

# Package ‘CorrBin’

July 2, 2014

**Title** Nonparametrics with clustered binary and multinomial data

**Version** 1.4

**Date** 2013-1-8

**Author** Aniko Szabo

**Maintainer** Aniko Szabo <aszabo@mcw.edu>

**Depends** R (>= 2.10), boot, combinat

**Suggests** geepack, lattice

**Description** This package implements non-parametric analyses for clustered binary data. The elements of the cluster are assumed exchangeable, and identical joint distribution (also known as marginal compatibility, or reproducibility) is assumed for clusters of different sizes. A trend test based on stochastic ordering is implemented.

**License** GPL (>= 2)

**LazyLoad** yes

**Repository** CRAN

**Repository/R-Forge/Project** corrbn

**Repository/R-Forge/Revision** 72

**Repository/R-Forge/DateTimeStamp** 2013-12-10 20:32:50

**Date/Publication** 2013-12-12 15:26:28

**NeedsCompilation** yes

**R topics documented:**

CorrBin-package	2
betabin.pdf	3
CBData	5
CMData	6
dehp	7
egde	8
GEE.trend.test	8
mc.est.CMData	10
mc.test.chisq.CMData	11
NOSTASOT	12
ran.CBData	14
read.CBData	15
read.CMData	16
RS.trend.test	17
shelltox	18
SO.LRT	19
SO.mc.est	20
SO.trend.test	21
soControl	23
trend.test	24
unwrap	25
<b>Index</b>	<b>27</b>

---

CorrBin-package	<i>Nonparameterics for correlated binary data</i>
-----------------	---

---

**Description**

This package implements nonparametric methods for analysing exchangeable binary data with variable cluster sizes with emphasis on trend testing. The input should specify the treatment group, cluster-size, and the number of responses (i.e. the number of cluster elements with the outcome of interest) for each cluster.

**Details**

Package:	CorrBin
Type:	Package
Version:	1.4
Date:	2013-01-13
License:	GPL 2
LazyLoad:	yes

- The `CBDData` and `read.CBDData` functions create a ‘`CBDData`’ object used by the analysis functions.
- `ran.CBDData` can be used to generate random data with prespecified mean response and within-cluster correlation.
- `mc.test.chisq` tests the assumption of marginal compatibility underlying all the methods, while `mc.est` estimates the distribution of the number of responses under marginal compatibility.
- Finally, `trend.test` performs three different tests for trend along the treatment groups.

### Author(s)

Aniko Szabo

Maintainer: Aniko Szabo <aszabo@mcw.edu>

### References

Szabo A, George EO. (2009) On the Use of Stochastic Ordering to Test for Trend with Clustered Binary Data. *Biometrika*

Stefanescu, C. & Turnbull, B. W. (2003) Likelihood inference for exchangeable binary data with varying cluster sizes. *Biometrics*, 59, 18-24

Pang, Z. & Kuk, A. (2007) Test of marginal compatibility and smoothing methods for exchangeable binary data with unequal cluster sizes. *Biometrics*, 63, 218-227

---

betabin.pdf

*Parametric distributions for correlated binary data*

---

### Description

`qpower.pdf` and `betabin.pdf` calculate the probability distribution function for the number of responses in a cluster of the q-power and beta-binomial distributions, respectively.

### Usage

```
betabin.pdf(p, rho, n)
```

```
qpower.pdf(p, rho, n)
```

### Arguments

<code>p</code>	numeric, the probability of success.
<code>rho</code>	numeric between 0 and 1 inclusive, the within-cluster correlation.
<code>n</code>	integer, cluster size.

**Details**

The pdf of the q-power distribution is

$$P(X = x) = \binom{n}{x} \sum_{k=0}^x (-1)^k \binom{x}{k} q^{(n-x+k)^\gamma},$$

$x = 0, \dots, n$ , where  $q = 1 - p$ , and the intra-cluster correlation

$$\rho = \frac{q^{2^\gamma} - q^2}{q(1 - q)}.$$

The pdf of the beta-binomial distribution is

$$P(X = x) = \binom{n}{x} \frac{B(\alpha + x, n + \beta - x)}{B(\alpha, \beta)},$$

$x = 0, \dots, n$ , where  $\alpha = p \frac{1-\rho}{\rho}$ , and  $\beta = (1 - p) \frac{1-\rho}{\rho}$ .

**Value**

a numeric vector of length  $n + 1$  giving the value of  $P(X = x)$  for  $x = 0, \dots, n$ .

**Author(s)**

Aniko Szabo, aszabo@mcw.edu

**References**

Kuk, A. A. (2004) litter-based approach to risk assesment in developmental toxicity studies via a power family of completely monotone functions *Applied Statistics*, 52, 51-61.

Williams, D. A. (1975) The Analysis of Binary Responses from Toxicological Experiments Involving Reproduction and Teratogenicity *Biometrics*, 31, 949-952.

**See Also**

[ran.CBDData](#) for generating an entire dataset using these functions

**Examples**

```
#the distributions have quite different shapes
#with q-power assigning more weight to the "all affected" event than other distributions
plot(0:10, betabin.pdf(0.3, 0.4, 10), type="o", ylim=c(0,0.34),
     ylab="Density", xlab="Number of responses out of 10")
lines(0:10, qpower.pdf(0.3, 0.4, 10), type="o", col="red")
```

---

CBData *Create a 'CBdata' object from a data frame.*

---

### Description

The CBData function creates an object of class *CBData* that is used in further analyses. It identifies the variables that define treatment group, clustersize and the number of responses.

### Usage

```
CBData(x, trt, clustersize, nresp, freq = NULL)
```

### Arguments

x	a data frame with one row representing a cluster or potentially a set of clusters of the same size and number of responses
trt	the name of the variable that defines treatment group
clustersize	the name of the variable that defines cluster size
nresp	the name of the variable that defines the number of responses in the cluster
freq	the name of the variable that defines the number of clusters represented by the data row. If NULL, then each row is assumed to correspond to one cluster.

### Value

A data frame with each row representing all the clusters with the same trt/size/number of responses, and standardized variable names:

Trt	factor, the treatment group
ClusterSize	numeric, the cluster size
NResp	numeric, the number of responses
Freq	numeric, number of clusters with the same values

### Author(s)

Aniko Szabo

### See Also

[read.CBData](#) for creating a CBData object directly from a file.

### Examples

```
data(shelltox)
sh <- CBData(shelltox, trt="Trt", clustersize="ClusterSize", nresp="NResp")
str(sh)
```

---

 CMDData

 Create a 'CMDdata' object from a data frame.
 

---

### Description

The CMDData function creates an object of class *CMDData* that is used in further analyses. It identifies the variables that define treatment group, clustersize and the number of responses for each outcome type.

### Usage

```
CMDData(x, trt, nresp, clustersize = NULL, freq = NULL)
```

### Arguments

x	a data frame with one row representing a cluster or potentially a set of clusters of the same size and number of responses for each outcome
trt	the name of the variable that defines treatment group
nresp	either a character vector with the names or a numeric vector with indices of the variables that define the number of responses in the cluster for each outcome type. If clustersize is NULL, then it will be calculated assuming that the nresp vector contains all the possible outcomes. If clustersize is given, then an additional category is created for the excess cluster members.
clustersize	the name or index of the variable that defines cluster size, or NULL. If NULL, its value will be calculated by adding the counts from the nresp variables. If defined, an additional response type will be created for the excess cluster members.
freq	the name or index of the variable that defines the number of clusters represented by the data row. If NULL, then each row is assumed to correspond to one cluster.

### Value

A data frame with each row representing all the clusters with the same trt/size/number of responses, and standardized variable names:

Trt	factor, the treatment group
ClusterSize	numeric, the cluster size
NResp.1--NResp.K+1	numeric, the number of responses for each of the K+1 outcome types
Freq	numeric, number of clusters with the same values

### Author(s)

Aniko Szabo

**See Also**

[read.CMData](#) for creating a CMData object directly from a file.

**Examples**

```
data(dehp)
dehp <- CMData(dehp, trt="Trt", nresp=c("NResp.1", "NResp.2", "NResp.3"))
str(dehp)
```

---

dehp

*Developmental toxicology study of DEHP in mice*

---

**Description**

This data set is based on a National Toxicology Program study on diethylhexyl phthalate, DEHP. Pregnant CD-1 mice were randomly assigned to receive 0, 250, 500, 1000, or 1500 ppm of DEHP in their feed during gestational days 0-17. The uterine contents of the mice were examined for toxicity endpoints prior to normal delivery. The possible outcomes are 1) malformation, 2) death or resorption, 3) no adverse event.

**Usage**

dehp

**Format**

A 'CMData' object, that is a data frame with the following variables

Trt	factor giving treatment group
ClusterSize	the size of the litter
NResp.1	the number of fetuses with a type 1 outcome (malformation)
NResp.2	the number of fetuses with a type 2 outcome (death or resorption)
NResp.3	the number of fetuses with a type 3 outcome (normal)
Freq	the number of litters with the given ClusterSize/NResp.1-NResp.3 combination

**Source**

National Toxicology Program, NTP Study TER84064.

**References**

Tyl, R. W., Price, C. J., Marr, M. C., and Kimmel, C. A. (1988). Developmental toxicity evaluation of dietary di(2-ethylhexy)phthalate in Fischer 344 rats and CD-1 mice. *Fundamental and Applied Toxicology* 10, 395-412.

---

 egde

*EGDE data*


---

### Description

The data set is based on a developmental toxicity experiment on the effect of ethylene glycol diethyl ether (EGDE) on fetal development of New Zealand white rabbits. In the study, four groups of pregnant does were randomly assigned to dose levels \$0, 25, 50\$, and \$100\$ milligrams per kilogram body weight of EGDE. For each litter and at each dose level, the adverse response used is the combined number of fetal malformation and fetal death.

### Usage

```
data(egde)
```

### Format

A 'CBData' object, that is a data frame with the following variables

Trt	factor giving treatment group
ClusterSize	the size of the litter
NResp	the number of affected fetuses
Freq	the number of litters with the given ClusterSize/NResp combination

### Source

Krewski, D., Zhu, Y. & Fung, K. (1995). Statistical analysis of overdispersed multinomial data from developmental toxicity studies. In *Statistics in Toxicology*, Ed. B. Morgan, pp. 151–179. New York: Oxford University Press.

### Examples

```
data(egde)
stripchart(I(NResp/ClusterSize)~Trt, cex=sqrt(egde$Freq), data=egde, pch=1,
           method="jitter", vertical=TRUE, ylab="Proportion affected")
```

---

 GEE.trend.test

*GEE-based trend test*


---

### Description

GEE.trend.test implements a GEE based test for linear increasing trend for correlated binary data.



## Usage

```
GEE.trend.test(cbddata, scale.method = c("fixed", "trend", "all"))
```

## Arguments

<code>cbddata</code>	a <a href="#">CBData</a> object
<code>scale.method</code>	character string specifying the assumption about the change in the overdispersion (scale) parameter across the treatment groups: "fixed" - constant scale parameter (default); "trend" - linear trend for the log of the scale parameter; "all" - separate scale parameter for each group.

## Details

The actual work is performed by the [geese](#) function of the [geepack](#) library. This function only provides a convenient wrapper to obtain the results in the same format as [RS.trend.test](#) and [S0.trend.test](#).

The implementation aims for testing for *increasing* trend, and a one-sided p-value is reported. The test statistic is asymptotically normally distributed, and a two-sided p-value can be easily computed if needed.

## Value

A list with components

<code>statistic</code>	numeric, the value of the test statistic
<code>p.val</code>	numeric, asymptotic one-sided p-value of the test

## Author(s)

Aniko Szabo, [aszabo@mcw.edu](mailto:aszabo@mcw.edu)

## See Also

[RS.trend.test](#), [S0.trend.test](#) for alternative tests; [CBData](#) for constructing a [CBData](#) object.

## Examples

```
data(shelltox)
GEE.trend.test(shelltox, "trend")
```

---

mc.est.CMData	<i>Distribution of the number of responses assuming marginal compatibility.</i>
---------------	---

---

### Description

The `mc.est` function estimates the distribution of the number of responses in a cluster under the assumption of marginal compatibility: information from all cluster sizes is pooled. The estimation is performed independently for each treatment group.

### Usage

```
## S3 method for class 'CMData'
mc.est(object, eps = 1e-06, ...)

mc.est(object, ...)

## S3 method for class 'CBData'
mc.est(object, ...)
```

### Arguments

<code>eps</code>	numeric; EM iterations proceed until the sum of squared changes fall below <code>eps</code>
<code>object</code>	a <code>CBData</code> or <code>CMData</code> object
<code>...</code>	other potential arguments; not currently used

### Details

The EM algorithm given by Stefanescu and Turnbull (2003) is used.

### Value

A data frame giving the estimated pdf for each treatment and clustersize. The probabilities add up to 1 for each `Trt/ClusterSize` combination. It has the following columns:

<code>Prob</code>	numeric, the probability of <code>NResp</code> responses in a cluster of size <code>ClusterSize</code> in group <code>Trt</code>
<code>Trt</code>	factor, the treatment group
<code>ClusterSize</code>	numeric, the cluster size
<code>NResp</code>	numeric, the number of responses

### Author(s)

Aniko Szabo

## References

Stefanescu, C. & Turnbull, B. W. (2003) Likelihood inference for exchangeable binary data with varying cluster sizes. *Biometrics*, 59, 18-24

## Examples

```
data(shelltox)
sh.mc <- mc.est(shelltox)

library(lattice)
xyplot(Prob~NResp|factor(ClusterSize), groups=Trt, data=sh.mc, subset=ClusterSize>0,
       type="l", as.table=TRUE, auto.key=list(columns=4, lines=TRUE, points=FALSE),
       xlab="Number of responses", ylab="Probability P(R=r|N=n)")
```

---

mc.test.chisq.CMData    *Test the assumption of reproducibility*

---

## Description

mc.test.chisq tests whether the assumption of reproducibility is violated in the data.

## Usage

```
## S3 method for class 'CMData'
mc.test.chisq(object, ...)

mc.test.chisq(object, ...)

## S3 method for class 'CBData'
mc.test.chisq(object, ...)
```

## Arguments

object	a <a href="#">CBData</a> or <a href="#">CMData</a> object
...	other potential arguments; not currently used

## Details

The assumption of marginal compatibility (AKA interpretability) implies that the marginal probability of response does not depend on clustersize. Stefanescu and Turnbull (2003), and Pang and Kuk (2007) developed a Cochran-Armitage type test for trend in the marginal probability of success as a function of the clustersize. mc.test.chisq implements a generalization of that test extending it to multiple treatment groups.

**Value**

A list with the following components:

<code>overall.chi</code>	the test statistic; sum of the statistics for each group
<code>overall.p</code>	p-value of the test
<code>individual</code>	a list of the results of the test applied to each group separately: <ul style="list-style-type: none"> <li>• <code>chi.sq</code> the test statistic for the group</li> <li>• <code>p</code> p-value for the group</li> </ul>

**Author(s)**

Aniko Szabo

**References**

- Stefanescu, C. & Turnbull, B. W. (2003) Likelihood inference for exchangeable binary data with varying cluster sizes. *Biometrics*, 59, 18-24
- Pang, Z. & Kuk, A. (2007) Test of marginal compatibility and smoothing methods for exchangeable binary data with unequal cluster sizes. *Biometrics*, 63, 218-227

**See Also**

[CBData](#) for constructing a CBData object, [mc.est](#) for estimating the distribution under marginal compatibility.

**Examples**

```
data(shelltox)
mc.test.chisq(shelltox)
```

---

NOSTASOT

*Finding the NOSTASOT dose*

---

**Description**

The NOSTASOT dose is the No-Statistical-Significance-Of-Trend dose – the largest dose at which no trend in the rate of response has been observed. It is often used to determine a safe dosage level for a potentially toxic compound.

**Usage**

```
NOSTASOT(cbddata, test = c("RS", "GEE", "GEEtrend", "GEEall", "SO"),
  exact = test == "SO", R = 100, sig.level = 0.05,
  control = soControl())
```

**Arguments**

<code>cbdata</code>	a <a href="#">CBData</a> object
<code>test</code>	character string defining the desired test statistic. See <a href="#">trend.test</a> for details.
<code>exact</code>	logical, should an exact permutation test be performed. See <a href="#">trend.test</a> for details.
<code>R</code>	integer, number of permutations for the exact test
<code>sig.level</code>	numeric between 0 and 1, significance level of the test
<code>control</code>	an optional list of control settings for the stochastic order ("SO") test, usually a call to <a href="#">soControl</a> . See there for the names of the settable control values and their effect.

**Details**

A series of hypotheses about the presence of an increasing trend overall, with all but the last group, all but the last two groups, etc. are tested. Since this set of hypotheses forms a closed family, one can test these hypotheses in a step-down manner with the same `sig.level` type I error rate at each step and still control the family-wise error rate.

The NOSTASOT dose is the largest dose at which the trend is not statistically significant. If the trend test is not significant with all the groups included, the largest dose is the NOSTASOT dose. If the testing sequence goes down all the way to two groups, and a significant trend is still detected, the lowest dose is the NOSTASOT dose. This assumes that the lowest dose is a control group, and this convention might not be meaningful otherwise.

**Value**

a list with two components

<code>NOSTASOT</code>	character string identifying the NOSTASOT dose.
<code>p</code>	numeric vector of the p-values of the tests actually performed.

The last element corresponds to all doses included, and will not be missing. p-values for tests that were not actually performed due to the procedure stopping are set to NA.

**Author(s)**

Aniko Szabo, [aszabo@mcw.edu](mailto:aszabo@mcw.edu)

**References**

Tukey, J. W.; Ciminera, J. L. & Heyse, J. F. (1985) Testing the statistical certainty of a response to increasing doses of a drug. *Biometrics* 41, 295-301.

**See Also**

[trend.test](#) for details about the available trend tests.

**Examples**

```
data(shelltox)
NOSTASOT(shelltox, test="RS")
```

---

ran.CBData	<i>Generate random correlated binary data</i>
------------	---

---

**Description**

ran.mc.CBData generates a random [CBData](#) object from a given two-parameter distribution.

**Usage**

```
ran.CBData(sample.sizes, p.gen.fun = function(g) 0.3,
           rho.gen.fun = function(g) 0.2, pdf.fun = qpwr.pdf)
```

**Arguments**

sample.sizes	a dataset with variables Trt, ClusterSize and Freq giving the number of clusters to be generated for each Trt/ClusterSize combination.
p.gen.fun	a function of one parameter that generates the value of the first parameter of pdf.fun ( $p$ ) given the group number.
rho.gen.fun	a function of one parameter that generates the value of the second parameter of pdf.fun ( $\rho$ ) given the group number.
pdf.fun	a function of three parameters ( $p$ , $\rho$ , $n$ ) giving the PDF of the number of responses in a cluster given the two parameters ( $p$ , $\rho$ ), and the cluster size ( $n$ ). Functions implementing two common distributions: the beta-binomial ( <a href="#">betabin.pdf</a> ) and q-power ( <a href="#">qpwr.pdf</a> ) are provided in the package.

**Details**

*p.gen.fun* and *rho.gen.fun* are functions that generate the parameter values for each treatment group; *pdf.fun* is a function generating the pdf of the number of responses given the two parameters  $p$  and  $\rho$ , and the cluster size  $n$ .

*p.gen.fun* and *rho.gen.fun* expect the parameter value of 1 to represent the first group, 2 - the second group, etc.

**Value**

a CBData object with randomly generated number of responses with sample sizes specified in the call.

**Author(s)**

Aniko Szabo, [aszabo@mcw.edu](mailto:aszabo@mcw.edu)

**See Also**

[betabin.pdf](#) and [qpower.pdf](#)

**Examples**

```
set.seed(3486)
ss <- expand.grid(Trt=0:3, ClusterSize=5, Freq=4)
#Trt is converted to a factor
rd <- ran.CBData(ss, p.gen.fun=function(g)0.2+0.1*g)
rd
```

---

read.CBData

*Read data from external file into a CBData object*

---

**Description**

A convenience function to read data from specially structured file directly into a CBData object.

**Usage**

```
read.CBData(file, with.freq = TRUE, ...)
```

**Arguments**

file	name of file with data. The first column should contain the treatment group, the second the size of the cluster, the third the number of responses in the cluster. Optionally, a fourth column could give the number of times the given combination occurs in the data.
with.freq	logical indicator of whether a frequency variable is present in the file
...	additional arguments passed to <a href="#">read.table</a>

**Value**

a CBData object

**Author(s)**

Aniko Szabo

**See Also**

[CBData](#)

---

`read.CMData`*Read data from external file into a CMData object*

---

### Description

A convenience function to read data from specially structured file directly into a CMData object. There are two basic data format options: either the counts of responses of all categories are given (and the cluster size is the sum of these counts), or the total cluster size is given with the counts of all but one category. The first column should always give the treatment group, then either the counts for each category (first option, chosen by setting `with.clustersize = FALSE`), or the size of the cluster followed by the counts for all but one category (second option, chosen by setting `with.clustersize = TRUE`). Optionally, a last column could give the number of times the given combination occurs in the data.

### Usage

```
read.CMData(file, with.clustersize = TRUE, with.freq = TRUE, ...)
```

### Arguments

<code>file</code>	name of file with data. The data in the file should be structured as described above.
<code>with.clustersize</code>	logical indicator of whether a cluster size variable is present in the file
<code>with.freq</code>	logical indicator of whether a frequency variable is present in the file
<code>...</code>	additional arguments passed to <a href="#">read.table</a>

### Value

a CMData object

### Author(s)

Aniko Szabo

### See Also

[CMData](#)



---

RS.trend.test	<i>Rao-Scott trend test</i>
---------------	-----------------------------

---

### Description

RS.trend.test implements the Rao-Scott adjusted Cochran-Armitage test for linear increasing trend with correlated data.

### Usage

```
RS.trend.test(cbddata)
```

### Arguments

cbddata            a [CBData](#) object

### Details

The test is based on calculating a *design effect* for each cluster by dividing the observed variability by the one expected under independence. The number of responses and the cluster size are then divided by the design effect, and a Cochran-Armitage type test statistic is computed based on these adjusted values.

The implementation aims for testing for *increasing* trend, and a one-sided p-value is reported. The test statistic is asymptotically normally distributed, and a two-sided p-value can be easily computed if needed.

### Value

A list with components

statistic            numeric, the value of the test statistic

p.val                numeric, asymptotic one-sided p-value of the test

### Author(s)

Aniko Szabo, aszabo@mcw.edu

### References

Rao, J. N. K. & Scott, A. J. A (1992) Simple Method for the Analysis of Clustered Data *Biometrics*, 48, 577-586.

### See Also

[SO.trend.test](#), [GEE.trend.test](#) for alternative tests; [CBData](#) for constructing a CBData object.

**Examples**

```
data(shelltox)
RS.trend.test(shelltox)
```

---

shelltox	<i>Shell Toxicology data</i>
----------	------------------------------

---

**Description**

This is a classical developmental toxicology data set. Pregnant banded Dutch rabbits were treated with one of four levels of a chemical. The actual doses are not known, instead the groups are designated as Control, Low, Medium, and High. Before term the animals were sacrificed, and the total number of fetuses, as well as the number affected by the treatment was recorded.

**Usage**

```
data(shelltox)
```

**Format**

A 'CBData' object, that is a data frame with the following variables

Trt	factor giving treatment group
ClusterSize	the size of the litter
NResp	the number of affected fetuses
Freq	the number of litters with the given ClusterSize/NResp combination

**Source**

Paul, S. R. (1982) Analysis of proportions of affected foetuses in teratological experiments. *Biometrics*, 38, 361-370.

This data set has been analyzed (and listed) in numerous papers, including

Rao, J. N. K. & Scott, A. J. (1992) A Simple Method for the Analysis of Clustered Data. *Biometrics*, 48, 577-586.

George, E. O. & Kodell, R. L. (1996) Tests of Independence, Treatment Heterogeneity, and Dose-Related Trend With Exchangeable Binary Data. *Journal of the American Statistical Association*, 91, 1602-1610.

Lee, S. (2003) Analysis of the Binary Littermate Data in the One-Way Layout. *Biometrical Journal*, 45, 195-206.

**Examples**

```
data(shelltox)
stripchart(I(NResp/ClusterSize)~Trt, cex=sqrt(shelltox$Freq), data=shelltox, pch=1,
           method="jitter", vertical=TRUE, ylab="Proportion affected")
```

---

SO.LRT	<i>Likelihood-ratio test statistic</i>
--------	--

---

**Description**

SO.LRT computes the likelihood ratio test statistic for stochastic ordering against equality assuming marginal compatibility for both alternatives. Note that this statistic does not have a  $\chi^2$  distribution, so the p-value computation is not straightforward. The [SO.trend.test](#) function implements a permutation-based evaluation of the p-value for the likelihood-ratio test.

**Usage**

```
SO.LRT(cbddata, control = soControl())
```

**Arguments**

cbdata	a CBData object
control	an optional list of control settings, usually a call to <a href="#">soControl</a> . See there for the names of the settable control values and their effect.

**Value**

The value of the likelihood ratio test statistic is returned with two attributes:

ll0	the log-likelihood under $H_0$ (equality)
ll1	the log-likelihood under $H_a$ (stochastic order)

**Author(s)**

Aniko Szabo

**See Also**

[SO.trend.test](#), [soControl](#)

**Examples**

```
data(shelltox)
LRT <- SO.LRT(shelltox, control=soControl(max.iter = 100, max.directions = 50))
LRT
```

SO.mc.est

*Order-restricted MLE assuming marginal compatibility***Description**

SO.mc.est computes the nonparametric maximum likelihood estimate of the distribution of the number of responses in a cluster  $P(R = r|n)$  under a stochastic ordering constraint. Umbrella ordering can be specified using the turn parameter.

**Usage**

```
SO.mc.est(cpdata, turn = 1, control = soControl())
```

**Arguments**

cpdata	an object of class <a href="#">CBData</a> .
turn	integer specifying the peak of the umbrella ordering (see Details). The default corresponds to a non-decreasing order.
control	an optional list of control settings, usually a call to <a href="#">soControl</a> . See there for the names of the settable control values and their effect.

**Details**

Two different algorithms: EM and ISDM are implemented. In general, ISDM (the default) should be faster, though its performance depends on the tuning parameter `max.directions`: values that are too low or too high slow the algorithm down.

SO.mc.est allows extension to an umbrella ordering:  $D_1 \geq^{st} \dots \geq^{st} D_k \leq^{st} \dots \leq^{st} D_n$  by specifying the value of  $k$  as the turn parameter. This is an experimental feature, and at this point none of the other functions can handle umbrella orderings.

**Value**

A list with components:

Components Q and D are unlikely to be needed by the user.

MLEst	data frame with the maximum likelihood estimates of $P(R_i = r n)$
Q	numeric matrix; estimated weights for the mixing distribution
D	numeric matrix; directional derivative of the log-likelihood
loglik	the achieved value of the log-likelihood
converge	a 2-element vector with the achieved relative error and the performed number of iterations

**Author(s)**

Aniko Szabo, [aszabo@mcw.edu](mailto:aszabo@mcw.edu)

## References

Szabo A, George EO. (2009) On the Use of Stochastic Ordering to Test for Trend with Clustered Binary Data. *Biometrika*

## See Also

[soControl](#)

## Examples

```
data(shelltox)
ml <- SO.mc.est(shelltox, control=soControl(eps=0.01, method="ISDM"))
attr(ml, "converge")

require(lattice)
panel.cumsum <- function(x,y,...){
  x.ord <- order(x)
  panel.xyplot(x[x.ord], cumsum(y[x.ord]), ...)}

xyplot(Prob~NResp|factor(ClusterSize), groups=Trt, data=ml, type="s",
       panel=panel.superpose, panel.groups=panel.cumsum,
       as.table=TRUE, auto.key=list(columns=4, lines=TRUE, points=FALSE),
       xlab="Number of responses", ylab="Cumulative Probability R(R>=r|N=n)",
       ylim=c(0,1.1), main="Stochastically ordered estimates\n with marginal compatibility")
```

---

SO.trend.test

*Likelihood ratio test of stochastic ordering*

---

## Description

Performs a likelihood ratio test of stochastic ordering versus equality using permutations to estimate the null-distribution and the p-value. If only the value of the test statistic is needed, use [SO.LRT](#) instead.

## Usage

```
SO.trend.test(cbdata, R = 100, control = soControl())
```

## Arguments

cbdata	a <a href="#">CBData</a> object.
R	an integer – the number of random permutations for estimating the null distribution.
control	an optional list of control settings, usually a call to <a href="#">soControl</a> . See there for the names of the settable control values and their effect.

**Details**

The test is valid only under the assumption that the cluster-size distribution does not depend on group. During the estimation of the null-distribution the group assignments of the clusters are permuted keeping the group sizes constant; the within-group distribution of the cluster-sizes will vary randomly during the permutation test.

The default value of R is probably too low for the final data analysis, and should be increased.

**Value**

A list with the following components

LRT	the value of the likelihood ratio test statistic. It has two attributes: l10 and l11 - the values of the log-likelihood under $H_0$ and $H_a$ respectively.
p.val	the estimated one-sided p-value.
boot.res	an object of class "boot" with the detailed results of the permutations. See <a href="#">boot</a> for details.

**Author(s)**

Aniko Szabo, aszabo@mcw.edu

**References**

Szabo A, George EO. (2009) On the Use of Stochastic Ordering to Test for Trend with Clustered Binary Data.

**See Also**

[SO.LRT](#) for calculating only the test statistic, [soControl](#)

**Examples**

```
data(shelltox)
set.seed(45742)
sh.test <- SO.trend.test(shelltox, R=10, control=soControl(eps=0.1, max.directions=25))
sh.test

#a plot of the resampled LRT values
#would look better with a reasonable value of R
null.vals <- sh.test$boot.res$t[,1]
hist(null.vals, breaks=10, freq=FALSE, xlab="Test statistic", ylab="Density",
      main="Simulated null-distribution", xlim=range(c(0,20,null.vals)))
points(sh.test$LRT, 0, pch="*", col="red", cex=3)
```

---

soControl *Control values for order-constrained fit*

---

### Description

The values supplied in the function call replace the defaults and a list with all possible arguments is returned. The returned list is used as the control argument to the [mc.est](#), [SO.LRT](#), and [SO.trend.test](#) functions.

### Usage

```
soControl(method = c("ISDM", "EM"), eps = 0.005, max.iter = 5000,
  max.directions = 0, start = ifelse(method == "ISDM", "H0", "uniform"),
  verbose = FALSE)
```

### Arguments

method	a string specifying the maximization method
eps	a numeric value giving the maximum absolute error in the log-likelihood
max.iter	an integer specifying the maximal number of iterations
max.directions	an integer giving the maximal number of directions considered at one step of the ISDM method. If zero or negative, it is set to the number of non-empty cells. A value of 1 corresponds to the VDM algorithm.
start	a string specifying the starting setup of the mixing distribution; "H0" puts weight only on constant vectors (corresponding to the null hypothesis of no change), "uniform" puts equal weight on all elements. Only a "uniform" start can be used for the "EM" algorithm.
verbose	a logical value; if TRUE details of the optimization are shown.

### Value

a list with components for each of the possible arguments.

### Author(s)

Aniko Szabo [aszabo@mcw.edu](mailto:aszabo@mcw.edu)

### See Also

[mc.est](#), [SO.LRT](#), [SO.trend.test](#)

### Examples

```
# decrease the maximum number iterations and
# request the "EM" algorithm
soControl(method="EM", max.iter=100)
```

---

trend.test	<i>Test for increasing trend with correlated binary data</i>
------------	--

---

### Description

The `trend.test` function provides a common interface to the trend tests implemented in this package: [SO.trend.test](#), [RS.trend.test](#), and [GEE.trend.test](#). The details of each test can be found on their help page.

### Usage

```
trend.test(cbdata, test = c("RS", "GEE", "GEEtrend", "GEEall", "SO"),
  exact = test == "SO", R = 100, control = soControl())
```

### Arguments

<code>cbdata</code>	a <a href="#">CBData</a> object
<code>test</code>	character string defining the desired test statistic. "RS" performs the Rao-Scott test ( <a href="#">RS.trend.test</a> ), "SO" performs the stochastic ordering test ( <a href="#">SO.trend.test</a> ), "GEE", "GEEtrend", "GEEall" perform the GEE-based test ( <a href="#">GEE.trend.test</a> ) with constant, linearly modeled, and freely varying scale parameters, respectively.
<code>exact</code>	logical, should an exact permutation test be performed. Only an exact test can be performed for "SO". The default is to use the asymptotic p-values except for "SO".
<code>R</code>	integer, number of permutations for the exact test
<code>control</code>	an optional list of control settings for the stochastic order ("SO") test, usually a call to <a href="#">soControl</a> . See there for the names of the settable control values and their effect.

### Value

A list with two components and an optional "boot" attribute that contains the detailed results of the permutation test as an object of class [boot](#) if an exact test was performed.

<code>statistic</code>	numeric, the value of the test statistic
<code>p.val</code>	numeric, asymptotic one-sided p-value of the test

### Author(s)

Aniko Szabo, [aszabo@mcw.edu](mailto:aszabo@mcw.edu)

### See Also

[SO.trend.test](#), [RS.trend.test](#), and [GEE.trend.test](#) for details about the available tests.



**Examples**

```
data(shelltox)
trend.test(shelltox, test="RS")
set.seed(5724)
#R=50 is too low to get a good estimate of the p-value
trend.test(shelltox, test="RS", R=50, exact=TRUE)
```

---

unwrap

*Unwrap a clustered object*


---

**Description**

unwrap.CBData is a utility function that reformats a CBData object so that each row is one observation (instead of one or more clusters). A new 'ID' variable is added to indicate clusters. This form can be useful for setting up the data for a different package.

unwrap.CMData is a utility function that reformats a CMData object so that each row is one observation (instead of one or more clusters). A new 'ID' variable is added to indicate clusters. This form can be useful for setting up the data for a different package.

**Usage**

```
unwrap(object, ...)

## S3 method for class 'CBData'
unwrap(object, ...)

## S3 method for class 'CMData'
unwrap(object, ...)
```

**Arguments**

object	a <a href="#">CBData</a> object
...	other potential arguments; not currently used

**Value**

For unwrap.CBData: a data frame with one row for each cluster element (having a binary outcome) with the following standardized column names

Trt	factor, the treatment group
ClusterSize	numeric, the cluster size
ID	factor, each level representing a different cluster
Resp	numeric with 0/1 values, giving the response of the cluster element

For unwrap.CMData: a data frame with one row for each cluster element (having a multinomial outcome) with the following standardized column names

Trt	factor, the treatment group
ClusterSize	numeric, the cluster size
ID	factor, each level representing a different cluster
Resp	numeric with integer values giving the response type of the cluster element

**Author(s)**

Aniko Szabo

**Examples**

```
data(shelltox)
ush <- unwrap(shelltox)
head(ush)
data(dehp)
dehp.long <- unwrap(dehp)
head(dehp.long)
```

# Index

## \*Topic **IO**

read.CBData, 15  
read.CMData, 16

## \*Topic **classes**

CBData, 5  
CMData, 6

## \*Topic **datasets**

dehp, 7  
egde, 8  
shelltox, 18

## \*Topic **distribution**

betabin.pdf, 3  
ran.CBData, 14

## \*Topic **file**

read.CBData, 15  
read.CMData, 16

## \*Topic **htest**

GEE.trend.test, 8  
mc.test.chisq.CMData, 11  
NOSTASOT, 12  
RS.trend.test, 17  
SO.LRT, 19  
SO.trend.test, 21  
trend.test, 24

## \*Topic **manip**

CBData, 5  
CMData, 6  
unwrap, 25

## \*Topic **models**

GEE.trend.test, 8  
mc.est.CMData, 10  
SO.mc.est, 20  
soControl, 23

## \*Topic **nonparametric**

CorrBin-package, 2  
mc.est.CMData, 10  
NOSTASOT, 12  
RS.trend.test, 17  
SO.LRT, 19

SO.mc.est, 20

SO.trend.test, 21

trend.test, 24

## \*Topic **package**

CorrBin-package, 2

betabin.pdf, 3, 14, 15

boot, 22, 24

CBData, 3, 5, 9–15, 17, 20, 21, 24, 25

CMData, 6, 10, 11, 16

CorrBin (CorrBin-package), 2

CorrBin-package, 2

dehp, 7

egde, 8

GEE.trend.test, 8, 17, 24

geese, 9

mc.est, 3, 12, 23

mc.est (mc.est.CMData), 10

mc.est.CMData, 10

mc.test.chisq, 3

mc.test.chisq (mc.test.chisq.CMData), 11

mc.test.chisq.CMData, 11

NOSTASOT, 12

qpower.pdf, 14, 15

qpower.pdf (betabin.pdf), 3

ran.CBData, 3, 4, 14

read.CBData, 3, 5, 15

read.CMData, 7, 16

read.table, 15, 16

RS.trend.test, 9, 17, 24

shelltox, 18

SO.LRT, 19, 21–23

SO.mc.est, [20](#)

SO.trend.test, [9](#), [17](#), [19](#), [21](#), [23](#), [24](#)

soControl, [13](#), [19–22](#), [23](#), [24](#)

trend.test, [3](#), [13](#), [24](#)

unwrap, [25](#)