Package 'influence.SEM'

September 3, 2013

Type Package

Title Case Influence in Structural Equation Models

Version 1.0

Date 2013-09-07

Author Massimiliano Pastore & Gianmarco Altoe'

Depends lavaan

Maintainer Massimiliano Pastore <massimiliano.pastore@unipd.it>

Description influence.SEM provides a set of tools for evaluating several measures of case influence for Structural Equation Models following the approach of Pek and MacCallum (2011).

License GPL (>=2)

R topics documented:

---

influence.SEM-package   *Case Influence in Structural Equation Models*

Description

influence.SEM provides a set of tools for evaluating several measures of case influence for Structural Equation Models following the approach of Pek and MacCallum (2011). It analyses models that were estimated using the function sem of the package lavaan.

Details

Package:	influence.SEM
Type:	Package
Version:	1.0
Depends:	lavaan
Date:	2013-09-07
License:	GPL (>=2)

Author(s)

Massimiliano Pastore & Gianmarco Altoe'

Maintainer: Massimiliano Pastore <massimiliano.pastore@unipd.it>

References

Pek, J., MacCallum, R.C. (2011). Sensitivity Analysis in Structural Equation Models: Cases and Their Influence. *Multivariate Behavioral Research*, 46, 202-228.

Rosseel, Y. (2013). The lavaan tutorial. URL: http://lavaan.ugent.be/tutorial/tutorial. pdf.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Deltachi   *Chi-square difference.*

Description

Quantifies case influence on overall model fit by change in the test statistic

$$\Delta_{\chi_i^2} = \chi^2 - \chi_{(i)}^2$$

where χ^2 and $\chi_{(i)}^2$ are the test statistics obtained from original and deleted i samples.

Usage

```
Deltachi(model, data, ..., scaled = FALSE)
```

Arguments

model	A description of the user-specified model using the lavaan model syntax. See [lavaan](#) for more information.
data	A data frame containing the observed variables used in the model. If any variables are declared as ordered factors, this function will treat them as ordinal variables.
...	Additional parameters for [sem](#) function.
scaled	Logical, if TRUE the function uses the scaled χ^2 (Rosseel, 2013).

Value

Returns a vector of $\Delta_{\chi_i^2}$.

Note

If for observation i model does not converge or yelds a solution with negative estimated variances, the associated value of $\Delta_{\chi_i^2}$ is set to NA.

This function is a particular case of [fitinfluence](#), see example below.

Author(s)

Massimiliano Pastore

References

Pek, J., MacCallum, R.C. (2011). Sensitivity Analysis in Structural Equation Models: Cases and Their Influence. *Multivariate Behavioral Research*, 46, 202-228.

Rosseel, Y. (2013). The lavaan tutorial. URL: http://lavaan.ugent.be/tutorial/tutorial.pdf.

Examples

```
## not run: this example take several minutes
data("PDII")
model <- "
  F1 =~ y1+y2+y3+y4
"

# fit0 <- sem(model, data=PDII)
# Dchi <- Deltachi(model,data=PDII)
# plot(Dchi,pch=19,xlab="observations",ylab="Delta chisquare")

## not run: this example take several minutes
## an example in which the deletion of a case yelds a solution
## with negative estimated variances
```

```
model <- "
  F1 =~ x1+x2+x3
  F2 =~ y1+y2+y3+y4
  F3 =~ y5+y6+y7+y8
"

# fit0 <- sem(model, data=PDII)
# Dchi <- Deltachi(model,data=PDII)
# plot(Dchi,pch=19,xlab="observations",ylab="Delta chisquare")

## the case that produces negative estimated variances
# sem(model,data=PDII[-which(is.na(Dchi)),])

## same results
# Dchi <- fitinfluence("chisq",model,data=PDII)
# plot(Dchi,pch=19,xlab="observations",ylab="Delta chisquare")
```

---

explore.influence   *Explores case influence.*

Description

It explores case influence. Cases with extreme values of the considered measure of influence are reported. Extreme values are determined using the boxplot criterion (Tukey, 1977) or user-defined cut-offs. Cases for which deletion leads to a model that does not converge or yelds a solution with negative estimated variances are also reported. In addition, explore.influence provides a graphical representation of case influence.

Usage

```
explore.influence(x, cut.offs = 'default', plot = 'TRUE', cook = 'FALSE')
```

Arguments

x	A vector containing the influence of each case as returned by Deltachi, fitinfluence, genCookDist, Likedist or parinfluence functions.
cut.offs	A vector of two numeric elements containing the lower and the upper cut-offs to be considered. If default, the cut-offs are calculated according to the boxplot criterion for outliers (see also, cook).
plot	If TRUE (the default) a graphical representation of case influence is given.
cook	If TRUE, x is interpreted as a vector containing Cook's distances, and so the lower cut-off is forced to be greater or equal to zero.

Value

A list with the following components:

n	number of cases.
cook	logical, indicating if x is treated as a vector of Cook's distances.
cut.low	the lower cut-off.
cut.upp	the upper cut-off.
not.allowed	a vector containing cases with negative variance or not converging models.
less.cut.low	a vector containing cases with influence value less than the lower cut-off.
greater.cut.low	
	a vector containing cases with influence value greater than the upper cut-off.

Author(s)

Gianmarco Altoe'

References

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Examples

```
data("PDII")
model <- "
F1 =~ y1+y2+y3+y4
"
fit0 <- sem(model, data=PDII,std.lv=TRUE)
gCD <- genCookDist(model,data=PDII,std.lv=TRUE)
explore.influence(gCD,cook=TRUE)

##
## not run: this example take several minutes
model <- "
F1 =~ x1+x2+x3
F2 =~ y1+y2+y3+y4
F3 =~ y5+y6+y7+y8
"

# fit0 <- sem(model, data=PDII)
# FI <- fitinfluence('rmsea',model,PDII)
# explore.influence(FI)
```

---

fitinfluence   *Case influence on model fit.*

Description

This function evaluate the case's effect on a user-defined fit index.

Usage

```
fitinfluence(index, model, data, ...)
```

Arguments

index	A model fit index.
model	A description of the user-specified model using the lavaan model syntax. See [lavaan](#) for more information.
data	A data frame containing the observed variables used in the model. If any variables are declared as ordered factors, this function will treat them as ordinal variables.
...	Additional parameters for [sem](#) function.

Value

Returns a vector of case influence.

Note

If for observation i model does not converge or yelds a solution with negative estimated variances, the associated value of influence is set to NA.

Author(s)

Massimiliano Pastore

References

Pek, J., MacCallum, R.C. (2011). Sensitivity Analysis in Structural Equation Models: Cases and Their Influence. *Multivariate Behavioral Research*, 46, 202-228.

Examples

```
## not run: this example take several minutes
data("PDII")
model <- "
  F1 =~ y1+y2+y3+y4
"

# fit0 <- sem(model, data=PDII)
```

```
# FI <- fitinfluence("cfi",model,data=PDII)
# plot(FI,pch=19)

## not run: this example take several minutes
## an example in which the deletion of a case yelds a solution
## with negative estimated variances
model <- "
  F1 =~ x1+x2+x3
  F2 =~ y1+y2+y3+y4
  F3 =~ y5+y6+y7+y8
"

# fit0 <- sem(model, data=PDII)
# FI <- fitinfluence("rmsea",model,PDII)
# explore.influence(FI)
```

---

genCookDist	*Generalized Cook Distance.*

Description

Case influence on a vector of parameters may be quantified by generalized Cook's Distance (gCD; Cook 1977, 1986):

$$gCD_i = (\hat{\theta} - \hat{\theta}_{(i)})'_a \hat{\mathbf{\Sigma}}(\hat{\theta}_{(i)})^{-1}(\hat{\theta} - \hat{\theta}_{(i)})$$

where $\hat{\theta}$ and $\hat{\theta}_{(i)}$ are $l \times 1$ vectors of parameter estimates obained from the original and delete i samples, and $_a\hat{\mathbf{\Sigma}}(\hat{\theta}_{(i)})$ is the estimated asymptotic covariance matrix of the parameter estimates obtained from reduced sample.

Usage

```
genCookDist(model, data, ...)
```

Arguments

model	A description of the user-specified model using the lavaan model syntax. See lavaan for more information.
data	A data frame containing the observed variables used in the model. If any variables are declared as ordered factors, this function will treat them as ordinal variables.
...	Additional parameters for [sem](sem) function.

Value

Returns a vector of gCD_i.

Note

If for observation i model does not converge or yelds a solution with negative estimated variances, the associated value of gCD_i is set to NA.

Author(s)

Massimiliano Pastore

References

Cook, R.D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15-18.

Cook, R.D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society B*, 48, 133-169.

Pek, J., MacCallum, R.C. (2011). Sensitivity Analysis in Structural Equation Models: Cases and Their Influence. *Multivariate Behavioral Research*, 46, 202-228.

Examples

```
## not run: this example take several minutes
data("PDII")
model <- "
  F1 =~ y1+y2+y3+y4
"
# fit0 <- sem(model, data=PDII)
# gCD <- genCookDist(model,data=PDII)
# plot(gCD,pch=19,xlab="observations",ylab="Cook distance")


## not run: this example take several minutes
## an example in which the deletion of a case produces solution
## with negative estimated variances
model <- "
  F1 =~ x1+x2+x3
  F2 =~ y1+y2+y3+y4
  F3 =~ y5+y6+y7+y8
"

# fit0 <- sem(model, data=PDII)
# gCD <- genCookDist(model,data=PDII)
# plot(gCD,pch=19,xlab="observations",ylab="Cook distance")
```

Likedist   *Likelihood Distance.*

Description

A general model-based measure of case influence on model fit is likelihood distance (Cook, 1977, 1986; Cook & Weisberg, 1982) defined as

$$LD_i = 2[L(\hat{\theta}) - L(\hat{\theta}_{(i)})]$$

where $\hat{\theta}$ and $\hat{\theta}_{(i)}$ are the $k \times 1$ vectors of estimated model parameters on the original and deleted i samples, respectively, where $i = 1, \ldots, N$. The subscript (i) indicates that the estimate was computed on the sample excluding case i. $L(\hat{\theta})$ and $L(\hat{\theta}_{(i)})$ are the log-likelihoods based on the original and the deleted i samples, respectively.

Usage

```
Likedist(model, data, ...)
```

Arguments

model	A description of the user-specified model using the lavaan model syntax. See lavaan for more information.
data	A data frame containing the observed variables used in the model. If any variables are declared as ordered factors, this function will treat them as ordinal variables.
...	Additional parameters for sem function.

Value

Returns a vector of LD_i.

Note

If for observation i model does not converge or yelds a solution with negative estimated variances, the associated value of LD_i is set to NA.

Author(s)

Massimiliano Pastore, Gianmarco Altoe'

References

Cook, R.D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15-18.

Cook, R.D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society B*, 48, 133-169.

Cook, R.D., Weisberg, S. (1986). *Residuals and influence in regressions.* New York, NY: Chapman & Hall.

Pek, J., MacCallum, R.C. (2011). Sensitivity Analysis in Structural Equation Models: Cases and Their Influence. *Multivariate Behavioral Research*, 46, 202-228.

Examples

```
## not run: this example take several minutes
data("PDII")
model <- "
  F1 =~ y1+y2+y3+y4
"
# fit0 <- sem(model, data=PDII)
# LD <-Likedist(model,data=PDII)
# plot(LD,pch=19,xlab="observations",ylab="Likelihood distances")

## not run: this example take several minutes
## an example in which the deletion of a case yelds a solution
## with negative estimated variances
model <- "
  F1 =~ x1+x2+x3
  F2 =~ y1+y2+y3+y4
  F3 =~ y5+y6+y7+y8
"

# fit0 <- sem(model, data=PDII)
# LD <-Likedist(model,data=PDII)
# plot(LD,pch=19,xlab="observations",ylab="Likelihood distances")
```

---

parinfluence   *Case influence on model parameters.*

Description

Computes direction of change in parameter estimates with

$$\Delta\hat{\theta}_{ji} = \frac{\hat{\theta}_j - \hat{\theta}_{j(i)}}{[VAR(\hat{\theta}_{j(i)})]^{1/2}}$$

where $\hat{\theta}_j$ and $\hat{\theta}_{j(i)}$ are the parameter estimates obtained from original and deleted i samples.

Usage

```
parinfluence(parm, model, data, ...)
```

Arguments

parm	Single parameter or vector of parameters.
model	A description of the user-specified model using the lavaan model syntax. See [lavaan](#) for more information.
data	A data frame containing the observed variables used in the model. If any variables are declared as ordered factors, this function will treat them as ordinal variables.
...	Additional parameters for [sem](#) function.

Value

Return a list:

Dparm   Direction of change in parameter estimates.

THi   Parameter estimates.

Note

If for observation i model does not converge or yelds a solution with negative estimated variances, the associated values of $\Delta\hat{\theta}_{ji}$ are set to NA.

Author(s)

Massimiliano Pastore

References

Pek, J., MacCallum, R.C. (2011). Sensitivity Analysis in Structural Equation Models: Cases and Their Influence. *Multivariate Behavioral Research*, 46, 202-228.

Examples

```
## not run: this example take several minutes
data("PDII")
model <- "
  F1 =~ y1+y2+y3+y4
"
# fit0 <- sem(model, data=PDII)
# PAR <- c("F1=~y2","F1=~y3","F1=~y4")
# LY <- parinfluence(PAR,model,PDII)
# str(LY)
# explore.influence(LY$Dparm[,1])

## not run: this example take several minutes
## an example in which the deletion of a case yelds a solution
## with negative estimated variances
model <- "
  F1 =~ x1+x2+x3
  F2 =~ y1+y2+y3+y4
  F3 =~ y5+y6+y7+y8
"

# fit0 <- sem(model, data=PDII)
# PAR <- c("F2=~y2","F2=~y3","F2=~y4")
# LY <- parinfluence(PAR,model,PDII)

## not run: this example take several minutes
## dealing with ordinal data
data(Q)
model <- "
 F1 =~ it1+it2+it3+it4+it5+it6+it7+it8+it9+it10
```

```
"
# fit0 <- sem(model, data=Q, ordered=colnames(Q))
# LY <- parinfluence("F1=~it4",model,Q)
# explore.influence(LY$Dparm[,1])
```

PDII   *Industrialization and Democracy indicators.*

Description

Simulated data set from covariance matrix reported in Bollen (1989).

Usage

```
data(PDII)
```

Format

This data frame contains 75 obs. of 11 variables:

- x1: num, gross national product per capita.
- x2: num, consumption per capita.
- x3: num, percentage of the labor force in industrial occupations.
- y1: num, freedom of the press in 1960.
- y2: num, freedom of group opposition in 1960.
- y3: num, fairness of elections in 1960.
- y4: num, elective nature and effectiveness of the legislative body in 1960.
- y5: num, freedom of the press in 1965.
- y6: num, freedom of group opposition in 1965.
- y7: num, fairness of elections in 1965.
- y8: num, elective nature and effectiveness of the legislative body in 1965.

References

Bollen, K.A. (1989). *Structural Equations with latent Variables*, Wiley, NY.

Examples

```
data(PDII)
```

Q	*Simulated data set.*

Description

Simulated data set.

Usage

```
data(Q)
```

Format

This data frame contains 919 obs. of 10 ordinal discrete variables.

Examples

```
data(Q)
```

Index