Package 'NADA2'

March 23, 2026

Type Package

Title Data Analysis for Censored Environmental Data

Version 2.0.2

Date 2026-03-23

Maintainer Paul Julian <pauljulianphd@gmail.com>

URL <https://github.com/SwampThingPaul/NADA2>

BugReports <https://github.com/SwampThingPaul/NADA2/issues>

Description Contains methods described by Dennis Helsel in
his book ``Statistics for Censored Environmental Data
using Minitab and R'' (2011) and courses and videos at
<<https://practicalstats.com>>. This package incorporates functions
of NADA and adds new functionality.

License MIT + file LICENSE

Depends R (>= 3.6), EnvStats (>= 2.4)

Imports grDevices, graphics, stats, utils, methods, fitdistrplus,
multcomp, survminer, mgcv, cenGAM, vegan, coin, survival,
NbClust

Suggests knitr, rmarkdown, bestglm, car, nlme, rms

Encoding UTF-8

LazyLoad yes

LazyData yes

RoxygenNote 7.3.3

NeedsCompilation no

Author Paul Julian [aut, cre],
Dennis Helsel [aut, cph],
Lopaka Lee [aut, cph]

Repository CRAN

Date/Publication 2026-03-23 13:20:02 UTC

Contents

---

anosimPlot   *Permutation Analysis of Similarity (anosim) for Censored Data*

Description

Plots the permutation histogram and test statistic produced by an anosim (nonparametric multivariate) test of differences between groups.

Usage

```
anosimPlot(
  ano.out,
  hcol = "light blue",
  title = "Histogram of anosim permutations"
)
```

Arguments

ano.out	an anosim object. See details and example
hcol	color of histogram
title	title of histogram

Value

Plots a histogram of the permutation test statistics representing the null hypothesis along with the observed test statistic from the data. The p-value is the proportion of test statistics equal to or more extreme than the observed test statistic.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Oksanen, J., Guillaume, F., 2018. Vegan: ecological diversity. CRAN R-Project. https://cran.r-project.org/package=vegan

See Also

vegan::anosim

Examples

```
data(PbHeron)

# ROS model for each group
PbHeron.high <- with(subset(PbHeron,DosageGroup=="High"),ros(Blood,BloodCen))
PbHeron.high <- data.frame(PbHeron.high)
PbHeron.high$DosageGroup <- "High"

PbHeron.low <- with(subset(PbHeron,DosageGroup=="Low"),ros(Blood,BloodCen))
PbHeron.low <- data.frame(PbHeron.low)
PbHeron.low$DosageGroup <- "Low"

PbHeron.ros=rbind(PbHeron.high,PbHeron.low)

# ANOSIM analysis
library(vegan)
PbHeron.anosim <- with(PbHeron.ros,anosim(modeled,DosageGroup))
summary(PbHeron.anosim)

# Plot
anosimPlot(PbHeron.anosim)
```

---

atrazine   *Atrazine concentrations in Nebraska ground water*

Description

From the NADA R-Package.

Atrazine concentrations in a series of Nebraska wells before (June) and after (September) the growing season.

Objective is to determine if concentrations increase from June to September. There is one detection limit, at 0.01 ug/L. Used in Chapters 4, 5, and 9 of the Helsel (2011) book.

Usage

```
data(atrazine)
```

Format

An object of class `data.frame` with 24 rows and 4 columns.

Source

Junk et al., 1980. Areal, vertical, and temporal differences in ground water chemistry: II. Journal of Environmental Quality. 9(3) 479 - 483.

References

Helsel, D.R., 2011. 2011. Statistics for Censored Environmental Data Using Minitab and R. 2nd ed. John Wiley and Sons, USA, N.J.

ATS   *Akritas-Theil-Sen line for censored data*

Description

Computes Kendall's tau and the Akritas-Theil-Sen (ATS) line for censored data, along with the test that the slope (and Kendall's tau) equal zero. For one x variable regression.

Usage

```
ATS(
  y.var,
  y.cen,
  x.var,
  x.cen = rep(0, length(x.var)),
  LOG = TRUE,
  retrans = FALSE,
  xlabel = NULL,
  ylabel = NULL,
  printstat = TRUE,
  drawplot = TRUE
)
```

Arguments

y.var	The column of y (response variable) values plus detection limits
y.cen	The y-variable indicators, where 1 (or TRUE) indicates a detection limit in the y.var column, and 0 (or FALSE) indicates a detected value in y.var.
x.var	The column of x (explanatory variable) values plus detection limits

x.cen	The x-variable indicators, where 1 (or TRUE) indicates a detection limit in the x.var column, and 0 (or FALSE) indicates a detected value in x.var.
LOG	Indicator of whether to compute the ATS line in the original y units, or for their logarithms. The default is to use the logarithms (LOG = TRUE). To compute in original units, specify the option LOG = FALSE (or LOG = 0).
retrans	Indicator of whether to retransform the plot and line back to original Y-variable units. Not needed when LOG = FALSE. When retrans = FALSE & LOG = TRUE the plot is drawn with logY units (default). When retrans = TRUE & LOG = TRUE the plot is drawn with original Y units.
xlabel	Custom label for the x axis of plots. Default is x variable column name.
ylabel	Custom label for the y axis of plots. Default is y variable column name.
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.
drawplot	Logical TRUE/FALSE option of whether to draw plots or not. Default is TRUE

Value

Coefficients (intercept and slope) for the ATS line are printed, along with Kendall's tau correlation coefficient, test statistic S, and the (single) p-value for the test that tau and slope both equal zero. A scatterplot with the fitted trend-line superimposed is also drawn.

References

Akritas, M.G., Murphy, S.A., LaValley, M.P., 1995. The Theil-Sen Estimator With Doubly Censored Data and Applications to Astronomy. Journal of the American Statistical Association 90, 170–177. https://doi.org/10.2307/2291140

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Examples

```
## Not run:
# Both y and x are censored
data(PbHeron)
with(PbHeron, ATS(Blood, BloodCen, Kidney, KidneyCen))

# x is not censored
data(Brumbaugh)
with(Brumbaugh,ATS(Hg, HgCen, PctWetland))

## End(Not run)
```

Description

Computes Kendall's tau and the Akritas-Theil-Sen (ATS) line for censored data. Is called by censeaken because it is much faster than the ATS function. It is not expected to be of much use to users on its own. The x variable (time) may not be censored.

Usage

```
ATSmini(y.var, y.cen, x.var)
```

Arguments

y.var   The column of y (response variable) values plus detection limits.

y.cen   The y-variable indicators, where 1 (or TRUE) indicates a detection limit in the
            y.var column, and 0 (or FALSE) indicates a detected value in y.var.

x.var   The column of x (time for a trend test) values.

Value

Returns the intercept and slope (ATS line), tau (Kendall's tau), p-value and S-value (test statistic).

References

Akritas, M.G., Murphy, S.A., LaValley, M.P., 1995. The Theil-Sen Estimator With Doubly Censored Data and Applications to Astronomy. Journal of the American Statistical Association 90, 170–177. https://doi.org/10.2307/2291140

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Examples

```
## Not run:
# x may not be censored.  Use the ATS function when x is censored.
data(Brumbaugh)

with(Brumbaugh, ATSmini(Hg, HgCen, SedLOI))

## End(Not run)
```

---

bestaic   *Find the lowest AIC multiple regression model*

Description

Computes (2^k-1) censored regression models and their AIC statistics. Prints out the lowest AIC models and the terms used.

Usage

```
bestaic(y.var, cen.var, x.vars, LOG = TRUE, n.models = 10)
```

Arguments

y.var	The column of y (response variable) values plus detection limits.
cen.var	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y.var column, and 0 (or FALSE) indicates a detected value is in y.var.
x.vars	One or more uncensored explanatory variable(s). See Details
LOG	Indicator of whether to compute the regression in the original y units, or on their logarithms. The default is to use the logarithms (LOG = TRUE). To compute in original units, specify the option LOG = FALSE (or LOG = 0).
n.models	The number of models with their AIC values to be printed in the console window. All (2^k-1) models are computed internally. This sets how many "best" (lowest AIC) models have output printed to the console.

Details

x.vars: If 1 x variable only, enter its name. If multiple x variables, enter the name of a data frame of columns of the x variables. No extra columns unused in the regression allowed. Create this by x.frame <- data.frame (Temp, Flow, Time) for 3 variables (temperature, flow and time).

AIC of each model is printed from lowest to highest AIC to help evaluate the 'best' regression model. n.models determines how many lines of model info is printed.

LOG: The default is that the Y variable will be log transformed (LOG = TRUE).

Value

Prints number of x.vars, lists x.vars and AIC values.

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

See Also

survival::survreg

Examples

```
## Not run:
data(Brumbaugh)

# Multiple regression
bestaic(Brumbaugh$Hg, Brumbaugh$HgCen, Brumbaugh[, c("SedMeHg","PctWetland", "SedAVS")])

## End(Not run)
```

---

binaryClust	*Cluster Matrix of Binary Censored Data*

Description

Performs clustering of a matrix of 0s and 1s, ie. the censoring indicator columns for multiple variables. If there are multiple detection limits within a column first convert the 0/1 designated to be Below vs Above the highest detection limit in the column. Detection limits may differ among columns.

Usage

```
binaryClust(
  data,
  method = "ward.D2",
  group = NULL,
  ncluster = NULL,
  plotncluster = TRUE,
  clustIndex = "all"
)
```

Arguments

data	A data frame containing only the 0/1 columns, one column per chemical parameter.
method	Method of forming clusters. The default is `"ward.D2"`, which is appropriate for a variety of types of data. Another appropriate option is "average" – average distances between cluster centers. See the vegan package for other possible clustering methods.
group	Optional grouping variable. If used, sites being clustered will be represented by their group name, rather than by the row number.
ncluster	Optional number of clusters to be differentiated on the graph. Clusters are fenced off with rectangles.
plotncluster	default is `TRUE` logical flags to add identification of clusters on dendrogram
clustIndex	Optional, if not specified, potential number of clusters will be determined based on the mean best number of clusters across all indicies. For a specific index, see details

Details

If a specific index is desired to determine the best number of clusters see NbClust::NbClust for index values.

Value

Prints a cluster dendrogram based on clustering method and outputs a list of clusters and hierarchical cluster analysis results

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Examples

```
data(PbHeron)

# without group specified
binaryClust(PbHeron[,4:15])

# With Group argument
binaryClust(PbHeron[,4:15],group=PbHeron$DosageGroup)
```

---

binaryDiss   *Binary dissimilarity coefficient matrix*

Description

Computes a simple matching dissimilarity coefficient

Usage

```
binaryDiss(dat.frame)
```

Arguments

dat.frame   A data frame containing only the 0/1 columns.

Value

Returns a binary dissimilarity matrix.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

See Also

vegan::designdist

Examples

```
data(PbHeron)

binaryDiss(PbHeron$LiverCen)
```

---

binaryMDS   *Plot Nonmetric Multidimensional Scaling of binary censored data*

Description

Plots an NMDS of a matrix of 0s and 1s, the censoring indicator columns for multiple variables, to discern the pattern of data below vs. above the detection limit. With multiple detection limits within a column, re-censoring to the highest limit in the column must be done prior to running this function. May have different censoring levels in different columns.

Usage

```
binaryMDS(dat.frame, group = NULL, title = NULL, legend.pos = "bottomleft")
```

Arguments

dat.frame	A data frame containing only the columns of 0/1 values.
group	Optional grouping variable. Sites will be represented by different colored symbols for each group.
title	Optional title for the NMDS graph.
legend.pos	When group is specified, determines the location of the legend on the graph showing the colors representing each group's data. Default is "bottomleft". Alternatives are "topright" and "centerleft", etc.

Details

Binary data may not provide sufficient information to discern differences in location on the plot if sample size is small. Prior to running this analysis it is suggested to consult best analysis practice when performing NMDS. As a rule of thumb, an NMDS ordination with a stress value around or above 0.2 is deemed suspect and a stress value approaching 0.3 indicates that the ordination is arbitrary. Stress values equal to or below 0.1 are considered fair, while values equal to or below 0.05 indicate good fit.

Value

Plots an NMDS of censored data represented as the binary Above vs Below a detection limit for each parameter.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

See Also

vegan::metaMDS

Examples

```
## Not run:
data(PbHeron)

# without group specified
binaryMDS(PbHeron[,4:15])

# With Group argument
binaryMDS(PbHeron[,4:15],group=PbHeron$DosageGroup)

## End(Not run)
```

---

binarySim   *Binary similarity coefficient matrix*

Description

Computes a simple matching similarity coefficient

Usage

```
binarySim(dat.frame)
```

Arguments

dat.frame   A data frame containing only the 0/1 columns.

Value

Returns a binary similarity matrix.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

See Also

vegan::designdist

Examples

```
data(PbHeron)

binarySim(PbHeron$LiverCen)
```

Brumbaugh   *Brumbaugh*

Description

From the NADA R-Package.

Mercury concentrations in fish across the United States.

Objective is to determine if mercury concentrations differ by watershed land use. Can concentrations be related to water and sediment characteristics of the streams?

There are three detection limits, at 0.03, 0.05, and 0.10 ug/g wet weight. Used in Chapters 10, 11 and 12 of the Helsel (2011) book.

Usage

```
data(Brumbaugh)
```

Format

An object of class data.frame with 133 rows and 14 columns.

Source

Brumbaugh et al., 2001, USGS Biological Science Report BSR-2001-0009.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Cadmium   *Cadmium*

Description

From the NADA R-Package.

Cadmium concentrations in fish for two regions of the Rocky Mountains.

Objective is to determine if concentrations are the same or different in fish livers of the two regions. There are four detection limits, at 0.2, 0.3, 0.4, and 0.6 ug/L. Used in Chapter 9 of the NADA book.

Usage

```
data(Cadmium)
```

Format

An object of class data.frame with 19 rows and 3 columns.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

---

cboxplot   *Draws censored boxplots*

Description

Draws boxplots for left-censored data with one ore more detection limit(s). Portions below the maximum detection limit(s) are not shown by default, as their percentiles are not known.

Usage

```
cboxplot(
  x1,
  x2,
  xgroup = NULL,
  LOG = FALSE,
  show = FALSE,
  ordr = NULL,
  Ylab = yname,
  Xlab = gname,
  Title = NULL,
  dl.loc = "topright",
  dl.col = "red",
```

```
      dl.lwd = 2,
      dl.lty = "longdash",
      bxcol = "white",
      Ymax = NULL,
      minmax = FALSE,
      printstat = FALSE,
      Hlines = NULL,
      ...
   )
```

Arguments

x1	The column of x (response variable) values plus detection limits.
x2	The x-variable censoring indicators, where 1 (or TRUE) indicates a detection limit in the y1 column, and 0 (or FALSE) indicates a detected value in y1.
xgroup	An optional column of a grouping variable. Draws side-by-side boxplots if this variable is present.
LOG	TRUE/FALSE indicator of whether to plot the Y axis data on the original scale (FALSE) or log scale (TRUE).
show	TRUEVFALSE indicator of whether to show estimated values computed using ROS for the portion of the box below the maximum DL (TRUE), or just leave the lower portion blank (FALSE).
ordr	A vector indicating the order of boxes to be drawn on the boxplot, if not in alphabetical order (the default). Example: for 4 boxplots for groups A, B, C, D, to change the order to the reverse type ordr = c(4, 3, 2, 1). Example 2: To change the order to A, C, D, B, type ordr = c(1, 3, 4, 2)
Ylab	Y axis label text, if something is wanted other than the Y variable name in the dataset.
Xlab	X axis label text, if something is wanted other than the group variable name in the dataset.
Title	Text to show as the graph title. Default is blank.
dl.loc	Location indicator of where to plot the "MaxDL=" text on some versions of the plot. Possible entries are "topleft", "topright", "topcenter", and the corresponding "bottom" text.
dl.col	Color of the max detection limit line(s), and the legend text stating the max DL. Default is "red", but all standard R colors may be used.
dl.lwd	line wide of max detection limit line(s).
dl.lty	line type of max detection limit line(s).
bxcol	Color for interior of boxplots. Specify just one color if all boxes are to be the same color. If a different color is desired for each of three boxplots, as one example, use bxcol = c("red", "white", "blue") etc.
Ymax	Maximum Y value to be shown on the plot. Used to cut off high outliers on plot and better show the bulk of the boxplots.

minmax   TRUE/FALSE indicator of whether to draw outliers individually. Default is to
                show outliers. Setting minmax = TRUE will draw the whiskers out to the max and
                min of the dataset.

printstat   Logical TRUE/FALSE option of whether to print the resulting statistics in the con-
                sole window, or not. Default is TRUE.

Hlines   Data to add horizontal reference lines to the boxplot. Required input is a data
                frame of 4 columns. See Details.

...   Additional arguments passed to the default boxplot function

Details

If maximum detection limits vary among groups, separate maxDL lines will be drawn for each
group's boxplot. If one group has fewer than 3 detected observations its boxplot will not be drawn.
Its detection limits will not count when computing the maximum limit. However, if only one box-
plot is drawn for the entire dataset by not specifying a group variable, the detection limits from the
portion that is the mostly ND group will be used when computing the maximum limit.

The reuired input to draw additional horizontal lines (Hlines option) is a data frame with 4 columns
of input, one row per horizontal line. More than one line may be drawn. Column one is the Y
axis value for the line. Column 2 is the line color, column 3 is the line type (lty) and column 4 is
the text to be added just above the line. To add one line at a value of 40, for example, use Hlines
= yline, after defining yline = data.frame(c(40, "purple", "dotted", "New Health Std")). To draw
two lines, define yline as yline = data.frame(matrix(c(40, "purple", "dotted", "New Health Std", 70,
"blue", "longdash", "Old Health Std"), ncol = 4, byrow=TRUE))) . If no text is wanted use " " for
the column 4 entry for that line. See ?par under lty for standard line types.

Value

Prints a boxplot with detection limit identified and a concatenated list of the maximum detection
limit for each group.

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John
Wiley & Sons, USA, N.J.

Examples

```
data(PbHeron)
cboxplot(PbHeron$Liver,PbHeron$LiverCen,PbHeron$Group)
```

---

cen1way   *Peto-Peto one-factor test*

Description

Performs a Peto-Peto nonparametric test of differences in cdfs between groups. If more than two
groups, the test is followed by a nonparametric multiple comparison test. Uses the BH method of
adjusting p-values.

Usage

```
cen1way(x1, x2, group, mcomp.method = "BH", printstat = TRUE)
```

Arguments

x1	The column of data values plus detection limits
x2	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y1 column, and 0 (or FALSE) indicates a detected value in y1.
group	Grouping or factor variable. Can be either a text or numeric value indicating the group assignment.
mcomp.method	One of the standard methods for adjusting p-values for multiple comparisons. Type ?p.adjust for the list of possible methods. Default is Benjamini-Hochberg "BH" false discover rate.
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Value

A list of summary statistics for each group evaluated containing the following components:

- N Number of samples
- PctND Percentage of non-detects
- KMmean Kaplan-Meier estimate of the mean
- KMsd Kaplan-Meier estimate of standard deviation
- KMmedian Kaplan-Meier estmate of the median

Peto-Peto test results including Chi-Squared value, degrees of freedom and p-value of the test.

If more than two groups, p-values of the pairwise multiple comparisons, adjusted using the BH false-discovery rate, are reported.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Peto, R., Peto, J., 1972. Asymptotically Efficient Rank Invariant Test Procedures. Journal of the Royal Statistical Society. Series A (General) 135, 185. doi:10.2307/2344317

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological), 57, 289-300.

Examples

```
data(PbHeron)

# Two Groups
cen1way(PbHeron$Liver,PbHeron$LiverCen,PbHeron$DosageGroup)
```

```
# More than two groups
cen1way(PbHeron$Liver,PbHeron$LiverCen,PbHeron$Group)
```

---

cen2means   *Censored data two-group test for difference in means*

Description

Performs a parametric test of differences in means between two groups of censored data, either in original or in log units (the latter becomes a test for difference in geometric means).

Usage

```
cen2means(x1, x2, group, LOG = TRUE, printstat = TRUE)
```

Arguments

x1	The column of data values plus detection limits
x2	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y1 column, and 0 (or FALSE) indicates a detected value in y1.
group	Grouping or factor variable. Can be either a text or numeric value indicating the group assignment.
LOG	Indicator of whether to compute tests in the original units, or on their logarithms. The default is to use the logarithms (LOG = TRUE). To compute in original units, specify the option LOG = FALSE (or LOG = 0).
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Details

Because this is an MLE procedure, when a normal distribution model is used (LOG=FALSE) values may be modeled as below zero. When this happens the means may be too low and the p-values may be unreal (often lower than they should be). Because of this, testing in log units is preferable and is the default.

Value

Q-Q Plot with Shapiro-Francia test for normality W and p-values. Returns the Maximum Likelihood Estimation (MLE) test results including Chi-Squared value, degrees of freedom and p-value of the test.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Shapiro, S.S., Francia, R.S., 1972. An approximate analysis of variance test for normality. Journal of the American Statistical Association 67, 215–216.

Examples

```
data(PbHeron)
cen2means(PbHeron$Liver,PbHeron$LiverCen,PbHeron$DosageGroup)
```

---

cen2way   *Parametric Two Factor Fixed Effects ANOVA for censored data*

Description

Computes tests for each of the two factors and optionally for their interaction using likelihood ratio tests. p-values will not be identical to the usual method of moments ANOVA tests but will be similar.

Usage

```
cen2way(y1, y2, fac1, fac2, LOG = TRUE, interact = TRUE)
```

Arguments

y1	The column of data values plus detection limits
y2	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y1 column, and 0 (or FALSE) indicates a detected value in y1.
fac1	The first grouping or factor variable. Can be either a text or numeric value indicating the group assignment.
fac2	The second grouping or factor variable. Can be either a text or numeric value indicating the group assignment.
LOG	A logical variable indicating whether natural logs are to be taken of the 'y1' column data. Default is TRUE.
interact	A logical variable indicating whether to compute an interaction term between the two variables. Default is TRUE. #' @keywords two-way two-factor factorial ANOVA analysis of variance censored

Details

Tests are computed using Maximum Likelihood Estimation. When a gaussian distribution model is used (LOG=FALSE) modeled values may fall below zero, producing unreal p-values (often lower than they should be). Because of this, testing in log units is preferable and is the default unless you are transforming the y values prior to running the function (such as taking cube roots to approximate a gamma distribution). The 'delta.lr0x2' stat output is the -2loglikehood for the test of the model versus an intercept-only model.

Value

Q-Q plots of residuals. Likelihood ratio test statistics ("chisquare"), degrees of freedom ("df") and p-values (pval) for two factors and optionally the interaction. Data on the underlying models, including AIC and R^2 are also provided.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J. Millard, S.P., 2013. EnvStats: An R Package for Environmental Statistics. Springer-Verlag, New York.

See Also

survival::survreg

Examples

```
data(Gales_Creek)
Gales_Creek$Period <- c(rep("early", 35), rep("middle", 12), rep("late", 16))
with(Gales_Creek,cen2way(TCr, CrND, Season, Period))
```

---

cenanova   *ANOVA for censored data*

Description

Performs a parametric test of differences in means between groups of censored data, followed by a parametric Tukey's multiple comparison test.

Usage

```
cenanova(x1, x2, group, LOG = TRUE, printstat = TRUE)
```

Arguments

x1	The column of data values plus detection limits
x2	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y1 column, and 0 (or FALSE) indicates a detected value in y1.
group	Grouping or factor variable. Can be either a text or numeric value indicating the group assignment.
LOG	Indicator of whether to compute tests in the original units, or on their logarithms. The default is to use the logarithms (LOG = TRUE). To compute in original units, specify the option LOG = FALSE (or LOG = 0).
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Details

Test is computed using Maximum Likelihood Estimation. When a gaussian distribution model is used (LOG=FALSE) modeled values may fall below zero, producing unreal p-values (often lower than they should be). Because of this, testing in log units is preferable and is the default.

Value

Returns the Maximum Likelihood Estimation (MLE) comparison results including Chi-Squared value, degrees of freedom and `p-value` of the test. Test assumes log-normal(`LOG=TRUE`) or normal(`LOG=FALSE`) distribution of residuals from group means.

Tukey's multiple comparison p-values of pairwise differences in group means are also printed.

- Group Names of groups (NOTE: `== 0` indicates null hypothesis of "equals zero").
- `Estimate` Estimated difference between group means.
- `Std. Error` Standard error of estimate.
- `z value` Test statistic.
- `Pr(>|z|)` P-values for test that difference in means equals zero.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

#' @examples data(PbHeron) cenanova(PbHeron$Liver,PbHeron$LiverCen,PbHeron$DosageGroup)

cenanova(PbHeron$Liver,PbHeron$LiverCen,PbHeron$DosageGroup,LOG=FALSE)

See Also

survival::survreg

---

cenboxplot *Produces a censored boxplot*

Description

Draws a boxplot with the highest censoring threshold shown as a horizontal line. Any statistics below this line are invalid and must be estimated using methods for censored data.

Usage

```
cenboxplot(obs, cen, group, log = TRUE, range = 0, ...)
```

Arguments

obs	A numeric vector of observations.
cen	A logical vector indicating TRUE where an observation in obs is censored (a less-than value) and FALSE otherwise.
group	A factor vector used for grouping obs into subsets (each group will be a separate box).
log	A logical indicating if the y-axis should be in log units. Default is TRUE.

range	A numeric value determining how far the plot whiskers extend from the box. If positive, the whiskers extend to the most extreme data point which is no more than range times the interquartile range from the box. The default is 0, which extends whiskers to the min and max.
...	Additional arguments passed to `graphics::boxplot()`.

Value

The output of the default `graphics::boxplot()` method.

References

Helsel, Dennis R. (2005). *Nondetects and Data Analysis; Statistics for censored environmental data*. John Wiley and Sons, Hoboken, NJ.

---

cenCompareCdfs	*Comparison of empirical cdf of censored data*

Description

Plots the empirical cdf and cdfs of three theoretical distributions, fit by maximum likelihood estimation (MLE).

Usage

```
cenCompareCdfs(x.var, cens.var, dist3 = "norm", Yname = yname)
```

Arguments

x.var	The column of y (response variable) values plus detection limits
cens.var	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y.var column, and 0 (or FALSE) indicates a detected value in y.var.
dist3	Name of the third distribution to be plotted, default is norm (normal distribution). Alternate third distribution is weibull(for Weibull distribution). Log-normal and gamma distributions are always used.
Yname	Optional – input text in quotes to be used as the variable name. The default is the name of the y.var input variable.

Value

prints a plot of the empirical CDFs with BIC value for each distribution.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Delignette-Muller, M., Dutang, C., 2015. fitdistrplus : An R Package for Fitting Distributions. Journal of Statistical Software, 64, 1-34. http://www.jstatsoft.org/v64/i04/.

Examples

```
data(Brumbaugh)
cenCompareCdfs(Brumbaugh$Hg,Brumbaugh$HgCen)

# With Weibull distribution
cenCompareCdfs(Brumbaugh$Hg,Brumbaugh$HgCen,dist3="weibull")

# Using an distribution not supported by this function (yet)
# you will get an error message
## Not run: cenCompareCdfs(Brumbaugh$Hg,Brumbaugh$HgCen,dist3="beta")

# With Yname specified
cenCompareCdfs(Brumbaugh$Hg,Brumbaugh$HgCen,Yname="TCE Conc (ug/L)\nLong Island, NY USA")
```

---

cenCompareQQ   *Censored Q-Q Plot comparison*

Description

Produces three quantile-quantile (Q-Q) plots, also called probability plots, based on three distributions (normal, lognormal and gamma distributions).

Usage

```
cenCompareQQ(x.var, cens.var, Yname = yname, printrslt = TRUE, ...)
```

Arguments

x.var	The column of x (response variable) values plus detection limits
cens.var	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y.var column, and 0 (or FALSE) indicates a detected value in y.var.
Yname	Optional – input text in quotes to be used as the variable name on all plots. The default Yname is the name of the y.var input variable.
printrslt	Logical TRUE/FALSE option of whether to print the best distribution in the console window, or not. Default is TRUE.
...	further graphical parameters (from par), such as srt, family and xpd.

Details

Produces three Q-Q plots and reports which one has the highest Shapiro-Francia test statistic (W). The distribution with the highest W is the best fit of the three.

Value

Plots three Q-Q plots based on normal, lognormal and gamma distributions and prints the best-fit distribution.

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Millard, S.P., 2013. EnvStats: An R Package for Environmental Statistics. Springer-Verlag, New York.

Shapiro, S.S., Francia, R.S., 1972. An approximate analysis of variance test for normality. Journal of the American Statistical Association 67, 215–216.

Examples

```
data(Brumbaugh)
cenCompareQQ(Brumbaugh$Hg,Brumbaugh$HgCen)
```

---

cencorreg	*Correlation and Regression with censored data*

Description

Computes three parametric correlation coefficients for one X variable and the corresponding R squared for multiple X variables, and a regression equation for censored data.

Usage

```
cencorreg(
  y.var,
  cen.var,
  x.vars,
  LOG = TRUE,
  verbose = 2,
  pred.plot = FALSE,
  pred.col = "purple"
)
```

Arguments

y.var	The column of y (response variable) values plus detection limits.
cen.var	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y.var column, and 0 (or FALSE) indicates a detected value is in y.var.
x.vars	One or more uncensored explanatory variable(s). For multiple variables it must be a data frame of numeric, character and factor variables. See Details
LOG	Indicator of whether to compute the regression in the original y units, or on their logarithms. The default is to use the logarithms (LOG = TRUE). To compute in original units, specify the option LOG = FALSE (or LOG = 0).
verbose	default verbose=2, see details.

pred.plot	default is FALSE. Produces a plot of data and regression model predictions. To do this the first (or only) x variable in the X dataframe must be a continuous (not factor) variable, and it becomes the x variable in the plot.
pred.col	default is "purple". Changes the color of the predicted lines in the prediction plot.

Details

`x.vars`: If one x variable only, enter its name. If multiple x variables, enter the name of a data frame of columns of the x variables. Only columns used as X variables in the regression are allowed. Create this by `x.frame <- data.frame (Temp, Flow, Time)` for 3 variables (temperature, flow and time) used as the X variables in the regression. To produce a pred.plot plot of predicted values the first variable in the array must be a continuous (not a factor) variable.

AIC and BIC are printed to help evaluate the 'best' regression model. Lower values are better when comparing models with the same Y units and same data.Cannot be used to compare models with differing Y units (such as Y~X versus logY~X). Can be used to compare models with differing X units such as Y~X vs Y~logX.

The default Y units are that the Y variable will be log transformed. Change this with the LOG = option, setting LOG = FALSE.

`verbose` option. Default is 2 which provides full output in the console and qqplots in a graphics window. A value of 1 only provides partial results in the console and no plots. A value of 0 provides no output; the returning computations will be stored in the specified object.

The Y parameter in the model output (modelname$y) is -1 times the data that were input. This is due to the shift from left censored data to the required right censored data of the survreg function. The original data are not changed and are used to draw the pred.plot.

Value

When `x.vars` is one variable, likelihood, rescaled likelihood and McFaddens correlation coefficient (R) are printed. When `x.vars` is a `data.frame` of more than one variable, likelihood, rescaled likelihood and McFaddens coefficent of determination (R2) are printed.

Model coefficients (intercept and slopes), Chi-Squared statistic and p-value for the test that all slope coefficients equal zero (overall test), and model AIC and BIC are provided.

A Q-Q plot of model residuals with corresponding Shapiro-Francia W and p-value are plotted for evaluation of model distributional assumptions when `verbose=2` (the default).

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Helsel, D.R., 2005. Nondetects and Data Analysis: Statistics for Censored Environmental Data, 1st ed. John Wiley and Sons, USA, N.J.

See Also

survival::survreg

Examples

```
data(Brumbaugh)

# One variable
cencorreg(Brumbaugh$Hg,Brumbaugh$HgCen,Brumbaugh$SedMeHg)

# One variable with pred.plot=T
cencorreg(Brumbaugh$Hg,Brumbaugh$HgCen,Brumbaugh$SedMeHg,pred.plot=TRUE)

# More than one variable for demostration purposes
cencorreg(Brumbaugh$Hg,Brumbaugh$HgCen,Brumbaugh[,c("SedMeHg","PctWetland")])
```

---

cenken   *Compute Kendall's Tau Correlation and ATS Line for Censored Data*

Description

Computes Kendall's tau for singly (y only) or doubly (x and y) censored data. Also computes the Akritas-Theil-Sen (ATS) nonparametric regression line, with the Turnbull estimate of the intercept.

Usage

```
cenken(y, ycen = NULL, x = NULL, xcen = NULL)
```

Arguments

y	A numeric vector of observations or a formula.
ycen	A logical vector indicating TRUE where an observation in y is censored (a less-than value) and FALSE otherwise. Can be omitted for the case where y is not censored.
x	A numeric vector of observations.
xcen	default is NULL for trend analysis purposes. If included, a logical vector indicating TRUE where an observation in x is censored (a less-than value) and FALSE otherwise.

Details

If using the formula interface, the ycen, x, and xcen arguments are not specified. All information is instead provided through a formula via the y argument. The formula must use a Cen object as the response (on the left-hand side of ~), and predictors (optional) on the right-hand side separated by +. See examples below.

Kendall's tau is a nonparametric correlation coefficient that measures monotonic association between y and x. For left-censored data, concordant and discordant pairs are evaluated wherever possible. For example, with increasing x values, a change in y from <1 to 10 is considered an increase (concordant), while a change from <1 to 0.5 or <5 is considered a tie.

The ATS line is defined as the slope resulting in Kendall's tau of 0 between residuals (y - slope
* x) and x. The routine uses an iterative bisection search to find this slope. The intercept is the
median residual, computed using the Turnbull estimate for interval-censored data as implemented
in the **Icens** package.

Value

A list containing:

tau   Kendall's tau correlation coefficient.

slope   The estimated slope from the ATS line.

p.value   P-value for testing the null hypothesis of no association.

References

Helsel, D. R. (2005). *Nondetects and Data Analysis: Statistics for Censored Environmental Data*.
John Wiley & Sons.

Akritas, M. G., Murphy, S. A., & LaValley, M. P. (1995). The Theil-Sen Estimator with Doubly
Censored Data and Applications to Astronomy. *Journal of the American Statistical Association*, 90,
170–177.

Examples

```
# Both y and x are censored
data(PbHeron)
with(PbHeron,cenken(Blood,BloodCen,Kidney,KidneyCen))

# x is not censored
data(TCEReg)
with(TCEReg, cenken(log(TCEConc), TCECen, PopDensity))

# Synthetic time-series with trend analysis
set.seed(123)

## Parameters
n <- 15  # 15 years of data
time <- 1:n

## Components
trend <- 0.235 * time
noise <- rnorm(n, mean = 5, sd = 1.5)

syn_dat <- data.frame(Yr = 1989 + time, value = trend + noise)
syn_dat$censored <- syn_dat$value < quantile(syn_dat$value, 0.2)

with(syn_dat,cenken(value,censored,Yr))
## Not run:
plot(value~Yr,syn_dat,pch=21,bg=ifelse(syn_dat$censored==TRUE,"red","blue",cex=1.5))
abline(h=quantile(syn_dat$value, 0.2),lty=2,col="red")

## End(Not run)
```

---

cenperm2   *Censored two-group permutation test*

Description

Performs a permutation test of differences in means between two groups of censored data.

Usage

```
cenperm2(y1, y2, grp, R = 9999, alternative = "two.sided", printstat = TRUE)
```

Arguments

y1	The column of data values plus detection limits
y2	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y1 column, and 0 (or FALSE) indicates a detected value in y1.
grp	Grouping or factor variable. Can be either a text or numeric value indicating the group assignment.
R	The number of permutations used. Default is 9999
alternative	indicates the alternative hypothesis and must be one of "two.sided", "greater" or "less". You may also specify just the initial letter. Default is "two.sided".
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Details

Because this is a permutation test it avoids the problem with MLE tests (cen2means) that assume a normal distribution. No values are modeled as below zero and p-values are trustworthy. Ranges in means and p-values are due to interval-censoring of censored data means.

Value

Permutation test results with the number of permutations, range in group means and their difference, and range in p-value.

References

Good, P., 2000. Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses, 2nd ed, Springer Series in Statistics. Springer-Verlag, New York, NY. doi:10.1007/97814757-32351

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Shapiro, S.S., Francia, R.S., 1972. An approximate analysis of variance test for normality. Journal of the American Statistical Association 67, 215–216.

Examples

```
data(PbHeron)
cenperm2(PbHeron$Liver,PbHeron$LiverCen,PbHeron$DosageGroup,alternative="t")
```

---

cenpermanova	*Censored data one-factor permutation test*

Description

Performs a permutation test of differences in means between groups of censored data.

Usage

```
cenpermanova(y1, y2, grp, R = 9999, printstat = TRUE)
```

Arguments

y1	The column of data values plus detection limits.
y2	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y1 column, and 0 (or FALSE) indicates a detected value in y1.
grp	Grouping or factor variable. Can be either a text or numeric value indicating the group assignment.
R	The number of permutations used. Default is 9999.
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Details

Because this is a permutation test it avoids the problem with MLE tests (see cenanova) that assume a normal distribution. No values are modeled as below zero and group means and p-values are trustworthy.

Value

Permutation test results with the number of permutations, range in test statistics and p-value values through the various permutations. Group means are also listed.

References

Good, P., 2000. Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses, 2nd ed, Springer Series in Statistics. Springer-Verlag, New York, NY. doi:10.1007/97814757-32351

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Examples

```
data(PbHeron)
cenpermanova(PbHeron$Liver,PbHeron$LiverCen,PbHeron$DosageGroup)
```

---

cenPredInt   *Prediction interval for censored data*

Description

Computes prediction intervals for censored data assuming lognormal, gamma and normal distributions.

Usage

```
cenPredInt(
  x.var,
  cens.var,
  pi.type = "two-sided",
  conf = 0.95,
  newobs = 1,
  method = "mle",
  printstat = TRUE
)
```

Arguments

x.var	The column of x (response variable) detected values plus detection limits
cens.var	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y.var column, and 0 (or FALSE) indicates a detected value in y.var.
pi.type	Designation of either a "two-sided" interval (default) or a 1-sided "upper" or 1-sided "lower" interval.
conf	Confidence coefficient of the interval, 0.95 (default).
newobs	The number of new observations to be contained in the interval.
method	Character string specifying the method of estimation. Default is mle (maximum likelihood). See details.
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Details

Computes prediction intervals for three distributions. This is a front-end to the individual functions from the EnvStats package. By default all three are computed using maximum likelihood estimation (mle). The gamma distribution for censored data uses the Wilson-Hilferty approximation (normal distribution on cube roots of data). Other methods are available in EnvStats, but few methods are available for all three distributions. For info on other methods, see help for elnormCensored and enormCensored commands in EnvStats.

Value

A table of prediction limits based on user provided confidence coefficient (`conf`) and prediction invterval type (`pi.type`)

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Millard, S.P., 2013. EnvStats: An R Package for Environmental Statistics. Springer-Verlag, New York.

Krishnamoorthy, K., Mathew, T., Mukherjee, S., 2008. Normal-Based Methods for a Gamma Distribution, Technometrics, 50, 69-78.

See Also

EnvStats::enormCensored

Examples

```
data(PbHeron)

# Default
cenPredInt(PbHeron$Liver,PbHeron$LiverCen)

# User defined confidence coefficient
cenPredInt(PbHeron$Liver,PbHeron$LiverCen, conf=0.5)

# User defined confidence coefficient outside of acceptable range
# the procedure will stop and give an error.
# cenPredInt(PbHeron$Liver,PbHeron$LiverCen, conf=1.1)

# User defined prediction interval type
cenPredInt(PbHeron$Liver,PbHeron$LiverCen,pi.type="lower")
cenPredInt(PbHeron$Liver,PbHeron$LiverCen,pi.type="upper")
```

---

cenQQ *Q-Q Plot censored data*

Description

Plots a quantile-quantile (Q-Q) plot of censored data versus a fitted data distribution

Usage

```
cenQQ(x.var, cens.var, dist = "lnorm", Yname = yname)
```

Arguments

`x.var`	The column of x (response variable) values plus detection limits
`cens.var`	The column of indicators, where 1 (or `TRUE`) indicates a detection limit in the `y.var` column, and 0 (or `FALSE`) indicates a detected value in `y.var`.
`dist`	One of three distributional shapes to fit to your data: lognormal (`lnorm`), normal (`norm`) or gamma (`gamma`).
`Yname`	Optional – input text in quotes to be used as the variable name on the Q-Q plot. The default is the `Yname` name of the `y.var` input variable.

Value

A single Q-Q plot of data fitted by normal, lognormal or gamma distributions with Shapiro-Francia W value printed on plot.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Shapiro, S.S., Francia, R.S., 1972. An approximate analysis of variance test for normality. Journal of the American Statistical Association 67, 215–216.

Examples

```
data(Brumbaugh)
cenQQ(Brumbaugh$Hg,Brumbaugh$HgCen)

# User defined distribution
cenQQ(Brumbaugh$Hg,Brumbaugh$HgCen,dist="gamma")
```

---

cenregQQ	*Q-Q plot of censored regression residuals*

Description

Plots a quantile-quantile (Q-Q) plot of censored regression residuals for simple or multiple regression.

Usage

```
cenregQQ(y.var, cen.var, x.vars, LOG = TRUE, intcens = FALSE, main = NULL)
```

Arguments

y.var	The column of y (response variable) values plus detection limits. Alternatively, with interval-censord data, the column of the lower end of the interval.
cen.var	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y.var column, and 0 (or FALSE) indicates a detected value in y.var. Alternatively, with interval-censored data the column of the high end of the interval.
x.vars	One or more uncensored explanatory variable(s). See Details
LOG	Indicator of whether to compute the regression in the original y units, or on their logarithms. The default is to use the logarithms (LOG = TRUE). To compute in original units, specify the option LOG = FALSE (or LOG = 0).
intcens	a logical value indicating the input data is interval-censored instead of a column of values plus a column of indicators.
main	overall title for the plot. A default titel will be used if none is specified.

Value

Q-Q Plot of model residuals and Shapiro-Francia test results.

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Millard, S.P., 2013. EnvStats: An R Package for Environmental Statistics. Springer-Verlag, New York.

Shapiro, S.S., Francia, R.S., 1972. An approximate analysis of variance test for normality. Journal of the American Statistical Association 67, 215–216.

Examples

```
data(Brumbaugh)

# One variable
cenregQQ(Brumbaugh$Hg,Brumbaugh$HgCen,Brumbaugh$PctWetland)

# More than one variable for demostration purposes
cenregQQ(Brumbaugh$Hg,Brumbaugh$HgCen,Brumbaugh[,c("PctWetland","SedLOI","Weight")])
```

---

censeaken	*Seasonal Kendall permutation test on censored data*

Description

Seasonal Kendall permutation test on censored data

Usage

```
censeaken(
  time,
  y,
  y.cen,
  group,
  data = NULL,
  LOG = FALSE,
  R = 4999,
  nmin = 4,
  seaplots = FALSE,
  printstat = TRUE,
  ...
)
```

Arguments

time	Column of the time variable, either a sequence of days or decimal times, etc. Will be the scale used for time in the trend analysis.
y	The column of y (response variable) values plus detection limits
y.cen	The y-variable indicators, where 1 (or TRUE) indicates a detection limit in the y.var column, and 0 (or FALSE) indicates a detected value in y.var.
group	Column of the season classifications. A factor in R, so usually though not necessarily a text variable. If numeric, define as a factor before running the script.
data	an optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model.
LOG	Indicator of whether to compute the trend analysis in the original y units, or on their logarithms. The default is to use the logarithms (LOG = TRUE). To compute in original units, specify the option LOG = FALSE (or LOG = 0).
R	The number of repetitions in the permutation process. R is often between 999 and 9999 (adding +1 to represent the observed test statistic produces 1000 to 10000 repetitions). By default R=4999. Increasing R results in lower variation in the p-values produced between runs.
nmin	The minimum number of observations needed for the entire time period to be tested, per season. For example, with 1 sample per year per season over an 8-year period, you have 8 observations for each season. You can increase this number if you want a higher minimum. Don't decrease it below the default of 4. If there are fewer than nmin values that season is skipped and not included in the overall test and a note of this will be printed in the console.
seaplots	In addition to the plot of the overall Seasonal Kendall trend line, plots of the trend in individual seasons can also be drawn.
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.
...	other inputs associated with modifying plots produced by this function.

Details

For each season the ATS function is used to compute the season's Kendall-S statistic and p-value for the trend test. The test is the usual ATS procedure, not a permutation test. For the overall test there are R (default 4999) random rearrangements of data that are generated with no mixing of data from one season to another season within a permutation – data over time within a season are randomized. This retains any between-seasons differences while removing any trend from year to year to use as the permuted "representation of the null hypothesis" of no trend in the R Seasonal Kendall tests. For a 2-sided trend test the p-value is the number of the absolute value of the permutation S statistics that are equal to or greater than the absolute value of the observed S from the original data, plus 1 (for the observed S of the original data), divided by the total (R+1) S values.

The censeaken function differs from other R packages that do not incorporate censored data in their trend tests (EnvStats, Kendall, rkt and others) in that censeaken uses all of the data across the years without an option to equalize each year's or season's influence by using the overall mean or median of the period's data rather than the original observations. Seasons with more data will have more influence on the outcome, just as years with more data will have more influence. If the numbers of observations differ enough between seasons or years that this is of concern, it is up to the user to perhaps eliminate some data in data-rich periods that are most unlike the conditions or times of data collected in sparse periods. Or to compute summary statistics such as the median of each season/year combination and run the test on the medians, though this will result in a loss of power to detect trends and will require the user to use methods such as those in NADA2 to compute medians when there are censored data.

If `seaplots=TRUE` each season's trend line will be plotted seperately. A plot of the overall Seasonal Kendall (Akritas-Theil-Sen) line is always plotted. If `seaplots=FALSE` only the overall Seasonal Kendall (Akritas-Theil-Sen) line will be plotted on a data scatterplot.

Value

Prints the Kendall trend test results for each season individually. The overall Seasonal Kendall test and Theil-Sen line results are both printed and returned.

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Hirsch, R.M., Slack, J.R., Smith, R.A., 1982. Techniques of Trend Analysis for Monthly Water Quality Data, Water Res. Reseach 18, 107-121.

Examples

```
data(Brumbaugh)

# Artificial time and season variables for demonstration purposes
Brumbaugh$time=1:nrow(Brumbaugh)
Brumbaugh$sea=as.factor(round(runif(nrow(Brumbaugh),1,4),0))

with(Brumbaugh,censeaken(time,Hg,HgCen,sea,seaplots = TRUE))

censeaken(time,Hg,HgCen,sea,Brumbaugh,seaplots = TRUE)
```

---

censtats   *Summary Statistics for Censored Data using ROS, MLE, and K-M*

Description

A convenience function that produces a comparative table of summary statistics obtained using the `ros()`, `cenmle()`, and `cenfit()` routines. These methods are Regression on Order Statistics (ROS), Maximum Likelihood Estimation (MLE), and Kaplan-Meier (K-M).

Usage

```
censtats(...)
```

Arguments

`...`	to pass argument

Details

If the data do not fulfill the criteria for the application of any method, no summary statistics will be produced.

Value

A data frame with the summary statistics.

---

censummary   *Summary Statistics for Censored Data*

Description

Produces basic, and hopefully useful, summary statistics on censored data.

Usage

```
censummary(obs, censored, groups = NULL, ...)

censummary.default(obs, censored, ...)

censummary.factor(obs, censored, groups, ...)

censummary.numeric(obs, censored, groups, ...)
```

Arguments

obs	A numeric vector of observations.
censored	A logical vector indicating TRUE where an observation in obs is censored (a less-than value), and FALSE otherwise.
groups	Optional grouping variable (factor or numeric) used to compute summaries by subset.
...	Additional arguments passed to methods.

Details

This is a generic function with methods for numeric vectors and grouped data.

The generic function dispatches to methods based on the presence and type of groups:

censummary.default Basic summary for numeric vectors.

censummary.factor Summary grouped by a factor variable.

censummary.numeric Summary grouped by a numeric grouping variable (converted to factor).

Value

An object of class "censummary" containing summary statistics, or a list of such objects if grouped.

References

Helsel, Dennis R. (2005). *Nondetects and Data Analysis: Statistics for Censored Environmental Data*. John Wiley and Sons, USA, NJ.

---

cenTolInt	*Upper Tolerance interval for censored data*

Description

Computes a one-sided upper tolerance interval for censored data assuming log-normal, gamma and normal distributions.

Usage

```
cenTolInt(
  x.var,
  cens.var,
  conf = 0.95,
  cover = 0.9,
  method.fit = "mle",
  printstat = TRUE
)
```

Arguments

x.var	The column of x (response variable) values plus detection limits
cens.var	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y.var column, and 0 (or FALSE) indicates a detected value in y.var.
conf	Confidence coefficient of the interval, 0.95 (default).
cover	Coverage, the percentile probability above which the tolerance interval is computed. The default is 90, so a tolerance interval will be computed above the 90th percentile of the data.
method.fit	The method used to compute the parameters of the distribution. The default is maximum likelihood ("mle"). The alternative is robust ROS ("rROS"). See Details.
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Details

Computes upper one-sided tolerance intervals for three distributions. This is a front-end to the individual functions from the EnvStats package. By default all three are computed using maximum likelihood estimation (mle); robust ROS is available as an alternate method for all three distributions. The gamma distribution for censored data uses the Wilson-Hilferty approximation (normal distribution on cube roots of data). For more info on the relative merits of robust ROS versus mle, see Helsel (2011) and Millard (2013).

Value

Prints and returns the percentile (cover), upper tolerance limit (conf) and BIC of fit for lognormal, normal and approximated gamma distributions. Plots empirical and theoretical CDFs with BIC values in the legend.

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Millard, S.P., 2013. EnvStats: An R Package for Environmental Statistics. Springer-Verlag, New York.

Krishnamoorthy, K., Mathew, T., Mukherjee, S., 2008. Normal-Based Methods for a Gamma Distribution, Technometrics, 50, 69-78.

Examples

```
data(PbHeron)

# Default
cenTolInt(PbHeron$Liver,PbHeron$LiverCen)

# User defined conficence interval
cenTolInt(PbHeron$Liver,PbHeron$LiverCen,conf=0.75)
```

```
# User defined percentile
cenTolInt(PbHeron$Liver,PbHeron$LiverCen,cover=0.5)

# inputs outside acceptable ranges
# Will result in errors/warnings
# cenTolInt(PbHeron$Liver,PbHeron$LiverCen,cover=1.25)
# cenTolInt(PbHeron$Liver,PbHeron$LiverCen,conf=1.1)
# cenTolInt(PbHeron$Liver,PbHeron$LiverCen,method.fit="ROS")
```

---

centrend   *Trend analysis of censored data with a covariate*

Description

Computes the ATS (Mann-Kendall trend test for censored data) after adjustment of censored data for a covariate.

Usage

```
centrend(
  y.var,
  y.cens,
  x.var,
  time.var,
  link = "identity",
  Smooth = "cs",
  printstat = TRUE,
  stackplots = FALSE,
  drawplot = TRUE
)
```

Arguments

y.var	The column of y (response variable) values plus detection limits
y.cens	The column of indicators, where 1 (or TRUE) indicates a detectionlimit in the y.var column, and 0 (or FALSE) indicates a detected value in y.var.
x.var	Column of a covariate (not time). y.var will be smoothed versus x.var and residuals taken to subtract out the relationship between y and x.
time.var	Column of the time variable, either a sequence of days or decimal times, etc. Will be the scale used for time in ATS trend analysis.
link	Default = "identity" which means it uses data in the original units. See details.
Smooth	Type of smoother used in the GAM. Default is "cs", shrinkage cubic regression splines. See details for other options.

printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.
stackplots	logical TRUE/FALSE option to stack three plots that are output onto the same page instead of each on separate page. Default is 'FALSE', each separate.
drawplot	Logical TRUE/FALSE option of whether to draw plots or not. Default is TRUE

Details

Default link = identity. The y variables are then used in their original units. Other options are available see cenGAM::tobit1 for more options.

Default Smooth is "cs" for shrinkage cubic regression splines. See mgcv::smooth.terms for other types of smoothing algorithms. '"ts"' is a thin-plate regression spline and is also commonly used.

Value

Prints three plots: Y data vs time with GAM Smooth, Residuals from GAM Smooth vs time, and ATS trend line of residuals vs time.

Returns GAM residuals and ATS results on trend test of residuals (intercept, slope, Kendall's tau, p-value for trend)

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

See Also

mgcv::gam

Examples

```
data(Brumbaugh)

Brumbaugh$time=1:nrow(Brumbaugh)

with(Brumbaugh,centrend(Hg,HgCen,SedTotHg,time.var=time,drawplot=TRUE))
```

---

centrendsea	*Trend analysis of censored data with a covariate and seasonal blocks*

Description

Computes the Seasonal Kendall trend test for censored data after adjustment of censored data for a covariate. The adjustment is by subtracting off a censored GAM smooth, removing the effect of the covariate. Trend analysis is performed on the residuals from the GAM smooth.

Usage

```
centrendsea(
  y.var,
  y.cens,
  x.var,
  time.var,
  season,
  R = 4999,
  nmin = 4,
  link = "identity",
  Smooth = "cs",
  printstat = TRUE,
  seaplots = TRUE
)
```

Arguments

`y.var`	The column of y (response variable) values plus detection limits
`y.cens`	The column of indicators, where 1 (or `TRUE`) indicates a detectionlimit in the `y.var` column, and 0 (or `FALSE`) indicates a detected value in `y.var`.
`x.var`	Column of a covariate (not time). `y.var` will be smoothed versus `x.var` and residuals taken to subtract out the relationship between `y` and `x`.
`time.var`	Column of the numerical time variable, either a sequence of numbered days or decimal times, etc. Will be the scale used for time in ATS trend analysis.
`season`	Column of the seasonal variable. Usually a text variable but may be numeric. Will be converted to a factor..
`R`	The number of repetitions in the permutation process. R is often between 999 and 9999 (adding +1 to represent the observed test statistic produces 1000 to 10000 repetitions). By default R=4999. Increasing R results in lower variation in the p-values produced between runs.
`nmin`	The minimum number of observations needed for the entire time period to be tested, per season. For example, with 1 sample per year per season over an 8-year period, you have 8 observations for each season. You can increase this number if you want a higher minimum. Don't decrease it below the default of 4. If there are fewer than nmin values that season is skipped and not included in the overall test and a note of this will be printed in the console.
`link`	Default = `"identity"` which means it uses data in the original units. See details.
`Smooth`	Type of smoother used in the GAM. Default is `"cs"`, shrinkage cubic regression splines. See details for other options.
`printstat`	Logical `TRUE`/`FALSE` option of whether to print the resulting statistics in the console window, or not. Default is `TRUE`.
`seaplots`	logical 'TRUE'/'FALSE' option to print plots with trend line for each season. Default is 'TRUE'.

Details

Default `link` = identity. The y variables are then used in their original units. Other options are available see cenGAM::tobit1 for more options.

Default `Smooth` is `"cs"` for shrinkage cubic regression splines. See mgcv::smooth.terms for other types of smoothing algorithms. '"ts"' is a thin-plate regression spline and is also commonly used.

As with the censeaken function, observations are not edited when there are more data in some seasons than others. Seasons with more data will have more influence on the overall SK test than seasons with fewer data. To avoid this (as done with some Seasonal Kendall software) data in the seasons with more can be selectively deleted to better match the data frequency of the seasons with fewer data.

Value

Prints four plots: 1. Y data vs X covariate with GAM Smooth, 2. Residuals from GAM Smooth vs X covariate, 3. histogram of the SK test results illlustrating the p-value, and 4. Seasonal Kendall trend line of residuals vs time. Plots of data and SK trend line for each season are produced when the seaplots option is TRUE.

Returns the seasonal Kendall trend test results on residuals (intercept, slope, Kendall's tau, p-value for trend)

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

See Also

mgcv::gam

Examples

```
data(Gales_Creek)
with(Gales_Creek,centrendsea(TCr,CrND,discharge,dectime,Season))
```

---

  cenxyplot   *Produces a censored x-y scatter plot*

Description

Draws an x-y scatter plot with censored values represented by dashed lines spanning from the censored threshold to zero.

Usage

```
cenxyplot(x, xcen, y, ycen, log = "", lty = "dashed", ...)
```

Arguments

x	A numeric vector of observations.
xcen	A logical vector indicating TRUE where an observation in x is censored (a less-than value) and FALSE otherwise.
y	A numeric vector of observations.
ycen	A logical vector indicating TRUE where an observation in y is censored (a less-than value) and FALSE otherwise.
log	A character string specifying which axes are logarithmic: "x" for x-axis, "y" for y-axis, "xy" or "yx" for both axes, or "" for none (default).
lty	The line type for the lines representing censored-data ranges.
...	Additional arguments passed to [graphics::plot()]().

References

Helsel, Dennis R. (2005). *Nondetects and Data Analysis; Statistics for censored environmental data.* John Wiley and Sons, Hoboken, NJ.

---

cen_ecdf	*Censored Empirical Cumulative Distribution Function*

Description

Plots ecdfs of one or more groups of censored data. Illustrates the differences between groups for group tests such as those done using cen1way or cenanova.

Usage

```
cen_ecdf(
  x.var,
  cens.var,
  xgroup = NULL,
  xlim = c(0, max(y.var)),
  Ylab = varname
)
```

Arguments

x.var	The column of data values plus detection limits
cens.var	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y.var column, and 0 (or FALSE) indicates a detected value in y.var.
xgroup	Optional - grouping or factor variable. Can be either a text or numeric value indicating the group assignment.
xlim	Limits for the x (data) axis of the ecdf plot. Default is 0 to the maximum of the y.var variable. To change, use option xlim = c(0, 50) if 50 is to be the maximum on the plot.
Ylab	Optional – input text in quotes to be used as the variable name on the ecdf plot. The default is the name of the y.var input variable.

Value

Plots an empirical cumulative distribution function. If group=NULL the ECDF of the entire dataset
is produced. If group is identified then ECDFs are plotted for each group seperately and a legend
added.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John
Wiley & Sons, USA, N.J. Millard, S.P., 2013. EnvStats: An R Package for Environmental Statistics.
Springer-Verlag, New York.

Examples

```
data(PbHeron)

# with groups
with(PbHeron, cen_ecdf(Liver, LiverCen, DosageGroup))

# all data
with(PbHeron, cen_ecdf(Liver, LiverCen))
```

---

cen_paired   *Censored data paired t-test*

Description

Performs a parametric test of whether the mean difference between two columns of paired censored
data equals 0. Assumes that the paired differences follow a gaussian (normal) distribution.

Usage

```
cen_paired(xd, xc, yd, yc, alternative = "two.sided", printstat = TRUE)
```

Arguments

xd	The first column of data values plus detection limits
xc	The column of censoring indicators, where 1 (or TRUE) indicates a detection limit in the xd column, and 0 (or FALSE) indicates a detected value in xd.
yd	The second column of data values plus detection limits, or a single number representing a standard / guideline value.
yc	The column of censoring indicators for yd, where 1 (or TRUE) indicates a detection limit in the yd column, and 0 (or FALSE) indicates a detected value in yd. Not needed if yd is a single standard number.
alternative	The usual notation for the alternate hypothesis. Default is "two.sided". Options are "greater" or "less".
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Details

You may also test for whether the mean of the xd data exceeds a standard by entering the single number for the standard as yd. In that case no yc is required.

Value

A list of statistics containing the following components:

- n Number of observations
- Z The value of the test statistic
- p.value the p-value of the test
- Mean difference the mean difference between xd and yd

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

See Also

survival::survreg

Examples

```
data(atrazine)

cen_paired(atrazine$June,atrazine$JuneCen,atrazine$Sept,atrazine$SeptCen)

# Comparing standard/guieline value
cen_paired(atrazine$June, atrazine$JuneCen, 0.01, alternative = "greater")
```

---

cen_signedranktest   *Wilcoxcon Signed-Rank test for censored data*

Description

Performs a nonparametric Wilcoxon signed-rank test of whether the median difference between two columns of paired censored data equals 0. Uses the Pratt adjustment for pairs of equal or indistinguishable values.

Usage

```
cen_signedranktest(xd, xc, yd, yc, alternative = "two.sided", printstat = TRUE)
```

Arguments

xd	The first column of data values plus detection limits
xc	The column of censoring indicators for xd, where 1 (or TRUE) indicates a detection limit in the xd column, and 0 (or FALSE) indicates a detected value in xd.
yd	The second column of data values plus detection limits, or a single number representing a standard / guideline value.
yc	The column of censoring indicators for yd, where 1 (or TRUE) indicates a detection limit in the yd column, and 0 (or FALSE) indicates a detected value in yd. Not needed if yd is a single standard number.
alternative	The usual notation for the alternate hypothesis. Default is "two.sided". Options are "greater" or "less".
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Value

Prints a list of Wilcoxon Signed-Rank test with Pratt correction for ties statistics containing the following components:

- n Number of samples

- Z The value of the test statistic

- p.value the p-value of the test

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Page, E.B., 1963. Ordered Hypotheses for Multiple Treatments: A Significance Test for Linear Ranks. Journal of the American Statistical Association 58, 216–230. doi:10.2307/2282965

Pratt, J.W., 1959. Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedures. Journal of the American Statistical Association 54, 655–667. doi:10.2307/2282543

Examples

```
data(atrazine)

cen_signedranktest(atrazine$June,atrazine$JuneCen,atrazine$Sept,atrazine$SeptCen)
```

---

cen_signtest	*Sign test for censored data*

Description

Performs a nonparametric sign test of whether the median difference between two columns of paired censored data equals 0. Uses the Fong adjustment for pairs of equal values.

Usage

```
cen_signtest(xd, xc, yd, yc, alternative = "two.sided", printstat = TRUE)
```

Arguments

xd	The first column of data values plus detection limits
xc	The column of censoring indicators, where 1 (or TRUE) indicates a detection limit in the xd column, and 0 (or FALSE) indicates a detected value in xd.
yd	The second column of data values plus detection limits.
yc	The column of censoring indicators, where 1 (or TRUE) indicates a detection limit in the yd column, and 0 (or FALSE) indicates a detected value in yd.
alternative	The usual notation for the alternate hypothesis. Default is "two.sided". Options are "greater" or "less".
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Value

Returns the number of xd and yd values greater than, less than and tied. Corrected and uncorrected p-value for ties also displayed.

References

Fong, D.Y.T., Kwan, C.W., Lam, K.F., Lam, K.S.L., 2003. Use of the Sign Test for the Median in the Presence of Ties. The American Statistician 57, 237–240.

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J. #'

Examples

```
data(atrazine)

cen_signtest(atrazine$June,atrazine$JuneCen,atrazine$Sept,atrazine$SeptCen)
```

| cfit | *Compute an ECDF and Distribution Parameters for Censored Data* |

Description

Computes the empirical cumulative distribution function (ECDF) for censored data. Estimates parameters of the distribution, including the mean and quantiles.

Usage

```
cfit(
  y,
  cens,
  conf = 0.95,
  qtls = c(0.1, 0.25, 0.5, 0.75, 0.9),
  q.type = 6,
  Cdf = TRUE,
  ecdf.col = 1,
  km.orig = TRUE,
  printstat = TRUE,
  Ylab = NULL
)
```

Arguments

y	Concentrations plus detection limits for indicator formatted data.
cens	Censoring indicators (logical. 1 or TRUE = censored, 0 or FALSE = detected) for indicator formatted data.
conf	The confidence coefficient for confidence intervals around the Kaplan-Meier mean and median. Default = 0.95.
qtls	Probabilities for the quantiles to be estimated. Defaults are (0.10, 0.25, 0.50, 0.75, 0.90). You may add and/or substitute probabilities – all must be between and not including 0 to 1.
q.type	an integer between 1 and 9 selecting one of the nine quantile algorithms used only when km.orig = FALSE. See stats::quantile and Details below for more detail, default when km.orig = FLASE is set to 6.
Cdf	Logical TRUE/FALSE indicator of whether to plot the empirical cumulative distribution function (cdf).
ecdf.col	Color for the ecdf plotted step function line. Default is black.
km.orig	If TRUE (default), Kaplan-Meier results in the realm below detection limits reported as "NA". If FALSE, information in the detection limits is used and results in the realm of detection limits reported as "< DL", where DL is the appropriate detection limit.
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Ylab   Optional input text in quotes to be used as the variable name on the ecdf plot.
                The default is the name of the y1 input variable.

Details

Moment statistics are estimated using the enparCensored function of the EnvStats package. This avoids a small bias in the mean produced by the NADA package's cenfit function, which uses the reverse Kaplan-Meier procedure, converting left-censored to right-censored data prior to computing the ecdf and mean. See Gillespie et al.(2010) for more discussion on the bias of the estimate of the mean.

Quantiles and their two-sided confidence limits are estimated using the quantile function of the survfit command. See ?quantiles or Helsel et al. (2020) for choosing the q.type; default q.type = 4 (Kaplan-Meier; prob = i/n) when km.orig = TRUE. This is standard procedure in the survival analysis discipline and in the survival package of R, and is also used by the cenfit function in the NADA package. While this is 'industry standard' in medical applications it is a poor choice for observed sample (rather than census) data because it means that the largest observation is assigned a probability equal to 1, the 100th percentile. This implies that this value is never expected to be exceeded when more sample data are collected. It also means the largest observation is not plotted on the ecdf in most software because it is "off the chart". For small datasets in particular it is unlikely that the current largest observation is the largest value in the population and so the resulting ecdf quantiles are likely not opotimal. The default q.type = 6 (Weibull; prob = i/(n+1)) when km.orig = FALSE, though that may be changed by the user. the largest observation plots at a probability less than 1 on the ecdf. Differences in results when differing q.types are used will decrease as sample size increases.

All printed values will also be output to an object if saved. Confidence intervals on the quantiles are also output when data include nondetects. Values are character because of the possibility of a <1, but if no < symbol can be converted to numeric using the as.numeric(...) command. For data without censoring cfit will return numeric values. In that case it returns standard arithmetic mean, standard deviation and quantiles instead of K-M versions.

Value

If printstat=TRUE: Based on the provided conf value, Kaplan-Meier summary statistics (mean,sd,median), lower and upper confidence intervals around the mean and median value, sample size and percent of censored samples are returned. The specified quantile values are also printed and returned.

If Cdf=TRUE: The ecdf of censored data is plotted.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Gillespie, B.W., et al., 2010. Estimating Population Distributions When Some Data Are Below a Limit of Detection by Using a Reverse Kaplan-Meier Estimator. Epidemiology 21, 564-570.

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., and Gilroy, E.J., 2020, Statistical Methods in Water Resources: U.S. Geological Survey Techniques and Methods, book 4, chapter A3, 458 p., https://doi.org/10.3133/tm4a3.

Millard, S.P, 2013. EnvStats: An R Package for Environmental Statistics, 2nd ed. Springer Science+Business Media, USA, N.Y. DOI 10.1007/978-1-4614-8456-1© Springer Science+Business Media New York 2013"

Excerpt From: Steven P. Millard. "EnvStats." Apple Books.

See Also

survival::survfit

Examples

```
data(Brumbaugh)

cfit(Brumbaugh$Hg,Brumbaugh$HgCen)
```

---

computeS   *Kendall's S-statistic for permutations of censored data*

Description

Computes a Kendall rank correlation S-statistic for permutations of censored data. Collectively these represent the variation in S expected when the null hypothesis is true. Called by censeaken. computeS is not expected to be of much use to users on its own.

Usage

```
computeS(x, y, ycen, seas = NULL, R = R)
```

Arguments

x	Column of the time variable, either a sequence of days or decimal times, etc. Time data for one season.
y	The column of y (response variable) values plus detection limits for one season.
ycen	The y-variable indicators, where 1 (or TRUE) indicates a detection limit in the y column, and 0 (or FALSE) indicates a detected value in y.
seas	Name of a single season classification. Usually though not necessarily a text variable.
R	The number of repetitions in the permutation process. R is often between 999 and 9999 (+ the 1 observed test statistic produces 1000 to 10000 realizations).

Value

An Rx1 matrix containing an S-value for each of the R data permutations.

References

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., Gilroy, E.J., 2020. Statistical Methods in Water Resources. U.S. Geological Survey Techniques and Methods, book 4, chapter A3, 458p., https://doi.org/10.3133/tm4a3.

See Also

Kendall::Kendall

Examples

```
data(Brumbaugh)

#Artifical time and season variables for demonstration purposes
Brumbaugh$time=1:nrow(Brumbaugh)
Brumbaugh$sea=as.factor(round(runif(nrow(Brumbaugh),1,4),0))


with(Brumbaugh,computeS(time,Hg,HgCen,sea,R=100))
```

CuZn   *CuZn*

Description

Originally in the `NADA` R-Package.

Copper and zinc concentrations in ground waters from two zones in the San Joaquin Valley of California. The zinc concentrations were used.

Objective is to determine if zinc concentrations differ between the two zones. Zinc has two detection limits, at 3 and 10 ug/L. Used in Chapters 4, 5 and 9 of the NADA book.

Usage

```
data(CuZn)
```

Format

An object of class `data.frame` with 118 rows and 5 columns.

Source

Millard and Deverel, 1988, Water Resources Research 24, pp.2087-2098.

References

Helsel, D.R., 2005.Nondectects and Data Analysis; Statistics for censored environmental data. John Wiley & Sons, USA, N.J.

EM   *EM Algorithm for Interval-Censored Data*

Description

EM Algorithm for Interval-Censored Data

Usage

```
EM(A, pvec, maxiter = 500, tol = 1e-12)
```

Arguments

A	Either a logical matrix or an interval matrix (n x 2).
pvec	Optional initial probability vector.
maxiter	Maximum number of EM iterations (default 500).
tol	Convergence tolerance (default 1e-12).

Value

An object of class `"icsurv"` with elements:

pf Estimated probability mass function

numiter Number of iterations

converge Logical, whether EM converged

intmap Interval map (if applicable)

---

equivalent_n   *Censored data sample size*

Description

Computes the equivalent sample size of censored data. Observations at lower detection limits have a greater percent of the equivalent information of a detected value than observations at higher detection limits.

Usage

```
equivalent_n(y.var, y.cen, printstat = TRUE)
```

Arguments

`y.var`	The column of data values plus detection limits.
`y.cen`	The column of indicators, where 1 (or TRUE) indicates a detection limit in the `y.var` column, and 0 (or FALSE) indicates a detected value in `y.var`.
`printstat`	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Details

Based on "Method 2" of Dr. Brenda Gillespie's talk at ASA National Meeting 2019. This method differs from hers in how the percentile probabilities for the detection limits are computed. Probabilities here are computed using Regression on Order Statistics (ROS).

Computes the equivalent n, the number of observations including censored values, as a measure of information content for data with nondetects.

Value

Prints summary statistics including

- `n` sample size
- `n.cen` number of censored data
- `pct.cen` percent of data censored
- `min` minimum reported value
- `max` maximum reported value

Summary of censored data including

- `limit` detection limit
- `n` number of censored values per limit
- `uncen` number of detected values at or above the limit
- `pexceed` proportion of data that exceeds the limit

Summary of the equivalent sample size for detected and censored values.

- `n.equiv` the equivalent number of observations
- `n.cen.equiv` equivalent number of detected obs in the censored data
- `n.detected` number of uncensored values

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Gillespie, B.W., Dominguez, A., Li, Y., 2019. Quantifying the information in values below the detection limit (left-censored data). Presented at the 2019 Joint Statistical Meetings of the Amer. Stat. Assoc., Denver, CO., July 31, 2019.

Examples

```
data(Brumbaugh)

equivalent_n(Brumbaugh$Hg,Brumbaugh$HgCen)
```

Example1 *Example1*

Description

Arsenic concentrations (possibly artificial) in groundwater.

Objective – To test whether the mean concentration can be show to have exceeded the 10 microgram per liter health standard.

Usage

```
data(Example1)
```

Format

An object of class data.frame with 21 rows and 3 columns.

Source

unknown

References

unknown

Example2 *Example2*

Description

Methyl Isobutyl Ketone (MIBK) was measured in air above a medium-sized US city. Data were reported only as "ND" or as a detected concentration (no information on reporting limits).

Objective – To compute an upper confidence limit on the mean without having the value(s) of the reporting limit(s).

Usage

```
data(Example2)
```

Example3 55

Format

An object of class `data.frame` with 31 rows and 5 columns.

Source

unknown

References

unknown

Example3 *Example3*

Description

Arsenic concentrations in ground water from a private supply well. All 14 observations are below one of several reporting limits (1005 non-detects).

Objective – To show what type of comparison to a health standard can be made using a dataset without detected observations.

Usage

```
data(Example3)
```

Format

An object of class `data.frame` with 14 rows and 4 columns.

Source

unknown

References

unknown

Gales_Creek   *Gales_Creek*

Description

Total recoverable chromium concentrations in streamflows of Gales Creek in Oregon, USA.

Objective is to relate chromium concentrations including censored values to mean daily flows over time and by season (wet versus dry seasons). Two detection limits decreasing over time.

Usage

```
data(Gales_Creek)
```

Format

An object of class `tbl_df` (inherits from `tbl`, `data.frame`) with 63 rows and 11 columns.

Source

US Geological Survey. https://waterdata.usgs.gov/monitoring-location/453026123063401/

References

Publicly available data.

---

kenplot   *Plot robust median ATS line for censored data*

Description

Function used by other functions to plot the Akritas-Theil-Sen (ATS) line for censored data. Only one x variable allowed. Both Y and X variables may be censored.

Usage

```
kenplot(
  y1,
  ycen,
  x1,
  xcen,
  atsline = FALSE,
  xnam = NULL,
  ynam = NULL,
  Title = "Akritas - Theil - Sen line",
  ylim = NULL,
```

```
  xlim = NULL,
  pch = 19,
  cex = 0.7,
  xaxs = "r",
  yaxs = "r",
  ...
)
```

Arguments

y1	The column of y (response variable) values plus detection limits
ycen	The y-variable indicators, where 1 (or TRUE) indicates a detection limit in the y.var column, and 0 (or FALSE) indicates a detected value in y.var.
x1	The column of x (explanatory variable) values plus detection limits
xcen	The x-variable indicators, where 1 (or TRUE) indicates a detection limit in the x.var column, and 0 (or FALSE) indicates a detected value in x.var.
atsline	Indicator of whether to draw the ATS line or not. Default is FALSE.
xnam	Custom label for the x axis of plots. Default is x variable column name.
ynam	Custom label for the y axis of plots. Default is y variable column name.
Title	Custom title for plots. Default is "Akritas - Theil - Sen line".
ylim	argument consistent with base plotting to adjust y-axis limits; see par for more details
xlim	argument consistent with base plotting to adjust x-axis limits; see par for more details
pch	argument consistent with base plotting to adjust point type, see par for more details
cex	argument consistent with base plotting to adjust point size; see par for more details
xaxs	argument consistent with base plotting to adjust x-axis type; see par for more details
yaxs	argument consistent with base plotting to adjust y-axis type; see par for more details
...	argument to adjust other base plotting functions; see par for more details

Value

Scatterplot of data plus ATS line. Censored values are drawn for both X and Y variables as dashed lines up to the detection limits.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Examples

```
## Not run:
# Both y and x are censored
data(PbHeron)
with(PbHeron, kenplot(Blood, BloodCen, Kidney, KidneyCen))

# x is not censored
data(Brumbaugh)
with(Brumbaugh, kenplot(Hg, HgCen, PctWetland,rep(0, times=length(PctWetland))))

## End(Not run)
```

---

markers   *Markers*

Description

Three MST markers of human-origin contamination, and three of animal contamination, were evaluated along with a general fecal indicator to determine if the pattern of their levels differed among five coastal sites off the coast of Florida. Data are interval-censored.

Objective – Test whether the pattern of six microbial source tracking (MST) markers is related to the Entero1A indicator of general fecal pollution. Also determine if the pattern of markers plus Entero1A indicator differs among five coastal water sites using ANOSIM.

Usage

```
data(Markers)
```

Format

An object of class data.frame with 30 rows and 15 columns.

Source

Symonds et al. (2016) Journ Applied Microbio 121, p. 1469-1481.

References

Symonds et al. (2016) Journ Applied Microbio 121, p. 1469-1481.

NADAList   *NADA List Class*

Description

A simple S3 class used to group related NADA objects as a list.

Usage

```
NADAList(...)
```

Arguments

`...`	Objects to include in the list.

Value

An object of class `NADAList`, which is a list of the input components.

NADAprobs   *Default Probability Values for Quantile Estimation*

Description

A numeric vector of default probabilities used in quantile calculations, commonly used in censored data methods.

Usage

```
NADAprobs
```

Format

Numeric vector

Examples

```
## Not run:
NADAprobs
quantile(1:10, probs = NADAprobs)

## End(Not run)
```

---

ordranks   *Computes ranks of data with one or multiple detection limits*

Description

Computes the within-column ranks of data having one or more detection limits. If multiple limits are present in a column, data are first re-censored at the highest detection limit.

Usage

```
ordranks(dat.frame, paired = TRUE)
```

Arguments

dat.frame   A data frame. Default format is paired = TRUE, where for 3 chemical parameters the input format is C1 I1 C2 I2 C3 I3, a concentration column followed by its censoring indicator column.

paired   An option to specify paired = FALSE, where the input format would be C1 C2 C3 I1 I2 I3 where the C columns contain concentrations or a detection limit, and the I columns are their associated indicators, in the same order as the concentration columns.

Value

Returns columns of ranks of censored data in the same order as the paired columns of input data. For 3 chemical parameters, the data frame returned will be R1 R2 R3 where R represents the ranks of the C1 C2 C3 input data accounting for the censoring indicated by columns I1 I2 I3.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Examples

```
data(PbHeron)

ordranks(PbHeron[,4:15])
```

partplots	*Partial plots for censored MLE regression*

Description

Draws a partial plot for each X variable in regression of a censored Y variable against multiple X variables.

Usage

```
partplots(
  y.var,
  cen.var,
  x.vars,
  LOG = TRUE,
  smooth.method = "gam",
  gam.method = "tp",
  multiplot = TRUE,
  printstat = TRUE
)
```

Arguments

y.var	The column of y (response variable) values plus detection limits.
cen.var	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y.var column, and 0 (or FALSE) indicates a detected value in y.var.
x.vars	Multiple uncensored explanatory variable(s). See Details
LOG	Indicator of whether to compute the censored regression in the original y units, or on their logarithms. The default is to use the logarithms (LOG = TRUE). To compute in original Y units, specify the option LOG = FALSE (or LOG = 0).
smooth.method	Method for drawing a smooth on the partial plot. Options are c("gam", "none"). "gam" is a censored generalized additive model using the cenGAM and mgcv packages.
gam.method	Method for computing the gam smooth. See the mgcv package for options. Default is a thinplate ("tp") spline. "cs" is another good option.
multiplot	If TRUE, plots are drawn 6 per page. If FALSE, all plots are drawn on a separate page.
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Details

Partial plots for uncensored data often are drawn with superimposed smooths. At times looking only at the data values without a smooth can better enable the human eye to determine whether the overall pattern is linear or not. If this is the best method for you, use the smooth.method = "none"

option to not draw a smooth. The most common smooth used for uncensored data is loess, which does not recognize censored data and so uses the detection limit (DL) value itself. This results in biased-high smooths that incorrectly treat values at the DLs equal to uncensored (detected) data. The partplots function in NADA2 was written to provide a better alternative, smoothing the partial residual pattern with a censored generalized additive model (gam). The censored gam recognizes the nondetects as left-censored data with a maximum at the DL when computing the smooth. DLs may vary with each observation – multiple DLs in a dataset are not a problem in routines of the NADA2 package.

'y.var': The default is that the Y variable will be log transformed.

x.vars: Enter the name of a data frame of columns of the x variables. No extra columns un-used in the regression allowed. Create this by x.frame <- data.frame (Temp, Flow, Time) for 3 variables (temperature, flow and time).

Gray open circles represent censored data and are the residual between the detection limit and the predicted value from the censored regression. The GAM recognizes that the detection limit is an upper limit, predicted values on the regression line are most often below the detection limit, leading to positive residuals. Note that the true residual for censored data could be anywhere below the plotted value. That fact is recognized by the censored GAM but is difficult to represent on a plot.

AIC for regression models with un-transformed X, log and cube-root transforms of X are printed to evaluate which of the three transformations results in the 'best' model.

Value

When x.vars is one variable, a message is printed that partial plots are not possible with only one X variable and execution stops. When x.vars is a data frame of more than one variable, partial plots are drawn for each X variable and text is printed comparing AICs for regression using the untransformed X variable with log and cube-root transforms of the X variable, as a supplement to evaluating linearity on the partial plots alone.

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Cook, R.D., 1993. Exploring Partial Residual Plots, Technometrics 35, 351-362.

See Also

survival::survreg

Examples

```
data(Brumbaugh)

# For demostration purposes
partplots (Brumbaugh$Hg,Brumbaugh$HgCen,Brumbaugh[,c("SedMeHg","PctWetland")])
```

PbHeron   *PbHeron*

Description

From the NADA R-Package.

Lead concentrations in the blood and several organs of herons in Virginia.

Objective is to determine the relationships between lead concentrations in the blood and various organs. Do concentrations reflect environmental lead concentrations, as represented by dosing groups? There is one detection limit, at 0.02 ug/g. Used in Chapters 10 and 11 of the Helsel (2011).

Usage

```
data(PbHeron)
```

Format

An object of class data.frame with 27 rows and 15 columns.

Source

Golden et al., 2003, Environmental Toxicology and Chemistry 22, pp. 1517-1524.

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Pollen_Thia   *Thiamethoxam concentrations in pollen*

Description

Thiamethoxam concentrations in pollen from the Ontario Pollen Monitoring Network study https://data.ontario.ca/en/dataset/monitoring-network-study

Variables are:

Thiamethoxam: Thiamethoxam concentration in Concentrations in microgram per gram.

ThiaCens: Censoring indicator. 1 denotes that the value in column 1 is a reporting limit not a specific concentration.

SamplingEvent: A grouping variable from the sample design. A concentration is from 1 of 4 events in time.

ThiaAbvBelow: A binary variable denoting whether the Thiamethoxam concentration is above or below 0.05 ug/g.

Usage

```
data(Pollen_Thia)
```

Format

An object of class data.frame with 204 rows and 4 columns.

Source

Ontario Ministry of Agriculture, Food and Rural Affairs (Pollen Monitoring Network)

---

ppw.test	*Test for difference in left-censored samples*

Description

Performs a nonparametric Paired Prentice-Wilcoxon test of whether the median difference between two columns of paired censored data equals 0 (O'Brien and Fleming, 1987)

Usage

```
ppw.test(xd, xc, yd, yc, alternative = "two.sided", printstat = TRUE)
```

Arguments

xd	The first column of data values plus detection limits
xc	The column of censoring indicators, where 1 (or TRUE) indicates a detection limit in the xd column, and 0 (or FALSE) indicates a detected value in xd.
yd	The second column of data values plus detection limits
yc	The column of censoring indicators, where 1 (or TRUE) indicates a detection limit in the yd column, and 0 (or FALSE) indicates a detected value in yd
alternative	The usual notation for the alternate hypothesis. Default is "two.sided". Options are "greater" or "less".
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Value

Paired Prentice-Wilcoxon test results including Z-statistic, n (sample size), p-value and median difference

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

O'Brien, P.C., Fleming, T.R., 1987. A Paired Prentice-Wilcoxon Test for Censored Paired Data. Biometrics 43, 169–180. https://doi.org/10.2307/2531957

See Also

survival::survfit survival::Surv

Examples

```
data(PbHeron)
ppw.test(PbHeron$Liver,PbHeron$LiverCen,PbHeron$Bone,PbHeron$BoneCen)
```

---

print.NADAList	*Show Method for NADAList Objects*

Description

Prints named elements of a `NADAList` object.

Usage

```
## S3 method for class 'NADAList'
print(x, ...)
```

Arguments

x	A `NADAList` object.
...	further arguments passed to or from other methods. (not used)

ReconLogistic	*ReconLogistic*

Description

Atrazine exceedances (Y/N) above 1 ug/L in streams throughout the Midwestern United States. Original concentration data were censored to show either greater than (1) or less than (0) 1 ug/L.

Objective – determine if the pattern of atrazine concentrations above versus below 1 ug/L were related to seven environmental variables.

Usage

```
data(ReconLogistic)
```

Format

An object of class `data.frame` with 423 rows and 8 columns.

Source

Mueller et al., 1997, Journal of Environmental Quality 26, pp. 1223-1230.

References

Chapter 12 of Helsel, Dennis R. (2012). Statistics for Censored Environmental Data using Minitab and R. John Wiley and Sons, USA, NJ..

---

rescaleP   *Rescale probability vector*

Description

Rescale probability vector

Usage

```
rescaleP(pvec, tiny)
```

Arguments

pvec	A probability vector.
tiny	Threshold below which values are set to zero.

Value

A rescaled probability vector.

---

ros   *Regression on Order Statistics (ROS)*

Description

Perform regression on order statistics for left-censored data.

Usage

```
ros(
  obs,
  censored = NULL,
  data = NULL,
  forwardT = "log",
  reverseT = "exp",
  na.action = getOption("na.action")
)
```

```
## S3 method for class 'ros'
plot(
  x,
  plot.censored = FALSE,
  lm.line = TRUE,
  grid = TRUE,
  ylab = "Value",
  pch = 16,
  ...
)
```

Arguments

obs	Numeric vector of observations or formula of the form `response ~ censor`, where `response` is numeric values and `censor` is a logical or binary indicator (TRUE if censored).
censored	Logical vector of left-censored indicators.
data	A `data.frame` containing the variables in the formula.
forwardT	Name of transformation function (e.g., "log", "trueT").
reverseT	Name of back-transformation function (e.g., "exp", "trueT").
na.action	Function to handle missing values.
x	ros2 model object
plot.censored	default = FALSE, if set to true it will also plot censored data
lm.line	will plot linear model line
grid	will add grid
ylab	default is "Value" but custom text can be added
pch	default set to 16, codes consistent with `points` and `plot` functions
...	arguments passed to plot function

Details

Code for this function is originally from the NADA package developed by R. Lopaka Lee and Dennis Helsel. By default, ros performs a log transformation prior to, and after operations over the data. This can be changed by specifying a forward and reverse transformation function using the forwardT and reverseT parameters. No transformation will be performed if either forwardT or reverseT are set to NULL.

The procedure first computes the Weibull-type plotting positions of the combined uncensored and censored observations using a formula designed for multiply-censored data (see hc_ppoints). A linear regression is formed using the plotting positions of the uncensored observations and their normal quantiles. This model is then used to estimate the concentration of the censored observations as a function of their normal quantiles. Finally, the observed uncensored values are combined with modeled censored values to corporately estimate summary statistics of the entire population. By combining the uncensored values with modeled censored values, this method is more resistant of any non-normality of errors, and reduces any transformation errors that may be incurred.

Value

A list with:

modeled Numeric vector of uncensored + imputed censored values.

modeled.censored Imputed values only for censored observations.

uncensored Original uncensored values.

censored Original censored values.

censored.ranks Censored ranks used in estimation.

uncensored.ranks Uncensored ranks used in estimation.

model Fitted linear model object.

Examples

```
df <- data.frame(
  conc = c(0.2, 0.5, 1.0, 0.4, 2.0, 0.3),
  censored = c(TRUE, TRUE, FALSE, TRUE, FALSE, TRUE)
)
ros(conc ~ censored, data = df)
```

ROSci   *Computes confidence intervals on regression on order statistics (ROS)*
                                       mean

Description

Uses ROS model output, computes the Zhou and Gao 1997 modified Cox's method two-sided confidence interval around the mean for a lognormal distribution. Computes a t-interval for a gaussian ROS model output.

Usage

```
ROSci(cenros.out, conf = 0.95, printstat = TRUE)
```

Arguments

cenros.out	an ROS model output object (see details)
conf	Confidence coefficient of the interval (Default is 0.95)
printstat	Logical TRUE/FALSE option of whether to print the resulting statistics in the console window, or not. Default is TRUE.

Details

This function uses an ROS model output based on the ros function, prevuously in the NADA package, now in this package. The lognormal distribution is the default for the NADA package but a gaussian distribution is optional here.

For implementation of ROSci(...) see the examples below.

Value

Prints a lower (LCL) and upper (UCL) confidence interval based on the `conf` provided (Default is 95%)

References

Helsel, D.R., 2011. Statistics for censored environmental data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Lee, L., Helsel, D., 2005. Statistical analysis of water-quality data containing multiple detection limits: S-language software for regression on order statistics. Computers & Geosciences 31, 1241–1248. doi:10.1016/j.cageo.2005.03.012

Zhou, X.-H., Gao, S., 1997. Confidence Intervals for the Log-Normal Mean. Statistics in Medicine 16, 783–790. doi:10.1002/(SICI)10970258(19970415)16:7<783::AIDSIM488>3.0.CO;22

Examples

```
data(Brumbaugh)
myros <- ros(Brumbaugh$Hg,Brumbaugh$HgCen)

summary(myros)

# ROS Mean
mean(myros$modeled)

# 95% CI around the ROS mean
ROSci(myros)
```

ShePyrene *ShePyrene*

Description

Originally in the `NADA` R-Package.

Pyrene concentrations in milligrams per liter from 20 water-quality monitoring stations in the Puget Sound of Washington State, USA.

Used for characterizing priority pollutant concentrations in sediments of Puget Sound by computing summary statisitics. Contains eight detection limits with 11 nondetects out of 56 total measurements.

Usage

```
data(ShePyrene)
```

Format

An object of class `data.frame` with 56 rows and 2 columns.

Source

She, N., 1997, Analyzing censored water quality data using a nonparametric approach. Journal of the American Water Resources Association, 33, pp615–624.

References

Helsel, D.R., 2005.Nondectects and Data Analysis; Statistics for censored environmental data. John Wiley & Sons, USA, N.J.

TCE2   *TCE2*

Description

TCE concentrations (ug/L) in ground waters of Long Island, New York. Categorized by two of the three land use types (medium and high density residential) surrounding the wells and measured in the study.

Objective – determine if TCE concentrations differ in ground water under two land use types. There are four detection limits, at 1,2,4 and 5 ug/L.

Usage

```
data(TCE2)
```

Format

An object of class data.frame with 222 rows and 6 columns.

Source

Eckhardt et al., 1989, USGS Water Resources Investigations Report 86-4142.

References

The full dataset with three land use types is used in Chapter 10 of Statistics for Censored Environmental Data using Minitab and R, by D.R.Helsel; Wiley (2012).

TCEReg *TCE Ground Waters of Long Island — with Explanatory Variables*

Description

TCE concentrations (μg/L) in ground waters of Long Island, New York, along with several possible explanatory variables.

Usage

```
data(TCEReg)
```

Format

A data frame with variables related to TCE groundwater concentrations.

Details

The objective is to determine if concentrations are related to one or more explanatory variables.

There are four detection limits at 1, 2, 4, and 5 μg/L. One column indicates whether concentrations are above or below 5. Used in Chapter 12 of the NADA book.

Source

Eckhardt et al., 1989. USGS Water Resources Investigations Report 86-4142.

References

Helsel, Dennis R. (2005). *Nondetects and Data Analysis: Statistics for Censored Environmental Data*. John Wiley and Sons, USA, NJ.

---

uMDS *Plot U-score Nonmetric Multidimensional Scaling of censored data*

Description

Plots an NMDS of uscores output from the uscores or uscoresi functions.

Usage

```
uMDS(uscor, group = NULL, title = NULL, legend.pos = "bottomleft")
```

Arguments

uscor	A data frame of uscores or ranks of uscores produced by either the uscores(...) or uscoresi(...) functions
group	Optional grouping variable. Sites will be represented by different colored symbols for each group.
title	Optional title for the NMDS graph.
legend.pos	For when group is specified, the location of the legend on the graph showing the colors representing each group's data. Default is "bottomleft". Alternatives are "topright" and "centerleft", etc.

Value

Prints an NMDS plot of censored data groupings based on U-scores #' @references Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Examples

```
data(PbHeron)

PbHeron.u <- uscores(PbHeron[,4:15])
uMDS(PbHeron.u)

# With group specific
uMDS(PbHeron.u,group=PbHeron$DosageGroup)
```

Usc	*U-scores for (non-interval, sinle-column) Censored Data*

Description

Computes the column of uscores from 2 columns of data in the indicator value format. Multiple detection limits allowed. Called by the uscores function, Usc (this function) is not expected to be of much use to users on its own.

Usage

```
Usc(y, ind, rnk = TRUE)
```

Arguments

y	The column of data values plus detection limits
ind	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y column, and 0 (or FALSE) indicates a detected value in y.
rnk	A TRUE/FALSE variable on whether to compute the multivariate pattern on the uscores, or the ranks of the uscores. Default is rnk=TRUE, use the ranks. rnk = FALSE returns the uscores.

Value

Returns a single column of uscores or the ranks of uscores for a single pair of (concentration, indicator) censored data columns.

Examples

```
data(Brumbaugh)
uscore(Brumbaugh$Hg,Brumbaugh$HgCen)
```

Usci   *Interval-censored U-Score*

Description

Interval-censored computation of uscores and their ranks for 1 parameter. Called by uscoresi. Usci is not expected to be of much use to users on its own.

Usage

```
Usci(ylo, yhi, rnk = TRUE)
```

Arguments

ylo	The lower end of the concentration interval
yhi	The upper end of the concentration interval
rnk	A TRUE/FALSE variable on whether to compute the multivariate pattern on the uscores, or the ranks of the uscores. Default is rnk=TRUE, use the ranks. rnk = FALSE returns the uscores.

Value

Returns a single column of uscores or the ranks of uscores for a single pair of (low, high) interval-censored data columns.

Examples

```
data(Brumbaugh)

# for demonstration purposes create a lower end concentration interval
Brumbaugh$lowHg<-Brumbaugh$Hg*(1-Brumbaugh$HgCen)

with(Brumbaugh,Usci(lowHg,Hg))
```

---

uscore   *U-score (individual value)*

Description

Computes the uscore of data (required 2-columns) with one or more detection limits.

Usage

```
uscore(y, ind, rnk = TRUE)
```

Arguments

y	The column of data values plus detection limits
ind	The column of indicators, where 1 (or TRUE) indicates a detection limit in the y column, and 0 (or FALSE) indicates a detected value in y.
rnk	A TRUE/FALSE variable on whether to compute the multivariate pattern on the uscores, or the ranks of the uscores. Default is rnk=TRUE, use the ranks. rnk = FALSE returns the uscores.

Value

prints the uscore number of observations known to be lower - number of observations known to be higher, for each observation.

Examples

```
data(Brumbaugh)
uscore(Brumbaugh$Hg,Brumbaugh$HgCen)
```

---

uscores   *Uscores for multiple columns of censored data*

Description

Computes uscores or the ranks of uscores of censored data in the indicator format. Multiple DLs allowed.

Usage

```
uscores(dat.frame, paired = TRUE, rnk = TRUE)
```

Arguments

`dat.frame`	A data frame. Default format is paired = TRUE, where for 3 chemical parameters the input format is C1 I1 C2 I2 C3 I3, a concentration column followed by its corresponding indicator column.
`paired`	When paired = FALSE, the input format is C1 C2 C3 I1 I2 I3 where the C columns contain concentrations or a detection limit, and the I columns are their associated indicators, in the same order as the concentration columns.
`rnk`	A logical TRUE/FALSE variable on whether to compute the multivariate pattern on the uscores, or the ranks of the uscores. Default is rnk=TRUE, use the ranks. rnk = FALSE returns the uscores.

Value

A matrix of uscores or ranks of uscores, one column for each chemical parameter. If there is only one chemical parameter a vector of uscores or ranks of uscores is returned.

Examples

```
data(PbHeron)

uscores(PbHeron[,4:15])
```

---

uscoresi	*U-scores for interval-censored data (multiple columns)*

Description

Computes uscores or the ranks off uscores within columns of interval-censored data (the "i"). Data may have one or more detection limits.

Usage

```
uscoresi(dat.frame, paired = TRUE, rnk = TRUE, Cnames = 1)
```

Arguments

`dat.frame`	A data frame. Default format is: paired = TRUE, the default input format is ylo1 yhi1 ylo2 yhi2 ylo3 yhi3, where ylo is the low end of the interval of possible values, and yhi is the high end. There is a pair of columns for each censored parameter.
`paired`	An option to specify paired = FALSE, where the format would be ylo1 ylo2 ylo3 yhi1 yhi2 yhi3, low values for each parameter followed by the high values in the same order.
`rnk`	rnk=TRUE returns the ranks of uscores. rnk = FALSE returns the uscores themselves. Default is rnk = TRUE to return the ranks.
`Cnames`	Cnames =1 uses the "lo" column names to name the uscores columns (the default). Cnames = 2 uses the "hi" column names.

Details

Input is a data.frame of paired low and high possible range of values, in an interval-censored format. ylo = the lower end of the interval is the first (left) column in the pair. yhi is the upper end of the interval, at the second (right) column in the pair. For a detected value, ylo=yhi. For a ND, ylo != yhi. The uscore is the number of observations known to be lower minus the number of observations known to be higher. Ties, including those such as <1 vs <3 or 4 vs 4 or <3 vs 2, are given a 0 value in the uscore computation. The ranks of uscores provides a scale that is often more manageable than the uscores themselves.

Value

prints the uscore number of observations known to be lower - number of observations known to be higher, for each observation.

References

Helsel, D.R., 2011. Statistics for Censored Environmental Data using Minitab and R, 2nd ed. John Wiley & Sons, USA, N.J.

Index