution. We
have done this to save space. We suggest that you use a reasonable range here, and each solution will be shown in the output.
For example, if we had 2 4 at the end of the option, we would see the
two-factor, three-factor and four-factor solution in the output. The missing
option of the variable command informs Mplus that in

the data file any missing values are represented by -9999.

Data::
File is https://stats.idre.ucla.edu/wp-content/uploads/2016/02/fa_categorical.dat ;
Variable:
Names are
facsex facnat facrank salary yrsut nstud studrank grade;
Missing are all (-9999) ;
Categorical are facsex facnat facrank studrank grade;
Analysis: Type = efa 3 3;

The output for this model is shown below. The results of this analysis are interpreted in a manner similar to an exploratory factor analysis with all continuous variables.

SUMMARY OF ANALYSIS

Number of groups 1
Number of observations 1428

Number of dependent variables 8

Number of independent variables 0

Number of continuous latent variables 0

Observed dependent variables

Continuous

SALARY YRSUT NSTUD

Binary and ordered categorical (ordinal)

FACSEX FACNAT FACRANK STUDRANK GRADE

Estimator WLSM

Rotation GEOMIN

Row standardization CORRELATION

Type of rotation OBLIQUE

Epsilon value Varies

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)

https://stats.idre.ucla.edu/wp-content/uploads/2016/02/fa_categorical.dat

Input data format FREE

SUMMARY OF DATA

Number of missing data patterns 3

COVARIANCE COVERAGE OF DATA

Minimum covariance coverage value 0.100

PROPORTION OF DATA PRESENT

Covariance Coverage

FACSEX FACNAT FACRANK SALARY YRSUT

_____ _____ _____ _____ _____

FACSEX 1.000

FACNAT 1.000 1.000

FACRANK 1.000 1.000 1.000

SALARY 1.000 1.000 1.000 1.000

YRSUT 0.945 0.945 0.945 0.945 0.945

NSTUD 1.000 1.000 1.000 1.000 0.945

STUDRANK 0.992 0.992 0.992 0.992 0.937

GRADE 1.000 1.000 1.000 1.000 0.945

Covariance Coverage

NSTUD STUDRANK GRADE

_____ _____ _____

NSTUD 1.000

STUDRANK 0.992 0.992

GRADE 1.000 0.992 1.000

SUMMARY OF CATEGORICAL DATA PROPORTIONS

FACSEX

Category 1 0.595

Category 2 0.405

FACNAT

Category 1 0.840

Category 2 0.160

FACRANK

Category 1 0.230

Category 2 0.270

Category 3 0.343

Category 4 0.156

STUDRANK

Category 1 0.171

Category 2 0.212

Category 3 0.250

Category 4 0.242

Category 5 0.125

GRADE

Category 1 0.005

Category 2 0.023

Category 3 0.204

Category 4 0.476

Category 5 0.291

RESULTS FOR EXPLORATORY FACTOR ANALYSIS

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

1 2 3 4 5

_____ _____ _____ _____ _____

1 2.821 1.763 1.107 0.809 0.590

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

6 7 8

_____ _____ _____

1 0.448 0.329 0.135

EXPLORATORY FACTOR ANALYSIS WITH 3 FACTOR(S):

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

Value 64.604*
Degrees of Freedom 7
P-Value 0.0000
Scaling Correction Factor 0.373
for MLR

* The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used
for chi-square difference tests. MLM, MLR and WLSM chi-square difference
testing is described in the Mplus Technical Appendices at www.statmodel.com.
See chi-square difference testing in the index of the Mplus User's Guide.

Chi-Square Test of Model Fit for the Baseline Model

Value 3734.662

Degrees of Freedom 28

P-Value 0.0000

CFI/TLI

CFI 0.984

TLI 0.938

Number of Free Parameters 24

RMSEA (Root Mean Square Error Of Approximation)

Estimate 0.076

MINIMUM ROTATION FUNCTION VALUE 0.22117

GEOMIN ROTATED LOADINGS

1 2 3

FACSEX -0.447 -0.655 0.004

FACNAT -0.457 0.374 -0.007

FACRANK 1.009 -0.016 0.069

SALARY 0.756 0.067 0.114

YRSUT 0.668 -0.324 -0.029

NSTUD -0.002 0.650 -0.289

STUDRANK -0.005 -0.007 0.767

GRADE 0.007 -0.001 0.274

GEOMIN FACTOR CORRELATIONS

1 2 3

_____ _____ _____

1 1.000

2 -0.121 1.000

3 -0.023 -0.207 1.000

ESTIMATED RESIDUAL VARIANCES

FACSEX FACNAT FACRANK SALARY YRSUT

_____ _____ _____ _____ _____

1 0.440 0.609 -0.024 0.430 0.398

ESTIMATED RESIDUAL VARIANCES

NSTUD STUDRANK GRADE

_____ _____ _____

1 0.417 0.409 0.925

S.E. GEOMIN ROTATED LOADINGS

1 2 3

_____ _____ _____

FACSEX 0.043 0.064 0.002

FACNAT 0.030 0.040 0.021

FACRANK 0.013 0.005 0.066

SALARY 0.016 0.028 0.058

YRSUT 0.021 0.038 0.050

NSTUD 0.001 0.049 0.062

STUDRANK 0.006 0.012 0.110

GRADE 0.028 0.045 0.050

S.E. GEOMIN FACTOR CORRELATIONS

1 2 3

_____ _____ _____

1 0.000

2 0.054 0.000

3 0.063 0.063 0.000

S.E. ESTIMATED RESIDUAL VARIANCES

FACSEX FACNAT FACRANK SALARY YRSUT

_____ _____ _____ _____ _____

1 0.083 0.034 0.025 0.021 0.027

S.E. ESTIMATED RESIDUAL VARIANCES

NSTUD STUDRANK GRADE

_____ _____ _____

1 0.076 0.166 0.025

Est./S.E. GEOMIN ROTATED LOADINGS

1 2 3

_____ _____ _____

FACSEX -10.394 -10.166 1.969

FACNAT -15.178 9.421 -0.350

FACRANK 77.424 -2.962 1.054

SALARY 48.518 2.412 1.979

YRSUT 31.658 -8.535 -0.590

NSTUD -2.271 13.136 -4.637

STUDRANK -0.859 -0.609 6.961

GRADE 0.253 -0.012 5.434

Est./S.E. GEOMIN FACTOR CORRELATIONS

1 2 3

_____ _____ _____

1 0.000

2 -2.260 0.000

3 -0.365 -3.308 0.000

Est./S.E. ESTIMATED RESIDUAL VARIANCES

FACSEX FACNAT FACRANK SALARY YRSUT

_____ _____ _____ _____ _____

1 5.331 17.653 -0.952 20.375 14.900

Est./S.E. ESTIMATED RESIDUAL VARIANCES

NSTUD STUDRANK GRADE

_____ _____ _____

1 5.487 2.458 37.379

FACTOR STRUCTURE

1 2 3

_____ _____ _____

FACSEX -0.368 -0.602 0.150

FACNAT -0.502 0.431 -0.074

FACRANK 1.009 -0.153 0.049

SALARY 0.745 -0.049 0.083

YRSUT 0.708 -0.399 0.022

NSTUD -0.074 0.710 -0.423

STUDRANK -0.022 -0.165 0.769

GRADE 0.001 -0.058 0.274

FACTOR DETERMINACIES

1 2 3

_____ _____ _____

1 1.012 0.847 0.800

See Also

Mplus Annotated Output: Factor Analysis