Package 'tailTransform'

March 21, 2022

Type Package

Title Symmetric Transformation of Tails for Plotting Differences

Version 1.0.4

Author Paul R. Rosenbaum

Maintainer Paul R. Rosenbaum <rosenbaum@wharton.upenn.edu>

Description When plotting treated-minus-control differences, after-minus-before changes, or difference-in-differences, the ttrans() function symmetrically transforms the positive and negative tails to aid plotting. The package includes an observational study with three control groups and an unaffected outcome; see Rosenbaum (2020) <doi:10.1111/biom.13558>.

License GPL-2

Encoding UTF-8

LazyData true

Imports stats, graphics

Suggests sensitivitymw, sensitivitymult

Depends R (>= 3.5.0)

NeedsCompilation no

Repository CRAN

Date/Publication 2022-03-21 08:10:02 UTC

R topics documented:

---

tailTransform-package   *Symmetric Transformation of Tails for Plotting Differences*

Description

When plotting treated-minus-control differences, after-minus-before changes, or difference-in-differences, the ttrans() function symmetrically transforms the positive and negative tails to aid plotting. The package includes an observational study with three control groups and an unaffected outcome; see Rosenbaum (2020) <doi:10.1111/biom.13558>.

Details

The DESCRIPTION file:

Package:	tailTransform
Type:	Package
Title:	Symmetric Transformation of Tails for Plotting Differences
Version:	1.0.4
Author:	Paul R. Rosenbaum
Maintainer:	Paul R. Rosenbaum <rosenbaum@wharton.upenn.edu>
Description:	When plotting treated-minus-control differences, after-minus-before changes, or difference-in-differences, the t
License:	GPL-2
Encoding:	UTF-8
LazyData:	true
Imports:	stats, graphics
Suggests:	sensitivitymw, sensitivitymult
Depends:	R (>= 3.5.0)

Index of help topics:

```
aHDL                 Alcohol and HDL Cholesterol: An Observational
                     Study with 3 Control Groups
boxplotTT            Parallel Boxplots After Tail Transformation
tailTransform-package Symmetric Transformation of Tails for Plotting
                     Differences
ttrans               Symmetric Tail Transformation of Differences
                     for Graphical Display
```

The package contains three items: (i) a function ttrans() that symmetrically shortens the tails of pair differences for graphical display, (ii) a function boxplotTT() that aids in interpreting and displaying the transformed data, and (iii) an observational study aHDL with three control groups and an unaffected outcome that are intended to reveal unmeasured confounding if it is present.<doi:10.1111/biom.13558>

Author(s)

Paul R. Rosenbaum

Maintainer: Paul R. Rosenbaum <rosenbaum@wharton.upenn.edu>

References

Rosenbaum, P. R. (2022a). Sensitivity analyses informed by tests for bias in observational studies. Biometrics. <doi:10.1111/biom.13558>

Rosenbaum, P. R. (2022b). A new transformation of treated-control matched-pair differences for graphical display. Manuscript.

Examples

```
data(aHDL)
attach(aHDL)
d<-hdl[grpL=="D"]-hdl[grpL=="N"]  # pair differences
tcks<-c(-100,-60,-40,-20,0,20,40,60,200)
boxplotTT(d,p=-1,qu=.95,tcks=tcks)
detach(aHDL)
rm(tcks,aHDL)
```

---

aHDL   *Alcohol and HDL Cholesterol: An Observational Study with 3 Control Groups*

Description

A small observational study of light daily alcohol consumption and HDL cholesterol – so-called good cholesterol – derived from NHANES 2013-2014 and 2015-2016. There are 406 matched sets of four individuals, making 1624 individuals in total. Sets were matched for age, female and education in five ordered categories.

Usage

```
data("aHDL")
```

Format

A data frame with 1624 observations on the following 11 variables.

nh NHANES 2013-2014 is 1314, and NHANES 2015-2016 is 1516

SEQN NHANES ID number

age Age in years

female 1=female, 0=male

education 1 is <9th grade, 3 is high school, 5 is a BA degree

z 1=light almost daily alcohol, 0=little or no alcohol last year.

grp Treated group and control groups. Daily=light almost daily alcohol, Never=fewer than 12 drinks during entire life, Rarely=more than 12 drinks in life, but fewer than 12 in the past year, and never had a period of daily binge drinking, PastBinge = a past history of binge drinking on most days, but currently drinks once a week or less. For details, see Rosenbaum (2022a, Appendix).

grpL Short labels for plotting formed as the first letters of grp. D < N < R < B

hdl HDL cholesterol level mg/dL

mmercury Methylmercury level ug/L

mset Matched set indicator, 1, 2, ..., 406. The 1624 observations are in 406 matched sets, each of size 4.

Details

There is a debate about whether light daily alcohol consumption – a single glass of red wine – shortens or lengthens life. LoConte et al. (2018) emphasize that alcohol is a carcinogen. Suh et al. (1992) claim reduced cardiovascular mortality brought about by an increase in high density high-density lipoprotein (HDL) cholesterol, the so-called good cholesterol. There is on-going debate about whether there are cardiovascular benefits, and if they exist, whether they are large enough to offset an increased risk of cancer. This example looks at a small corner of the larger debate, namely the effect on HDL cholesterol.

The example contains several attempts to detect unmeasured confounding bias, if present. There is a secondary outcome thought to be unaffected by alcohol consumption, namely methylmercury levels in the blood, likely an indicator of the consumption of fish, not of alcohol; see Pedersen et al. (1994) and WHO (2021). There are also three control groups, all with little present alcohol consumption, but with different uses of alcohol in the past; see the definition of variable grp above.

The appendix to Rosenbaum (2022a) describes the data and matching in detail. It is used as an example in Rosenbaum (2022b).

The help file for boxplotTT() applies the tail transformation to this example, reproducing a plot from Rosenbaum (2022b).

Source

US National Health and Nutrition Examination Survey (NHANES), 2013-2014 and 2015-2016. <www.cdc.gov/nchs/nhanes>

References

LoConte, N. K., Brewster, A. M., Kaur, J. S., Merrill, J. K., and Alberg, A. J. (2018). Alcohol and cancer: a statement of the American Society of Clinical Oncology. Journal of Clinical Oncology 36, 83-93. <doi:10.1200/JCO.2017.76.1155>

Pedersen, G. A., Mortensen, G. K. and Larsen, E. H. (1994) Beverages as a source of toxic trace element intake. Food Additives and Contaminants, 11, 351–363. <doi:10.1080/02652039409374234>

Rosenbaum, P. R. (1987). The role of a second control group in an observational study. Statistical Science, 2, 292-306. <doi:10.1214/ss/1177013232>

Rosenbaum, P. R. (1989). The role of known effects in observational studies. Biometrics, 45, 557-569. <doi:10.2307/2531497>

Rosenbaum, P. R. (1989). On permutation tests for hidden biases in observational studies. The Annals of Statistics, 17, 643-653. <doi:10.1214/aos/1176347131>

Rosenbaum, P. R. (2014) Weighted M-statistics with superior design sensitivity in matched observational studies with multiple controls. Journal of the American Statistical Association, 109(507), 1145-1158 <doi:10.1080/01621459.2013.879261>

Rosenbaum, P. R. (2022a). Sensitivity analyses informed by tests for bias in observational studies. Biometrics. <doi:10.1111/biom.13558>

Rosenbaum, P. R. (2022b). A new transformation of treated-control matched-pair differences for graphical display. Manuscript.

Suh, I., Shaten, B. J., Cutler, J. A., and Kuller, L. H. (1992). Alcohol use and mortality from coronary heart disease: the role of high-density lipoprotein cholesterol. Annals of Internal Medicine 116, 881-887. <doi:10.7326/0003-4819-116-11-881>

World Health Organization (2021). Mercury and Health, <https://www.who.int/news-room/fact-sheets/detail/mercury-and-health>, (Accessed 30 August 2021).

Examples

```
data(aHDL)
table(aHDL$grp,aHDL$grpL) # Short labels for plotting
boxplot(aHDL$age~aHDL$grp,xlab="Group",ylab="Age")
boxplot(aHDL$education~aHDL$grp,xlab="Group",ylab="Education")
table(aHDL$female,aHDL$grpL)
table(aHDL$z,aHDL$grpL)

# The sets were also matched for is.na(aHDL$mmercury), for use
# in Rosenbaum (2022a).  About half of the matched sets
# have values for mmercury.
table(is.na(aHDL$mmercury),aHDL$grp)

# Sensitivity analysis in Rosenbaum (2022b); see
# also Rosenbaum (2014)
y<-t(matrix(aHDL$hdl,4,406))
y<-y[,c(1,3,2,4)]
colnames(y)<-c("D","N","R","B")
sensitivitymw::senmw(y,gamma=6,method="f")$pval
sensitivitymult::amplify(6,11)

# See also the informedSen package for additional analysis
```

| boxplotTT | *Parallel Boxplots After Tail Transformation* |

Description

Plots one or more boxplots for differences after applying the same tail transformation to all differences. See the help file for ttrans() for information about the transformation.

Usage

```
boxplotTT(y, p = -1, qu = 0.95, tcks = NULL, ylab = "", xlab = "",main = "")
```

Arguments

y	A vector, matrix, or dataframe of differences to be transformed and plotted. Each column becomes a different boxplot after transformation.
p	The power of the transformation. See the help file for the ttrans() function.
qu	One number strictly between 0 and 1, commonly 0.9 or 0.95. Define beta to be the qu quantile of abs(as.vector(y)). Then, values between -beta and beta are not transformed.
tcks	A vector of untransformed values to become tick marks for the y axis after transformation. Although the y axis was transformed, it is labeled with corresponding untransformed values given by tcks. If is.null(tcks), then the y axis has no tick marks.
ylab	Label for the y-axis.
xlab	Label for the x-axis.
main	Title for the plot.

Details

See the help file for ttrans() for an explanation of the transformation.

Value

A boxplot.

Author(s)

Paul R. Rosenbaum

References

Rosenbaum, P. R. (2022). A new transformation of treated-control matched-pair differences for graphical display. Manuscript.

See Also

ttrans

Examples

```
data(aHDL)
y<-t(matrix(aHDL$hdl,4,406))
y<-y[,c(1,3,2,4)]
colnames(y)<-c("D","N","R","B")
```

```
# 6 pairwise comparisons of 4 groups
o<-matrix(NA,dim(y)[1],6)
colnames(o)<-1:6
k<-0
for (i in 1:3)
  for (j in (i+1):4){
    k<-k+1
    colnames(o)[k]<-paste(colnames(y)[i],colnames(y)[j],sep="-")
    o[,k]<-y[,i]-y[,j]
  }
rm(i,j,k)

# Plotting tick marks.  Remember, the transformation compresses
# extremes, so unequally spaced tick marks are usually needed.
tcks<-c(-100,-60,-40,-20,0,20,40,60,200)

# tails transformed by the p=-1 power (i.e., reciprocal)
boxplotTT(o,p=-1,qu=.95,tcks=tcks,ylab="HDL",
     xlab="Pairwise Comparisons of 4 Groups",
     main="HDL Differences in 4 Alcohol Groups")
```

| ttrans | *Symmetric Tail Transformation of Differences for Graphical Display* |

Description

Performs a differentiable, strictly increasing, odd transformation of treated-minus-control pair differences d, or after-minus-before changes, or difference-in-differences. The transformation allows one to see the undistorted center of a distribution that contains extreme outliers, while also seeing the outliers. The transformation t(d) is: (i) odd, meaning t(d)= -t(-d), so positive and negative values of d are transformed symmetrically, (ii) for some number beta>0, the transformation leaves d untouched between -beta and beta, so t(d)=d for -beta < d < beta, (iii) the transformation has derivative 1 at -beta and beta, so it is smooth at the point where the nonlinear transformation begins to take effect.

Usage

```
ttrans(d, p = -1, qu = NULL, beta = NULL)
```

Arguments

d	A vector of differences to be transformed.
p	The power to be used in the transformation of the tails, with p=0 being the log, as in the Box-Cox-Tukey transformation.
qu	If qu is specified, it is a number strictly between 0 and 1, commonly 0.9 or 0.95. Then beta is set to be the qu quantile of abs(d). If qu=.95, then 95 percent of the differences in d are not transformed. You must specify either qu or beta, and you must not specify both qu and beta.

beta   The value beta mentioned in the description. You must specify either qu or beta,
                         and you must not specify both qu and beta.

Details

Recall that beta>0. Let y be one difference in d. If y=0, then t(y)=y=0. If 0<y<=beta, then t(y) =
y. If y>beta, then a Box-Cox-Tukey power transformation is applied to y, with the transformation
relocated and scaled so that t(y) has derivative 1 at beta. If y<0, then t(y) = -t(|y|). Properties of the
transformation are discussed in Rosenbaum (2022).

Although t(y) is nonlinear, it is exactly linear with slope 1 between -beta and beta, and t(y) is smooth
with slope 1 at -beta and beta. The nonlinear aspect of the transformation is barely visible near -beta
and beta.

If d is symmetric about zero, then the transformed values are also symmetric about zero. If there is
no effect on the differences, in the sense that they are symmetric about zero, then the transformed
differences also exhibit no effect.

The transformation does not alter Wilcoxon's signed rank statistic, or other signed rank statistics.
Specifically, the transformation does not alter the ranks of |d|, and it does not alter sign(d).

The p=-1 reciprocal transformation has an upper and lower asymptote, so it limits the range of the
d's, but it shows outliers clearly.

Value

A vector of transformed values of d.

Author(s)

Paul R. Rosenbaum

References

Box, George E. P. and David R. Cox. (1964) An analysis of transformations. Journal of the Royal
Statistical Society: Series B 26, 211-243. <doi:10.1111/j.2517-6161.1964.tb00553.x>

Rosenbaum, P. R. (2022). A new transformation of treated-control matched-pair differences for
graphical display. Manuscript.

Tukey, J. W. (1949). One degree of freedom for non-additivity. Biometrics, 5, 232-242. <doi:10.2307/3001938>

Tukey, J. W. (1957). On the comparative anatomy of transformations. Annals of Mathematical
Statistics, 28, 602-632. <doi:10.1214/aoms/1177706875>

Examples

```
data(aHDL)
attach(aHDL)
d<-hdl[grpL=="D"]-hdl[grpL=="N"]  # pair differences
detach(aHDL)
oldpar<-par()
par(mfrow=c(1,2))
boxplot(d) # untransformed
boxplot(ttrans(d,qu=.95,p=-1)) # reciprocal transformation of tails
```

```
par(mfrow=c(1,1))
# Label the transformed vertical axis with untransformed values
# Add -beta and beta on the right axis
tcks<-c(-100,-60,-40,-20,0,20,40,60,200)
boxplotTT(d,p=-1,qu=.95,tcks=tcks)

par<-oldpar
rm(aHDL,d,oldpar)
```

Index