Package 'weightedRank'

July 22, 2025

Type Package

Title Sensitivity Analysis Using Weighted Rank Statistics

Version 0.3.7

Description Performs a sensitivity analysis using weighted rank tests in observational studies with I blocks of size J; see Rosenbaum (2024) <doi:10.1080/01621459.2023.2221402>. The package can perform adaptive inference in block designs; see Rosenbaum (2012) <doi:10.1093/biomet/ass032>. The main functions are wgtRank(), wgtRankCI() and wgtRanktt().

License GPL-2

Encoding UTF-8

LazyData true

Imports stats, graphics, mvtnorm, sensitivitymv, senstrat

Suggests sensitivitymw, sensitivitymult, DOS2

Depends R (>= 3.5.0)

NeedsCompilation no

Author Paul Rosenbaum [aut, cre]

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

Repository CRAN

Date/Publication 2024-07-06 19:20:01 UTC

Contents

---

weightedRank-package   *Sensitivity Analysis Using Weighted Rank Statistics*

Description

Performs a sensitivity analysis using weighted rank tests in observational studies with I blocks of size J; see Rosenbaum (2024) <doi:10.1080/01621459.2023.2221402>. The package can perform adaptive inference in block designs; see Rosenbaum (2012) <doi:10.1093/biomet/ass032>. The main functions are wgtRank(), wgtRankCI() and wgtRanktt().

Details

The DESCRIPTION file:

Package:	weightedRank
Type:	Package
Title:	Sensitivity Analysis Using Weighted Rank Statistics
Version:	0.3.7
Authors@R:	person("Paul", "Rosenbaum", email = "rosenbaum@wharton.upenn.edu", role = c("aut", "cre"))
Description:	Performs a sensitivity analysis using weighted rank tests in observational studies with I blocks of size J; see Ro
License:	GPL-2
Encoding:	UTF-8
LazyData:	true
Imports:	stats, graphics, mvtnorm, sensitivitymv, senstrat
Suggests:	sensitivitymw, sensitivitymult, DOS2
Depends:	R (>= 3.5.0)
Author:	Paul Rosenbaum [aut, cre]
Maintainer:	Paul Rosenbaum <rosenbaum@wharton.upenn.edu>

Index of help topics:

```
aBP                 Binge Drinking and Blood Pressure
aHDL                Alcohol and HDL Cholesterol
aHDLe               HDL Cholesterol and Light Daily Alcohol
                    2013-2020
amplify             Amplification of sensitivity analysis in
                    observational studies.
dwgtRank            Weighted Rank Statistics for Evidence Factors
                    with Two Control Groups
ef2C                Evidence Factors For Matched Triples With Two
```

```
                            Control Groups
gwgtRank                    Generalized Sensitivity Analysis for Weighted
                            Rank Statistics in Block Designs
stepSolve                   Root of a Monotone Decreasing Step Function
weightedRank-package        Sensitivity Analysis Using Weighted Rank
                            Statistics
wgtRank                     Sensitivity Analysis for Weighted Rank
                            Statistics in Block Designs
wgtRankC                    Sensitivity Analysis for a Conditional Weighted
                            Rank Test
wgtRankCI                   Sensitivity Analysis for Confidence Intervals
                            and Point Estimates from Weighted Rank
                            Statistics in Block Designs
wgtRanktt                   Adaptive Inference Using Two Test Statistics in
                            a Block Design
```

The package conducts either fixed or adaptive sensitivity analyses for observational studies with
I blocks and J individuals in each block, one treated and J-1 controls. The two main functions
are wgtRank() for a fixed test statistic, and wgtRanktt() for an adaptive choice of one of two test
statistics. The function wgtRankCI() inverts the test to obtain confidence intervals and Hodges-
Lehmann point estimates. The function ef2C() is used to extract two evidence factors when a
treated group is compared to two different control groups.

Author(s)

NA

Maintainer: NA

References

Berk, R. H. and Jones, D. H. (1978) <https://www.jstor.org/stable/4615706> Relatively optimal
combinations of test statistics. Scandinavian Journal of Statistics, 5, 158-162.

Quade, D. (1979) <doi:10.2307/2286991> Using weighted rankings in the analysis of complete
blocks with additive block effects. Journal of the American Statistical Association, 74, 680-683.

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

Rosenbaum, P. R. (2011) <doi:10.1111/j.1541-0420.2010.01535.x> A new U-Statistic with superior
design sensitivity in matched observational studies. Biometrics, 67(3), 1017-1027.

Rosenbaum, P. R. (2012) <doi:10.1093/biomet/ass032> Testing one hypothesis twice in observa-
tional studies. Biometrika, 99(4), 763-774.

Rosenbaum, P. R. (2021) <doi:10.1201/9781003039648> Replication and Evidence Factors in Ob-
servational Studies. Chapman and Hall/CRC.

Rosenbaum, P. R. (2023) <doi:10.1111/biom.13921> A second evidence factor for a second control
group. Biometrics, 79(4), 3968-3980.

Rosenbaum, P. R. (2024) <doi:10.1080/01621459.2023.2221402> Bahadur efficiency of observa-
tional block designs. Journal of the American Statistical Association.

Tardif, S. (1987) <doi:10.2307/2289476> Efficiency and optimality results for tests based on weighted rankings. Journal of the American Statistical Association, 82(398), 637-644.

Examples

```
data(aHDL)
y<-t(matrix(aHDL$hdl,4,406))
wgtRank(y,phi="u878",gamma=6) # New U-statistic weights (8,7,8)
wgtRanktt(y,phi1="u868",phi2="u878",gamma=5.9)
```

---

aBP	*Binge Drinking and Blood Pressure*

Description

A matched observational study from NHANES with two control groups, examining the possible effects of on blood pressure of frequent binge drinking of alcohol.

Usage

```
data("aBP")
```

Format

A data frame with 621 observations on the following 13 variables.

SEQN NHANES identification number

age Age in years

female 1=female, 0=male

education Education, with levels: "<9th" = less than 9th grade, "9-11" = grades 9 to 11, "HS" = high school, "SomeCol" = Some College, ">=BA" = BA degree or more

bmi BMI or body-mass index

waisthip Waist-to-hip ratio

vigorR Engages in vigorous recreational exercise, 1=yes, 0=no

smokenow Do you smoke now? Answers: Everyday, Some days, No

bpRX Reports currently taking medication for high blood pressure

bpSystolic Systolic blood pressure, mm Hg. Average of up to three readings.

bpDiastolic Diastolic blood pressure, mm Hg. Average of up to three readings.

group Drinking group, B=currently engages in frequent binge drinking, N=never binged regularly, and drank at most one drink per week in the last year, P=binged on most days for some period in the past but stopped, nevery binged in the last year, and drank at most one drink per week in the last year. See Details.

mset Matched set indicator, 1, 2, ..., 207. There are 207 blocks of size 3, each containing one B, one N and one P.

Details

The data are from data from the 2017-2020 National Health and Nutrition Examination Survey (which was interrupted by COVID-19, so it is not a survey). There were 5624 people who were at least 20 years of age, with an alcohol use survey, blood pressure measurements and covariates used here. Blood pressure measurements are the average of up to three measurements. One question asked about binge drinking in the past, defined as 4 drinks for women or 5 drinks for men. Question ALQ151 asks (essentially): "Was there ever a time or times in your life when you drank 4/5 or more drinks of any kind of alcoholic beverage almost every day?" Another question ALQ142 asked about binge drinking last year: "During the past 12 months, about how often did you have 4/5 or more drinks of any alcoholic beverage?" Question ALQ121 about the overall frequency of alcohol consumption in the past 12 months. Proper use of ALQ121 and ALQ142 accounts for certain screening questions. By definition, group "binge" responded by saying that they engaged in binge drinking on 3 or more days each week in the past 12 months. By definition, group "never" responded to ALQ142 saying they never binged in the past 12 months, responded to ALQ151 saying they had no past time when they binged almost every day, and drank any alcohol on at most one day a week in the past 12 months. By definition, group "past" said yes to question ALQ151, so there was a period in their life of binge drinking almost every day, but they never binged in the past 12 months, and drank alcohol on at most one day a week in the past 12 months. There were 9232 people aged 20 or more. Of these, 9187 had covariate information, aside from BMI and waist/hip ratio. Of these, 7876 had an alcohol survey. Of these, 7281 had at least one measurement of diastolic and systolic blood pressure. Of these, 7076 had body measurements, namely BMI and waist/hip ratio. The three treatment groups — binge, never and past — are mutually exclusive but not exhaustive, and 5624 people fell in one of the groups. All 207 members of the binge group were matched to one control from each control group. Before matching, the never group had 3995 people and the past group had 505 people. The never group was large enough to closely match two or three controls to each member of the binge group, but that was not done in this illustrative example. Up to 3 repeated measures of blood pressure were often present, and the analysis uses their average.

Note

The data are used as an example in Rosenbaum (2023).

Source

US National Health and Nutrition Examination Survey (https://www.cdc.gov/nchs/nhanes/index.htm)

References

Roerecke, M., Kaczorowski, J., Tobe, S. W., Gmel, G., Hasan, O. S. and Rehm, J. (2017). <doi:10.1016/S2468-2667(17)30003-8> The effect of a reduction in alcohol consumption on blood pressure: a systematic review and meta-analysis. Lancet Public Health, 2, e108-e120.

Rosenbaum, P. R. (2023) <doi:10.1111/biom.13921> A second evidence factor for a second control group. Biometrics, 79(4), 3968-3980.

Examples

```
# The following code creates Figure 2 in Rosenbaum (2023)

data(aBP)
```

```
attach(aBP)

yD<-t(matrix(bpDiastolic,3,207))
yS<-t(matrix(bpSystolic,3,207))
vS<-c(yS[,1]-yS[,2],yS[,1]-yS[,3],yS[,2]-yS[,3])
vD<-c(yD[,1]-yD[,2],yD[,1]-yD[,3],yD[,2]-yD[,3])
y<-(yD/median(abs(vD)))+(yS/median(abs(vS)))


par(mfrow=c(1,3))
graphics::boxplot(yD[,1]-yD[,2],yD[,1]-yD[,3],yD[,2]-yD[,3],las=1,
        main="",ylab="Difference mm Hg",
        names=c("B-N","B-P","N-P"),cex.main=.9,
        cex.axis=.8,cex.lab=.9,xlab="Diastolic Difference")
graphics::abline(h=0)
wx<-round(stats::wilcox.test(yD[,1]-yD[,2],conf.int=TRUE)$conf.int,1)
graphics::segments(1,wx[1],1,wx[2],col="black",lwd=2)
wx<-round(stats::wilcox.test(yD[,1]-yD[,3],conf.int=TRUE)$conf.int,1)
graphics::segments(2,wx[1],2,wx[2],col="black",lwd=2)
wx<-round(stats::wilcox.test(yD[,2]-yD[,3],conf.int=TRUE)$conf.int,1)
graphics::segments(3,wx[1],3,wx[2],col="black",lwd=2)


graphics::boxplot(yS[,1]-yS[,2],yS[,1]-yS[,3],yS[,2]-yS[,3],las=1,
        main="",ylab="Difference mm Hg",
        names=c("B-N","B-P","N-P"),cex.main=.9,
        cex.axis=.8,cex.lab=.9,xlab="Systolic Difference")
graphics::abline(h=0)
wx<-round(stats::wilcox.test(yS[,1]-yS[,2],conf.int=TRUE)$conf.int,1)
graphics::segments(1,wx[1],1,wx[2],col="black",lwd=2)
wx<-round(stats::wilcox.test(yS[,1]-yS[,3],conf.int=TRUE)$conf.int,1)
graphics::segments(2,wx[1],2,wx[2],col="black",lwd=2)
wx<-round(stats::wilcox.test(yS[,2]-yS[,3],conf.int=TRUE)$conf.int,1)
graphics::segments(3,wx[1],3,wx[2],col="black",lwd=2)

graphics::boxplot(y[,1]-y[,2],y[,1]-y[,3],y[,2]-y[,3],las=1,
        main="",ylab="(Diastolic/10.7)+(Systolic/14.7)",
        names=c("B-N","B-P","N-P"),cex.main=.9,
        cex.axis=.8,cex.lab=.9,xlab="Combined Difference")
graphics::abline(h=0)
wx<-round(stats::wilcox.test(y[,1]-y[,2],conf.int=TRUE)$conf.int,1)
graphics::segments(1,wx[1],1,wx[2],col="black",lwd=2)
wx<-round(stats::wilcox.test(y[,1]-y[,3],conf.int=TRUE)$conf.int,1)
graphics::segments(2,wx[1],2,wx[2],col="black",lwd=2)
wx<-round(stats::wilcox.test(y[,2]-y[,3],conf.int=TRUE)$conf.int,1)
graphics::segments(3,wx[1],3,wx[2],col="black",lwd=2)
graphics::abline(h=0)
par(mfrow=c(1,1))
detach(aBP)
```

aHDL   *Alcohol and HDL Cholesterol*

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 (2023, 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 (2023) describes the data and matching in detail. The data are used as an example in Rosenbaum (2022). The help file for boxplotTT() applies the tail transformation to this example, reproducing a plot from Rosenbaum (2022).

This data set is also included in the tailTransform package. See also the informedSen package which contains a part of this data set.

Source

US National Health and Nutrition Examination Survey (NHANES), 2013-2014 and 2015-2016.

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

Rosenbaum, P. R. (2023) <doi:10.1111/biom.13558> Sensitivity analyses informed by tests for bias in observational studies. Biometrics 79, 475–487.

Rosenbaum, P. R. (2022). <doi:10.1080/00031305.2022.2063944> A new transformation of treated-control matched-pair differences for graphical display. American Statistician, 76(4), 346-352.

Rosenbaum, P. R. (2024) <doi:10.1080/01621459.2023.2221402> Bahadur efficiency of observational block designs. Journal of the American Statistical Association.

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$mmercury), for use
# in Rosenbaum (2023).  About half of the matched sets
# have values for mmercury.
table(is.na(aHDL$mmercury),aHDL$grp)

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

---

aHDLe   *HDL Cholesterol and Light Daily Alcohol 2013-2020*

Description

A blocked observational study of light daily alcohol consumption and HDL cholesterol from NHANES 2013-2020. This is an enlarged version of the aHDL data in this package, adding data from NHANES 2017-2020.

Usage

```
data("aHDLe")
```

Format

A data frame with 2888 observations on the following 10 variables.

nh NHANES 2013-2014 is 1314, NHANES 2015-2016 is 1516, and NHANES 2017-2020 is 1720

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.

grpL Treated group and control groups. D=Daily=light almost daily alcohol, Never=never drinks – see note, Rarely drinks – see note, B = PastBinge = a past history of binge drinking on most days, but currently drinks once a week or less. For details, see Rosenbaum (2023, Appendix).

hdl HDL cholesterol level mg/dL

mmercury Methylmercury level ug/L

mset Matched set indicator, 1, 2, ..., 722. The 2888 observations are in 722 matched sets, each of size 4.

Details

This data set enlarges the aHDL data in this package to include also data from NHANES 2017-2020. The aHDL data included 406 blocks of 4 people from NHANES 2013-2016, and they are included in aHDLe. From NHANES 2017-2020, an additional 316 blocks of 4 people were added, making 722 blocks of size 4 in total.

The alcohol questions changed slightly in NHANES 2017-2020, forcing small changes in the definitions of two of the control groups. See the Note for specifics.

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 (2023) describes the 2013-2016 data and matching in detail. See also the documentation in this package for the aHDL data.

Note

There is a treated group and three control groups, with one member of each group in each of 722 blocks. Of these, 406 blocks came from NHANES 2013-2016 and 316 blocks came from NHANES 2017-2020, making 722=406+316 blocks in total.

In all three NHANES surveys, 2013-2014, 2015-2016 and 2017-2020, groups D and B had the same definition, as described in the data appendix to Rosenbaum (2023). Specifically:

D=Daily alcohol. Drank alcohol on most days, meaning at least 260 = 5 x 52 days in the past year, drinking between 1 and 3 alcoholic drinks on drinking days.

B=Past binge drinker. An individual in this control group had a period in their life when they drank at least 4 or 5 drinks on most days, but quit, and now drinks on at most one day a week, or at most 52 days in the past year.

NHANES slightly changed its alcohol questions in 2017-2020, so that small changes were required in the definitions of the other two control groups. NHANES 2013-2016 had asked whether a person had "12 drinks in their life," but in 2017-2020 NHANES asked whether a person had "ever had a drink". These were screen questions, so little was asked of a person about alcohol consumption if the person said at the outset that they had never had much alcohol. As it turned out, a yes to the old question, "fewer than 12 drinks in their life," was fairly common, but a yes to the new question, "never had one drink," was much less common. This affects the definitions of groups N=never and R=rare drinking.

As described in greater detail in the data appendix to Rosenbaum (2023), for NHANES 2013-2016, group N=never was comprised of people who said they had fewer than 12 drinks in their life. In the expansion to include data from NHANES 2017-2020, group N=never was comprised of people

who either (i) never had a drink of alcohol, or (ii) had 0 alcoholic drinks in the past year and never in their life had a period of regular binge drinking on most days, defined to be at least 4 or 5 drinks per day.

As described in greater detail in the data appendix to Rosenbaum (2023), for NHANES 2013-2016, group R="rare drinking" was comprised of people who did have 12 or more drinks in their life, but had fewer than 12 drinks in the past year, and never in their life had a period of regular binge drinking on most days, defined to be at least 4 or 5 drinks per day. In the expansion to include data from NHANES 2017-2020, group R="rare drinking" drank between 1 and 11 alcoholic drinks in the past year and never in their life had a period of regular binge drinking on most days, defined to be at least 4 or 5 drinks per day.

In brief, everyone in control groups N and R reports drinking little or no alcohol in the past year. No one in groups N and R reports a period in their lives when they engaged in binge drinking on most days. As nearly as possible given the questions asked, group N appears to be abstaining from alcohol, at least in the past year. As nearly as possible given the questions asked, group R appears to be open to having a rare drink at a wedding or a wake, but drank a negligible amount of alcohol in the past year. Yet, the exact definitions of control groups N and R are slightly different for NHANES 2017-2020. In each block, everyone was given the same structured questions by NHANES, so differences between questions occur between blocks, not within blocks.

Source

US National Health and Nutrition Examination Survey, 2013-2014, 2015-2016 and 2017-2020.

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) <doi:10.1214/ss/1177013232> The role of a second control group in an observational study. Statistical Science, 2, 292-306. Discusses multiple control groups, as in this example.

Rosenbaum, P. R. (1989) <doi:10.2307/2531497> The role of known effects in observational studies. Biometrics, 45, 557-569. Discusses a known effect, such as the effect of alcohol on methylmercury.

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

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

Rosenbaum, P. R. (2023) <doi:10.1111/biom.13558> Sensitivity analyses informed by tests for bias in observational studies. Biometrics 79, 475–487. Uses the known effect of alchol on methylmercury. Data appendix contains detail information about the NHANES data.

Rosenbaum, P. R. (2024) <doi:10.1080/01621459.2023.2221402> Bahadur efficiency of observational block designs. Journal of the American Statistical Association. Discusses properties of weighted rank statistics.

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

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(aHDLe)
nh20172020<-aHDLe$nh==1720
table(nh20172020)
table(nh20172020,aHDLe$grpL)
par(mfrow=c(1,2))
boxplot(aHDLe$hdl[!nh20172020]~aHDLe$grpL[!nh20172020],main="NHANES 2013-2016",
    ylim=c(15,230),ylab="HDL Cholesterol",xlab="Group",las=1)
boxplot(aHDLe$hdl[nh20172020]~aHDLe$grpL[nh20172020],main="NHANES 2017-2020",
    ylim=c(15,230),ylab="HDL Cholesterol",xlab="Group",las=1)
par(mfrow=c(1,1))
```

---

amplify	*Amplification of sensitivity analysis in observational studies.*

Description

Uses the method in Rosenbaum and Silber (2009) to interpret a value of the sensitivity parameter gamma. Each value of gamma amplifies to a curve (lambda,delta) in a two-dimensional sensitivity analysis, the inference being the same for all points on the curve. That is, a one-dimensional sensitivity analysis in terms of gamma has a two-dimensional interpretation in terms of (lambda,delta).

Usage

```
amplify(gamma, lambda)
```

Arguments

gamma	gamma > 1 is the value of the sensitivity parameter, for instance the parameter in senmv. length(gamma)>1 will generate an error.
lambda	lambda is a vector of values > gamma. An error will result unless lambda[i] > gamma > 1 for every i.

Details

A single value of gamma, say gamma = 2.2 in the example, corresponds to a curve of values of (lambda, delta), including (3, 7), (4, 4.33), (5, 3.57), and (7, 3) in the example. An unobserved covariate that is associated with a lambda = 3 fold increase in the odds of treatment and a delta = 7 fold increase in the odds of a positive pair difference is equivalent to gamma = 2.2.

The curve is gamma = (lambda*delta + 1)/(lambda+delta). Amplify is given one gamma and a vector of lambdas and solves for the vector of deltas. The calculation is elementary.

This interpretation of gamma is developed in detail in Rosenbaum and Silber (2009), and it makes use of Wolfe's (1974) family of semiparametric deformations of an arbitrary symmetric distribuiton. See also Rosenbaum (2020, Section 3.6). For an elementary discussion, see Rosenbaum (2017, Table 9.1).

Strictly speaking, the amplification describes matched pairs, not matched sets. The senm function views a k-to-1 matched set with k controls matched to one treated individual as a collection of k correlated treated-minus-control matched pair differences; see Rosenbaum (2007). For matched sets, it is natural to think of the amplification as describing any one of the k matched pair differences in a k-to-1 matched set.

The curve has asymptotes that the function amplify does not compute: gamma corresponds with (lambda,delta) = (gamma, Inf) and (Inf, gamma).

A related though distict idea is developed in Gastwirth et al (1998). The two approaches agree when the outcome is binary, that is, for McNemar's test.

Value

Returns a vector of values of delta of length(lambda) with names lambda.

Note

The amplify function is also in the sensitivitymv package where a different example is used.

Author(s)

Paul R. Rosenbaum

References

Gastwirth, J. L., Krieger, A. M., Rosenbaum, P. R. (1998) <doi:10.1093/biomet/85.4.907> Dual and simultaneous sensitivity analysis for matched pairs. Biometrika, 85, 907-920.

Rosenbaum, P. R. and Silber, J. H. (2009) <doi:10.1198/jasa.2009.tm08470> Amplification of sensitivity analysis in observational studies. Journal of the American Statistical Association, 104, 1398-1405.

Rosenbaum, P. R. (2017) <doi:10.4159/9780674982697> Observation and Experiment: An Introduction to Causal Inference. Cambridge, MA: Harvard University Press. Table 9.1.

Rosenbaum, P. R. (2020) <doi:10.1007/978-3-030-46405-9> Design of Observational Studies (2nd ed.) NY: Springer. Section 3.6.

Wolfe, D. A. (1974) <doi:10.2307/2286025> A charaterization of population weighted symmetry and related results. Journal of the American Statistical Association, 69, 819-822.

Examples

```
#  Consider a treated-control match pair as the unit of measure,
#  analogous to one meter or one foot.   The calculation
#  amplify(4,7) says that, in a matched pair, gamma=4
#  is the same a bias that increases the odds of treatment
#  7-fold and increases the odds of positive matched-pair
#  difference in outcomes 9-fold.
amplify(4,7)
#  It is also true that, in a matched pair, gamma=4
#  is the same a bias that increases the odds of treatment
#  9-fold and increases the odds of positive matched-pair
#  difference in outcomes 7-fold.
amplify(4,9)
#  It is also true that, in a matched pair, gamma=4
#  is the same a bias that increases the odds of treatment
#  5-fold and increases the odds of positive matched-pair
#  difference in outcomes 19-fold.
amplify(4,5)
# The amplify function can produce the entire curve at once:
amplify(4,5:19)
```

---

dwgtRank	*Weighted Rank Statistics for Evidence Factors with Two Control Groups*

Description

In an observational complete block design, dwgtRank computes a sensitivity analysis for a weighted rank statistic designed to perform well when the comparison of a treated group and two control groups is conducted as two nearly independent evidence factors; see Rosenbaum (2023). For this task, a suggested setting of m, m1, m2, scores and range is given in a note in the documentation below. A simpler way to use the suggested settings is to use the function ef2C instead.

Usage

```
dwgtRank(y, gamma = 1, m = 2, m1 = 2, m2 = 2, phifunc = NULL,
    alternative = "greater", scores = NULL, range = TRUE)
```

Arguments

y	With I blocks and J individuals in each block, y is and I x J matrix or dataframe containing the outcomes. The first column of y is compared to columns 2, ..., J. J must be at least 2.
gamma	A real number >=1 giving the value of the sensitivity parameter. gamma=1 yields a randomization test.
m	One of three parameters that define the weights that attach to blocks. The three parameters are integers with 1 <= m1 <= m2 <= m. See Details.

m1	See m.
m2	See m.
phifunc	An optional function that can be used to substitute your own weights for the weights defined by (m, m1, m2). The function must map [0,1] into [0,1]. If phifunc is NULL, then the weight function is defined by (m, m1, m2). If phifunc is not NULL, then it defines the weights and (m, m1, m2) are ignored.
alternative	For an upper-tailed test, use the default, alternative="greater". For an lower-tailed test, use alternative="less". An error will result if alternative is something besides "greater" or "less". In this context, a two-sided test is best viewed as two one-sided tests with a Bonferroni correction, e.g., testing in both tails at level 0.025 to ensure overall level of 0.05; see Cox (1977). For more information, see the notes.
scores	If scores is NULL, the scores are 1, 2, ..., J. Otherwise, scores should specify the J scores for the J within-block ranks. If scores are specified, there must be J scores, but the J scores need not be distinct.
range	If range=TRUE, then the within-block ranges are calculated, ranked from 1 to I, and scored (m, m1, m2) or phifunc. If range=FALSE, then the within-block gap between the largest response and the average of the remaining J-1 responses is used instead.

Details

The method uses a weighted rank statistic to compare the first column of y to the rest; see Rosenbaum (2023, 2024). Weighted rank statistics generalize the methods of Quade (1979) and Tardif (1987). Quade (1979) applied unscored ranks to the I within block ranges, and used unscored ranks within-blocks. In contrast, here, the scores of ranks of ranges or gaps are based on expression (9) in Rosenbaum (2011a); see also Rosenbaum (2014) where weighted M-statistics are used instead of weighted rank statistics. If J=2, the method agrees exactly with the method for pairs in Rosenbaum (2011a).

Using m=1, m1=1, m2=1, is the same as the stratified Wilcoxon rank sum with I strata, ignoring the ranges or gaps; see Lehmann 1975, Chapter 3). Using m=2, m1=2, m2=2 applies unscored ranks to the I ranges or gaps. Using m=5, m1=5, m2=5 is the suggestion of Conover and Salsburg (1988), and m=8, m1=8, m2=8 is a more extreme version of the same theme. In pairs, J=2, m=8, m1=7, m2=8 performs well in Rosenbaum (2011a), as does m=8, m1=6, m2=8. Detailed evaluations in terms of design sensitivity and Bahadur efficiency are in Rosenbaum (2023, 2024).

Value

pval	Upper bound on the one-sided P-value.
detail	A vector with the standarized deviate, the statistic, its null expectation and variance and the value of gamma.

Note

SUGGESTED SETTINGS FOR m, m1, m2, range AND scores WHEN USED WITH TWO CONTROL GROUPS. These suggested settings are more conveniently implemented in the function ef2C. Rosenbaum (2023) considered a matched block design with I blocks of size 3, containing

one treated individual and one control from each of two control groups. The two evidence factors are: (1) compare treated to the first control group, and (2) compare the second control group to the pooled group that does not distinguish the treated individual and the control from the first control group. For the matched pair comparison (1) with one control group, the suggested settings are (m=8, m1=7, m2=8), together with the defaults of range=TRUE and scores=NULL; see Rosenbaum (2011a). For comparison (2), Rosenbaum (2023) evaluated 40 statistics, judging best the statistic with (m=8, m1=8, m2=8), range=FALSE, scores=c(1,2,5), as illustrated below. This statistic had good Bahadur efficiency of a sensitivity analysis against several simple alternative hypotheses involving a treatment effect and no unmeasured bias.

If we expect the treated group to have higher responses than controls, then comparison (1) sets alternative to greater and comparison (2) sets alternative to less. If we expect the treated group to have lower responses than controls, then comparison (1) sets alternative to less and comparison (2) sets alternative to greater. See also the note about alternatives.

Suppose that the data are initially in an Ix3 matrix with outcomes for treated in the first column, control group 1 in the second column, and control group 2 in the third column. The dwgtRank function always compares the first column to the remaining columns. So, the first factor applies the function to y[,1:2] and the second factor applies the function to y[,3:1]. Note carefully here that y[,3:1] has reversed the order of the columns, so column 3 is compared with the other two columns. If the treatment is expected to cause an increase in the response, then comparison (1) applies the function to y[,1:2] with alternative = greater, and comparison (2) applies the function to y[,3:1] with alternative = less. This is illustrated in the example below which reproduces analyses from Rosenbaum (2023). If the treatment is expected to cause a decrease in the response, then repeat these steps with y replaced by -y, so the treatment is expected to increase -y.

Note

ALTERNATIVE. Setting alternative to less is the same as changing the sign of the within-block scored rank, that is, changing phi(a_ij) to -phi(a_ij) in Rosenbaum (2023); see especially equation (1) in the on-line supplement to that paper. Note carefully that setting alternative to less does not change the between block ranks, even when range=FALSE. In general, in dwgtRank, changing the alternative will give a different answer from changing y to -y if range=FALSE because, unlike the range, the gap is not invariant to the sign change. I suggest using function ef2C – the recommended analysis – before trying out variations on that analysis using dwgtRank.

The dwgtRank function was designed for the evidence factor analysis with two different control groups in blocks of size 3, and this is reflected in the way alternative is defined when range=FALSE. For a simple analysis in the suggested form, use ef2C instead of dwgtRank; it calls dwgtRank with appropriate settings. If you wish to explore alternative settings for this problem, use dwgtRank. For several controls from a single control group, use wgtRank instead of dwgtRank.

Note

TIES WITHIN BLOCKS. If there are ties within blocks, then these are resolved as follows. If scores are not specified, so the within block ranks are intended to be 1, 2, ..., J, then average ranks are used for ties. If scores are specified, then ties are resolved by the ties.method="min" in the rank function in base R. This means that tied observations are all given the same rank, hence the same score, and that score corresponds with the smallest rank to which a tied group is entitled. Suppose the scores are scores=c(1,2,5) for J=3. If all three observations are different, then the smallest observation gets score 1, the middle gets 2, and the largest gets 5. If the three values are, say, 16, 14, 14 in a block,

they get ranks 3, 1, 1, with scores 5, 1, 1. If the three values are 16, 16, 14 in a block, they get ranks 2, 2, 1, with scores 2, 2, 1. This is in keeping with the idea that we want to emphasize those blocks in which one observation stands well above the rest. In the example, there are no within-block ties, so the issue does not arise.

TIES BETWEEN BLOCKS. If there are ties among the I blocks in the within-block ranges or gaps, then average ranks are used for ties.

Note

This function compares the first column of y to the other columns. To implement the second evidence factor analysis in Rosenbaum (2023), the second control group must be placed in the first column. See the examples, where y is y[,1:2] for the first evidence factor, but y becomes y[,3:1] for the second evidence factor. All of this is automated in the function ef2C.

Author(s)

Paul R. Rosenbaum

References

Conover, W. J. and Salsburg, D. S. (1988). <doi:10.2307/2531906> Locally most powerful tests for detecting treatment effects when only a subset of patients can be expected to "respond" to treatment. Biometrics, 44, 189-196.

Cox, D. R. (1977). The role of significance tests [with discussion and reply]. Scandinavian Journal of Statistics, 4, 49-70.

Gastwirth, J. L., Krieger, A. M., and Rosenbaum, P. R. (2000). <doi:10.1111/1467-9868.00249> Asymptotic separability in sensitivity analysis. Journal of the Royal Statistical Society B 2000, 62, 545-556.

Lehmann, E. L. (1975). Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day.

Quade, D. (1979). <doi:10.2307/2286991> Using weighted rankings in the analysis of complete blocks with additive block effects. Journal of the American Statistical Association, 74, 680-683.

Rosenbaum, P. R. (1987). <doi:10.2307/2336017> Sensitivity analysis for certain permutation inferences in matched observational studies. Biometrika, 74(1), 13-26.

Rosenbaum, P. R. (2010). <doi:10.1093/biomet/asq019> Evidence factors in observational studies. Biometrika, 97(2), 333-345.

Rosenbaum, P. R. (2011a). <doi:10.1111/j.1541-0420.2010.01535.x> A new U-Statistic with superior design sensitivity in matched observational studies. Biometrics, 67(3), 1017-1027.

Rosenbaum, P. R. (2011b). <doi:10.1198/jasa.2011.tm10422> Some approximate evidence factors in observational studies. Journal of the American Statistical Association, 106(493), 285-295.

Rosenbaum, P. R. (2013). <doi:10.1111/j.1541-0420.2012.01821.x> Impact of multiple matched controls on design sensitivity in observational studies. Biometrics, 2013, 69, 118-127.

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

Rosenbaum, P. R. (2015). <doi:10.1080/01621459.2014.960968> Bahadur efficiency of sensitivity analyses in observational studies. Journal of the American Statistical Association, 110(509), 205-217.

Rosenbaum, P. (2017). <doi:10.4159/9780674982697> Observation and Experiment: An Introduction to Causal Inference. Cambridge, MA: Harvard University Press.

Rosenbaum, P. R. (2018). <doi:10.1214/18-AOAS1153> Sensitivity analysis for stratified comparisons in an observational study of the effect of smoking on homocysteine levels. The Annals of Applied Statistics, 12(4), 2312-2334.

Rosenbaum, P. R. (2021) <doi:10.1201/9781003039648> Replication and Evidence Factors in Observational Studies. Chapman and Hall/CRC.

Rosenbaum, P. R. (2024) <doi:10.1080/01621459.2023.2221402> Bahadur efficiency of observational block designs. Journal of the American Statistical Association.

Rosenbaum, P. R. (2023) <doi:10.1111/biom.13921> A second evidence factor for a second control group. Biometrics 79, 3968-3980.

Tardif, S. (1987). <doi:10.2307/2289476> Efficiency and optimality results for tests based on weighted rankings. Journal of the American Statistical Association, 82(398), 637-644.

Examples

```
# The calculation below reproduce analyses from Rosenbaum (2023).

data(aBP)
attach(aBP)

yD<-t(matrix(bpDiastolic,3,207))
yS<-t(matrix(bpSystolic,3,207))
vS<-c(yS[,1]-yS[,2],yS[,1]-yS[,3],yS[,2]-yS[,3])
vD<-c(yD[,1]-yD[,2],yD[,1]-yD[,3],yD[,2]-yD[,3])
y<-(yD/median(abs(vD)))+(yS/median(abs(vS)))

# The following analysis contrasts the second P control
# with the pooled group consisting of the treated B
# binge drinker and the N control.  That contrast creates
# a second evidence factor, not redundant with the comparison
# of B and N, yet avoids dilution of the effect by using
# a statistic like that of Conover and Salsburg (1988)
# that allows for nonresponders.

# SECOND EVIDENCE FACTOR: CONTROL2 VS TREATED+CONTROL1
dwgtRank(y[,3:1],gamma=1.45,alternative="less",
     scores=c(1,2,5),range=FALSE,m=8,m1=8,m2=8)

amplify(1.45, 2.5)

# This is a much less sensitive result than is obtained
# from the stratified Wilcoxon rank sum statistic wiht
# I strata,

dwgtRank(y[,3:1],gamma=1.45,alternative="less",
```

```
              scores=c(1,2,3),m=1,m1=1,m2=1)

# and theory leads us to expect this difference
# in performance of the two statistics; see Rosenbaum (2023)

# EVIDENCE FACTOR ANALYSIS, COMBINING TWO FACTORS
# The evidence factor analysis compares treated to the first control group,
# then compares the second control group to the pooled group consisting of
# treated and first control, then combines the two analyses using meta-analysis.
# Treated/first-control matched pairs are compared using the method in
# Rosenbaum (2011).

p1<-dwgtRank(y[,1:2],gamma=2.3,alternative="greater",
         m=8,m1=7,m2=8)$pval
p2<-dwgtRank(y[,3:1],gamma=1.45,alternative="less",
             scores=c(1,2,5),range=FALSE,m=8,m1=8,m2=8)$pval
c(p1,p2)
sensitivitymv::truncatedP(c(p1,p2))
amplify(2.3,4)
amplify(1.45,2.5)

# THE COMBINED ANALYSIS IS INSENSITIVE TO LARGER BIASES
# The combined analysis is insensitive to larger biases
# than are its components
p1<-dwgtRank(y[,1:2],gamma=2.6,alternative="greater",
         m=8,m1=7,m2=8)$pval
p2<-dwgtRank(y[,3:1],gamma=1.7,alternative="less",
             scores=c(1,2,5),range=FALSE,m=8,m1=8,m2=8)$pval
c(p1,p2)
sensitivitymv::truncatedP(c(p1,p2))

amplify(2.6,5)
amplify(1.7,3)

# CONNECTION WITH OTHER PACKAGES
# Although dwgtRank() computes the matched pair P-value bound,
dwgtRank(y[,1:2],gamma=2.3,alternative="greater",
        m=8,m1=7,m2=8)$pval
# a simpler way to do it uses senU() in the DOS2 package
DOS2::senU(y[,1]-y[,2],m=8,m1=7,m2=8,gamma=2.3)
# where senU also provides bounds on point estimates and confidence
# intervals for each gamma
DOS2::senU(y[,1]-y[,2],m=8,m1=7,m2=8,gamma=1.5,conf.int=TRUE)

detach(aBP)
rm(p1,p2)
rm(y)

# USING SIMULATION TO GET THE GENERAL IDEAS OF DESIGN SENSITIVITY
# AND THE BAHADUR EFFICIENCY OF A SENSITIVITY ANALYSIS

# IN THIS LARGE SAMPLE SIZE, THE DESIGN SENSITIVITY PREDICTS
# U888 WILL HAVE MORE POWER THAN U555, AND IT DOES.
```

```
# SEE TABLE 2 OF ROSENBAUM (2023), NORMAL tau=1/2
# FOR U888/125/GAP AND U555/125/GAP
set.seed(1)
ss<-10000
ysim<-matrix(rnorm(3*ss),ss,3)
ysim[,1]<-ysim[,1]+sqrt(2)/2 # This is tau=1/2 for Normal errors
# Compare U888/125/gap and U555/125/gap
dwgtRank(ysim[,3:1],gamma=3,alternative="less",scores=c(1,2,5),
  range=FALSE,m=8,m1=8,m2=8)$pval
dwgtRank(ysim[,3:1],gamma=3,alternative="less",scores=c(1,2,5),
  range=FALSE,m=5,m1=5,m2=5)$pval
# IF YOU INCREASED ss FROM 10000, AS ABOVE, TO INFINITY, THE
# POWER FUNCTION WOULD TEND TO A STEP FUNCTION WITH A SINGLE
# STEP DOWN FROM POWER 1 TO POWER 0 AT THE DESIGN SENSITIVITY.

# IN THIS SMALLER SAMPLE SIZE, THE BAHADUR EFFICIENCY PREDICTS
# U555 WILL HAVE MORE POWER THAN U888, AND IT DOES.
# SEE TABLE 3 OF ROSENBAUM (2023), NORMAL tau=1/2
# FOR U888/125/GAP AND U555/125/GAP AT UPSILON = 1.5
set.seed(1)
ss<-100
ysim<-matrix(rnorm(3*ss),ss,3)
ysim[,1]<-ysim[,1]+sqrt(2)/2 # This is tau=1/2 for Normal errors
# Compare U888/125/gap and U555/125/gap
dwgtRank(ysim[,3:1],gamma=1.5,alternative="less",scores=c(1,2,5),
  range=FALSE,m=8,m1=8,m2=8)$pval
dwgtRank(ysim[,3:1],gamma=1.5,alternative="less",scores=c(1,2,5),
  range=FALSE,m=5,m1=5,m2=5)$pval
```

---

ef2C   *Evidence Factors For Matched Triples With Two Control Groups*

Description

In an observational complete block design, with bocks of size three, each containing a treated individual and one control from each of two control groups, ef2C (with the default settings) performs the evidence factor analysis suggested in Rosenbaum (2023). One factor compares the treated group to the first control group in a matched pairs analysis. The other factor pools the treated group and the first control group and compares it to the second control group.

Usage

```
ef2C(y, gamma=1, upsilon=1, alternative="greater", trunc=0.2,
    m=c(8,8), m1=c(7,8), m2=c(8,8), scores=c(1,2,5), range=FALSE)
```

Arguments

y	With I blocks and 3 individuals in each block, y is and I x 3 matrix or dataframe containing the outcomes. The first column is the response of the treated individual. The second response is the response of the control from the first control group. The third response is the response of the control from the second control group.
gamma	A real number >=1 giving the value of the sensitivity parameter for the comparison of the treated group and the first control group. gamma=1 yields a randomization test.
upsilon	A real number >=1 giving the value of the sensitivity parameter for the comparison of the second control group and the combination of the treated group plus the first control group. upsilon=1 yields a randomization test.
alternative	Use alternative=greater if the treatment is expect to cause an increase in the response in y. Use alternative=less if the treatment is expect to cause a decrease in the response in y. In this context, a two-sided test is best viewed as two one-sided tests with a Bonferroni correction, e.g., testing in both tails at level 0.025 to ensure overall level of 0.05; see Cox (1977). For more information, see the notes.
trunc	The two P-values from the two factors are combined using the trucated product of P-values due to Zaykin et al. (2002): it is the P-value derived from the product of those P-values that are less than trunc. For more information, see the notes.
m	A vector of two integers greater than or equal to 1. The first coordinate of m is for the first evidence factor, comparing the treated individual to the control from the first control group. The second coordinate of m is for the second evidence factor, comparing the control from the second control group to the set containing the treated individual and the first control. One of three parameters that define the weights that attach to blocks. The three parameters are integers with $1 <= m1[1] <= m2[1] <= m[1]$ and $1 <= m1[2] <= m2[2] <= m[2]$. The default settings are suggested in Rosenbaum (2023). See Details.
m1	A vector of two integers greater than or equal to 1. See m.
m2	A vector of two integers greater than or equal to 1. See m.
scores	A vector of three nonnegative integer scores used to rank the three responses in each block in the test concerning the second evidence factor. The default settings are suggested in Rosenbaum (2023). See Details.
range	A TRUE or a FALSE for the second evidence factor. If TRUE, blocks are ranked by their within-block ranges. If FALSE, blocks are ranked by the so-called gap, which is the difference between the maximum order statistic in a block and the average of the remaining order statistics. The default setting of FALSE is suggested in Rosenbaum (2023). See Details.

Details

The default settings produce the recommended evidence-factor analysis in Rosenbaum (2023). The example below reproduces some results from the example in that paper. That paper considered 40 test statistics in terms of the Bahadur efficiency of a sensitivity – all of these analyses can be

reproduced by the more flexible but more complicated dwgtRank function. The ef2C function calls dwgtRank twice and combines the resulting two analyses.

The comparison of the treated group and the first control group is equivalent to dwgtRank(y[,1:2],gamma=gamma,m=8,m1=7, and these settings are motivated by results in Rosenbaum (2011, 2015). Notice that y[,1:2] uses the first two columns of y.

The comparison of the second control group and the merger of the treated group with the first control group is equivalent to dwgtRank(y[,3:1], gamma=upsilon, m=8,m1=8,m2=8, range=FALSE, alternative="less", scores=c(1,2,5)), and these settings are motivated by results in Rosenbaum (2023). Notice that y[,3:1] compares the third column to the pooled group consisting of columns 1 and 2.

Value

pvals   Upper bounds on the one-sided P-values for the two factors and their combination.

detail   A matrix with some details of the computations that produced the P-values.

Note

The two P-values from the two factors are combined using the trucated product of P-values due to Zaykin et al. (2002): it is the P-value derived from the product of those P-values that are less than trunc. Taking trunc=1 yields Fisher's method for combining independent P-values. Fisher's method is not ideal when combining P-value bounds produced by sensitivity analyses; see Hsu et al. (2013). Reasonable values are trunc=.1, truc=.15 and trunc=.2. As illustrated in the example below, lower truncation values produce smaller combined P-values when the P-values are below the truncation point, but a P-value that barely exceeds the truncation point is effectively discarded. Hsu et al. (2013) compare truncation values when used in a sensitivity analysis. For discussion of combining sensitivity analyses as independent, see the required conditions in Rosenbaum (2011b, 2021). These conditions hold for the comparison performed by ef2C.

Note

The setting alternative = "less" simply replaces y by -y before testing in the upper tail.

Note

For a deeper understanding, see the documentation of dwgtRank.

Author(s)

Paul R. Rosenbaum

References

Cox, D. R. (1977). The role of significance tests [with discussion and reply]. Scandinavian Journal of Statistics, 4, 49-70.

Hsu, J. Y., Small, D. S., & Rosenbaum, P. R. (2013). <doi:10.1080/01621459.2012.742018> Effect modification and design sensitivity in observational studies. Journal of the American Statistical Association, 108(501), 135-148.

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

Rosenbaum, P. R. (2011a). <doi:10.1111/j.1541-0420.2010.01535.x> A new U-Statistic with superior design sensitivity in matched observational studies. Biometrics, 67(3), 1017-1027.

Rosenbaum, P. R. (2011a). <doi:10.1111/j.1541-0420.2010.01535.x> A new U-Statistic with superior design sensitivity in matched observational studies. Biometrics, 67(3), 1017-1027.

Rosenbaum, P. R. (2011b). <doi:10.1198/jasa.2011.tm10422> Some approximate evidence factors in observational studies. Journal of the American Statistical Association, 106(493), 285-295.

Rosenbaum, P. R. (2015). <doi:10.1080/01621459.2014.960968> Bahadur efficiency of sensitivity analyses in observational studies. Journal of the American Statistical Association, 110(509), 205-217.

Rosenbaum, P. R. (2021) <doi:10.1201/9781003039648> Replication and Evidence Factors in Observational Studies. Chapman and Hall/CRC.

Rosenbaum, P. R. (2023) <doi:10.1111/biom.13921> A second evidence factor for a second control group. Biometrics 79, 3968-3980.

Zaykin, D. V., Zhivotovsky, L. A., Westfall, P. H. and Weir, B. S. (2002) <doi:10.1002/gepi.0042> Truncated product method of combining P-values. Genetic Epidemiology, 22, 170-185.

Examples

```
# The calculation below reproduce analyses from Rosenbaum (2023).

data(aBP)
attach(aBP)

yD<-t(matrix(bpDiastolic,3,207))
yS<-t(matrix(bpSystolic,3,207))
vS<-c(yS[,1]-yS[,2],yS[,1]-yS[,3],yS[,2]-yS[,3])
vD<-c(yD[,1]-yD[,2],yD[,1]-yD[,3],yD[,2]-yD[,3])
y<-(yD/median(abs(vD)))+(yS/median(abs(vS)))

# EVIDENCE FACTOR ANALYSIS, COMBINING TWO FACTORS
# The evidence factor analysis compares treated to the first control group,
# then compares the second control group to the pooled group consisting of
# treated and first control, then combines the two analyses using meta-analysis.
# Treated/first-control matched pairs are compared using the method in
# Rosenbaum (2011).

ef2C(y,gamma=2.3,upsilon=1.45)
amplify(2.3,4)
amplify(1.45,2.5)

# THE COMBINED ANALYSIS IS INSENSITIVE TO LARGER BIASES
# The combined analysis is insensitive to larger biases
# than are its components
ef2C(y,gamma=2.6,upsilon=1.7)
amplify(2.6, 5)
amplify(1.7,c(2.7,3))
```

```
# The calculations above are also produced in the
# example for dwgtRank, where alternative
# analyses from Rosenbaum (2023) are compared.

####################################################

# Comparing trucation points to understand trunc:
ef2C(y,gamma=2.6,upsilon=1.7,trunc=.2) # Default
ef2C(y,gamma=2.6,upsilon=1.7,trunc=1) # Fisher's method
ef2C(y,gamma=2.6,upsilon=1.7,trunc=.1)
ef2C(y,gamma=2.5,upsilon=1.6,trunc=.2)
ef2C(y,gamma=2.5,upsilon=1.6,trunc=.1)
# See Hsu et al. (2013) for discussion of the
# truncation point for a sensitivity analysis.
```

---

gwgtRank   *Generalized Sensitivity Analysis for Weighted Rank Statistics in Block*
                              Designs

Description

Uses a weighted rank statistic to perform a sensitivity analysis for an I x J observational block
design in which each of I blocks contain J individuals, some of whom are treated individuals and
others are controls.

Usage

```
gwgtRank(y, z, phi = "u868", phifunc = NULL, gamma = 1, detail=FALSE)
```

Arguments

y
: A matrix or data frame of responses with I rows and J columns. An error will result if y contains NAs.

z
: A matrix or data frame of treatment indicators with I rows and J columns. In z, the ith row and jth column is a 1 if this individual is treated, or a 0 if this individual is a control. If a row of z contains J ones or J zeros, that block will be removed from further computations as there is no one to compare, and a warning will appear. An error will result if z contains NAs.

phi
: The weight function to be applied to the ranks of the within block ranges. The options are: (i) "wilc" for the stratified Wilcoxon test, which gives every block the same weight, (ii) "quade" which ranks the within block ranges from 1 to I, and is closely related to Quade's (1979) statistic; see also Tardif (1987), (iii) "u868" based on Rosenbaum (2011), (iv) u878 based on Rosenbaum (2011). Note that phi is ignored if phifunc is not NULL.

phifunc
: If not NULL, a user specified weight function for the ranks of the within block ranges. The function should map [0,1] into [0,1]. The function is applied to the ranks divided by the sample size. See the example.

gamma	A single number greater than or equal to 1. gamma is the sensitivity parameter. Two individuals with the same observed covariates may differ in their odds of treatment by at most a factor of gamma; see Rosenbaum (1987; 2017, Chapter 9).
detail	If detail=FALSE, then brief output is given, similar to that provided by wgtRank(). In both cases, the asymptotically separable approximation of Gastwirth et al. (2000) is used. If detail=TRUE, then the separable approximation is compared with a bound in Rosenbaum (2018), and the output from the senstrat package is given. For moderate J and large I, the bound is likely to agree exactly with the separable approximation.

Details

The function gwgtRank() differs from wgtRank() in that wgtRank requires one treated individual and J-1 controls, whereas gwgtRank permits 1 to J-1 treated indivuals with the rest as controls.

This function uses the senstrat() function in the senstrat package to perform the calculations.

To test in the lower tail – to test against the alternative that treated responses are lower than control responses, apply the function to -y. For a two-sided test, do both one-sided tests and apply the Bonferroni inequality, doubling the smaller of the two one-sided P-value bounds; see Cox (1977, Section 4.2).

Value

pval	Upper bound on the one-sided P-value when testing the null hypothesis of no treatment effect against the alternative hypothesis that treated responses are higher than control responses.
detail	Details of the computation of pval: the standardized deviate, the test statistic, its null expectation, its null variance and the value of gamma.

Author(s)

Paul R. Rosenbaum

References

Brown, B. M. (1981). <doi:10.1093/biomet/68.1.235> Symmetric quantile averages and related estimators. Biometrika, 68(1), 235-242.

Cox, D. R. (1977). The role of significance tests [with discussion and reply]. Scandinavian Journal of Statistics, 4, 49-70.

Gastwirth, J. L., Krieger, A. M., and Rosenbaum, P. R. (2000). <doi:10.1111/1467-9868.00249> Asymptotic separability in sensitivity analysis. Journal of the Royal Statistical Society B 2000, 62, 545-556.

Lehmann, E. L. (1975). Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day.

Quade, D. (1979). <doi:10.2307/2286991> Using weighted rankings in the analysis of complete blocks with additive block effects. Journal of the American Statistical Association, 74, 680-683.

Rosenbaum, P. R. (1987). <doi:10.2307/2336017> Sensitivity analysis for certain permutation inferences in matched observational studies. Biometrika, 74(1), 13-26.

Rosenbaum, P. R. (2011). <doi:10.1111/j.1541-0420.2010.01535.x> A new U-Statistic with superior design sensitivity in matched observational studies. Biometrics, 67(3), 1017-1027.

Rosenbaum, P. R. (2013). <doi:10.1111/j.1541-0420.2012.01821.x> Impact of multiple matched controls on design sensitivity in observational studies. Biometrics, 2013, 69, 118-127.

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

Rosenbaum, P. R. (2015). <doi:10.1080/01621459.2014.960968> Bahadur efficiency of sensitivity analyses in observational studies. Journal of the American Statistical Association, 110(509), 205-217.

Rosenbaum, P. (2017). <doi:10.4159/9780674982697> Observation and Experiment: An Introduction to Causal Inference. Cambridge, MA: Harvard University Press.

Rosenbaum, P. R. (2018). <doi:10.1214/18-AOAS1153> Sensitivity analysis for stratified comparisons in an observational study of the effect of smoking on homocysteine levels. The Annals of Applied Statistics, 12(4), 2312-2334.

Rosenbaum, P. R. (2024) <doi:10.1080/01621459.2023.2221402> Bahadur efficiency of observational block designs. Journal of the American Statistical Association.

Tardif, S. (1987). <doi:10.2307/2289476> Efficiency and optimality results for tests based on weighted rankings. Journal of the American Statistical Association, 82(398), 637-644.

See Also

An alternative approach avoids rank tests and uses weighted M-statistics instead, as in the sensitivitymw package and Rosenbaum (2014). However, Bahadur efficiency calculations are available for weighted rank statistics; see Rosenbaum (2024).

Examples

```
data(aHDL)
y<-t(matrix(aHDL$hdl,4,406))
z<-matrix(0,dim(y)[1],dim(y)[2])
z[,1]<-1
gwgtRank(y,z,phi="quade",gamma=3.5) # Quade's test
gwgtRank(y,z,phi="quade",gamma=3.5,detail=TRUE) # Alternative output

# A user defined weight function, brown, analogous to Brown (1981).
brown<-function(v){((v>=.333)+(v>=.667))/2}
gwgtRank(y,z,phifunc=brown,gamma=4.7)
```

---

stepSolve   *Root of a Monotone Decreasing Step Function*

Description

Of limited interest to most users, stepSolve finds a root of a monotone decreasing step function by bisection search. It is called by other functions in this package, including wgtRankCI. If the function f passes but never equals zero, then the argument where f passes zero is returned. If the function equals zero for an interval, the average of the endpoints of the interval are returned. The function is used to obtain confidence intervals and point estimates by inverting a rank test, in the spirit of the Hodges-Lehmman estimate.

Usage

```
stepSolve(f, int, eps = 1e-05)
```

Arguments

f	A nonincreasing function of a single argument.
int	A vector of length 2 with int[1]<int[2]. A search for a root of f will be done on the interval int.
eps	A small number. In searching for a root, a difference of eps will be regarded as negligible. A larger eps will produce an answer more quickly, while a smaller eps will produce a more accurate answer.

Value

average	The average of low and high, defined below.
low	See high, defined below.
high	If the monotone decreasing step function f equals zero for an interval, then that interval is approximately [low, high]. If f crosses zero but never equals zero, then the crossing point is between low and high, which should be close. Reducing eps will search longer to improve the accuracy of [low, high].

Author(s)

Paul R. Rosenbaum

References

Hodges Jr, J. L. and Lehmann, E. L. (1963) <doi:10.1214/aoms/1177704172> Estimates of location Bbased on rank tests. The Annals of Mathematical Statistics, 34(2), 598-611.

Lehmann, E. L. (1963). Nonparametric confidence intervals for a shift parameter. The Annals of Mathematical Statistics, 34(4), 1507-1512.

Rosenbaum, P. R. (1993) <doi:10.1080/01621459.1993.10476405> Hodges-Lehmann point estimates of treatment effect in observational studies. Journal of the American Statistical Association, 88 (424), 1250-1253.

Examples

```
#  Uses stepSolve to find the one-sample Hodges-Lehmann estimate
## Not run:
set.seed(1)
y<-rnorm(4)+7
wilcox.test(y,conf.int=TRUE,conf.level=.5)$estimate
f<-function(hl){wilcox.test(y-hl)$statistic-(length(y)*(1+length(y)))/4}
stepSolve(f,c(-100,100))

## End(Not run)
```

---

wgtRank   *Sensitivity Analysis for Weighted Rank Statistics in Block Designs*

Description

Uses a weighted rank statistic to perform a sensitivity analysis for an I x J observational block design in which each of I blocks contains one treated individual and J-1 controls.

Usage

```
wgtRank(y, phi = "u868", phifunc = NULL, gamma = 1)
```

Arguments

y
: A matrix or data frame with I rows and J columns. Column 1 contains the response of the treated individuals and columns 2 throught J contain the responses of controls in the same block. An error will result if y contains NAs.

phi
: The weight function to be applied to the ranks of the within block ranges. The options are: (i) "wilc" for the stratified Wilcoxon test, which gives every block the same weight, (ii) "quade" which ranks the within block ranges from 1 to I, and is closely related to Quade's (1979) statistic; see also Tardif (1987), (iii) "u858", "u868", "u878", or "u888" based on Rosenbaum (2011). Note that phi is ignored if phifunc is not NULL.

phifunc
: If not NULL, a user specified weight function for the ranks of the within block ranges. The function should map [0,1] into [0,1]. The function is applied to the ranks divided by the sample size. See the example.

gamma
: A single number greater than or equal to 1. gamma is the sensitivity parameter. Two individuals with the same observed covariates may differ in their odds of treatment by at most a factor of gamma; see Rosenbaum (1987; 2017, Chapter 9).

Details

This method is developed and evaluated in Rosenbaum (2024).

To test in the lower tail – to test against the alternative that treated responses are lower than control responses, apply the function to -y. For a two-sided test, do both one-sided tests and apply the Bonferroni inequality, doubling the smaller of the two one-sided P-value bounds; see Cox (1977, Section 4.2).

Value

pval   Upper bound on the one-sided P-value when testing the null hypothesis of no treatment effect against the alternative hypothesis that treated responses are higher than control responses.

detail   Details of the computation of pval: the standardized deviate, the test statistic, its null expectation, its null variance and the value of gamma.

Note

The computations use the separable approximation discussed in Gastwirth et al. (2000) and Rosenbaum (2018). Compare with the method in Rosenbaum (2014) and the R package sensitivitymw.

Author(s)

Paul R. Rosenbaum

References

Brown, B. M. (1981). <doi:10.1093/biomet/68.1.235> Symmetric quantile averages and related estimators. Biometrika, 68(1), 235-242.

Cox, D. R. (1977). The role of significance tests [with discussion and reply]. Scandinavian Journal of Statistics, 4, 49-70.

Gastwirth, J. L., Krieger, A. M., and Rosenbaum, P. R. (2000). <doi:10.1111/1467-9868.00249> Asymptotic separability in sensitivity analysis. Journal of the Royal Statistical Society B 2000, 62, 545-556.

Lehmann, E. L. (1975). Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day.

Quade, D. (1979). <doi:10.2307/2286991> Using weighted rankings in the analysis of complete blocks with additive block effects. Journal of the American Statistical Association, 74, 680-683.

Rosenbaum, P. R. (1987). <doi:10.2307/2336017> Sensitivity analysis for certain permutation inferences in matched observational studies. Biometrika, 74(1), 13-26.

Rosenbaum, P. R. (2011). <doi:10.1111/j.1541-0420.2010.01535.x> A new U-Statistic with superior design sensitivity in matched observational studies. Biometrics, 67(3), 1017-1027.

Rosenbaum, P. R. (2013). <doi:10.1111/j.1541-0420.2012.01821.x> Impact of multiple matched controls on design sensitivity in observational studies. Biometrics, 2013, 69, 118-127.

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

Rosenbaum, P. R. (2015). <doi:10.1080/01621459.2014.960968> Bahadur efficiency of sensitivity analyses in observational studies. Journal of the American Statistical Association, 110(509), 205-217.

Rosenbaum, P. (2017). <doi:10.4159/9780674982697> Observation and Experiment: An Introduction to Causal Inference. Cambridge, MA: Harvard University Press.

Rosenbaum, P. R. (2018). <doi:10.1214/18-AOAS1153> Sensitivity analysis for stratified comparisons in an observational study of the effect of smoking on homocysteine levels. The Annals of Applied Statistics, 12(4), 2312-2334.

Rosenbaum, P. R. (2024) <doi:10.1080/01621459.2023.2221402> Bahadur efficiency of observational block designs. Journal of the American Statistical Association.

Tardif, S. (1987). <doi:10.2307/2289476> Efficiency and optimality results for tests based on weighted rankings. Journal of the American Statistical Association, 82(398), 637-644.

See Also

An alternative approach avoids rank tests and uses weighted M-statistics instead, as in the sensitivitymw package and Rosenbaum (2014). However, Bahadur efficiency calculations are available for weighted rank statistics; see Rosenbaum (2024).

Examples

```
data(aHDL)
y<-t(matrix(aHDL$hdl,4,406))
wgtRank(y,phi="wilc",gamma=3.5) # Stratified Wilcoxon rank sum test
wgtRank(y,phi="quade",gamma=3.5) # Quade's test
wgtRank(y,phi="quade",gamma=4.5) # Quade's test, larger gamma
wgtRank(y,phi="quade",gamma=4.6) # Quade's test, larger gamma
wgtRank(y,phi="u868",gamma=5.4) # New U-statistic weights (8,6,8)
wgtRank(y,phi="u878",gamma=6) # New U-statistic weights (8,7,8)

# As an aid to interpreting gamma, see the amplify function.
amplify(3.5,8)
amplify(4.6,8)
amplify(5.4,8)
amplify(6,8)

# A user defined weight function, brown, analogous to Brown (1981).
brown<-function(v){((v>=.333)+(v>=.667))/2}
wgtRank(y,phifunc=brown,gamma=4.7)
```

---

wgtRankC   *Sensitivity Analysis for a Conditional Weighted Rank Test*

Description

Uses a conditional weighted rank statistic to perform a sensitivity analysis for an I x J observational block design in which each of I blocks contains one treated individual and J-1 controls. The size J of a block must be at least 3.

Usage

```
wgtRankC(y, phi = "u878", phifunc = NULL, gamma = 1, alternative = "greater")
```

Arguments

y	A matrix or data frame with I rows and J columns, where J is at least 3. Column 1 contains the response of the treated individuals and columns 2 throught J contain the responses of controls in the same block. An error will result if y contains NAs. An error will result if y has fewer than 3 columns.
phi	The weight function to be applied to the ranks of the within block ranges. The options are: (i) "wilc" for the stratified Wilcoxon test, which gives every block the same weight, (ii) "quade" which ranks the within block ranges from 1 to I, and is closely related to Quade's (1979) statistic; see also Tardif (1987), (iii) "u868", "u878", or "u888" based on Rosenbaum (2011). Note that phi is ignored if phifunc is not NULL.
phifunc	If not NULL, a user specified weight function for the ranks of the within block ranges. The function should map [0,1] into [0,1]. The function is applied to the ranks divided by the sample size. See the example.
gamma	A single number greater than or equal to 1. gamma is the sensitivity parameter. Two individuals with the same observed covariates may differ in their odds of treatment by at most a factor of gamma; see Rosenbaum (1987; 2017, Chapter 9).
alternative	The null hypothesis asserts that the treatment has no effect. If alternative is "greater", then the null hypothesis is tested against the alternative that the treatment increases the response of treated individuals. If alternative is "less", then the null hypothesis is tested against the alternative that the treatment decreases the response of treated individuals.

Details

The conditional test restricts attention to blocks in which the treated individual has either the highest or lowest response in a block. This tactic may be shown to increase design sensitivity; see Rosenbaum (2024).

Value

pval	The upper bound on the one-sided P-value
detail	Details of the computation of the P-value
block.information	
	The conditional test uses a block only if the treated individual has either the highest or lowest respones in the block. This vector indicates: (i) the number of blocks that were used, (ii) in those blocks, the number in which the treated individual had the highest response, and (iii) in those block, the number of blocks in which there was a tie for the maximum or minimum response.

Note

Do a two-sided test by testing in both tails and rejecting at level alpha if the smaller P-value is less than alpha/2; see Cox (1977, section 4.2). Aside from labeling the output, setting alternative to "less" is the same as applying the test to -y with alternative set to "greater".

Author(s)

Paul R. Rosenbaum

References

Brown, B. M. (1981) <doi:10.1093/biomet/68.1.235> Symmetric quantile averages and related estimators. Biometrika, 68, 235-242.

Cox, D. R. (1977). The role of significance tests (with discussion and reply). Scandinavian Journal of Statistics, 4, 49-70.

Noether, G. E. (1963) <doi:10.2307/2283321> Efficiency of the Wilcoxon two-sample statistic for randomized blocks. Journal of the American Statistical Association, 58, 894-898.

Noether, G. E. (1973) <doi:10.2307/2284805> Some simple distribution-free confidence intervals for the center of a symmetric distribution. Journal of the American Statistical Association, 68, 716-719.

Quade, D. (1979) <doi:10.2307/2286991> Using weighted rankings in the analysis of complete blocks with additive block effects. Journal of the American Statistical Association, 74, 680-683.

Rosenbaum, P. R. (1987). <doi:10.2307/2336017> Sensitivity analysis for certain permutation inferences in matched observational studies. Biometrika, 74(1), 13-26.

Rosenbaum, P. R. (2004) <doi:10.1093/biomet/91.1.153> Design sensitivity in observational studies. Biometrika, 91, 153-164.

Rosenbaum, P. R. (2011). <doi:10.1111/j.1541-0420.2010.01535.x> A new U-Statistic with superior design sensitivity in matched observational studies. Biometrics, 67(3), 1017-1027.

Rosenbaum, P. (2017). <doi:10.4159/9780674982697> Observation and Experiment: An Introduction to Causal Inference. Cambridge, MA: Harvard University Press.

Rosenbaum, P. R. (2024) <doi:10.1080/01621459.2023.2221402> Bahadur efficiency of observational block designs. Journal of the American Statistical Association.

Rosenbaum, P. R. (2024) A conditioning tactic that increases design sensitivity in observational block designs. Manuscript.

Tardif, S. (1987) <doi:10.2307/2289476> Efficiency and optimality results for tests based on weighted rankings. Journal of the American Statistical Association, 82, 637-644.

Examples

```
data(aHDLe)
y<-t(matrix(aHDLe$hdl,4,722))
colnames(y)<-c("D","R","N","B")
y<-y[,c(1,3,2,4)]
boxplot(y,ylab="HDL Cholesterol",las=1,xlab="Group")
wgtRankC(y,gamma=6.65,phi="u878")
wgtRankC(y,gamma=6.24,phi="u868")
```

```
wgtRankC(y,phi="quade",gamma=5.28)
wgtRankC(y,phi="wilc",gamma=4.9)
#
#  Examples with a user-defined phi-function
#  Rank scores of Brown (1981) and Noether (1973)
brown<-function(v){(v>=(1/3))+(v>=(2/3))}
wgtRankC(y,phifunc=brown,gamma=5.5)
noether<-function(v){v>=(2/3)}
wgtRankC(y,phifunc=noether,gamma=6.5)
```

---

wgtRankCI	*Sensitivity Analysis for Confidence Intervals and Point Estimates from Weighted Rank Statistics in Block Designs*

Description

Uses a weighted rank statistic to perform a sensitivity analysis for an I x J observational block design in which each of I blocks contains one treated individual and J-1 controls. Inverts the test to obtain an estimate and a confidence interval for an additive treatment effect, tau; see Rosenbaum (1993, 2002, 2007).

Usage

```
wgtRankCI(y, phi = "u868", phifunc = NULL, gamma = 1,
     alternative="greater", alpha=0.05, eps = 0.00001)
```

Arguments

y
: A matrix or data frame with I rows and J columns. Column 1 contains the response of the treated individuals and columns 2 throught J contain the responses of controls in the same block. An error will result if y contains NAs.

phi
: The weight function to be applied to the ranks of the within block ranges. The options are: (i) "wilc" for the stratified Wilcoxon test, which gives every block the same weight, (ii) "quade" which ranks the within block ranges from 1 to I, and is closely related to Quade's (1979) statistic; see also Tardif (1987), (iii) "u868" based on Rosenbaum (2011), (iv) u878 based on Rosenbaum (2011). Note that phi is ignored if phifunc is not NULL.

phifunc
: If not NULL, a user specified weight function for the ranks of the within block ranges. The function should map [0,1] into [0,1]. The function is applied to the ranks divided by the sample size.

gamma
: A single number greater than or equal to 1. gamma is the sensitivity parameter. Two individuals with the same observed covariates may differ in their odds of treatment by at most a factor of gamma; see Rosenbaum (2002, Chapter 4; 2017, Chapter 9).

alternative
: Must equal "greater", "less", or "twosided". This determines whether the confidence interval is one-sided or two-sided.

alpha	Coverage of the confidence interval is 1-alpha.
eps	A small number. In searching for a root, a difference of eps will be regarded as negligible. A larger eps will produce an answer more quickly, while a smaller eps will produce a more accurate answer.

Details

This test is developed and evaluated in Rosenbaum (2024), and it is inverted for point estimates and confidence intervals in the usual way. Understand the test first.

The two-sided interval is the intersection of two one-sided intervals, each with coverage 1-alpha/2; see Cox (1977, Section 4.2).

Value

estimate	The interval of point estimates, reducing to a single number when gamma=1.
confidence	The one-sided or two-sided confidence interval.

Note

The computations use the separable approximation discussed in Gastwirth et al. (2000) and Rosenbaum (2018). Compare with the method in Rosenbaum (2014) and the R package sensitivitymw.

Author(s)

Paul R. Rosenbaum

References

Brown, B. M. (1981). <doi:10.1093/biomet/68.1.235> Symmetric quantile averages and related estimators. Biometrika, 68(1), 235-242.

Cox, D. R. (1977). The role of significance tests [with discussion and reply]. Scandinavian Journal of Statistics, 4, 49-70.

Gastwirth, J. L., Krieger, A. M., and Rosenbaum, P. R. (2000). <doi:10.1111/1467-9868.00249> Asymptotic separability in sensitivity analysis. Journal of the Royal Statistical Society B 2000, 62, 545-556.

Hodges, J. L. and Lehmann, E. L. (1963). <doi:10.1214/aoms/1177704172> Estimates of Location Based on Rank Tests. The Annals of Mathematical Statistics, 34(2), 598-611.

Lehmann, E. L. (1975). Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day.

Quade, D. (1979). <doi:10.2307/2286991> Using weighted rankings in the analysis of complete blocks with additive block effects. Journal of the American Statistical Association, 74, 680-683.

Rosenbaum, P. R. (1993). <doi:10.2307/2291264> Hodges-Lehmann point estimates of treatment effect in observational studies. Journal of the American Statistical Association, 88(424), 1250-1253.

Rosenbaum, P. R. (2002). <doi:10.1007/978-1-4757-3692-2> Observational Studies (2nd Edition). New York: Springer.

Rosenbaum, P. R. (2007). <doi:10.1111/j.1541-0420.2006.00717.x> Sensitivity analysis for M-estimates, tests, and confidence intervals in matched observational studies. Biometrics, 63(2), 456-464.

Rosenbaum, P. R. (2011). <doi:10.1111/j.1541-0420.2010.01535.x> A new U-Statistic with superior design sensitivity in matched observational studies. Biometrics, 67(3), 1017-1027.

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

Rosenbaum, P. R. (2015). <doi:10.1080/01621459.2014.960968> Bahadur efficiency of sensitivity analyses in observational studies. Journal of the American Statistical Association, 110(509), 205-217.

Rosenbaum, P. R. (2017). <doi:10.4159/9780674982697> Observation and Experiment: An Introduction to Causal Inference. Cambridge, MA: Harvard University Press.

Rosenbaum, P. R. (2018). <doi:10.1214/18-AOAS1153> Sensitivity analysis for stratified comparisons in an observational study of the effect of smoking on homocysteine levels. The Annals of Applied Statistics, 12(4), 2312-2334.

Rosenbaum, P. R. (2024) <doi:10.1080/01621459.2023.2221402> Bahadur efficiency of observational block designs. Journal of the American Statistical Association.

Tardif, S. (1987). <doi:10.2307/2289476> Efficiency and optimality results for tests based on weighted rankings. Journal of the American Statistical Association, 82(398), 637-644.

See Also

An alternative approach avoids rank tests and uses weighted M-statistics instead, as in the sensitivitymw package and Rosenbaum (2014). However, Bahadur efficiency calculations are available for weighted rank statistics; see Rosenbaum (2024).

Examples

```
## Not run:
data(aHDL)
y<-t(matrix(aHDL$hdl,4,406))

# Without sensitivity analysis
wgtRankCI(y,phi="wilc",gamma=1) # Stratified Wilcoxon rank sum test

# With sensitivity analysis

# The CI excludes 0 if the test rejects 0.
wgtRank(y,phi="wilc",gamma=3.7) # Stratified Wilcoxon rank sum test
wgtRankCI(y,phi="wilc",gamma=3.7)

wgtRank(y,phi="u878",gamma=3.7) # U-statistic with weights (8,7,8)
wgtRankCI(y,phi="u878",gamma=3.7)

# A user defined weight function, brown, analogous to Brown (1981).
brown<-function(v){((v>=.333)+(v>=.667))/2}
wgtRankCI(y,phifunc=brown,gamma=3.7)
```

```
#
# COMPARISONS WITH STANDARD METHODS
#
y2<-c(-17,  -3,  -1,   2,   8,   9,  10,  11,  12,  16,  20)
y2a<-cbind(y2,-y2)/2
#
# Compare with the usual Hodges-Lehmann estimate,
# namely the median of the Walsh averages (yi+yj)/2
#
hl<-function(y){
  w<-NULL
  I<-length(y)
  for (i in 1:I) for (j in i:I) w<-c(w,(y[i]+y[j])/2)
  median(w)
}
hl(y2)
wgtRankCI(y2a,phi="quade")$estimate
#
# Compare with wilcox.test confidence interval in the stats package
#
stats::wilcox.test(y2,conf.int=TRUE,correct=FALSE,exact=FALSE,
        alternative = "greater")$conf.int
wgtRankCI(y2a,phi="quade",alternative="greater")$confidence
stats::wilcox.test(y2,conf.int=TRUE,correct=FALSE,exact=FALSE,
        alternative = "less")$conf.int
wgtRankCI(y2a,phi="quade",alternative="less")$confidence
stats::wilcox.test(y2,conf.int=TRUE,correct=FALSE,exact=FALSE,
        alternative = "two.sided")$conf.int
wgtRankCI(y2a,phi="quade",alternative="twosided")$confidence
#
# Compare with senWilcox in the DOS2 package
#
# Note: senWilcox reports only one-sided estimates with 1-sided tests
#
DOS2::senWilcox(y2,conf.int=TRUE,alternative="greater",gamma=1.25)
wgtRankCI(y2a,phi="quade",gamma=1.25,alternative = "greater")
DOS2::senWilcox(y2,conf.int=TRUE,alternative="less",gamma=1.25)
wgtRankCI(y2a,phi="quade",gamma=1.25,alternative = "less")
DOS2::senWilcox(y2,conf.int=TRUE,alternative="twosided",gamma=1.25)
wgtRankCI(y2a,phi="quade",gamma=1.25,alternative = "twosided")
#
# Compare with senU in the DOS2 package
#
# Note: senU reports only one-sided estimates with 1-sided tests
#
DOS2::senU(y2,gamma=1.25,m=8,m1=6,m2=8,alternative = "greater",
          exact=FALSE,conf.int = TRUE)
wgtRankCI(y2a,phi="u868",gamma=1.25,alternative = "greater")
DOS2::senU(y2,gamma=1.25,m=8,m1=6,m2=8,alternative = "less",
          exact=FALSE,conf.int = TRUE)
wgtRankCI(y2a,phi="u868",gamma=1.25,alternative = "less")
```

```
DOS2::senU(y2,gamma=1.25,m=8,m1=6,m2=8,alternative = "twosided",
        exact=FALSE,conf.int = TRUE)
wgtRankCI(y2a,phi="u868",gamma=1.25,alternative = "twosided")

## End(Not run)
```

---

wgtRanktt   *Adaptive Inference Using Two Test Statistics in a Block Design*

Description

Tests twice, using the better of two test statistics; see Rosenbaum (2012, 2024).

Usage

```
wgtRanktt(y, phi1 = "u868", phi2 = "u878", phifunc1 = NULL, phifunc2 = NULL, gamma = 1)
```

Arguments

y	A matrix or data frame with I rows and J columns. Column 1 contains the response of the treated individuals and columns 2 through J contain the responses of controls in the same block. An error will result if y contains NAs.
phi1	The weight function to be applied to the ranks of the within block ranges. The options are: (i) "wilc" for the stratified Wilcoxon test, which gives every block the same weight, (ii) "quade" which ranks the within block ranges from 1 to I, and is closely related to Quade's (1979) statistic; see also Tardif (1987), (iii) "u868" based on Rosenbaum (2011), (iv) u878 based on Rosenbaum (2011). Note that phi is ignored if phifunc is not NULL.
phi2	See phi1.
phifunc1	If not NULL, a user specified weight function for the ranks of the within block ranges. The function should map [0,1] into [0,1]. The function is applied to the ranks divided by the sample size. See the example.
phifunc2	See phifunc1.
gamma	A single number greater than or equal to 1. gamma is the sensitivity parameter. Two individuals with the same observed covariates may differ in their odds of treatment by at most a factor of gamma; see Rosenbaum (1987; 2017, Chapter 9).

Value

jointP	Upper bound on the one-sided joint P-value obtained from two test statistics in the presence of a bias of at most gamma.
cor12	Correlation of the two test statistics at the treatment assignment distribution that provides the joint upper bound. Often, this correlation is high, so the joint distribution that is used here is much less conservative than use of the Bonferroni inequality when testing twice.
detail	Details about the two statistics separately. Equivalent to the result from wgtRank() run twice with different test statistics.

Note

For discussion of testing twice in matched pairs, see Rosenbaum (2012).

Testing twice is also possible in block designs using weighted rank statistics because the same value of the unobserved covariate provides the upper bound for both statistics when using the separable approximation in Gastwirth et al. (2000) and Rosenbaum (2018, Remarks 4 and 5). See also Rosenbaum (2022) where the Bahadur efficiency of such tests is computed.

Other packages that use testing twice in a different way are "sensitivity2x2xk"" and "testtwice". The "testtwice" package is restricted to matched pairs, and "sensitivity2x2xk"" is for binary outcomes. With some attention to detail (e.g., the handling of zero pair differences), in the case of matched pairs, the "testtwice" package and the wgtRanktt() function will yield identical results. In that sense, wgtRanktt() extends the method to blocks designs.

Testing twice achieves the larger Bahadur efficiency of the two component statistics; see Berk and Jones (1978).

Author(s)

Paul R. Rosenbaum

References

Berk, R. H. and Jones, D. H. (1978) <https://www.jstor.org/stable/4615706> Relatively optimal combinations of test statistics. Scandinavian Journal of Statistics, 5, 158-162.

Brown, B. M. (1981) <doi:10.1093/biomet/68.1.235> Symmetric quantile averages and related estimators. Biometrika, 68(1), 235-242.

Gastwirth, J. L., Krieger, A. M., and Rosenbaum, P. R. (2000) <doi:10.1111/1467-9868.00249> Asymptotic separability in sensitivity analysis. Journal of the Royal Statistical Society B 2000, 62, 545-556.

Lehmann, E. L. (1975) Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day.

Quade, D. (1979) <doi:10.2307/2286991> Using weighted rankings in the analysis of complete blocks with additive block effects. Journal of the American Statistical Association, 74, 680-683.

Rosenbaum, P. R. (1987) <doi:10.2307/2336017> Sensitivity analysis for certain permutation inferences in matched observational studies. Biometrika, 74(1), 13-26.

Rosenbaum, P. R. (2011) <doi:10.1111/j.1541-0420.2010.01535.x> A new U-Statistic with superior design sensitivity in matched observational studies. Biometrics, 67(3), 1017-1027.

Rosenbaum, P. R. (2012) <doi:10.1093/biomet/ass032> Testing one hypothesis twice in observational studies. Biometrika, 99(4), 763-774.

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

Rosenbaum, P. R. (2015) <doi:10.1080/01621459.2014.960968> Bahadur efficiency of sensitivity analyses in observational studies. Journal of the American Statistical Association, 110(509), 205-217.

Rosenbaum, P. (2017) <doi:10.4159/9780674982697> Observation and Experiment: An Introduction to Causal Inference. Cambridge, MA: Harvard University Press.

Rosenbaum, P. R. (2018) <doi:10.1214/18-AOAS1153> Sensitivity analysis for stratified comparisons in an observational study of the effect of smoking on homocysteine levels. The Annals of Applied Statistics, 12(4), 2312-2334.

Rosenbaum, P. R. (2024) <doi:10.1080/01621459.2023.2221402> Bahadur efficiency of observational block designs. Journal of the American Statistical Association.

Tardif, S. (1987) <doi:10.2307/2289476> Efficiency and optimality results for tests based on weighted rankings. Journal of the American Statistical Association, 82(398), 637-644.

Examples

```
data(aHDL)
y<-t(matrix(aHDL$hdl,4,406))

# This is the simplest example of a general property.  The
# example simply illustrates, but does not fully exploit
# the property.  In this case, use of the stratified
# Wilcoxon statistic is a mistake, because Quade's
# statistic correctly reports insensitivity to a bias
# of gamma=4.5, but the stratified Wilcoxon statistic
# is sensitive at gamma=3.5.  The adaptive procedure
# that does both tests and corrects for multiple testing
# is insensitive to gamma=4.4; so, it is almost as good
# as knowing what you cannot know, namely that Quade's
# statistic is the better choice in this one example.
# The price paid for testing twice is very small;
# see Berk and Jones (1978) and Rosenbaum (2012, 2022).
wgtRank(y,phi="wilc",gamma=3.5)
wgtRank(y,phi="quade",gamma=3.5)
wgtRank(y,phi="wilc",gamma=4.5)
wgtRank(y,phi="quade",gamma=4.5)
wgtRanktt(y,phi1="wilc",phi2="quade",gamma=4.4)

# Sensitivity to gamma=3.5 is very different from
# sensitivity to gamma=4.4; see documentation for amplify.
amplify(3.5,8)
amplify(4.4,8)


# In this example, u878 exhibits greater insensitivity to bias
# than u868.  However, adaptive inference using both is almost
# as good as the better statistic, yet it strongly controls the
# family-wise error rate despite testing twice;
# see Rosenbaum (2012,2022).
wgtRank(y,phi="u868",gamma=6) # New U-statistic weights (8,6,8)
wgtRank(y,phi="u878",gamma=6) # New U-statistic weights (8,7,8)
wgtRanktt(y,phi1="u868",phi2="u878",gamma=5.9)

# A user defined weight function, brown, analogous to Brown (1981).
brown<-function(v){((v>=.333)+(v>=.667))/2}
# In this example, the joint test rejects based on u878
wgtRanktt(y,phi1="u878",phifunc2=brown,gamma=5.8)
```

Index