

# Package: submax (via r-universe)

May 16, 2026

**Type** Package

**Title** Effect Modification in Observational Studies Using the Submax Method

**Version** 1.1.5

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

**Description** Effect modification occurs if a treatment effect is larger or more stable in certain subgroups defined by observed covariates. The submax or subgroup-maximum method of Lee et al. (2018) <doi:10.1111/biom.12884> does an overall test and separate tests in subgroups, correcting for multiple testing using the joint distribution.

**Imports** stats, mvtnorm, sensitivityfull

**License** GPL-2

**Encoding** UTF-8

**LazyData** true

**NeedsCompilation** no

**Depends** R (>= 3.5.0)

**Author** Paul R. Rosenbaum [aut, cre]

**Repository** <https://rosenbap.r-universe.dev>

**Date/Publication** 2025-04-30 18:43:31 UTC

**RemoteUrl** <https://github.com/cran/submax>

**RemoteRef** HEAD

**RemoteSha** 531ad780c71c4403fd039e97aace908977272b4c

## Contents

submax-package	2
Active	3
amplify	5
mercury	7

mscorev . . . . .	8
score . . . . .	9
separable1fc . . . . .	13
submax . . . . .	15
tbmetaphase . . . . .	18

<b>Index</b>	<b>20</b>
--------------	-----------

---

submax-package	<i>Effect Modification in Observational Studies Using the Submax Method</i>
----------------	---

---

## Description

Effect modification occurs if a treatment effect is larger or more stable in certain subgroups defined by observed covariates. The submax or subgroup-maximum method of Lee et al. (2018) <doi:10.1111/biom.12884> does an overall test and separate tests in subgroups, correcting for multiple testing using the joint distribution.

## Details

The DESCRIPTION file:

```
Package:      submax
Type:        Package
Title:       Effect Modification in Observational Studies Using the Submax Method
Version:     1.1.5
Authors@R:   person(given = c("Paul", "R."), family = "Rosenbaum", role = c("aut", "cre"), email = "rosenbaum@wharton.upenn.edu")
Maintainer:  Paul R. Rosenbaum <rosenbaum@wharton.upenn.edu>
Description: Effect modification occurs if a treatment effect is larger or more stable in certain subgroups defined by observed covariates.
Imports:     stats, mvtnorm, sensitivityfull
License:     GPL-2
Encoding:    UTF-8
LazyData:   true
NeedsCompilation: no
Depends:    R (>= 3.5.0)
Author:     Paul R. Rosenbaum [aut, cre]
```

Index of help topics:

Active	Physical Activity and Survival in NHANES
amplify	Amplification of sensitivity analysis in observational studies.
mercury	NHANES Mercury/Fish Data
mscorev	Computes M-scores for Permutational M-tests.
score	Creates M-scores for Use by submax().
separable1fc	Computes the Separable Approximation.

submax	Effect Modification Using the Submax Method in Observational Studies
submax-package	Effect Modification in Observational Studies Using the Submax Method
tbmetaphase	Genetic damage from drugs used to treat TB

The main function is `submax()`. Also helpful is `score()`. See their documentation.

### Author(s)

Paul R. Rosenbaum [aut, cre]

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

### References

Lee, K., Small, D. S., & Rosenbaum, P. R. (2018) <doi:10.1111/biom.12884> A powerful approach to the study of moderate effect modification in observational studies. *Biometrics*, 74(4), 1161-1170.

### Examples

```
#Reproduces parts of Table 2 of Lee et al. (2018)
data(Active)
submax(Active$delta,Active[,1:7],gamma=1.70,alternative="less")
```

---

Active

*Physical Activity and Survival in NHANES*

---

### Description

Physical activity and survival in the NHANES I Epidemiologic Follow-up Study or NHEFS. This is the example in Lee et al. (2018). It is patterned after a study by Davis et al. (1994). The NHEFS combined the NHANES I study with follow-up for survival. There are 470 matched pairs consisting of a treated group who were quite inactive at the time of NHANES I and a matched control group who were very active.

### Usage

```
data("Active")
```

### Format

A data frame with 470 observations on the following 12 variables.

All 470 ones, one for each pair

male 1 for male, 0 for female

female 1 for female, 0 for male

poor 1 for income less than 2x poverty level, 0 otherwise

notpoor 1 for income greater than 2x poverty level, 0 otherwise  
 smoker 1 for current smoker, 0 otherwise  
 nonsmoker 1 for current nonsmoker, 0 otherwise  
 delta O'Brien and Fleming scores for censored matched pairs. See details.  
 treated.followup.time Death or censoring time for inactive individual  
 treated.censored.time Equals Inf if the inactive individual was censored  
 control.followup.time Death or censoring time for active individual  
 control.censored.time Equals Inf if the active individual was censored

### Details

Pairs were exactly matched for male/female, poor/notpoor and smoker/nonsmoker. Additionally, pairs were matched for age, white/nonwhite, years of education, employed or not during the previous 3 months, marital status, alcohol consumption and dietary quality. The covariates that were not exactly matched were matched by minimizing the total Mahalanobis distance within matched pairs. Table 1 of Lee et al. (2018) shows covariate balance before and after matching. These matching techniques are described in Chapter 8 and Section 9.2 of Rosenbaum (2010). The matching is the usual kind in epidemiology and biostatistics, that is, so-called without-replacement matching, in which no person appears twice.

The example reproduces a row from Table 2 of Lee et al. (2018).

The values in delta in Active are the Prentice-Wilcoxon scores for censored paired data proposed by O'Brien and Fleming (1987). Specifically, delta is the  $\Delta$  in section 2 of O'Brien and Fleming (1987). Following their suggestion at the end of their section 2, data are considered censored at the earlier censoring time in a matched pair. These deltas are computed separately in  $8 = 2 \times 2 \times 2$  subgroups defined by male/female  $\times$  poor/notpoor  $\times$  smoker/nonsmoker. Separate computation of  $\Delta$  in subgroups is not needed for the global test of no treatment effect at all in section 3.2 of Lee et al. (2018), but it is an aspect of the simultaneous inference by closed testing in section 4 of Lee et al. (2018).

The NHEFS was, essentially, the NHANES I snapshot survey combined with follow-up for survival. Data on mortality and time of death were collected in four follow-up surveys in 1982-1984, 1986, 1987 and 1992. Tracing of subjects which enable determination of whether the subject was alive or had died was high. Ninety six percent of the study population had been successfully traced at some point through the 1992 follow-up. Tracing rates for each follow-up ranged from 90 to 94 percent. See Cox et al. (1997).

### Source

The data set was constructed by Kwonsang Lee from the NHEFS; see Lee et al. (2018). The original NHEFS data are publicly available at the NHANES web-page at CDC.

### References

Cox, C. S., Mussolino, M. E., Rothwell, S. T., Lane, M. A., Golden, C. D., Madans, J. H., and Feldman, J. J. (1997). Plan and operation of the NHANES I Epidemiologic Followup Study, 1992. Vital and health statistics. Ser. 1, Programs and collection procedures, (35), 1-231.

Davis, M. A., Neuhaus, J. M., Moritz, D. J., Lein, D., Barclay, J. D., and Murphy, S. P. (1994). Health behaviors and survival among middle aged and older men and women in the NHANES I Epidemiologic Follow-Up Study. *Preventive Medicine*, 23, 369-376.

Lee, K., Small, D. S., & Rosenbaum, P. R. (2018) <doi:10.1111/biom.12884> A powerful approach to the study of moderate effect modification in observational studies. *Biometrics*, 74(4), 1161-1170.

O'Brien, P. C. and Fleming, T. R. (1987). A paired Prentice-Wilcoxon test for censored paired data. *Biometrics*, 43, 169-180. The variable delta in Active is the delta in section 2 of this paper. Following their suggestion at the end of their section 2, data are considered censored at the earlier censoring time in a matched pair.

Rosenbaum, P. R. (2010). *Design of Observational Studies*. New York: Springer.

### Examples

```
# The example is from Lee et al. (2018).
data(Active)
submax(Active$delta, Active[,1:7], gamma=1, alternative="less")
```

---

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 <code>senmv</code> . <code>length(gamma)&gt;1</code> will generate an error.
lambda	lambda is a vector of values > gamma. An error will result unless <code>lambda[i] &gt; gamma &gt; 1</code> 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  $\text{gamma} = (\text{lambda} * \text{delta} + 1) / (\text{lambda} + \text{delta})$ . 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 distribution.

Strictly speaking, the amplification describes matched pairs, not matched sets. 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  $(\lambda, \delta) = (\gamma, \infty)$  and  $(\infty, \gamma)$ .

A related though distinct 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` and the `sensitivitymult` packages. The calculations are elementary.

### Author(s)

Paul R. Rosenbaum

### References

Gastwirth, J. L., Krieger, A. M., Rosenbaum, P. R. (1998) Dual and simultaneous sensitivity analysis for matched pairs. *Biometrika*, 85, 907-920.

Lee, K., Small, D. S., & Rosenbaum, P. R. (2018) <doi:10.1111/biom.12884> A powerful approach to the study of moderate effect modification in observational studies. *Biometrics*, 74(4), 1161-1170.

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

Rosenbaum, P. R. (2015). Two R packages for sensitivity analysis in observational studies. *Observational Studies*, v. 1. (Free on-line.)

Wolfe, D. A. (1974) A characterization of population weighted symmetry and related results. *Journal of the American Statistical Association*, 69, 819-822.

### Examples

```
# The following is the calculation in Section 3.1 of Lee et al. (2018).
amplify(1.77,c(2,3,4))
```

---

mercury

*NHANES Mercury/Fish Data*

---

### Description

Data from NHANES 2009-2010. 397 treated people who ate at least 15 servings of fish or shellfish during the previous month are matched to two controls who ate at most one serving of fish or shellfish. The values in methylmercury record the level of methylmercury in blood in mu-g/dl.

### Usage

```
data("mercury")
```

### Format

A data frame with 1191 observations on the following 6 variables.

SEQN NHANES 2009-2010 id number

methylmercury Methylmercury in blood in mu-g/dl

fish 1 if ate  $\geq$  15 servings of fish or shellfish, 0 if  $\leq$  1 serving

mset Matched set indicator, 1,...,397.

female 1 if female, 0 if male

black 1 if black, 0 otherwise

### Details

Sets were matched 2-to-1 for for age, sex, ratio of household income to the poverty level, education, ethnic group (black, Hispanic, or other), and cigarette smoking. A table showing covariate balance after matching is in Rosenbaum (2014, Table 1).

### Source

From NHANES 2009-2010, publicly available at the NHANES web page at CDC.

### References

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

### Examples

```
data(mercury)
boxplot(mercury$methylmercury~mercury$fish)
```

---

mscorev

*Computes M-scores for Permutational M-tests.*


---

### Description

Of limited interest to most users, function `mscorev()` computes M-scores. A similar function `func-`  
tion is in the package `sensitivitymv`.

### Usage

```
mscorev(ymat, inner = 0, trim = 3, lambda = 0.5)
```

### Arguments

<code>ymat</code>	<p>If there are <math>I</math> matched sets and the largest matched set contains <math>J</math> individuals, then <math>y</math> is an <math>I</math> by <math>J</math> matrix with one row for each matched set. If matched set <math>i</math> contains one treated individual and <math>k</math> controls, where <math>k</math> is at least 1 and at most <math>J-1</math>, then <math>y[i,1]</math> is the treated individual's response, <math>y[i,2], \dots, y[i,k+1]</math> are the responses of the <math>k</math> controls, and <math>y[i,k+2], \dots, y[i,J]</math> are equal to NA.</p> <p>Although <math>y</math> can contain NA's, <math>y[i,1]</math> and <math>y[i,2]</math> must not be NA for every <math>i</math>. That is, every matched set must have at least one treated subject and one control.</p>
<code>inner</code>	<p><code>inner</code> and <code>trim</code> together define the <math>\psi</math>-function for the M-statistic. The default values yield a version of Huber's <math>\psi</math>-function, while setting <code>inner = 0</code> and <code>trim = Inf</code> uses the mean within each matched set. The <math>\psi</math>-function is an odd function, so <math>\psi(w) = -\psi(-w)</math>. For <math>w \geq 0</math>, the <math>\psi</math>-function is <math>\psi(w) = 0</math> for <math>0 \leq w \leq \text{inner}</math>, is <math>\psi(w) = \text{trim}</math> for <math>w \geq \text{trim}</math>, and rises linearly from 0 to <code>trim</code> for <code>inner</code> &lt; <math>w</math> &lt; <code>trim</code>.</p> <p>If uncertain about <code>inner</code>, <code>trim</code> and <code>lambda</code>, then use the defaults.</p> <p>An error will result unless <math>0 \leq \text{inner} \leq \text{trim}</math>.</p> <p>Taking <code>trim</code> &lt; <code>Inf</code> limits the influence of outliers; see Huber (1981). Taking <code>trim</code> &lt; <code>Inf</code> and <code>inner = 0</code> uses Huber's <math>\psi</math> function. Taking <code>trim = Inf</code> does no trimming and is similar to a weighted mean; see <code>TonT</code>. Taking <code>inner</code> &gt; 0 often increases design sensitivity; see Rosenbaum (2013).</p>
<code>trim</code>	<p><code>inner</code> and <code>trim</code> together define the <math>\psi</math>-function for the M-statistic. See <code>inner</code>.</p>
<code>lambda</code>	<p>Before applying the <math>\psi</math>-function to treated-minus-control differences, the differences are scaled by dividing by the <code>lambda</code> quantile of all within set absolute differences. Typically, <code>lambda = 1/2</code> for the median. The value of <code>lambda</code> has no effect if <code>trim=Inf</code> and <code>inner=0</code>. See Maritz (1979) for the paired case and Rosenbaum (2007) for matched sets.</p> <p>An error will result unless <math>0 &lt; \text{lambda} &lt; 1</math>.</p>

### Value

Generally, a matrix with the same dimensions as `ymat` containing the M-scores.

**Note**

The example reproduces Table 3 in Rosenbaum (2007).

Matched sets of unequal size are weighted using weights that would be efficient in a randomization test under a simple model with additive set and treatment effects and errors with constant variance; see Rosenbaum (2007). Specifically, the total score in set (row)  $i$  is divided by the number  $n_i$  of individuals in row  $i$ , as in expression (8) in Rosenbaum (2007).

**Author(s)**

Paul R. Rosenbaum

**References**

- Huber, P. (1981) Robust Statistics. New York: John Wiley. (M-estimates based on M-statistics.)
- Maritz, J. S. (1979). A note on exact robust confidence intervals for location. *Biometrika* 66 163–166. (Introduces exact permutation tests based on M-statistics by redefining the scaling parameter.)
- Rosenbaum, P. R. (2007) Sensitivity analysis for m-estimates, tests and confidence intervals in matched observational studies. *Biometrics*, 2007, 63, 456-464. <doi:10.1111/j.1541-0420.2006.00717.x>
- Rosenbaum, P. R. (2013). Impact of multiple matched controls on design sensitivity in observational studies. *Biometrics* 69 118-127. (Introduces inner trimming.) <doi:10.1111/j.1541-0420.2012.01821.x>
- Rosenbaum, P. R. (2015). Two R packages for sensitivity analysis in observational studies. *Observational Studies*, v. 1. (Free on-line.)

**Examples**

```
# The example reproduces Table 3 in Rosenbaum (2007).
data(tbmetaphase)
mscorev(tbmetaphase, trim=1)
```

---

score

*Creates M-scores for Use by submax().*

---

**Description**

The score() function is an optional aid in using the submax() function. The submax() function may be used on its own, but the user must then create the y matrix and the cmat matrix. The score() function can be used to create the y matrix and cmat matrix for use in the submax() function. It creates Huber-Maritz M-scores.

**Usage**

```
score(y, z, mset, x, expandx = FALSE, scale = "closed", inner = 0, trim = 3,
lambda = 1/2, xnames=NULL)
```

**Arguments**

y	A vector of responses with no missing data.
z	Treatment indicator, z=1 for treated, z=0 for control with $\text{length}(z)=\text{length}(y)$ .
mset	Matched set indicator, 1, 2, ..., sum(z) with $\text{length}(mset)=\text{length}(y)$ . Matched set indicators should be either integers or a factor.
x	A matrix of binary covariates. These are the potential effect modifiers. If x is a binary vector for one covariate, it will be reshaped into a matrix. An error will result if x does not have 1 or 0 coordinates. The matrix x should have $\text{length}(y)$ rows.
expandx	If <code>expandx=FALSE</code> , then x is used as is. If <code>expandx=TRUE</code> , then a column of 1's is added to x, as well as $1-x[:,j]$ for each column j. For example, if x is a single column, 1 for female, 0 for male, then with <code>expandx=TRUE</code> , x will be replaced by 3 columns, all 1's, female, and $1-\text{female}=\text{male}$ .
scale	scale determines how the observations are scaled in computing M-scores. scale must equal "closed" or "global" or "interaction". If you use the mean (or equivalently the total) as your test statistic (by setting <code>inner=0</code> and <code>trim=Inf</code> ), then scaling is not needed and the scale parameter is ignored. If <code>scale=global</code> , then all observations are used in computing a single scale factor. This is appropriate when testing Fisher's global hypothesis of no treatment effect at all. If <code>scale=interaction</code> , then the scale is determined separately in each of the unique groups formed from the interaction of all of the columns of x. This can be reasonable if the interaction groups are not too small. If x had a column for men, a column for women, a column for people over 50, a column for people under 50, then there would be 4 interaction groups, such as men under 50. If <code>scale=closed</code> , then a single scale factor is computed using every $y[i]$ such that $x[i,j]=1$ for at least one j. With <code>scale=closed</code> , the <code>score()</code> function will return matrices y and cmat that only contain the rows i such that $x[i,j]=1$ for at least one j. For instance, if your hypotheses are confined to women, then the men will be excluded. <code>scale=closed</code> is useful in closed testing, as described in Lee et al. (2017). If x contains a column of 1's, then <code>scale=closed</code> and <code>scale=global</code> are equivalent. Of course, x will contain a column of 1's if you set <code>expandx=TRUE</code> . You can use <code>scale=interaction</code> only if sets are exactly matched for all effect modifiers in x, but this is not required for <code>scale=global</code> or <code>scale=closed</code> . If you set <code>scale=interaction</code> when sets are not exactly matched, you will receive a warning and scale will be reset to the default of closed.
inner	inner and trim together define the $\psi$ -function for the M-statistic. The default values yield a version of Huber's $\psi$ -function, while setting <code>inner = 0</code> and <code>trim = Inf</code> uses the mean or total within each matched set. The $\psi$ -function is an odd function, so $\psi(w) = -\psi(-w)$ . For $w \geq 0$ , the $\psi$ -function is $\psi(w) = 0$ for $0 \leq w \leq \text{inner}$ , is $\psi(w) = \text{trim}$ for $w \geq \text{trim}$ , and rises linearly from 0 to trim for $\text{inner} < w < \text{trim}$ .  If uncertain about inner, trim and lambda, then use the defaults. An error will result unless $0 \leq \text{inner} \leq \text{trim}$ .
trim	inner and trim together define the $\psi$ -function for the M-statistic. See inner.
lambda	Before applying the $\psi$ -function to treated-minus-control differences, the differences are scaled by dividing by the lambda quantile of all within set absolute

differences. Typically,  $\lambda = 1/2$  for the median. The value of  $\lambda$  has no effect if  $\text{trim}=\text{Inf}$  and  $\text{inner}=0$ . See Maritz (1979) for the paired case and Rosenbaum (2007) for matched sets.

An error will result unless  $0 < \lambda < 1$ .

`xnames` If `xnames=NULL` and `x` is a matrix or `data.frame` with column names, then those names are used to label output. If `xnames` is not null, and `x` is a matrix, then `xnames` must be a vector of `dim(x)[2]` names, and these names are used to label output. If `x` is a vector, not a matrix, then the output will be easier to read if you give it a name, `xnames="a.name"`. This is particularly true if `x` is a vector and `expandx=TRUE`.

## Details

Taking  $\text{trim} < \text{Inf}$  limits the influence of outliers; see Huber (1981).

Taking  $\text{trim} < \text{Inf}$  and  $\text{inner} = 0$  uses Huber's  $\psi$  function.

Taking  $\text{trim} = \text{Inf}$  and  $\text{inner} = 0$  does no trimming and is similar to a mean or a weighted mean.

Taking  $\text{inner} > 0$  often increases design sensitivity; see Rosenbaum (2013). This is most evident with matched pairs, where  $\text{inner}=0.5$  may be a good choice.

An M-statistic similar to a trimmed mean is obtained by setting  $\text{inner}=0$ ,  $\text{trim}=1$ , and  $1-\lambda$  to the total amount to be trimmed from both tails. For example,  $\text{inner}=0$ ,  $\text{trim}=1$ ,  $\lambda=0.9$  trims 10 percent, perhaps 5 percent from each tail. Arguably,  $\text{inner}=0$ ,  $\text{trim}=1$ , and  $\lambda=0.99$  is very much like a mean, but also safer than a mean.

In general, each call to `submax()` tests a global null hypothesis of no treatment effect, but does this by looking in several subgroups defined by pretreatment covariates, that is, by potential effect modifiers. We often want to say something about the subgroups themselves, not the global null hypothesis. This is possible by using closed testing (Marcus et al. 1976), which may entail several calls to `submax()`. Before using closed testing, it is suggested that you read the section in Lee et al. (2017) discussing closed testing.

Matched sets of unequal size are weighted with a view to efficiency. See the documentation for `mscorev()` and the references mentioned there.

## Value

`y` A matrix of M-scores suitable for use as the `y` matrix in the `submax()` function.

`cmat` A matrix of comparisons suitable for use as the `cmat` matrix in the `submax` function.

`detail` A `data.frame` reminding you of the settings that produced the M-scores. It contains `inner`, `trim`, `lambda`, `scale`, `permutational.t`, and `anyinexact`, where `permutational.t=TRUE` if you set  $\text{inner}=0$  and  $\text{trim}=\text{Inf}$ , and `anyinexact=TRUE` if some of the potential effect modifiers are not exactly matched.

## Note

Several other packages use M-scores, so their examples and documentation may be helpful. These include `sensitivitymv`, `sensitivitymw`, `sensitivitymult`, `sensitivityfull`, and `senstrat`. In particular, `sensitivitymw` and `sensitivitymult` compute sensitivity analyses for confidence intervals.

If you have inexactly matched sets, but you want to use `scale=interaction`, then you must discard the inexactly matched sets before using `score()`.

### Author(s)

Paul R. Rosenbaum.

### References

- Huber, P. (1981) *Robust Statistics*. New York: John Wiley. (M-estimates based on M-statistics.)
- Lee, K., Small, D. S., & Rosenbaum, P. R. (2017). A new, powerful approach to the study of effect modification in observational studies. [arXiv:1702.00525](https://arxiv.org/abs/1702.00525).
- Marcus, R., Eric, P., & Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63, 655-660.
- Maritz, J. S. (1979). A note on exact robust confidence intervals for location. *Biometrika* 66 163–166. (Introduces exact permutation tests based on M-statistics by redefining the scaling parameter.)
- 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* 63 456-64. (R package `sensitivitymv`)
- 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. (2014). Weighted M-statistics with superior design sensitivity in matched observational studies with multiple controls. *J. Am. Statist. Assoc.* 109 1145-1158. (R package `sensitivitymw`) <doi:10.1080/01621459.2013.879261>
- Rosenbaum, P. R. (2015). Two R packages for sensitivity analysis in observational studies. *Observational Studies*, v. 1. (Free on-line.)

### Examples

```
data(mercury)
attach(mercury)
# The mercury data has two binary covariates, black and female,
# that will be considered as potential effect modifiers.
# Both black and female are not exactly matched. Of 397
# matched sets, 72 contain three blacks, 319 contain no
# blacks, 3 contain one black, and 3 contain 2 blacks.
table(table(mset,black)[,2])
# A similar situation arises with females.
table(table(mset,female)[,2])
# When considering females as an effect modifier, only
# sets exactly matched for female are used, etc. A
# set that is inexact for black may be used when looking
# at females, providing that set is exactly matched for female.

male<-1-female
nonblack<-1-black
```

```

everyone<-rep(1,dim(mercury)[1])
x<-cbind(everyone, female,male,black,nonblack)
sc<-score(methylmercury,fish,mset,x)

# At gamma=4, the global null of no effect is
# rejected at alpha=0.05 by every subgroup test
submax(sc$y,sc$cmat,gamma=4,fast=TRUE)

# What does expandx do?
sc<-score(methylmercury,fish,mset,cbind(female,black))
head(sc$cmat)
sc<-score(methylmercury,fish,mset,cbind(female,black),expandx=TRUE)
head(sc$cmat)

# Using exandx with a vector: remember to give it a name.
sc<-score(methylmercury,fish,mset,female,xnames="Female",expandx=TRUE)
head(sc$cmat)

## Not run:
# For closed testing, the process is repeated with fewer columns.
# In general, if cmat has L columns, closed testing may require
# up to (2^L)-1 tests. Here are two of those tests.
sc<-score(methylmercury,fish,mset,cbind(female,black))
submax(sc$y,sc$cmat,gamma=4)
# Note that the critical.constant has become smaller, making it
# easier to reject a component hypothesis when fewer hypotheses
# are tested.
sc<-score(methylmercury,fish,mset,female,xnames="Female")
submax(sc$y,sc$cmat,gamma=4,fast=TRUE)
# Use of closed testing is discussed in Lee et al. (2017).

# For a two-sided test, change alpha and do 2 tests.
submax(sc$y,sc$cmat,gamma=4,alpha=0.025,alternative = "greater")
submax(sc$y,sc$cmat,gamma=4,alpha=0.025,alternative = "less")
# So we reject in the positive direction in all 5 component tests.

## End(Not run)
detach(mercury)

```

---

separable1fc

*Computes the Separable Approximation.*


---

### Description

Of limited interest to most users, `separable1fc()` is called by the main function, `submax()`.

### Usage

```
separable1fc(yamat, gamma = 1)
```

**Arguments**

ymat	A matrix of scores produced by mscoref.
gamma	The sensitivity parameter $\Gamma \geq 1$ .

**Details**

See Gastwirth, Krieger and Rosenbaum (2000) and Rosenbaum (2007, section 4) for discussion of the separable approximation.

**Value**

tstat	Vector of length $I = \dim(\text{ymat})[1]$ giving the values of the test statistic in the $I$ matched sets.
expect	Vector of length $I$ giving the maximum expectations in the $I$ matched sets.
vari	Vector of length $I$ giving the maximum variances at the maximum expectations in the $I$ matched sets.

**Note**

This function is similar to the separable1f() function in the sensitivityfull package. Unlike that function, separable1fc() returns the  $I$  components for the  $I$  matched sets, rather than computing a summary statistic from them.

**Author(s)**

Paul R. Rosenbaum

**References**

- Gastwirth, J. L., Krieger, A. M. and Rosenbaum, P. R. (2000). Asymptotic separability in sensitivity analysis. *J. Roy. Statist. Soc. B.* 62 545-555. <doi:10.1111/1467-9868.00249>
- 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* 63 456-64. (See section 4.)
- 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.

**Examples**

```
# The following artificial example computes mscores for a
# full matching, then applies separable1fc() to
# perform a sensitivity analysis. Compare with
# the example below from the sensitivityfull package.

# The artificial example that follows has I=9
# matched sets. The first 3 sets have one treated
# individual and two controls with treated subjects
```

```

# in column 1. The next 3 sets are
# matched pairs, with treated subjects in column 1.
# The next 3 sets have one control and two treated
# subjects, with the control in column 1. Simulated
# from a Normal distribution with an additive effect
# of tau=1.

y<-c(2.2, 1.4, 1.6, 2.4, 0.7, 1.3, 1.2, 0.6, 0.3,
0.5, -0.1, -1.3, -0.3, 0.1, 0.4, 3.0, 1.1, 1.4, -0.8,
0.1, 0.8, NA, NA, NA, 1.1, 0.5, 1.8)
y<-matrix(y,9,3)
treated1<-c(rep(TRUE,6),rep(FALSE,3))

s<-separable1fc(sensitivityfull::mscoref(y,treated1),gamma=2)
1-pnorm((sum(s$tstat)-sum(s$expect))/sqrt(sum(s$vari)))
sensitivityfull::senfm(y,treated1,gamma=2)
s

```

submax

*Effect Modification Using the Submax Method in Observational Studies*

## Description

Effect modification means that the magnitude or stability of a treatment effect varies with observed covariates. When there is effect modification, causal conclusions may be less sensitive to unmeasured biases in a subgroup in which the treatment effect is larger or more stable. The submax or subgroup maximum method looks at an overall test and subgroup tests, correcting for multiple testing using the joint distribution of the tests. The submax method was proposed by Lee et al. (2018).

## Usage

```
submax(y, cmat, gamma = 1, alternative = "greater", alpha = 0.05, rnd = 2, fast=FALSE)
```

## Arguments

**y** In general, **y** is either a matrix with **I** rows for **I** matched sets or a vector of length **I** for **I** matched pairs. The **y** values are outcomes, perhaps after scoring (e.g., Huber-Maritz M-scores) or ranking (e.g., Wilcoxon) to produce a robust test. If **y** contains the outcomes themselves, then the outcomes are permuted in the manner of a permutational t-test.

More precisely, **y** contains the scores  $q[gi\ j]$  discussed in section 2.1 of Lee et al. (2018). If every matched set  $g_i$  is a pair, then **y** may be a vector of treated-minus-control pair differences,  $q[gi\ 1] - q[gi\ 2]$ .

If there are **I** matched pairs, then **y** can be either a vector of length **I** giving the **I** treated-minus-control pair differences, or **y** can be a 2-column matrix with treated responses in the first column and control responses in the second column.

If there are  $I$  matched sets and the largest matched set contains  $J$  individuals, then  $y$  is an  $I$  by  $J$  matrix with one row for each matched set. If matched set  $i$  contains one treated individual and  $k$  controls, where  $k$  is at least 1 and at most  $J-1$ , then  $y[i,1]$  is the treated individual's response,  $y[i,2], \dots, y[i,k+1]$  are the responses of the  $k$  controls, and  $y[i,k+2], \dots, y[i,J]$  are equal to NA.

Although  $y$  can contain NA's,  $y[i,1]$  and  $y[i,2]$  must not be NA for every  $i$ . That is, every matched set must have at least one treated subject and one control.

cmat	A matrix with one row for each matched set in $y$ . For each column of $cmat$ , a statistical test is done. Typically, the first column is $(1, \dots, 1)$ and refers to a test that uses all of the matched sets. If the second column is 1 for matched sets of women and 0 for matched sets of men, then the second test is restricted to matched sets of women.
gamma	gamma is the sensitivity parameter $\Gamma$ , where $\Gamma \geq 1$ . Setting $\Gamma = 1$ is equivalent to assuming ignorable treatment assignment given the matched sets, and it performs a within-set randomization test.
alternative	If <code>alternative="greater"</code> , the null hypothesis of no treatment effect is tested against the alternative of a treatment effect larger than zero. If <code>alternative="less"</code> , the null hypothesis of no treatment effect is tested against the alternative of a treatment effect smaller than 0. In particular, <code>alternative="less"</code> is equivalent to: (i) <code>alternative="greater"</code> , (ii) $y$ replaced by $-y$ . See the note for discussion of two-sided sensitivity analyses.
alpha	The global null hypothesis of no effect is tested at simultaneous level $\alpha$ in the presence of a bias of at most $\Gamma$ .
rnd	The correlation matrix of the <code>dim(cmat)[2]</code> test statistics is returned, rounded to <code>rnd</code> digits. The <code>critical.constant</code> is also rounded to <code>rnd</code> digits.
fast	Determines the speed and accuracy of the determination of the <code>critical.constant</code> used in testing. <code>fast=TRUE</code> is faster but less precise, less stable. <code>fast=FALSE</code> is slower but more precise, more stable. See details.

## Details

The submax procedure is developed by Lee et al. (2018), and the example reproduces analyses from that paper.

The `submax()` function rejects the null hypothesis at level  $\alpha$  in the presence of a bias in treatment assignment of at most  $\Gamma$  if `maxdeviate` is greater than or equal to `critical.constant`.

The global null hypothesis of no effect is tested at simultaneous level  $\alpha$  in the presence of a bias of at most  $\Gamma$ . If the global null is true, and the bias in treatment assignment is at most  $\Gamma$ , then the probability of falsely rejecting the global null hypothesis is at most  $\alpha$ . The test looks at the largest of `dim(cmat)[2]` standardized test statistics and corrects for multiple testing using their joint distribution. The joint distribution is approximated by a multivariate Normal distribution, so the entire procedure is a large sample approximation.

The function `score()` in this package may be helpful in creating the matrices  $y$  and  $cmat$  that are arguments of the `submax` function.

The sensitivity bound is based on the separable approximation described in Gastwirth, Krieger and Rosenbaum (2000); see also Rosenbaum (2007).

The `critical.constant` is determined rather precisely by a call to the `qmvnorm()` function in the `mvt-norm` package. The `qmvnorm()` function uses random numbers, but this should produce negligible variability in the `critical.constant`. However, if you call `submax()` twice, there will be a tiny change in the `critical.constant`. There is a trade-off between precision and speed, and you can alter that trade-off – make `submax` faster or more precise – by editing the call to `qmvnorm()` in the `rcode` for `submax()`. For almost all users, there will be no need to alter the code.

### Value

<code>maxdeviate</code>	The <code>submax</code> or subgroup maximum statistic. It is the maximum standardized deviate for the several columns of <code>cmat</code> . In Lee et al. (2018), this is $D[\Gamma_{max}]$ .
<code>critical.constant</code>	If <code>maxdeviate</code> $\geq$ <code>critical.constant</code> , the global null hypothesis of no treatment effect is rejected at level $\alpha$ in the presence of a bias of at most $\Gamma$ .
<code>deviates</code>	There is one standardized deviate for each column of <code>cmat</code> , and the maximum of these is <code>maxdeviate</code> . If <code>deviate[j] &gt; critical.constant</code> , then <code>deviate[j]</code> would lead to rejection of the global null hypothesis.
<code>correlation</code>	The correlation matrix of the deviates. By default, it is printed to two digits for easy viewing. By changing <code>rnd=2</code> , it can be produced with additional digits, perhaps for use in further computations.
<code>detail</code>	Reminds the user of the value of $\alpha$ and $\Gamma$ .

### Note

In general, each call to `submax()` tests a global null hypothesis of no treatment effect, but does this by looking in several subgroups defined by pretreatment covariates, that is, by potential effect modifiers. We often want to say something about the subgroups themselves, not the global null hypothesis. This is possible by using closed testing (Marcus et al. 1976), which may entail several calls to `submax()`. Before using closed testing, it is suggested that you read the section in Lee et al. (2018) discussing closed testing.

A 2-sided,  $\alpha$ -level test may be obtained by performing two one-sided tests, each at level  $\alpha/2$ . This is a safe approach, but for  $\Gamma > 1$  it is slightly conservative. Cox (1977) suggests that we view a two-sided test as two one-sided tests with a Bonferroni correction for testing two hypotheses.

### Author(s)

Paul R. Rosenbaum

### References

- Cox, D. R. (1977). The role of significance tests (with Discussion). *Scand. J. Statist.* 4, 49-70.
- Gastwirth, J. L., Krieger, A. M. and Rosenbaum, P. R. (2000). Asymptotic separability in sensitivity analysis. *J. Roy. Statist. Soc. B.* 62 545-555. <doi:10.1111/1467-9868.00249>
- Lee, K., Small, D. S., & Rosenbaum, P. R. (2018) <doi:10.1111/biom.12884> A powerful approach to the study of moderate effect modification in observational studies. *Biometrics*, 74(4), 1161-1170.

Marcus, R., Eric, P. and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63, 655-660.

Rosenbaum, P. R. (2007). Sensitivity analysis for m-estimates, tests and confidence intervals in matched observational studies. *Biometrics* 63 456-64. (See section 4.) <doi:10.1111/j.1541-0420.2006.00717.x>

### Examples

```
# Reproduces parts of Table 2 of Lee et al. (2018).
data(Active)
submax(Active$delta,Active[,1:7],gamma=1.77,alternative="less",fast=TRUE)
amplify(1.77,c(2,3,4))

# Reproduces the closed-testing analysis in
#Section 4 of Lee et al. (2018)
submax(Active$delta,Active[,c(3,5)],gamma=1.4,alternative="less",fast=TRUE)

# See also the examples for the score() function.
```

---

tbmetaphase

*Genetic damage from drugs used to treat TB*

---

### Description

This is a matched comparison of the effects of two drug sequences, namely HRZ and H2R2Z2, for the treatment of tuberculosis. HRZ is a higher dose sequence than H2R2Z2. The outcome is a measure of genetic damage, namely the frequency of aberrant metaphases two months after treatment. Individuals were matched for the frequency of aberrant metaphases before treatment. 15 individuals treated with HRZ are matched to 1 or 2 controls treated with H2R2Z2. Each row is one matched set. If a set is a pair, the third element in a row is NA. The data are originally from Rao, Gupta and Thomas (1991) and were used as an example in Rosenbaum (2007, Table 3). Data are used to illustrate the `senmv` function in the `sensitivitymv` package.

### Usage

```
data(tbmetaphase)
```

### Format

A data frame with 15 observations on the following 3 variables.

HRZ Aberrant metaphases for individual treated with HRZ.

H2R2Z2.1 Aberrant metaphases for first matched individual treated with H2R2Z2.

H2R2Z2.2 Aberrant metaphases for second matched individual treated with H2R2Z2. For matched pairs, this is NA.

**References**

Rao, V. V. N. G., Gupta, E. V. V and Thomas, I. M. Chromosomal aberrations in tuberculosis patients before and after treatment with short-term chemotherapy. *Mutation Research* 1991, 259, 13-19.

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

**Examples**

```
# The example reproduces Table 3 in Rosenbaum (2007).  
data(tbmetaphase)  
mscorev(tbmetaphase,trim=1)
```

# Index

- \* **Amplification**
  - amplify, 5
- \* **Amplify**
  - amplify, 5
- \* **Causal inference**
  - amplify, 5
  - submax, 15
- \* **Effect modification**
  - submax, 15
- \* **M-estimate**
  - score, 9
- \* **M-test**
  - score, 9
- \* **Matched pairs**
  - Active, 3
- \* **Matched study**
  - mercury, 7
- \* **Observational studies**
  - submax, 15
- \* **Observational study**
  - Active, 3
  - amplify, 5
  - mercury, 7
- \* **Paired survival data**
  - Active, 3
- \* **Sensitivity analysis**
  - amplify, 5
  - submax, 15
- \* **Simultaneous inference**
  - submax, 15
- \* **Subgroup analyses**
  - submax, 15
- \* **datasets**
  - Active, 3
  - mercury, 7
  - tbmetaphase, 18
- \* **htest**
  - score, 9
  - separable1fc, 13
  - submax, 15
- \* **package**
  - submax-package, 2
- \* **rbounds**
  - submax, 15
- \* **robust**
  - mscorev, 8
  - score, 9
  - separable1fc, 13
  - submax, 15
- Active, 3
- amplify, 5
- mercury, 7
- mscorev, 8
- score, 9
- separable1fc, 13
- submax, 15
- submax-package, 2
- tbmetaphase, 18