Package 'alcoholSurv'

May 27, 2025

Type Package

Title Light Daily Alcohol and Longevity

Version 0.7.0

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

Description Contains data from an observational study concerning possible effects of light daily alcohol consumption on survival and on HDL cholesterol. It also replicates various simple analyses in Rosenbaum (2025a) <doi:10.1080/09332480.2025.2473291>. Finally, it includes new R code in wgtRankCef() that implements and replicates a new method for constructing evidence factors in observational block designs.

License GPL-3

Encoding UTF-8

LazyData true

Suggests survival, iTOS, coin

Depends R (>= 3.5.0), stats, sensitivitymv

NeedsCompilation no

Author Paul Rosenbaum [aut, cre]

Repository CRAN

Date/Publication 2025-05-27 09:10:16 UTC

Contents

---

alcoholSurv-package   *Light Daily Alcohol and Longevity*

Description

Contains data from an observational study concerning possible effects of light daily alcohol consumption on survival and on HDL cholesterol. It also replicates various simple analyses in Rosenbaum (2025a) <doi:10.1080/09332480.2025.2473291>. Finally, it includes new R code in wgtRankCef() that implements and replicates a new method for constructing evidence factors in observational block designs.

Details

This package contains a dataframe, alcoholSurv, used to study the effects of light daily alcohol on survival. It also contains a function, wgtRankDef() that conducts evidence factor analyses using decisive pairs.

Author(s)

Paul Rosenbaum [aut, cre]

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

References

Rosenbaum, P. R. (2025a) <doi:10.1080/09332480.2025.2473291> Does a Daily Glass of Wine Lengthen Life? Insight from a Second Control Group. Chance, 38(1), 25-30.

Rosenbaum, P. R. (2025b) A new construction of evidence factors in an observational study of light daily alcohol consumption and longevity. Manuscript.

Examples

```
data(alcoholSurv)
# Three treatment groups
table(alcoholSurv$gDrinks)
# In 1130 matched blocks of size 5
table(table(alcoholSurv$mset))
```

---

alcoholSurv   *Light Alcohol Consumption and Survival*

Description

Data from 6 NHANES surveys with follow-up for mortality, as a matched comparison of: (i) light daily alcohol, (ii) rare alcohol, and (iii) no alcohol.

Usage

```
data("alcoholSurv")
```

Format

A data frame with 5650 observations on the following 21 variables.

SEQN NHANES ID Number

nh Identifies the NHANES years. The data are from six NHANES, 2005-2006 to 2015-2016.

female 1=female, 0=male

age Age in years at the time of the NHANES survey.

education Level of education in five categories. 1 is less than 9th grade, 3 is high school, 5 is at least a BA degree.

hdl HDL cholesterol in mg/dL

bmi BMI or body-mass index

GH Glycohemoglobin as a percent

smoke Smoking status at the NHANES interview. An ordered factor with levels Everyday < Somedays < NotNow < Never

z Treatment indicator, 1 if consumes light daily alcohol, 0 if control.

gDrinks Daily means: consumes light daily alcohol, between 1 and 3 drinks on at least 260=5x52 days each year. Rare means rarely consumes alcohol. None means consumed no alcohol in the past year. An ordered factor with levels Daily < Rare < None

aDays Days consumed alcohol in the past year.

aDrinks Typical number of alcoholic drinks on drinking days.

a12life 1=consumed at least 12 alcoholic drinks in life, 0=other. Based on NHANES question ALQ110.

aEverBinge Was there ever a time in your life when you drank 5 or more drinks almost every day? 1=yes, 0=no. The wording of this question changed slightly from one NHANES to another, sometimes asking about 4 drinks for a woman rather than 5. See the NHANES documentation for details.

time Time to death or censoring in months from the date of the NHANES examination. Public data file using the National Death Index.

mortstat Death/censoring indicator, 1=dead, 0=censored.

cod Cause of death on the death certificate. See codf.

codf Cause of death as a factor. An ordered factor with levels Alive < Heart < Cancer < ChronicLung < Accident < Cerebrovascular < Alzheimer < Diabetes < FluPneumonia < Kidney < Other

mset Matched set indicator, 1 to 1130.

treated The SEQN for the treated individual in a matched set. Same information as mset, but in a different format.

Details

This is a matched data set, one treated, 2 rare controls plus 2 none controls in each of 1130 blocks of size 5. See the description of the gDrinks variable above. For details, see Rosenbaum (2025). The examples below replicate analyses from Rosenbaum (2025).

The mortality data is from the public use linked morality files. NHANES also has a restricted use version of the mortality files; it is not used here. The public use file masks identity in various ways; see its web-page referenced below.

Source

Data are from the NHANES webpage www.cdc.gov/nchs/nhanes/index.htm.

Also, 2019 Public-Use Linked Mortality Files are from www.cdc.gov/nchs/data-linkage/mortality-public.htm

References

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

Public-Use Linked Mortality Files. www.cdc.gov/nchs/data-linkage/mortality-public.htm

Rosenbaum, P. R. (2025) <doi:10.1080/09332480.2025.2473291> Does a Daily Glass of Wine Lengthen Life? Insight from a Second Control Group. Chance, 38 (1), 25-30.

Examples

```
#
# The example replicates results from Rosenbaum (2025)
#
oldpar <- par(no.readonly = TRUE)
data(alcoholSurv)
# Three treatment groups
table(alcoholSurv$gDrinks)
# In 1130 matched blocks of size 5
table(table(alcoholSurv$mset))
attach(alcoholSurv)
# Alcohol groups
table(gDrinks,aDays>0)
table(gDrinks,z)
table(gDrinks,aDrinks)
table(gDrinks,a12life)
table(gDrinks,aDays>24)
table(gDrinks,aDays>0)
# Alcohol groups are matched for covaiates
tapply(age,gDrinks,mean)
tapply(female,gDrinks,mean)
tapply(aEverBinge,gDrinks,mean)
tapply(education,gDrinks,mean)
prop.table(table(smoke,gDrinks),2)

library(survival)
```

```
par(bg="moccasin")

#  Make Figure 1
par(mfrow=c(1,3))
boxplot(age~gDrinks,las=1,cex.lab=1,cex.axis=1,xlab="Age",
        ylab="Age in Years")
axis(3,at=1:3,labels=round(tapply(age,gDrinks,mean)),cex.axis=1)
boxplot(education~gDrinks,las=1,cex.lab=1,cex.axis=1,xlab="Education",
        ylab="Education")
axis(3,at=1:3,labels=round(tapply(education,gDrinks,mean),2),cex.axis=1)

boxplot((aDays*aDrinks)~gDrinks,las=1,cex.lab=1,cex.axis=1,
        xlab="Alcoholic Drinks", ylab="Drinks Per Year")
axis(3,at=1:3,labels=round(tapply((aDays*aDrinks),gDrinks,mean)),cex.axis=1)

# Make Table 1

Female<-tapply(female,gDrinks,mean)*100
Age<-tapply(age,gDrinks,mean)
Education<-tapply(education,gDrinks,mean)
EverBinged<-tapply(aEverBinge,gDrinks,mean)*100
NeverSmoked<-tapply(smoke=="Never",gDrinks,mean)*100
NoLongerSmoke<-tapply(smoke=="NotNow",gDrinks,mean)*100
SmokeSomeDays<-tapply(smoke=="Somedays",gDrinks,mean)*100
SmokeEveryDay<-tapply(smoke=="Everyday",gDrinks,mean)*100
tabBal<-rbind(Female,Age,Education,EverBinged,NeverSmoked,NoLongerSmoke,SmokeSomeDays,
              SmokeEveryDay)
rm(Female,Age,Education,EverBinged,NeverSmoked,NoLongerSmoke,SmokeSomeDays,
   SmokeEveryDay)
tabBal2<-rbind(tabBal,prop.table(table(nh,gDrinks),2)*100)



# Make Figure 2

par(mfrow=c(1,2))

xlim<-c(0,150)  # Restrict plots to first 150 months, after which data are thin

coln<-c("blue","red","black")

st<-Surv(time,mortstat)
plot(survfit(st~(gDrinks=="Daily")),col=c("darkgreen","blue"),lty=c(4,1),lwd=2,ylim=c(.5,1),las=1,
     ylab="Probability of Survival",xlab="Months",cex.axis=.9,cex.lab=.9,
     main="(i)  All, I=1130, J=5", cex.main=.8,xlim=xlim)
legend(0.5,.63,c("Daily","Control"),col=c("blue","darkgreen"),lty=c(1,4),lwd=rep(2,2),cex=.8)

plot(survfit(st~gDrinks),col=coln,lty=1:3,lwd=2,ylim=c(.5,1),las=1,
     ylab="Probability of Survival",xlab="Months",cex.axis=.9,cex.lab=.9,
     main="(ii)  All, I=1130, J=5", cex.main=.8,xlim=xlim)
legend(0.5,.66,levels(gDrinks),col=coln,lty=1:3,lwd=rep(2,3),cex=.8)

# Make Figure 3
```

```
who<-smoke=="Never"
plot(survfit(st[who]~gDrinks[who]),col=coln,lty=1:3,lwd=2,ylim=c(.5,1),las=1,
     ylab="Probability of Survival",xlab="Months",cex.axis=.9,cex.lab=.9,xlim=xlim,
     main=paste("Never Smoker, I=",sum(z[who]),", J=5",sep=""),cex.main=.8)
legend(0.5,.66,levels(gDrinks),col=coln,lty=1:3,lwd=rep(2,3),cex=.8)

who<-(aEverBinge==0)
plot(survfit(st[who]~gDrinks[who]),col=coln,lty=1:3,lwd=2,ylim=c(.5,1),las=1,
     ylab="Probability of Survival",xlab="Months",cex.axis=.9,xlim=xlim,cex.lab=.9,
     main=paste("Never a Binge Drinker, I=",sum(z[who]),", J=5",sep=""),cex.main=.8)
legend(0.5,.66,levels(gDrinks),col=c(4,2,1),lty=1:3,lwd=rep(2,3),cex=.8)

rm(who)

#  Do formal analyses in footnote 1 using the Cox's stratified  proportional
#  hazards model

coxph(st~z+strata(treated))
confint(coxph(st~z+strata(treated)))
exp(confint(coxph(st~z+strata(treated))))

noDrinks<-1*(gDrinks=="None")
coxph(st~z+noDrinks+strata(treated))
confint(coxph(st~z+noDrinks+strata(treated)))
exp(confint(coxph(st~z+noDrinks+strata(treated))))
rm(coln,xlim,noDrinks)
detach(alcoholSurv)

par(oldpar)
```

---

wgtRankCef   *Conditional Evidence Factors in Observational Block Designs*

Description

In an observational block design, there are I blocks of size J, where each block has individuals from K groups, where $2 <= K <= J$. If K=2, then there is one type of control, and the analysis in Rosenbaum (2025a) is performed; however, this analysis is also available in the weightedRank package using the wgtRankC() function. If K>2, then the evidence factor analysis in Rosenbaum (2025b) is performed.

Usage

```
wgtRankCef(y, colGroups = NULL, phi = "u878",
    phifunc = NULL, gammas = 1, alternative = "greater",
    trunc = 1, seed = NULL, random = FALSE)
```

Arguments

y	In a block design with I blocks of size J, y is an IxJ matrix or dataframe of numeric outcomes or scores. Treated responses are in column 1 of y, whereas control responses are in columns 2 through J.
colGroups	If colGroups is not NULL, then it is a vector of group labels, where the length of colGroups equals the block size, J, which in turn equals the number of columns of y. In colGroups, the label for a control group may be repeated; so, in the example, colGroups = c("D","R","R","N","N") signifies that the treated group is D in column 1 of y, columns 2 and 3 contain controls of type R, and columns 4 and 5 contain controls of type N, and K=3 groups are represented in blocks of size J=5. If colGroups is NULL but y has column names, then the function uses the column names of y as colGroups. If colGroups is NULL and y lacks column names, then colNames is set to 1, 2, ..., J. The program will stop with an error if the treated group, the first group in colGroups, appears more than once in colGroups.
phi	If phi = "none" and phifunc is NULL, then the values in y are used in permutation inference, without ranking or scoring. Other acceptable values of phi are "u868", "u878", "u888", "quade", "u858", "wilc", and "noether". See Details. If phifunc in not NULL, then phi is ignored. phi can be a vector of length K-1, say phi=c("u878","quade") for K=3, signifying that "u878" should be used to compare treated individuals to the first of two control groups, but "quade" should be used to compare treated individuals to the second of two control groups.
phifunc	phifunc is a user supplied function that replaces phi. See Details.
gammas	The sensitivity parameter. gammas is either a number greater than or equal to 1, or a vector of K-1 such numbers. If gammas is of length K-1, then gamma[j] is used when comparing the treated group to control group j. If gammas is a single number, then that number is used for all K-1 comparisons.
alternative	The direction of the alternative hypothesis, either "greater" or "less". alternative = "less" is equivalent to replacing y by -y and using the default of alternative = "greater".
trunc	The P-values for the K-1 comparisons are combined using the truncated product of P-values developed by Zaykin et al. (2002), where trunc is the truncation point. If trunc=1, then this is Fisher's method for combining P-values.
seed	If seed is not NULL and random=TRUE, then within-block ties in y for the maximum or minimum are broken at random after setting the random number generator seed to seed (using set.seed(seed). If random=TRUE, then it is wise to set the seed, so that the analysis can be reproduced.)
random	If random is TRUE, then within-block ties in y for the maximum or minimum are broken at random after setting the random number generator seed to seed (using set.seed(seed). If random is FALSE, then a block (i.e., row of y) that lacks a unique maximum and a unique minimu is not used. See Details.

Details

A good all-around choice for phi is u868.

In the example, the survival analyses first computes log-rank scores, then gives these scores to the wgtRankCef as y with phi="none". This means that wgtRankCef is accepting rank scores computed by another program.

wgtRankCef implements the method in Rosenbaum (2025b) and replicates its analysis. Rosenbaum (2025b) creates K-1 evidence factors when there are K-1 control groups, and it extends the method in Rosenbaum (2025a) where there is only one control group and hence no evidence factors.

Let v be the range of responses in the row i of y, that is, in block i. Suppose that a block or row of y has a unique maximum and a unique minimum; then, of course, v>=0 equals that maximum-minus-minimum response. If the treated response in that block is either that maximum or minimum response, then the block is said to contain a decisive pair. In a decisive pair, the decisive pair difference is the difference is the treated response minus the response of the one control with the maximum or minimum response, so the decisve pair difference is v or -v. Each decisive pair belongs to one control group; so, the decisive pairs create K-1 essentially independent tests (i.e., their P-values under the null hypothesis of no effect are jointly stochastically larger than the uniform distribution on the K-1 dimensional unit cube, $[0,1]^{(K-1)}$.) This creates K-1 evidence factors, one for each control group.

Each evidence factor is essentially a matched pair permutation test using the decisive pairs for the relevant control group. Often, it is a general signed rank test for those decisive pairs, and in this case, the ranges, v, are ranked from 1 to I, the ranks are divided by I, and then scored using the nonnegative bounded function phi defined on [0,1]. The functions "u868", "u878", "u888", "quade", "u858", "wilc", and "noether" are particular functions used in these signed rank tests; see Rosenbaum (2011b) and Rosenbaum (2015a) for properties of these functions, and Rosenbaum (2025a, section 5.1) for discussion of their use with decisive pairs.

In particular, phi(v)=1 is "wilc" and phi(v)=v is "quade"; also, phi(v)=0 for v<(2/3), phi(v)=1 for v>=(2/3) is "noether". phifunc is a user-defined bounded, nonnegative function mapping a vector of v of length I into a vector of nonnegative real numbers also of length I.

Providing there is unique maximum and a unique minimum in a block i, then ties among other individuals in block i have no effect on the test. In a matched pair, J=2, a within-block tie means that the pair contains no information – permuting the treatment assignments in such a pair does not change the value of a test statistic – so, the common practice with pairs, J=2, effectively ignores the pairs; see, for instance, Pratt (1959). This is the default option, with random=FALSE; blocks with a tie for the maximum or minimum are not used. In fact, however, with J>2, a block with a tie for the maximum or minimum may contain some information, albeit less than if the tie were not present. The option random=TRUE allocates a tie for the maximum or minimum at random to one of the tied individuals, and then continues as before.

For general discussion of evidence factors, see: Rosenbaum (2010, 2011, 2021, 2023).

Value

tProd.pval	The upper bound on the combined one-sided P-value from all K-1 evidence factors using the truncated product of P-values or Fisher's method for trunc=1.
pvals	The K-1 bounds on the individual P-values, comparing treated individuals to each of the K-1 control groups.
blocks.used	The number of blocks containing a decisive pair for each of the K-1 comparions.
treated.max	The number of blocks containing a decisive pair in which the treated individual had the largest, not the smallest, response in the block.

Note

For examples using R in sensitivity analysis with evidence factors, see Rosenbaum (2015b) above. For an interactive introduction, see <https://rosenbap.shinyapps.io/learnsenShiny/>

Author(s)

Paul R. Rosenbaum

References

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.

Pratt, J. W. (1959). Remarks on zeros and ties in the Wilcoxon signed rank procedures. Journal of the American Statistical Association, 54(287), 655-667.

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

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

Rosenbaum, P. R. (2011b) <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. (2015a) <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. (2015b) <doi:10.1353/obs.2015.0000> Two R packages for sensitivity analysis in observational studies. Observational Studies, 1(2), 1-17. Available free on-line.

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 observational block designs. Journal of the American Statistical Association.

Rosenbaum, P. R. (2025a) <doi:10.1093/jrsssb/qkaf007> A conditioning tactic that increases design sensitivity in observational block designs. Journal of the Royal Statistical Society B.

Rosenbaum, P. R. (2025b) A new construction of evidence factors in an observational study of light daily alcohol consumption and longevity. 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.

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.

See Also

weightedRank, **senstrat**, **sensitivitymv**, **evident**, **sensitivitymw**

Examples

```
#
# The example reproduces analyses from Rosenbaum (2015b).
#
data("alcoholSurv")

library(survival)
sv<-survival::Surv(alcoholSurv$time,alcoholSurv$mortstat)
svlogrank<-coin::logrank_trafo(sv, ties.method = "average-scores")
names(svlogrank)<-alcoholSurv$treated

selectBlock<-function(grps,mortstat){
  # Return a logical vector picking out blocks with at least one
  # death in among alcoholSurv$dGroups in grps
  colnums<-NULL
  if (is.element("Daily",grps)) colnums<-c(colnums,1)
  if (is.element("Rare",grps)) colnums<-c(colnums,c(2,3))
  if (is.element("None",grps)) colnums<-c(colnums,c(4,5))
  mort<-t(matrix(mortstat,5,1130))
  who<-apply(mort[,colnums],1,sum)>0 # at least one death
  who<-as.vector(rbind(who,who,who,who,who))
  who&is.element(alcoholSurv$gDrinks,grps)
}

who<-selectBlock(c("Daily","Rare","None"),alcoholSurv$mortstat)
#
#  As suggested by O'Brien and Fleming (1987) Biometrics 43, 169-180,
#  a block without a death is omitted from the permutation test.
#  In a block without a death, the log-rank scores vary only
#  because of unequal censoring times.
#
lr<-svlogrank[who]
names(lr)<-alcoholSurv$SEQN[who]
mlr<-t(matrix(lr,5,length(lr)/5))
mwho<-t(matrix(as.numeric(names(lr)),5,length(lr)/5))
rownames(mlr)<-mwho[,1]



hdl<-t(matrix(alcoholSurv$hdl,5,1130))



#
#  Evidence factor analysis for HDL-C
#
wgtRankCef(hdl,colGroups = c("D","R","R","N","N"),gammas=5.5,
  trunc=1,phi="u878")
wgtRankCef(hdl,colGroups = c("D","R","R","N","N"),gammas=8.25,
  trunc=1,phi="u878")
#
# The example below repeats the one immediately above, but
# breaks ties at random.  The P-value is just slightly smaller
```

```
# but now depends upon random numbers.
#
wgtRankCef(hdl,colGroups = c("D","R","R","N","N"),gammas=8.25,
  phi="u878",random=TRUE,seed=12345)
#
#  Unwisely, the following example permutes the hdl data, not
#  its rank scores, and the results are sensitive to smaller biases.
#
wgtRankCef(hdl,colGroups = c("D","R","R","N","N"),gammas=8.25,
    trunc=1,phi="none")
#
#  Evidence factor analysis for survival using logrank scores
#  that were computed above.  The phi="none" option lets you
#  compute your own scores.
#
wgtRankCef(mlr,colGroups = c("D","R","R","N","N"),gammas=5/3,phi="none",
  trunc=1)
```

Index