simsem

R package for Simulated Structural Equation Modeling

simsem.org

www.crmda.ku.edu

Schoemann, A. M., Pornprasertmanit, S., & Miller, P. (2012)

simsem: SIMulated Structural Equation Modeling

The R package *simsem* has been developed to facilitate simulation and analysis of data within a structural equation modeling (SEM) framework. This package aims to help analysts create simulated data from hypotheses or analytic results from obtained data. The simulated data can be used for different purposes, such as power analysis, model fit evaluation, and methodological investigations.

simsem's capabilites can be broken down into four main categories: data generation, data modification, data analysis, and results summaries.

1. **Data generation:** Data can be generated from a normal or non-normal distribution. Both manifest and latent variables can have a non-normal distributions. Data can be generated from either a covariance matrix or through a series of linear equations. Generating and analysis models are specified using LISREL matrices.

2. **Data modification:** After data is generated, it can be modified prior to analyses. Creating missing data is a key form of modification. Missing data can be generated as MCAR, MAR, or based on two planned missing designs (the 2-method design and the 3 forms design). Attrition can be specified for longitudinal models as a MAR or MCAR process. Additionally, users can supply a function that modifies data in any other manner.

3. **Data analysis:** All models are fit using the lavaan package (Rosseel, 2012). Models can be estimated with robust ML estimators, and equality constraints can be included in models. *simsem* has several methods of handling missing data. FIML estimation with missing data is handled in lavaan; auxiliary variables can be easily included with FIML estimation. Multiple imputation is also available to handle missing data. Data can be imputed using either the Amelia (Honaker, King, & Blackwell, 2011) or mice (van Buuren & Groothuis-Oudshoorn, 2011) packages. Once data is imputed, each imputed data set is fit with lavaan, and results are combined using Rubin's Rules.

4. **Results summaries:** Results from a simulation can be easily summarized. Summaries include relative bias in parameters and standard errors, confidence interval coverage, power and much more.

simsem incorporates several advanced simulation methods, such as continuously varying parameters and population misspecification. With continuously varying parameters, parameters of the simulation (such as population parameters or sample size) can take on different values for each replication in a simulation. Population misfit can be specified such that the population has a given RMSEA.

Installing *simsem*

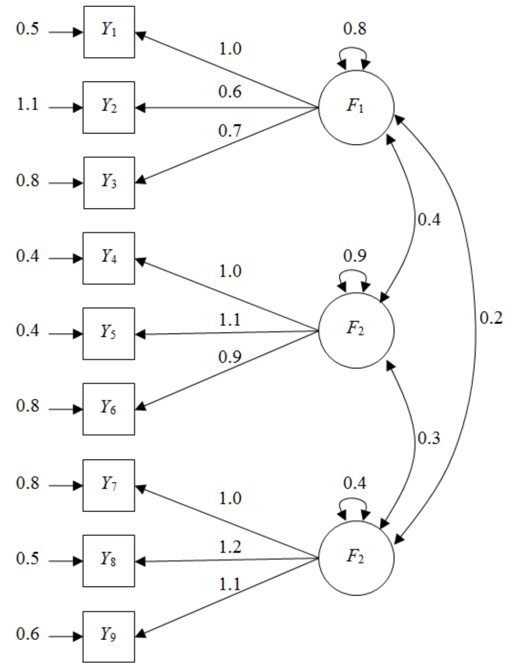
From CRAN:

```
install.packages("simsem")
```

The development version (from KRAN, KU R Archive Network)

```
install.packages("simsem", repos="http://rweb.quant.ku.edu/kran")
```

Example: Three factor CFA




```
library(simsem)

##Create two matrices for factor loadings
#First matrix is fixed and free parameters
loading <- matrix(0, 9, 3)
loading[1:3, 1] <- c(1, NA, NA)
loading[4:6, 2] <- c(1, NA, NA)
loading[7:9, 3] <- c(1, NA, NA)
#Second matrix is population values
loadingVal <- matrix(0, 9, 3)
loadingVal[2:3, 1] <- c(0.6, 0.7)
loadingVal[5:6, 2] <- c(1.1, 0.9)
loadingVal[8:9, 3] <- c(1.2, 1.1)
#bind two matrices together
LY <- bind(loading, loadingVal)

##Create two matrices for factor variances and
covariances.
facCov <- matrix(NA, 3, 3)
facCovVal <- diag(c(0.8, 0.9, 0.4))
facCovVal[lower.tri(facCovVal)] <- c(0.4, 0.2, 0.3)
facCovVal[upper.tri(facCovVal)] <- c(0.4, 0.2, 0.3)
#bind two matrices together. binds is for symmetric
matrices
PS <- binds(facCov, facCovVal)

#Create two matrices of error variances and covariances
errorCov <- diag(NA, 9)
errorCovVal <- diag(c(0.5, 1.1, 0.8, 0.4, 0.4, 0.8, 0.8, 0.5, 0.6))
#bind two matrices together. binds is for symmetric matrices
TE <- binds(errorCov, errorCovVal)

#Combine the relevant matrices together into a model specification
HS.Model <- model(LY=LY, PS=PS, TE=TE, modelType="CFA")

#Generate and analyze 1000 datasets based on the model with n=200
Output <- sim(1000, HS.Model, n=200)
#summarize results from the simulation
summary(Output)
```

References

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A Program for Missing Data. *Journal of Statistical Software, 45,* 1-47.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48,* 1-36.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45,* 1-67.