Package ‘gendata’

March 1, 2015

Type Package

Title Generate and Modify Synthetic Datasets

Version 1.1

Date 2015-02-27

Author Francis Huang <`flh3@hotmail.com`>

Maintainer Francis Huang <`flh3@hotmail.com`>

Description Set of functions to create datasets using a correlation matrix.

License GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2015-03-01 08:20:17

R topics documented:

`gendata-package` *Generate Synthetic Datasets*

Description

Create synthetic datasets based on a correlation table. Additional functions can be used to rescale, transform, and reverse code variables.

Details

Package:	gendata
Type:	Package
Version:	1.1
Date:	2012-02-27
License:	GPL-3

Additional functions are for modifying the dataset.

genmvnorm: creates the dataset (generates a multivariate normal dataset).
recalib : for rescaling the dataset
dtrans : for giving a variable a new mean and standard deviation
revcode : for reverse coding a variable

Author(s)

Francis Huang

Maintainer: Francis Huang <flh3@hotmail.com>

References

Fan, X., Felsovalyi, A., Sivo, S., & Keenan, S. (2002). SAS for Monte Carlo studies: A guide for quantitative researchers. SAS Institute.

See Also

genmvnorm revcode dtrans recalib

`dtrans` *Data Transform*

Description

Transforms variables in a dataset with a specified mean and standard deviation.

Usage

```
dtrans(data, m, sd, rnd = F)
```

Arguments

`data`	name of your dataset
`m`	indicate a vector of desired means
`sd`	indicate a vector of desired standard deviations
`rnd`	Indicates if you want to round the numbers (no decimals). T or F.

Author(s)

Francis Huang

Examples

```
sdata<-genmvnorm(cor=c(.7,.2,.3),k=3,n=500,seed=12345)
cor(sdata)
summary(sdata)
#note: data are in z scores

s2<-dtrans(sdata,c(0,100,50),c(1,15,10),rnd=FALSE)
summary(s2)
sd(s2[,2])
sd(s2[,3])
#note: variables X2 and X3 are now rescaled with the appropriate means and standard deviations.
head(s2)

s2<-dtrans(sdata,c(0,100,50),c(1,15,10),rnd=TRUE)
#at times, you may want a dataset to not have decimals. use rnd=T.
head(s2)
```

`genmvnorm` *Genmvnorm*

Description

Generates a multivariate normal dataset based on a specified correlation matrix.

Usage

```
genmvnorm(cor, k, n, seed = F)
```

Arguments

`cor`	Can be a correlation matrix– e.g., data<-cor(xyz)– or the lower half of a correlation matrix, e.g., for a 3 variable dataset, data<-c(.7,.3,.2)– useful for creating datasets without having to specify both halves of the correlation matrix.
`k`	Indicate the number of variables in your dataset.
`n`	Indicate the number of observations in your new synthetic dataset.
`seed`	For reproducability of results, set a specific seed number.

Details

For creating synthetic datasets. Based on the SAS chapter by Fan et al. (2002).

Author(s)

Francis Huang

References

Based on:

Fan, X., Felsovalyi, A., Sivo, S., & Keenan, S. (2002). SAS for Monte Carlo studies: A guide for quantitative researchers. SAS Institute.

See Also

revcode dtrans recalib

Examples

```
sdata<-genmvnorm(cor=c(.7,.2,.3),k=3,n=500,seed=12345)
cor(sdata)
#dataset above uses the lower half of a correlation table
#     1  .7  .2
#     .7  1  .3
#     .2 .3   1
# Can also use a correlation table

data(iris)
dat<-cor(iris[,1:3])
dat
sdata<-genmvnorm(cor=dat,k=3,n=100,seed=123)
cor(sdata)

#example above uses the IRIS dataset.
```

`recalib` *Recalibrate (rescale) Variables*

Description

Rescale variables (one at a time) to have a new minimum and maximum value.

Usage

```
recalib(data, var, low, high)
```

Arguments

`data`	the dataset to use.
`var`	indicate the variable number (or variable name).
`low`	Indicate the new minimum value.
`high`	Indicate the new maximum value.

Details

Specify the rescaling of variables one at a time.

Author(s)

Francis Huang

See Also

genmvnorm revcode dtrans

Examples

```
sdata<-genmvnorm(cor=c(.7,.2,.3),k=3,n=500,seed=12345)
cor(sdata)
summary(sdata[,1])
#note the min and max of variable X1
#changes variable one to have a minimum of 10 and a maximum of 50
#correlations remain the same

s2<-recalib(sdata,1,10,50)
cor(s2)
summary(s2[,1])
#note revised values of variable X1
```

`revcode` *Reverse Coding Variables*

Description

Reverse codes variables

Usage

```
revcode(data, vars)
```

Arguments

`data`	indicates your dataset.
`vars`	indicates the variable number or name to reverse code.

Author(s)

Francis Huang

See Also

genmvnorm dtrans recalib

Index