Package ‘EFAutilities’

December 3, 2016

Type Package

Title Utility Functions for Exploratory Factor Analysis

Version 0.1.0

Date 2016-11-27

Author Guangjian Zhang, Ge Jiang, Minami Hattori, Lauren Trichtinger

Maintainer Guangjian Zhang <gzhang3@nd.edu>

Depends R (>= 2.10)

Description A number of utility function for exploratory
factor analysis are included in this package. In particular, it computes standard errors for parameter estimates and factor correlations under a variety of conditions.

License GPL-2

LazyLoad yes

Imports GPArotation, mvtnorm

RoxygenNote 5.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2016-12-03 01:11:01

R topics documented:

BFI228 *Ordinal Data of the Big Five Inventory (BFI)*

Description

The BFI228 is part of the study on personality and relationship satisfication (Luo, 2005). The participants were 228 undergraduate students at a large public unverisity in the US. The data were participants' self ratings on the 44 items of the Big Five Inventory (John, Donahue, & kentle, 1991). These items are Likert variables: disagree strongly (1), disagree a little (2), neither agree nor diagree (3), agree a little (4), and agree strongly (5).

Usage

```
data(BFI228)
```

Format

The format is a n by p matrix of ordinal variables, where n is the number of participants (228) and p is the number of manifest variables (44).

Details

The variables were ordered such that indicators of the same factor are grouped together. Note that reverse-coded items are denoted by '_R'.

V01 to V08 are variables for the factor extraversion: talkative, researved_R, fullenergy, enthusiastic, quiet_R, assertive, shy_R, and outgoing.

V09 to V17 are variables for the factor agreeableness: findfault_R, helpful, quarrels_R, forgiving, trusting, cold_R, considerate, rude_R, and cooperative.

V18 to V26 are variables for the factor conscientiousness are: thorough, careless_R, reliable, disorganized_R, lazy_R, persevere, efficient, plans, and distracted_R.

V27 to V34 are variables for the factor neuroticism: blue, relaxed_R, tense, worries, emostable_R, moody, calm_R, and nervous.

V35 to V44 are variables for the factor openness: ideas, curious, ingenious, imagination, inventive, artistic, routine_R, reflect, nonartistic, and sophisticated.

References

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory versions 4a and 54. Berkeley, CA: University of California,Berkeley, Institute of Personality and Social Research.

Luo, S. (2005): unpublished study on personality traits and relationship satisfaction.

CPAI537 *Composite Scores of the Chinese Personality Assessment Inventory (CAPI)*

Description

CPAI537 is part of a big survey study on martial satisification (Luo et al., 2008). The participants were 537 urban Chinese couples in the first year of their marriage. Included here are 28 composite scores of the CPAI (Cheung et al., 1996) for the 537 wives.

Usage

```
data(CPAI537)
```

Format

The format is a n by p matrix, where n is the number of participants (537) and p is the number of manifest variables (28).

Details

The column names stand for the following variable names:
Nov - Novelty
Div - Diversity
Dit - Diverse thinking
LEA - Leadership
L_A - Logical orientation vs affective orientation
AES - Aesthetics
E_I - Extroversion-Introversion
ENT - Enterprise
RES - Responsibility
EMO - Emotionality
I_S - Inferiority vs. self-acceptance
PRA - Practical mindedness
O_P - Optimistic vs. pessimistic
MET - Meticulousness
FAC - Face
I_E - Internal control vs. external control
FAM - Family orientation
DEF - Defensiveness
G_M - Graciousness vs. meanness
INT - Interpersonal tolerance
S_S - Self orientation vs. social orientation
V_S - Veraciousness vs. slickness
T_M - Traditionalism vs. modernity
REN - Relationship orientation
SOC - Social sensitivity

DIS - Discipline
HAR - Harmony
T_E - Thrift vs. extravagence

References

Cheung, F. M., Leung, K., Fan, R., Song, W., Zhang, J., & Zhang, J. (1996). Development of the Chinese Personality Assessment Inventory (CPAI). Journal of Cross-Cultural Psychology, 27 ,181-199.

Luo, S., Chen, H., Yue, G., Zhang, G., Zhaoyang, R., & Xu, D. (2008). Predicting marital satisfaction from self, partner, and couple characteristics: Is it me, you, or us? Journal of Personality, 76 ,1231-1266.

---

efa *Exploratory Factor Analysis*

Description

Performs exploratory factor analysis under a variety of conditions. In particular, it provides standard errors for rotated factor loadings and factor correlations for normal variables, nonnormal continuous variables, and Likert scale variables with and without model error.

Usage

```
efa(x = NULL, factors = NULL, covmat = NULL, n.obs = NULL,
dist = "normal", fm = "ols", rtype = "oblique",
rotation = "CF-varimax", normalize = FALSE,
geomin.delta = NULL, MTarget = NULL, MWeight = NULL,
PhiWeight = NULL, PhiTarget = NULL, useorder = FALSE,
se = "information", Ib = 2000,
mnames = NULL, fnames = NULL, merror = "YES")
```

Arguments

`x`	The raw data: an n-by-p matrix where n is number of participants and p is the number of manifest variables.
`covmat`	A p-by-p manifest variable correlation matrix.
`n.obs`	The number of participants used in calculating the correlation matrix. This is not required when the raw data (x) is provided.
`factors`	The number of factors m: specified by a researcher; the default one is the Kaiser rule which is the number of eigenvalues of covmat larger than one.
`fm`	Factor extraction methods: 'ols' (default) and 'ml'
`dist`	Manifest variable distributions: 'normal'(default), 'continuous', 'ordinal' and 'ts'. 'normal' stands for normal distribution. 'continuous' stands for nonnormal continuous distributions. 'ordinal' stands for Likert scale variable. "ts" stands for distributions for time-series data.

merror	Model error: 'YES' (default) or 'NO'. In general, we expect our model is a parisonmious representation to the complex real world. Thus, some amount of model error is unavailable. When merror = 'NO', the efa model is assumed to fit perfectly in the population.
rtype	Factor rotation types: 'oblique' (default) and 'orthogonal'. Factors are correlated in 'oblique' rotation, and they are uncorrelated in 'orthogonal' rotation.
rotation	Factor rotation criteria: 'CF-varimax' (default), 'CF-quartimax', 'target', and 'geomin'. These rotation criteria can be used in both orthogonal and oblique rotation.
normalize	Row standardization in factor rotation: FALSE (default) and TRUE (Kaiser standardization).
se	Methods for estimating standard errors for rotated factor loadings and factor correlations, 'information', 'sandwich', 'bootstrap', and 'jackknife'. For normal variables and ml estimation, the default method is 'information'. For all other situations, the default method is 'sandwich'. In addition, the 'bootstrap' and 'jackknife' methods can be used for raw data.
geomin.delta	The controlling parameter in Geomin rotation, 0.01 as the default value.
MTarget	The p-by-m target matrix for the factor loading matrix in target rotation.
MWeight	The p-by-m weight matrix for the factor loading matrix in target rotation.
PhiWeight	The m-by-m target matrix for the factor correlation matrix in xtarget rotation
PhiTarget	The m-by-m weight matrix for the factor correlation matrix in xtarget rotation
useorder	Whether an order matrix is used for factor alignment: FALSE (default) and TRUE
Ib	The Number of bootstrap samples when se='bootstrap': 2000 (default)
mnames	Names of p manifest variables: Null (default)
fnames	Names of m factors: Null (default)

Details

The function `efa` conducts exploratory factor analysis (EFA) (Gorsuch, 1983) in a variety of conditions. Data can be normal variables, non-normal continuous variables, and Likert variables. Our implementation of EFA includes three major steps: factor extraction, factor rotation, and estimating standard errors for rotated factor loadings and factor correlations.

Factors can be extracted using two methods: maximum likelihood estimation (ml) and ordinary least squares (ols). These factor loading matrices are referred to as unrotated factor loading matrices. The ml unrotated factor loading matrix is obtained using `factanal`. The ols unrotated factor loading matrix is obtained using `optim` where the residual sum of squares is minimized. The starting values for communalities are squared mutiple correlations (SMCs).

Four rotation criteria (CF-varimax, CF-quartimax, geomin, and target) are available for both orthogonal rotation and oblique rotation (Browne, 2001). The factor rotation methods are achieved by calling functions in the package GPArotation. CF-varimax and CF-quartimax are members of the Crawford-Fugersion family (Crawford, & Ferguson, 1970) whose kappa = 1/p and kappa = 0, respectively. They are equivalent to varimax and quartimax rotation in orthogonal rotation. The equivalence does not carry over to oblique rotation, however. Althogh variamx and quartimax often

fail to give satisfactory results in oblique rotation, CF-varimax and CF-quartimax do give satisfactory results in many oblique rotation applications. CF-quartimax rotation is equivalent to direct oblimin rotation for oblique rotation. The target matrix in target rotation can either be a fully specified matrix or a partially specified matrix. Target rotation can be considered as a procedure which is located between EFA and CFA. In CFA, if a factor loading is specified to be zero, its value is fixed to be zero; if target rotation, if a factor loading is specified to be zeor, it is made to zero as close as possible.

Standard errors for rotated factor loadings and factor correlations are computed using a sandwich method (Ogasawara, 1998; Yuan, Marshall, & Bentler, 2002), which generalizes the augmented information method (Jennrich, 1974). The sandwich standard error are consistent estimates even when the data distribution is non-normal and model error exists in the population. Sandwich standard error estimates require a consistent estimate of the asymptotic covariance matrix of manifest variable correlations. Such estimates are described in Browne & Shapiro (1986) for non-normal continuous variables and in Yuan & Schuster (2013) for Likert variables. Estimation of the asymptotic covariance matrix of polychoric correlations is slow if the EFA model involves a large number of Likert variables.

When manifest variables are normally distributed (dist = 'normal') and model error does not exist (merror = 'NO'), the sandwich standard errors are equivalent to the usual standard error estimates, which come from the inverse of the information matrix. The information standard error estimates in EFA is available CEFA (Browne, Cudeck, Tateneni, & Mels, 2010) and SAS Proc Factor. Mplus (Muthen & Muthen, 2015) also implemented a version of sandwich standard errors for EFA, which are robust against non-normal distrubiton but not model error. Sandwich standard errors computed in `efa` tend to be larger than those computed in Mplus. Sandwich standard errors for non-normal distributions and with model error are equivalent to the infinitesimal jackknife standard errors described in Zhang, Preacher, & Jennrich (2012). Two computationally intensive standard error methods (se='bootstrap' and se='jackknife') are also implmented. More details on standard error estimation methods in EFA are documented in Zhang (2014).

Author(s)

Guangjian Zhang, Ge Jiang, Minami Hattori, and Lauren Trichtinger

References

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. Multivariate Behavioral Research, 36, 111-150.

Browne, M. W., Cudeck, R., Tateneni, K., & Mels, G. (2010). CEFA 3.04: Comprehensive Exploratory Factor Analysis. Retrieved from http://faculty.psy.ohio-state.edu/browne/.

Browne, M. W., & Shapiro, A. (1986). The asymptotic covariance matrix of sample correlation coefficients under general conditions. Linear Algebra and its applications, 82, 169-176.

Crawford, C. B., & Ferguson, G. A. (1970). A general rotation criterion and its use in orthogonal rotation. Psychometrika, 35 , 321-332.

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Jennrich, R. I. (1974). Simplified formulae for standard errors in maximum-likelihood factor analysis. British Journal of Mathematical and Statistical Psychology, 27, 122-131.

Jennrich, R. I. (2002). A simple general method for oblique rotation. Psychometrika, 67, 7-19.

Muthen, L. K., & Muthen, B. O. (1998-2015). Mplus user's guide (7th ed.). Los Angeles, CA: Muthen & Muthen.

Ogasawara, H. (1998). Standard errors of several indices for unrotated and rotated factors. Economic Review, Otaru University of Commerce, 49(1), 21-69.

Yuan, K., Marshall, L. L., & Bentler, P. M. (2002). A unified approach to exploratory factor analysis with missing data, nonnormal data, and in the presence of outliers. Psychometrika , 67 , 95-122.

Yuan, K.-H., & Schuster, C. (2013). Overview of statistical estimation methods. In T. D. Little (Ed.), The Oxford handbook of quantitative methods (pp. 361-387). New York, NY: Oxford University Press.

Zhang, G. (2014). Estimating standard errors in exploratory factor analysis. Multivariate Behavioral Research, 49, 339-353.

Zhang, G., Preacher, K. J., & Jennrich, R. I. (2012). The in

nitesimal jackknife with exploratory factor analysis. Psychometrika, 77 , 634-648.

Examples

```
#Examples using the data sets included in the packages:

data("CPAI537")    # Chinese personality assessment inventory (N = 537)

#1) normal, ml, oblique, CF-varimax, information, merror='NO'
efa(x=CPAI537,factors=4, fm='ml')

#2) continuous, ml, oblique, CF-quartimax, sandwich, merror='YES'
#efa(x=CPAI537, factors=4, dist='continuous',fm='ml',rotation='CF-quartimax', merror='YES')

#3) continuous, ols, orthogonal, geomin, sandwich, merror='Yes'
#efa(x=CPAI537, factors=4, dist='continuous',rtype= 'orthogonal',rotation='geomin', merror='YES')

#4) ordinal, ols, oblique, CF-varimax, sandwich, merror='Yes'
#data("BFI228")      # Big-five inventory (N = 228)
# For ordinal data, estimating SE with the sandwich method
#   can take time with a dataset with 44 variables
#reduced2 <- BFI228[,1:17] # extracting 17 variables corresponding to the first 2 factors
#efa(x=reduced2, factors=2, dist='ordinal', merror='YES')

#5) continuous, ml, oblique, Cf-varimax, jackknife
#efa(x=CPAI537,factors=4, dist='continuous',fm='ml', merror='YES', se= 'jackknife')
```

Index