Package ‘evident’

July 6, 2020

Type Package

Title Evidence Factors in Observational Studies

Version 1.0.2

Author Paul R. Rosenbaum

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

Description Contains a collection of examples of evidence factors in observational studies; e.g., Rosenbaum (2017) <doi:10.1214/17-STS621>. The examples are collected to aid readers of a book in preparation, ``Replication and Evidence Factors in Observational Studies''.

License GPL-2

Encoding UTF-8

LazyData true

Imports stats, MASS, sensitivity2x2xk, graphics, sensitivitymult, sensitivitymv, senstrat, DOS2

Suggests optmatch, DiPs, approxmatch

Depends R (>= 3.5.0)

NeedsCompilation no

Repository CRAN

Date/Publication 2020-07-06 13:30:02 UTC

R topics documented:

antineoplastic *Biomonitoring of Workers Exposed to Antineoplastic Drugs*

Description

Two groups of exposed workers compared to unexposed controls, where exposed workers prepared antineoplasitic drugs, protected either only by gloves or by gloves and a laminar hood with vertical air flow. The outcome is the comet assay applied to blood lymphocytes. The comet assay is a measure of damage to DNA. Data from Kopjar and Garaj-Vrhovac (2001). These data illustrate two evidence factors formed when comparing three treatment groups.

Usage

```
data("antineoplastic")
```

Format

A data frame with 59 observations on the following 9 variables.

`id` ID number. See Tables I and III of Kopjar and Garaj-Vrhovac (2001).

`age` Age in years

`str` Three age strata, cut at the thirds of age.

`grp` Gloves = exposed nurses/doctors protected only by gloves, Hood = exposed nurses/doctors protected by both gloves and a laminar hood with vertical air flow, Control = students and office workers not exposed to antineoplastic drugs.

`tailmoment` Tail moment of the comet assay.

`taillength` Tail length of the comet assay mu-m.

`z1` 1 if exposed, 0 if control

`z2` 1 if Gloves, 0 if Hood, NA if control

`f2` TRUE if in factor 2, FALSE otherwise

Details

The data set is intended to illustrate evidence factors, comparing exposed workers to controls, and workers protected only by gloves to workers with the additional protecting of a laminar hood with vertical air flow.

Source

From Tables I and III of Kopjar and Garaj-Vrhovac (2001).

References

Kopjar, N. and Garaj-Vrhovac, V. (2001) <doi:10.1093/mutage/16.1.71> "Application of the alkaline comet assay in human biomonitoring for genotoxicity: a study on Croation medical personnel handling antineoplastic drugs". Mutagenesis, 16, 71-78.

Examples

```
data(antineoplastic)
attach(antineoplastic)
table(str)
table(grp)
oldpar<-par(mfrow=c(1,2))
boxplot(tailmoment~grp,ylab="Tail Moment",
       main="3 Groups",ylim=c(8,20))
boxplot(tailmoment~z1,names=c("Control","Exposed"),
       main="Factor 1",ylab="Tail Moment",ylim=c(8,20))
boxplot(tailmoment[f2]~z2[f2],names=c("Hood","Gloves"),
       ylab="Tail Moment",main="Factor 2",ylim=c(8,20))
y<-senstrat::hodgeslehmann(tailmoment,z1,str)
# First factor
senstrat::senstrat(y,z1,str,gamma=20)
# Second factor
senstrat::senstrat(y[f2],z2[f2],str[f2],gamma=2.75)
detach(antineoplastic)
par(oldpar)
```

benzene *Chromosome Damage from Exposure to Benzene*

Description

Examines chromosome aberrations among shoe workers exposed to benzene and unexposed controls.

Usage

```
data("benzene")
```

Format

A data frame with 78 observations on the following 5 variables.

`age` Age in years

`exposure` Years employed as a shoe worker. For controls, exposure = 0

`alcohol` 0=no, 1=yes

`smoking` Cigarette smoking, packs per day

`totalplus` Total chromosome aberrations, including gaps.

Source

The data are frome Tunca, B.T. and Egeli, U. (1996). Paz-y-Mino (2002) discuss whether to include gaps in chromosome aberrations. Used as an example in Rosenbaum (2001).

References

Paz-y-Mino, C., Davalos, M.V., Sanchez, M.E., Arevalo, M. and Leone, P.E. (2002). <doi:10.1016/S1383-5718(02)00021-9> "Should gaps be included in chromosomal aberration analysis? Evidence based on the comet assay". Mutation Research: Genetic Toxicology and Environmental Mutagenesis, 516, 57-61.

Rosenbaum, P.R. (2001). <doi:10.1093/biomet/88.1.219> "Effects attributable to treatment: Inference in experiments and observational studies with a discrete pivot". Biometrika, 88, 219-231.

Tunca, B.T., Egeli, U. (1996). <doi:10.1289/ehp.961041313> "Cytogenetic findings on shoe workers exposed long-term to benzene". Environmental Health Perspectives, 104, 1313-1317.

Examples

```
data(benzene)
attach(benzene)
boxplot(totalplus[exposure==0],totalplus[exposure>0],names=c("C","W"),ylab="Chromosome aberrations",
       main=c("Controls vs Workers"),xlab="C=control, W=worker",ylim=c(0,36))
plot(exposure[exposure>0],totalplus[exposure>0],ylim=c(0,36),main="Shoe Workers Only",
       xlab="Years of Exposure",ylab="Chromosome aberrations",pch=16)
wilcox.test(totalplus[exposure>0],totalplus[exposure==0],conf.int=TRUE,alternative = "greater")
cor.test(exposure[exposure>0],totalplus[exposure>0],method="k",alternative = "greater")
detach(benzene)
```

ck *Minimum Wages and Employment*

Description

Data from Card and Krueger's study of minimum wages and employment in New Jersey and Pennsylvania.

Usage

```
data("ck")
```

Format

A data frame with 198 observations on the following 7 variables.

`mset` Matched set indicator, 1, 2, ..., 66

`grp` Group indicator `NJhigh` `NJlow` `PA`

`fte` Full time employees, Februrary 1992, before the increase in the minimum wage in New Jersey (NJ).

`fte2` Full time employees, November 1992, after the increase in the minimum wage in New Jersey (NJ).

`CHAIN` Restaurant chain, Burger King, Wendy's, KFC and Roy Rogers

`HRSOPEN` Number of hours the restaurant was open in February 1992

`SHEET` Identifier that may be used to link the limited data here to the full data set provided by Card and Krueger.

Details

On 1 April 1992, the state of New Jersy increased its minimum wage from 4.25 dollars per hour to 5.05 dollars per hour, while adjacent Pennsylvania made no change. David Card and Alan Krueger studied the change in employment at fast food restaurants, from before the wage increase to after, based on survey data they collected.

There are three groups matched for HRSOPEN and indicators of the CHAIN. PA are restaurants in eastern Pennsylvania, adjacent to western New Jersey. NJlow are New Jersey restaurants with a starting wage of at most 4.50 dollars in February 1992, and NJhigh are New Jersey restaurants with a starting wage of more than 4.50 dollars. The NJlow restaurants were required to increase the starting wage by substantially more than the NJhigh restaurants.

The match was built using the algorithm of Karmakar et al. (2019) and the approximatch package in R.

Using a different matched sample, an analysis of two evidence factors in Card and Krueger's study is given in Rosenbaum (2010).

Source

Card, D. and Krueger, A. B. (1994) <doi:10.1080/10618600.2019.1584900> "Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania". American Economic Review, 84, 772-793.

References

Card, D. and Krueger, A. B. (1994) <doi:10.1080/10618600.2019.1584900> "Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania". American Economic Review, 84, 772-793.

Karmakar, B., Small, D. S. and Rosenbaum, P. R. (2019) <doi:10.1080/10618600.2019.1584900> "Using approximation algorithms to build evidence factors and related designs for observational studies". Journal of Computational and Graphical Statistics, 28, 698-709.

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

Examples

```
data(ck)
boxplot((ck$fte2-ck$fte)~ck$grp)
```

---

ckA *Matching the Minimum Wage Data*

Description

The data are Card and Krueger's (1994) minimum wage case study, prior to matching for CHAIN and HRSOPEN. The data illustrate matching to form three groups using Bikram Karmakar's approxmatch package. A similar matched data set after matching is in ck in this package: ck is a matched subset of all of the data in ckA.

Usage

```
data("ckA")
```

Format

A data frame with 351 observations on the following 9 variables.

`SHEET` Identifier that may be used to link the limited data here to the full data set provided by Card and Krueger.

`HRSOPEN` Number of hours the restaurant was open in February 1992

`CHAIN` Restaurant chain, Burger King, Wendy's, KFC and Roy Rogers

`chain1` Binary indicator, chain 1

`chain2` Binary indicator, chain 2

`chain3` Binary indicator, chain 3

`FTE` Full time employees, Februrary 1992, before the increase in the minimum wage in New Jersey (NJ).

`FTE2` Full time employees, November 1992, after the increase in the minimum wage in New Jersey (NJ).

`grp` Group indicator `NJhigh` `NJlow` `PA`

Details

These data are used to illustrate matching to form matched triples from three groups, Pennsylvania restaurants, New Jersey restaurants with low starting wages before the wage increase, and New Jersey restaurants with high starting wages before the wage increase. The low/high wage cut is at 4.50 dollars per hour. The method illustrated is from Karmakar et al. (2019) as implemented in Karmakar's approxmatch package.

To use the approxmatch package, you must first install and load the optmatch package, which an academic license. The optmatch packages uses the relax code of Bertsekas and Tseng (1988).

Using a different matched sample, an analysis of two evidence factors in Card and Krueger's study is given in Rosenbaum (2010).

Source

Card, D. and Krueger, A. B. (1994) <doi:10.1080/10618600.2019.1584900> "Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania". American Economic Review, 84, 772-793.

References

Bertsekas, D. P. and Tseng, P. (1988) <doi:10.1007/BF02288322> "The Relax codes for linear minimum cost network flow problems". Annals of Operations Research, 13(1), 125-190.

Card, D. and Krueger, A. B. (1994) <doi:10.1257/aer.90.5.1397> "Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania". American Economic Review, 84, 772-793.

Karmakar, B., Small, D. S. and Rosenbaum, P. R. (2019) <doi:10.1080/10618600.2019.1584900> "Using approximation algorithms to build evidence factors and related designs for observational studies". Journal of Computational and Graphical Statistics, 28, 698-709.

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

Examples

```
data(ckA)
table(ckA$CHAIN)
table(ckA$grp)

## Not run:
#  To run this example, you must install and load
#  the optmatch package and accepts its academic license.

dist<-approxmatch::multigrp_dist_struc(ckA,as.character(ckA$grp),
                            list(mahal=c("chain1","chain2","chain3",
                                          "HRSOPEN")),wgts=1)
mtch<-approxmatch::tripletmatching(dist,as.character(ckA$grp),
                        indexgroup="PA",ckA,"CHAIN",design=c(1,1,1))
## End(Not run)
```

`hsmoke` *Smoking and Homocysteine*

Description

Data from NHANES 2005-2006 concerning homocysteine levels in daily smokers (z=1) and never smokers (z=0), aged 20 and older. Daily smokers smoked every day for the last 30 days, smoking an average of at least 10 cigarettes per day. Never smokers smoked fewer than 100 cigarettes in their lives, do not smoke now, and had no tobacco use in the previous 5 days. There are 2475 individuals in 104 strata defined by gender, age, education, income, and BMI.

Usage

```
data("hsmoke")
```

Format

A data frame with 2475 observations on the following 15 variables.

`SEQN` 2005-2006 NHANES ID number.

`homocysteine` Homocysteine level, umol/L. Based on LBXHCY.

`z` z=1 for a daily smoker, z=0 for a never smoker. Based on SMQ020, SMQ040, SMD641, SMD650, SMQ680.

`female` 1=female, 0=male. Based on RIAGENDR `age` Age in years. Based on RIDAGEYR.

`education` Education. Based on DMDEDUC2.

`povertyr` Ratio of family income to the poverty level, capped at five times the poverty level. Based on INDFMPIR.

`bmi` BMI = body-mass-index. Based on BMXBMI.

`cotinine` Cotinine level (ng/mL), a marker for recent smoking. Based on LBXCOT.

`st` Numeric strata indicating female, age, education, BMI and poverty.

`stf` A factor for strata indicating female, age, education, BMI and poverty.

`age3` Three age categories, 20-39, 40-50, >=60. Used to define the strata.

`ed3` Three education categories, <High School, High School, at least some College. Used to define the strata.

`bmi3` Three categories of the body-mass-index, BMI, <30, [30,35), >= 35. Used to define the strata.

`pov2` TRUE=income at least twice the poverty level, FALSE otherwise. Used to define the strata.

Details

Bazzano et al. (2003) noted higher homocysteine levels in the blood of smokers. This data set reexamines this issue using more recent data from NHANES 2005-2006, the most recent NHANES that measured homocysteine.

The example reproduces analyses from Rosenbaum (2018). The example illustrates use a sensitivity analysis with 104 strata defined by covariates, with and without additional adjustment for continuous versions of these covariates using robust covariance adjustment; see Rosenbaum (2002).

Other analyses of smoking and homocysteine are in Karmakar et al. (2020) and Pimentel et al. (2015).

Source

NHANES, the US National Health and Nutrition Examination Survey, 2005-2006. Publicly available from the US National Center for Health Statistics, US Centers for Disease Control.

References

Bazzano, L. A., He, J., Muntner, P., Vupputuri, S. and Whelton, P. K. (2003) <doi:10.7326/0003-4819-138-11-200306030-00010> "Relationship between cigarette smoking and novel risk factors for cardiovascular disease in the United States". Annals of Internal Medicine, 138, 891-897.

Karmakar, B., Small, D. S. and Rosenbaum, P. R. (2020) <doi:10.1093/aje/kwz263> "Using evidence factors to clarify exposure biomarkers". American Journal of Epidemiology, 189, 243–249.

Pimentel, S. D., Small, D. S. and Rosenbaum, P. R. (2016) <doi:10.1080/01621459.2015.1076342> "Constructed second control groups and attenuation of unmeasured biases". Journal of the American Statistical Association, 111, 1157-1167.

Rosenbaum, P. R. (2002). <doi:10.1214/ss/1042727942> "Covariance adjustment in randomized experiments and observational studies". Statistical Science, 17(3), 286-327.

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". Annals of Applied Statistics 12(4):2312–2334.

Examples

```
data(hsmoke)
attach(hsmoke)
# Calculations discussed in Rosenbaum (2018).

# Stratified analysis of log2(homocysteine)
# Uses Hodges-Lehmann aligned ranks
l2h<-log2(homocysteine)
hll2h<-senstrat::hodgeslehmann(l2h,z,st,align="hl")
senstrat::senstrat(hll2h,z,st,gamma=1.95)
senstrat::senstrat(hll2h,z,st,gamma=2.1)

# Stratification + Covariance Adjustment
# Covariance model does not include the treatment, z
mod<-MASS::rlm(l2h~female+age+povertyr+bmi+education)
l2hr<-as.vector(mod$residual)
hll2hr<-senstrat::hodgeslehmann(l2hr,z,st,align="hl")
senstrat::senstrat(hll2hr,z,st,gamma=2.1)

# Uses M-scores in place of aligned ranks
msr<-senstrat::mscores(l2hr,z,st=st)
senstrat::senstrat(msr,z,st,gamma=2.35)

# Evidence factor analysis from Section 6.2 of Rosenbaum (2018)
# Among smokers, is homocysteine higher with high cotinine?
summary(cotinine[z==1]) # Cotinine among smokers
table(cotinine[z==1]<=185.750)
table(cotinine[z==1]>=342)
use<-(z==1)&((cotinine<=185.750)|(cotinine>=342))
cot<-1*((cotinine>=342))
# Figure 4 in Rosenbaum (2018)
graphics::boxplot(l2hr[use]~cot[use],names=c("Low","High"),
      ylab="Residual",xlab="Cotinine")
cotl2hr<-senstrat::mscores(l2hr[use],cot[use],st=st[use])
senstrat::senstrat(cotl2hr,cot[use],st[use],gamma=2.5)
detach(hsmoke)
```

`lead` *Lead in Children*

Description

Data from Morton et al. (1982) concerning exposed children whose fathers worked in a battery plant where lead was used in the manufacture of batteries. Exposed children were matched to controls for age and neighborhood. For exposed children, also given are the father's level of exposure to lead at work (level) and the father's hygiene upon leaving the battery plant at the end of the day.

Usage

```
data("lead")
```

Format

A data frame with 33 observations on the following 6 variables.

`control` Blood lead level for the control, micrograms of lead per decaliter of whole blood.

`exposed` Blood lead level for the exposed/treated child, micrograms of lead per decaliter of whole blood.

`level` Father's level of exposure to lead: a factor with levels `high` `low` `medium`

`hyg` Father's hygiene before going home: a factor with levels `good` `mod` `poor`

`both` A factor built from level and hyg: a factor with levels `high.ok` `high.poor` `low` `medium`

`dif` Exposed-minus-control pair difference in blood lead levels.

Details

The data were assembled from two published tables in Morton et al. (1982). One matched pair with no control is omitted here. Small ambiguities in assembling a complete data set from two tables were resolved by a throw of the dice; however, it is a reasonable example to illustrate statistical methods. One table described the exposed-versus-control matched pairs, and these are as in the paper. The second table described the exposed individuals, their level of exposure and their hygiene, and again these are as in the paper. The two tables were linked using the blood lead level of the exposed children, and a couple of ties in these lead levels made for small ambiguities about which control responses belong with which hygienes and exposure levels for the exposed children. See, for instance, rows 18 and 22, where both exposed children have blood lead level 34.

Source

Data are from Morton et al. (1982). They were used as an example in Rosenbaum (1991, 2002, 2011, 2017).

References

Morton, D. E., Saah, A. J., Silberg, S. L., Owens, W. L., Roberts, M. A., & Saah, M. D. (1982) <doi:10.1093/oxfordjournals.aje.a113336> "Lead absorption in children of employees in a lead-related industry". American Journal of Epidemiology, 115(4), 549-555.

Rosenbaum, P. R. (1991) <doi:10.1214/aos/1176348141> "Some poset statistics". The Annals of Statistics, 19(2), 1091-1097.

Rosenbaum, P. R. (2002) <doi:10.1007/978-1-4757-3692-2_3> "Observational Studies" (2nd edition). New York: Springer. Section 4.3.

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

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. (2017) <https://www.hup.harvard.edu/catalog.php?isbn=9780674975576> "Observation and Experiment: An Introduction to Causal Inference". Cambridge, MA: Harvard University Press. Chapters 7 and 9.

Examples

```
data(lead)
# Reproduces parts of Table 2 in Rosenbaum (2011)
DOS2::senU(lead$dif,gamma=5.8,m=8,m1=5,m2=8)
DOS2::senU(lead$dif,gamma=5,m=5,m1=4,m2=5)

# m=2, m1=2, m2=2 is the U-statistic that closely
# resembles Wilcoxon's signed rank test.  Note
# that the results are almost the same.
DOS2::senWilcox(lead$dif,gamma=5) # In Table 2
DOS2::senU(lead$dif,gamma=5,m=2,m1=2,m2=2)
```

leadworker *DNA Damage in Lead Workers*

Description

Data from Table 1 of Wu et al. (2002) concerning DNA damage among lead workers compared to controls. In this example, the data have been matched for age and smoking, making 11 matched pairs.

Usage

```
data("leadworker")
```

Format

A data frame with 22 observations on the following 7 variables.

`mset` Matched set indicator, 1, 2, ..., 11

`group` 1=lead worker, 0=control

`age` Age in years

`smoking` Smoking, pack-years

`bll` Blood lead level, mu-g/dl

`dpc` DNA-protein cross-links, percent

`id` ID number in the original study, before matching

Details

The outcome, a measure of DNA damage is dpc.

Source

Fang-Yang Wu, Pao-Wen Chang, Chin-Ching Wu and Hsien-Wen Kuo (2002). "Correlations of Blood Lead with DNA-Protein Cross-Links and Sister Chromatid Exchanges in Lead Workers". Cancer Epidemiology, Biomarkers and Prevtion. March 1 2002 11 (3) 287-290.

Examples

```
data(leadworker)
boxplot(leadworker$dpc~leadworker$group)
```

periodontal *Smoking and Periodontal Disease*

Description

Data from NHANES 2011-2012 containing 441 matched pairs of a daily cigarette smoker and a never smoker, recording the extent of periodontal disease. See Rosenbaum (2017).

Usage

```
data("periodontal")
```

Format

A data frame with 882 observations on the following 12 variables.

`SEQN` NHANES 2011-2012 sequence number

`female` =1 for female, 0 for male

`age` Age in years

`black` =1 for black, 0 for other

`educf` Education, in five categories. An ordered factor with levels `<9` for less than 9th grade, `9 to 11` for 9th to 11th grade, `HS/GED` for high school or GED degree, `SomeCol` for some college, `College` for college degree.

`income` Ratio of family income to the poverty level, capped at 5 times the poverty level.

`cigsperday` Cigarettes smoked per day for daily smokers, 0 for never smokers

`either` Number of periodonal measurements indicative of periodontal disease.

`neither` Number of periodonal measurements

`pcteither` Percent indicative of periodontal disease, =100*either/neither.

`z` Treatment indicator, 1=daily smoker, 0=never smoker

`mset` Matched set indicator, 1 to 441.

Details

Excluding wisdom teeth, 6 measurements are taken for each tooth that is present, up to 28 teeth. Following Tomar and Asma (2000), a measurement indicates periodontal disease if either there is a loss of attachment of at least 4mm or a pocket depth of at least 4mm. The first individual has 11 measurements indicative of periodontal disease, out of 106 measurements, so pcteither is 100*11/106 = 10.38 percent.

Source

Data are from the National Health and Nutrition Examination Survey 2011-2012 and were used as an example in Rosenbaum (2017).

References

Rosenbaum, P. R. (2015) <https://obsstudies.org/two-r-packages-for-sensitivity-analysis-in-observational-studies/> "Two R packages for sensitivity analysis in observational studies". Observational Studies, 1(1), 1-17.

Rosenbaum, P. R. (2017) <doi:10.1214/17-STS621> "The general structure of evidence factors in observational studies". Statistical Science 32, 514-530.

Tomar, S. L. and Asma, S. (2000) <doi:10.1902/jop.2000.71.5.743> "Smoking attributable periodontitis in the US: Findings from NHANES III". J Periodont 71, 743-751.

"US National Health and Nutrition Examination Survey 2011-2012". www.cdc.gov/nchs/nhanes/index.htm

Examples

```
# Figure 1 in Rosenbaum (2017)
data(periodontal)
attach(periodontal)
oldpar<-par()
m<-matrix(1:2,1,2)
layout(m,widths=c(1,2))
boxplot(pcteither[z==1]-pcteither[z==0],ylab="Smoker-Control Difference",
        main="(i)",xlab="Matched Pairs",ylim=c(-100,100))
abline(h=0,lty=2)
DOS2::crosscutplot(cigsperday[z==1],
    pcteither[z==1]-pcteither[z==0],
    ylab="Smoker-Control Difference",
    xlab="Cigarettes per Day",main="(ii)",
    ylim=c(-100,100))
abline(h=0,lty=2)

# Sensitivity analysis in Section 2.3 of Rosenbaum (2017)
y<-pcteither[z==1]-pcteither[z==0]
x<-cigsperday[z==1]
DOS2::senWilcox(y,gamma=2.76)
# The following is the same as sensitivitymw::senmw(y,gamma=2.77,method="p")
sensitivitymult::senm(pcteither,z,mset,gamma=2.77,inner=.5,trim=2)
# The following is the same as sensitivitymw::senmw(y,gamma=3.5,method="p")
sensitivitymult::senm(pcteither,z,mset,gamma=3.5,inner=.5,trim=2)
# Second evidence factor
DOS2::crosscut(x,y)
DOS2::crosscut(x,y,gamma=1.6)

# Note, however, that other statistics report greater insensitivity to
# bias by virtue of having larger design sensitivity:
sensitivitymult::senm(pcteither,z,mset,gamma=3.5,inner=1,trim=4)
sensitivitymult::senm(pcteither,z,mset,gamma=4.2,inner=1,trim=4)
DOS2::senU(y,m1=4,m2=5,m=5,gamma=2.77)
```

```
DOS2::senU(y,m1=6,m2=8,m=8,gamma=2.77)
DOS2::senU(y,m1=6,m2=8,m=8,gamma=3.5)
detach(periodontal)
par(oldpar)
```

tannery *DNA Damage Among Tannery Workers*

Description

Data are from Zhang et al. (2008, Table 2) who studied DNA damage among tannery workers often exposed to trivalent chromium. The outcome is the mean tail moment (mtm) of the comet assay, a standard measure of DNA damage, with higher values signifying greater damage. The study describes 90 males in 30 blocks of 3 individuals. There are three groups, each with 30 individuals. The three groups had a simlar distribution of ages, and the blocks control for smoking as closely as possible. Group e1 consists of 30 exposed workers at the tannery who worked in the tannery department, where the highest exposures to trivalent chromium are expected. Group e2 consists of 30 workers at the tannery who worked in the finishing department, where exposure to trivalent chromium is expected to be much lower. Group c consists of 30 controls who did not work at the tannery. This example is discussed in Chapter 20 of "Design of Observational Studies", 2nd ed.

Usage

```
data("tannery")
```

Format

A data frame with 30 observations on the following 4 variables.

`block` Block indicator, 1 to 30.

`e1mtm` mtm for the tannery worker from the tannery department.

`e2mtm` mtm for the tannery worker from the finishing department.

`cmtm` mtm for the control who did not work in the tannery

Details

The comet assay is described by Collins (2004). It is thought to measure DNA strand breaks, producing an image that resembles the tail of a comet, a larger, longer tail suggesting more extensive strand breaks. This example was discussed in Rosenbaum (2011).

Source

Data from Zhang et al. (2008, Table 2).

References

Collins, A. R. (2004) <doi:10.1385/MB:26:3:249> "The comet assay for DNA damage and repair: principles, applications, and limitations". Molecular Biotechnology 26(3), 249-261.

Rosenbaum, P. R. (2011) <doi:10.1198/jasa.2011.tm10422> "Some approximate evidence factors in observational studies". Journal of the American Statistical Association, 106, 285-293.

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 69 118-127. (Introduces inner trimming.)

Rosenbaum, P. R. (2015) <https://obsstudies.org/two-r-packages-for-sensitivity-analysis-in-observational-studies/> "Two R packages for sensitivity analysis in observational studies". Observational Studies, 1(1), 1-17.

Zhang, M., Chen, Z., Chen, Q., Zou, H., Lou, J. He, J. (2008) <doi:10.1016/j.mrgentox.2008.04.011> "Investigating DNA damage in tannery workers occupationally exposed to trivalent chromium using comet assay". Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 654(1), 45-51.

Examples

```
data(tannery)
boxplot(tannery[,2:4],names=c("Tannery E1","Finishing E2",
    "Control C"),ylab="Mean Tail Moment")
oldpar<-par(mfrow=c(1,2))
boxplot(tannery[,2:3],names=c("E1","E2"),ylab="Mean Tail Moment",
     main="Tannery vs. Finishing",ylim=c(0,12))
boxplot(as.vector(unlist(tannery[,2:3])),tannery[,4],
     names=c("E1+E2","C"),ylab="Mean Tail Moment",
     main="Exposed vs. Control",ylim=c(0,12))

# Stratified Wilcoxon analysis from the chapter Evidence Factors
# of Design of Observational Studies, Second Edition
# Also reproduces the F1, F2 example in Rosenbaum (2011, sec 6).
y<-tannery[,2:4]
rkc<-t(apply(y,1,rank)) # Ranks for (E1,E2,C)
sum(rkc[,3]) # Stratified rank sum for C in (E1, E2, C)
(35-60)/sqrt(20)

y<-tannery[,2:3]
rkc<-t(apply(y,1,rank)) # Ranks for (E1,E2)
sum(rkc[,2]) # Stratified rank sum for E2 in (E1, E2)

# Reorganize y for input to 'separable1v' from 'sensitivitymult'
# 'separable1v' is one-sided, looking for a large rank sum

# Factor 1
y<-tannery[,4:2]*(-1)
rkc<-t(apply(y,1,rank)) # Ranks for -y for (C,E2,E1)
sensitivitymult::separable1v(rkc,gamma=1)
# Test for C in (E1, E2, C)
(85-60)/sqrt(20)
(35-60)/sqrt(20)
```

```
sensitivitymult::separable1v(rkc,gamma=6)
# Test for C in (E1, E2, C)
p1<-sensitivitymult::separable1v(rkc,gamma=7)$pval

#Factor 2
y<-tannery[,3:2]*(-1)
rkc<-t(apply(y,1,rank)) # Ranks for -y for (E2,E1)
sensitivitymult::separable1v(rkc,gamma=1)
# Test for E2 in (E2, E1)
# Combine P-values using Fisher's method
sensitivitymv::truncatedP(c(1.134237e-08,0.001743502),trunc=1)

# Larger gammas
sensitivitymult::separable1v(rkc,gamma=1.7)
p2<-sensitivitymult::separable1v(rkc,gamma=2)$pval
# Combine P-values using Fisher's method
c(p1,p2)
sensitivitymv::truncatedP(c(p1,p2),trunc=1)

# Nearly reproduces calculations from Section 6 of Rosenbaum (2011)
# However, in Rosenbaum (2011), the second factor
# uses a pooled scale factor, whereas senm does not,
# so the result is very slightly different.
attach(tannery)
mset<-rep(block,3)
zC<-c(rep(0,60),rep(1,30))
z12<-c(rep(1,30),rep(0,30),rep(NA,30))
y<-c(e1mtm,e2mtm,cmtm)
detach(tannery)
use<-!is.na(z12)
# Factor 1
sensitivitymult::senm(y,zC,mset,gamma=1,
   alternative=”less”,trim=1)
sensitivitymult::senm(y,zC,mset,gamma=11.7,
   alternative=”less”,trim=1)
# Factor 2
sensitivitymult::senm(y[use],z12[use],mset[use],
   gamma=2,alternative=”greater”,trim=1)

# Combine two evidence factors
p1<-sensitivitymult::senm(y,zC,mset,gamma=12,
    alternative=”less”,trim=1)$pval
p2<-sensitivitymult::senm(y[use],z12[use],mset[use],gamma=3,
    alternative=”greater”,trim=1)$pval
c(p1,p2)
sensitivitymv::truncatedP(c(p1,p2),trunc=1)
# Combine p-values using Fisher's method

# Other psi-functions often have higher design
# sensitivity; see Rosenbaum (2013)
par(oldpar)
```

Index