Package ‘clikcorr’

June 3, 2016

Type Package
Title Censoring Data and Likelihood-Based Correlation Estimation
Version 1.0
Date 2016-6-1
Author Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie
Maintainer Yanming Li <liyanmin@umich.edu>
Description A profile likelihood based method of estimation and inference on the correlation coeffi-
cient of bivariate data with different types of censoring and missingness.
License GPL (>= 2)
Imports mvtnorm, stats
Repository CRAN
NeedsCompilation no
Date/Publication 2016-06-03 21:24:03

R topics documented:

clikcorr ................................................................. 2
est ................................................................. 4
lrt ................................................................. 5
ND ................................................................. 7
plot.clikcorr ....................................................... 8
splot .............................................................. 9
splot2 ........................................................... 10
sv ................................................................. 11

Index 13
clikcorr  
*Censoring data and LIKelihood-based CORRelation estimation and inference*

**Description**

A profile likelihood based method of estimation and hypothesis testing on the correlation coefficient of bivariate data with different types of censoring.

**Usage**

```r
clikcorr(data, lower1, upper1, lower2, upper2, cp = 0.95, dist = "n", df = 4, sv = NA, nlm = FALSE, ...)
```

---

**Arguments**

- `data`: a data frame name.
- `lower1`: the lower bound of the first of the two variables whose correlation coefficient to be calculated.
- `upper1`: the upper bound of the first of the two variables whose correlation coefficient to be calculated.
- `lower2`: the lower bound of the second of the two variables whose correlation coefficient to be calculated.
- `upper2`: the upper bound of the second of the two variables whose correlation coefficient to be calculated.
- `cp`: confidence level for the confidence interval.
- `dist`: working distribution. By default, `dist="n"` assuming the data from a bivariate normal distribution. Set `dist="t"` if the data are assumed generated from a bivariate t-distribution.
- `df`: degree of freedom of the bivariate t-distribution when `dist="t"`. By default `df=4`.
- `sv`: user specified starting values for the vector of (mean1, mean2, var1, corr, var2).
- `nlm`: use `nlm` as the optimization method to minimize the negative log (profile) likelihood. By default `nlm=FALSE` and `optim` is used to maximize the log (profile) likelihood.
- `x`: an object of class "clikcorr", i.e., a fitted model.
- `object`: an object of class "clikcorr", i.e., a fitted model.
- `...`: not used.
clikcorr

Details

clikcorr conducts point estimation and hypothesis testing on the correlation coefficient of bivariate data with different types of censoring.

Value

A list with components:

- `pairName`: variable names for the input paired data structure in the `clikcorr` class.
- `pairData`: a paired data structure in the `clikcorr` class.
- `dist`: Normal or t distribution.
- `df`: degree of freedom for t distribution.
- `coefficients`: maximum likelihood estimate (MLE) of the correlation coefficient.
- `Cov`: estimated variance covariance matrix.
- `Mean`: estimated means.
- `CI`: unsymmetric profile confidence interval for the estimated correlation coefficient.
- `P0`: p-value for likelihood ratio test with null hypothesis says that the true correlation coefficient equals zero.
- `logLik`: the value of the log likelihood at MLE.

Author(s)

Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie.

References

Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie (2016). Calculating Profile Likelihood Estimates of the Correlation Coefficient in the Presence of Left, Right or Interval Censoring and Missing Data.

Examples

data(ND)
logND <- log(ND)
logND1 <- logND[51:90,]

obj <- clikcorr(logND1, "t1_OCDD", "t2_OCDD", "t1_HxCDF_234678", "t2_HxCDF_234678")

# Not run:
clikcorr(logND, "t1_OCDD", "t2_OCDD", "t1_HxCDF_234678", "t2_HxCDF_234678")
clikcorr(logND, "t1_OCDD", "t2_OCDD", "t1_HxCDF_234678", "t2_HxCDF_234678", nlm=TRUE)
clikcorr(logND, "t1_OCDD", "t2_OCDD", "t1_HxCDF_234678", "t2_HxCDF_234678", method="BFGS")
est
censoring data and likelihood-based correlation estimation

Description

Provides point estimation and confidence interval for the correlation coefficient.

Usage

est(data, lower1, upper1, lower2, upper2, cp = 0.95, dist = "n", df = 4, sv = NA, nlm = FALSE, ...)

Arguments

data data frame name.
lower1 the lower bound of the first of the two variables whose correlation coefficient to be calculated.
upper1 the upper bound of the first of the two variables whose correlation coefficient to be calculated.
lower2 the lower bound of the second of the two variables whose correlation coefficient to be calculated.
upper2 the upper bound of the second of the two variables whose correlation coefficient to be calculated.
cp confidence level for the confidence interval.
dist working distribution. By default, dist="n" assuming the data from a bivariate normal distribution. Set dist="t" if the data are assumed generated from a bivariate t-distribution.
df degree of freedom of the bivariate t-distribution when dist="t". By default df=4.
sv user specified starting values for the vector of (mean1, mean2, var1, corr, var2).
nlm use nlm as the optimization method to minimize the negative log (profile) likelihood. By default nlm=FALSE and optim is used to maximize the log (profile) likelihood.
... not used.
Value

- Cor: maximum likelihood estimate (MLE) of the correlation coefficient.
- Cov: estimated variance covariance matrix.
- Mean: estimated means.
- LCL: lower bound of the profile confidence interval.
- UCL: upper bound of the profile confidence interval.

Author(s)

Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie.

References

Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie (2016). Calculating Profile Likelihood Estimates of the Correlation Coefficient in the Presence of Left, Right or Interval Censoring and Missing Data.

Examples

```r
data(ND)
logND <- log(ND)
logND1 <- logND[51:90,]
est(logND1, "t1_OCDD", "t2_OCDD", "t1_HxCDF_234678", "t2_HxCDF_234678")

## Not run:
est(logND, "t1_TCDD", "t2_TCDD", "t1_PeCDD", "t2_PeCDD")
est(logND, "t1_TCDD", "t2_TCDD", "t1_PeCDD", "t2_PeCDD", dist="t", nlm=TRUE)

## End(Not run)
```

Description

Provides likelihood ratio tests for making statistical inference about the correlation coefficient from bivariate censored/missing data.

Usage

```r
lrt(data, lower1, upper1, lower2, upper2, dist = "n", df = 4,
sv = NA, r0 = 0, nlm = FALSE, ...)
```
**Arguments**

- **data**: a data frame name.
- **lower1**: the lower bound of the first of the two variables whose correlation coefficient to be calculated.
- **upper1**: the upper bound of the first of the two variables whose correlation coefficient to be calculated.
- **lower2**: the lower bound of the second of the two variables whose correlation coefficient to be calculated.
- **upper2**: the upper bound of the second of the two variables whose correlation coefficient to be calculated.
- **dist**: working distribution. By default, dist="n" assuming the data from a bivariate normal distribution. Set dist="t" if the data are assumed generated from a bivariate t-distribution.
- **df**: degree of freedom of the bivariate t-distribution when dist="t". By default df=4.
- **sv**: user specified starting values for the vector of (mean1, mean2, var1, corr, var2).
- **r0**: correlation coefficient value under the null hypothesis. By default is 0.
- **nlm**: use nlm as the optimization method to minimize the negative log (profile) likelihood. By default nlm=FALSE and optim is used to maximize the log (profile) likelihood.

**Value**

- **Cor**: maximum likelihood estimate (MLE) of the correlation coefficient.
- **m11k**: value of the log likelihood function evaluated at the MLE.
- **m01k**: value of the log likelihood function evaluated at the r0.
- **p0**: p-value for likelihood ratio test with null hypothesis says that the true correlation coefficient equals r0.

**Author(s)**

Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie.

**References**

Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie (2016). Calculating Profile Likelihood Estimates of the Correlation Coefficient in the Presence of Left, Right or Interval Censoring and Missing Data.
Examples

data(ND)
logND <- log(ND)

lrt(logND, "t1_TCDD", "t2_TCDD", "t1_PeCDD", "t2_PeCDD")

## Not run:
lrt(logND, "t1_TCDD", "t2_TCDD", "t1_PeCDD", "t2_PeCDD", dist="t")

## End(Not run)

ND is an example data set extracted from National Health and Nutrition Examination Survey (NHANSE). The data set contains 100 samples and IDs and upper and lower bounds for 22 chemical compounds, including 7 dioxins, 9 furans, and 6 PCBs.

Usage

data(ND)

Format

A data frame with 1643 observations and 45 variables. Variables contain SEQ: ID; t1_TCDD: lower bound for dioxin TCDD; t2_TCDD: upper bound for dioxin TCDD; ... t1_PCB_189: lower bound for PCB_189 and t2_PCB_189: upper bound for PCB_189.

References

Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie (2016). Calculating Profile Likelihood Estimates of the Correlation Coefficient in the Presence of Left, Right or Interval Censoring and Missing Data.

Examples

data(ND)
plot.clikcorr

Graphical function for visualizing bivariate profile likelihood.

Description

Produces a plot of the profile log likelihood function.

Usage

## S3 method for class 'clikcorr'
plot(x, type = "l", lwd = 2, col = "red", ...)

Arguments

- `x`: a "clikcorr" object.
- `type`: line type.
- `lwd`: line weight.
- `col`: line color.
- `...`: not used.

Details

produces a plot of the profile log likelihood function.

Author(s)

Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie.

References

Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie (2016). Calculating Profile Likelihood Estimates of the Correlation Coefficient in the Presence of Left, Right or Interval Censoring and Missing Data.

Examples

data(ND)
logND <- log(ND)
logND1 <- logND[51:90,]

obj <- clikcorr(logND1, "t1_OCDD", "t2_OCDD", "t1_HxCDF_234678", "t2_HxCDF_234678")
plot(obj, type="o")

## Not run:
obj <- clikcorr(logND, "t1_OCDD", "t2_OCDD", "t1_HxCDF_234678", "t2_HxCDF_234678")
plot(obj, type="o", col="blue", lwd=1)

## End(Not run)
splot

Graphical function for visualizing bivariate censored and/or missing data

Description

Generates matrix of scatter plots for bivariate data with different types of censoring and missing.

Usage

splot(data, lower.list, upper.list, ti = ifelse(length(lower.list)>2,
paste("Scatter plots of", lower.list[1], "to", lower.list[length(lower.list)]),
paste("Scatter plot of", lower.list[1], "and", lower.list[2])),
legend = TRUE, cex = 1.5, ...)

Arguments

data a data frame name.
lower.list the lower bounds names in the data frame of the variables between which the scatter plots are to be generated.
upper.list the upper bounds names in the data frame of the variables between which the scatter plots are to be generated.
ti figure title.
legend figure legend.
cex symbol sizes.
... not used.

Details

Generates matrix of scatter plots for bivariate data with different types of censoring and missing.

Author(s)

Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie.

References

Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie (2016). Calculating Profile Likelihood Estimates of the Correlation Coefficient in the Presence of Left, Right or Interval Censoring and Missing Data.
Examples

data(ND)
logND <- log(ND)

splot(logND, c("t1_OCDD", "t1_TCDF", "t1_HxCDF_234678"),
     c("t2_OCDD", "t2_TCDF", "t2_HxCDF_234678"), ti="scatter plot matrix")
splot(logND, c("t1_OCDD", "t1_TCDF", "t1_HxCDF_234678"),
     c("t2_OCDD", "t2_TCDF", "t2_HxCDF_234678"), ti="scatter plot matrix", bg="gold")

splot2

Graphical function 2 for visualizing bivariate censored and/or missing data.

Description

Generates scatter plot for bivariate data with different types of censoring and missing.

Usage

splot2(data, lower1, upper1, lower2, upper2, pch = 21, bg = "cyan",
       xlab = lower1, ylab = lower2, ...)

Arguments

data a data frame name.
lower1 the lower bound name in the data frame of the first of the two variables for whose pairwise correlation to be calculated.
upper1 the upper bound name in the data frame of the first of the two variables for whose pairwise correlation to be calculated.
lower2 the lower bound name in the data frame of the second of the two variables for whose pairwise correlation to be calculated.
upper2 the upper bound name in the data frame of the second of the two variables for whose pairwise correlation to be calculated.
pch point character.
bg point background color.
xlab x axis label.
ylab y axis label.
... not used.

Details

Generates scatter plot for bivariate data with different types of censoring and missing.
Author(s)
Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie.

References
Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie (2016). Calculating Profile Likelihood Estimates of the Correlation Coefficient in the Presence of Left, Right or Interval Censoring and Missing Data.

Examples

data(ND)
logND <- log(ND)

splot2(logND, "t1_OCDD", "t2_OCDD", "t1_HxCDF_234678",
   "t2_HxCDF_234678", xlab="OCDD", ylab="HxCDF234678")

x <- logND[which(!is.na(logND[,14]) & !is.na(logND[,15])),14]
y <- logND[which(!is.na(logND[,26]) & !is.na(logND[,27])),26]
xhist = hist(x, plot=FALSE, breaks=10)
yhist = hist(y, plot=FALSE, breaks=10)

zones=matrix(c(2,0,1,3), ncol=2, byrow=TRUE)
layout(zones, widths=c(5/6,1/6), heights=c(1/6,5/6))
top = max(c(xhist$counts, yhist$counts))
par(mar=c(5,5,1,1))
splot2(logND, "t1_OCDD", "t2_OCDD", "t1_HxCDF_234678",
   "t2_HxCDF_234678", xlab="OCDD", ylab="HxCDF234678", cex=1.5)

par(mar=c(0,6,2,4))
barplot(xhist$counts, axes=FALSE, ylim=c(0, max(xhist$counts)), space=0)
par(mar=c(6,0,4,2))
barplot(yhist$counts, axes=FALSE, xlim=c(0, max(yhist$counts)), space=0, horiz=TRUE)

---
sv Calculating starting values for the vector of (mean1, mean2, var1, corr, var2) from completely observed data.

Description
Calculates starting values for the vector of (mean1, mean2, var1, corr, var2) from completely observed data.

Usage
sv(data, lower1, upper1, lower2, upper2)
Arguments

- **data**
  - a data frame name.

- **lower1**
  - the lower bound of the first variable of the two variables whose correlation coefficient to be calculated.

- **upper1**
  - the upper bound of the first variable of the two variables whose correlation coefficient to be calculated.

- **lower2**
  - the lower bound of the second variable of the two variables whose correlation coefficient to be calculated.

- **upper2**
  - the upper bound of the second variable of the two variables whose correlation coefficient to be calculated.

Details

Function `sv` calculates starting values for the vector of \((\text{mean1}, \text{mean2}, \text{var1}, \text{corr}, \text{var2})\) from completely observed data.

Value

- **mu1**
  - starting value for the mean parameter of the first variable.

- **mu2**
  - starting value for the mean parameter of the second variable.

- **var1**
  - starting value for the variance parameter of the first variable.

- **cor**
  - starting value for the correlation coefficient.

- **var2**
  - starting value for the variance parameter of the second variable.

Author(s)

Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie.

References

Yanming Li, Kerby Shedden, Brenda W. Gillespie and John A. Gillespie (2016). Calculating Profile Likelihood Estimates of the Correlation Coefficient in the Presence of Left, Right or Interval Censoring and Missing Data.

Examples

```r
data(ND)
logND <- log(ND)
sv(logND, "t1_TCDD", "t2_TCDD", "t1_PeCDD", "t2_PeCDD")
```
Index

*Topic dataset
   ND, 7
clikcorr, 2
est, 4
lrt, 5
ND, 7
plot.clikcorr, 8
print.clikcorr (clikcorr), 2
splot, 9
splot2, 10
summary.clikcorr (clikcorr), 2
sv, 11