

# Package ‘PF’

July 2, 2014

**Type** Package

**Title** Prevented fraction.

**Version** 9.5

**Date** 2013-08-28

**Author** Dave Siev

**Maintainer** Marie Vendettuoli <marie.c.vendettuoli@aphis.usda.gov>

**Description** Functions related to PF (prevented fraction). Calculate incidence density ratio, confidence interval, and Rao-Scott weights of PF by multiple methods. See <http://goo.gl/eJ6Rxi> for definition of PF.

**License** MIT + file LICENSE

**LazyLoad** yes

**LazyData** yes

**Depends** methods, grid, xtable, RColorBrewer, gdata

**Collate** 'aaa.r' 'aab.r' 'class-rrl\_tables.r' 'class-rror\_tables.r' 'class-rrsc\_tables.r' 'class-rrsi\_tables.r' 'class-rrstr\_tables.r' 'class.r' 'generics.r' 'getPFTable\_generics.r' 'PF-package.r' 'IDRlsi.r' 'IDRsc.r' 'phiWt.r' 'RRlsi.r' 'RRmh.r' 'RRmpWald.r' 'RRor.r' 'RRotsst.r' 'RRsc.r' 'RRStr.r' 'RRtosst.r' 'rsbWt.r' 'tauWt.r'

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2013-09-04 07:52:25

**R topics documented:**

PF-package	2
bird	3
birdm	4
getPFTable	4
IDRlsi	5
IDRsc	7
New	8
pf	9
phiWt	9
print.rr1	11
rat	11
rr1	12
RRlsi	12
RRmh	14
rrmp	16
RRmpWald	17
RRor	19
rror	20
RRotsst	21
RRsc	23
rrsc	25
rrsi	25
RRstr	26
rrstr	28
RRtosst	29
rsb	31
rsbWt	32
Set1	33
set1	34
Table6	34
table6	34
tauWt	35
<b>Index</b>	<b>37</b>

---

 PF-package

*Package for PF.*


---

**Description**

Includes functions related to prevented fraction.

**Details**

Package: pf-package  
Type: Package  
Version: 9.5  
Date: 2013-08-29  
License: MIT  
LazyLoad: yes  
LazyData: yes

**Author(s)**

David Siev <David.Siev@aphis.usda.gov>

**Examples**

```
#-----  
# Checking PF package  
#-----  
example(RRsc)  
example(RRstr)  
example(RRmh)  
example(RRor)  
example(phiWt)  
example(tauWt)  
example(rsbWt)  
example(rsb)  
example(RRlsi)  
example(IDRsc)  
example(IDRlsi)  
#-----  
# The next two take a moment to run  
#-----  
example(RRtosst)  
example(RRotsst)  
#-----  
# End examples  
#-----  
invisible()
```

---

bird

*bird dataset*

---

**Description**

bird dataset

**Format**

a data.frame with 6 observations of the following 4 variables, no NAs

**y** number positive

**n** total number in group tx x all

**tx** treatment 'vac' or 'con'

**all** all?

**References**

we need some references

---

birdm	<i>birdm dataset</i>
-------	----------------------

---

**Description**

birdm dataset

**Format**

a data.frame with 6 observations of the following 4 variables, no NAs

**y** number positive

**n** total number in group tx x all

**tx** treatment 'vac' or 'con'

**all** all?

**References**

we need some references

---

getPFTable	<i>Write html and latex table from package PF objects</i>
------------	---

---

**Description**

Write html and latex table from package PF objects

**Usage**

```
getPFTable(x, type = "summary", out = "dev", file = "",
  ...)
```

**Arguments**

x	object of class rr1, rror, rrsc, rrstr, or rrsi
type	table type.
out	output type. dev for eye-readable to active device. tex for latex. texdoc for latex with preamble and footer. doc for word document via R2wd. html for html file
file	name to save file to. if file == '', display on active device.
...	Additional argument list

**Author(s)**

Marie Vendettuoli <marie.c.vendettuoli@aphis.usda.gov>

---

IDRlsi	<i>IDR likelihood support interval.</i>
--------	---

---

**Description**

Estimates likelihood support interval for the incidence density ratio or prevented fraction based on it.

**Usage**

```
IDRlsi(y, alpha = 0.05, k = 8, use.alpha = FALSE,
       pf = TRUE, converge = 1e-08, rnd = 3, start = NULL,
       trace.it = FALSE, iter.max = 24)
```

**Arguments**

y	Data vector c(y1, n1, y2, n2) where y are the positives, n are the total, and group 1 is compared to group 2.
k	Likelihood ratio criterion.
alpha	Complement of the confidence level.
use.alpha	Base choice of k on its relationship to alpha?
pf	Estimate <i>IDR</i> or its complement <i>PF</i> ?
trace.it	Verbose tracking of the iterations?
iter.max	Maximum number of iterations
converge	Convergence criterion
rnd	Number of digits for rounding. Affects display only, not estimates.
start	describe here.

**Details**

Estimates likelihood support interval for the incidence density ratio based on orthogonal factoring of reparameterized likelihood. The incidence density is the number of cases per subject-time; its distribution is assumed Poisson.

Likelihood support intervals are usually formed based on the desired likelihood ratio, often 1/8 or 1/32. Under some conditions the log likelihood ratio may follow the chi square distribution. If so, then  $\alpha = 1 - F(2\log(k), 1)$ , where  $F$  is a chi-square CDF. `RRsc()` will make the conversion from  $\alpha$  to  $k$  if `use.alpha = TRUE`.

The data may also be a matrix. In that case `y` would be entered as `matrix(c(y1, n1 - y1, y2, n2 - y2), 2, 2, byrow =`

**Value**

A `rrsi` object with the following elements.

<code>estimate</code>	vector with point and interval estimate
<code>estimator</code>	either <i>PF</i> or <i>IDR</i>
<code>y</code>	data vector
<code>k</code>	Likelihood ratio criterion
<code>rnd</code>	how many digits to round the display
<code>alpha</code>	complement of confidence level

**Note**

Level tested: Low.

**Author(s)**

David Siev <david.siev@aphis.usda.gov>

**References**

Royall R. *Statistical Evidence: A Likelihood Paradigm*. Chapman & Hall, Boca Raton, 1997. Section 7.2.

**See Also**

[IDRsc](#)

**Examples**

```
IDRlsi(c(26, 204, 10, 205), pf = FALSE)

# 1/8 likelihood support interval for IDR

# corresponds to 95.858% confidence
# (under certain assumptions)
```

```
# IDR
# IDR LL UL
# 2.61 1.26 5.88
```

---

IDRsc	<i>IDR confidence interval.</i>
-------	---------------------------------

---

### Description

Estimates confidence interval for the incidence density ratio or prevented fraction based on it.

### Usage

```
IDRsc(y, alpha = 0.05, pf = TRUE, rnd = 3)
```

### Arguments

y	Data vector c(y1, n1, y2, n2) where y are the positives, n are the total, and group 1 is compared to group 2.
alpha	Complement of the confidence level.
pf	Estimate <i>IDR</i> , or its complement <i>PF</i> ?
rnd	Number of digits for rounding. Affects display only, not estimates.

### Details

The incidence density is the number of cases per subject-time; its distribution is assumed Poisson. IDRsc estimates a confidence interval for the incidence density ratio using Siev's formula based on the Poisson score statistic.  $IDR = \widehat{IDR} \left\{ 1 + \left( \frac{1}{y_1} + \frac{1}{y_2} \right) \frac{z_{\alpha/2}^2}{2} \pm \frac{z_{\alpha/2}^2}{2y_1y_2} \sqrt{y_{\bullet} \left( y_{\bullet} z_{\alpha/2}^2 + 4y_1y_2 \right)} \right\}$

The data may also be a matrix. In that case y would be entered as matrix(c(y1, n1 - y1, y2, n2 - y2), 2, 2, byrow =

### Value

A `rr1` object with the following elements.

estimate	vector with point and interval estimate
estimator	either <i>PF</i> or <i>IDR</i>
y	data vector
rnd	how many digits to round the display
alpha	complement of confidence level

### Note

Level tested: High.

**Author(s)**

David Siev <david.siev@aphis.usda.gov>

**References**

Siev D, 1994. Estimating vaccine efficacy in prospective studies. *Preventive Veterinary Medicine* 20:279-296, Appendix 1.

Graham PL, Mengersen K, Morton AP, 2003. Confidence limits for the ratio of two rates based on likelihood scores:non-iterative method *Statistics in Medicine* 22:2071-2083.

Siev D, 2004. Letter to the editor. *Statistics in Medicine* 23:693. (Typographical error in formula: replace the two final minus signs with subscript dots.)

**See Also**

[IDR1si](#)

**Examples**

```
IDRsc(c(26, 204, 10, 205),pf = FALSE)

# IDR
# 95% interval estimates

# IDR  LL  UL
# 2.61 1.28 5.34
```

---

New

*New dataset*

---

**Description**

New dataset

**Format**

a data frame with 52 observations of the following 3 variables, no NAs

**cage** cage ID. 1 - 26

**tx** treatment. one of 'con' or 'vac'

**pos** numeric indicator of positive response. 0 = FALSE or 1 = TRUE

**References**

We need some references



---

pf *Data class pf*

---

### Description

Data class pf

### Fields

estimator either "PF" or "IDR"

rnd how many digits to round display

alpha complement of c.i.

### Author(s)

Marie Vendettuoli <marie.c.vendettuoli@aphis.usda.gov>

### See Also

[rr1](#), [rrsi](#), [rrsc](#), [rrstr](#)

---

phiWt *Binomial dispersion parameter.*

---

### Description

MME estimate of dispersion parameter phi.

### Usage

```
phiWt(fit, subset.factor = NULL, fit.only = TRUE,
      show.warns = FALSE)
```

### Arguments

fit	A <a href="#">glm</a> object.
subset.factor	Factor for estimating phi by subset.
fit.only	Return only the new fit? If FALSE, also returns the weights and phi estimates.
show.warns	Show warnings

### Details

Estimates binomial dispersion parameter  $\phi$  by the method of moments. Refits the model, weighting the observations by  $1/\phi$ . Uses quasibinomial family in `glm()`.

**Value**

A list with the following elements.

fit	the new model fit, updated by the estimated weights
weights	vector of weights
phi	vector of phi estimates

**Note**

Level tested: Moderate. (Extensive in S+, not as much in R. They have different glm() functions.)

**Author(s)**

David Siev <david.siev@aphis.usda.gov>

**References**

Wedderburn RWM, 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61:439-447.

**See Also**

[tauWt](#), [RRor](#). See the package vignette for more examples.

**Examples**

```
birdm.fit <- glm(cbind(y, n - y)~tx-1, binomial, birdm)
RRor(phiWt(birdm.fit))
#
# 95% t intervals on 4 df
#
# PF
#   PF      LL      UL
# 0.479 -0.537  0.823
#
#      mu.hat  LL      UL
# txcon  0.768 0.95 0.367
# txvac  0.400 0.78 0.111
#
# See the package vignette for more examples
```

---

print.rr1	<i>Print values for PF data obhects.</i>
-----------	--

---

**Description**

Print values for PF data obhects.

**Usage**

```
## S3 method for class 'rr1'
print(x, ...)

## S3 method for class 'rror'
print(x, ...)

## S3 method for class 'rrsi'
print(x, ...)

## S3 method for class 'rrmp'
print(x, ...)

## S3 method for class 'rrsc'
print(x, ...)

## S3 method for class 'rrstr'
print(x, ...)
```

**Arguments**

x	object of class rr1, rror, rrsi, rrmp, rrstr, rrsc
...	other arguments not used by this method

---

rat	<i>rat dataset</i>
-----	--------------------

---

**Description**

rat dataset

**Format**

a data.frame with 32 observations of the following 3 variables, no NAs

**y** number positive

**n** total number

**group** treatment group: 'control' or 'treated'

**References**

Weil's rat data (Table 1 of Rao and Scott)

---

rr1	<i>Data class rr1</i>
-----	-----------------------

---

**Description**

Data class rr1

**Fields**

estimate vector with point and interval estimate  
 estimator either "PF" or "IDR"  
 y data vector  
 rnd how many digits to round display  
 alpha complement of c.i.

**Author(s)**

Marie Vendettuoli <marie.c.vendettuoli@aphis.usda.gov>

**See Also**

[IDRsc](#), [RRotsst](#), [RRtosst](#)

---

RRlsi	<i>RR likelihood support interval.</i>
-------	--

---

**Description**

likelihood support interval for the risk ratio or prevented fraction by the likelihood profile.

**Usage**

```
RRlsi(y, alpha = 0.05, k = 8, use.alpha = FALSE,
      pf = TRUE, iter.max = 50, converge = 1e-06, rnd = 3,
      start = NULL, track = FALSE, full.track = FALSE)
```

**Arguments**

y	Data vector $c(y1, n1, y2, n2)$ where y are the positives, n are the total, and group 1 is compared to group 2.
k	Likelihood ratio criterion.
alpha	Complement of the confidence level (see details).
use.alpha	Base choice of k on its relationship to alpha?
pf	Estimate <i>RR</i> or its complement <i>PF</i> ?
iter.max	Maximum number of iterations
converge	Convergence criterion
rnd	Number of digits for rounding. Affects display onlyRR, not estimates.
start	Optional starting value.
track	Verbose tracking of the iterations?
full.track	Verbose tracking of the iterations?

**Details**

Estimates a likelihood support interval for *RR* or *PF* by the profile likelihood method using the DUD algorithm.

Likelihood support intervals are usually formed based on the desired likelihood ratio, often 1/8 or 1/32. Under some conditions the log likelihood ratio may follow the chi square distribution. If so, then  $\alpha = 1 - F(2\log(k), 1)$ , where  $F$  is a chi-square CDF. `RRlsi()` will make the conversion from  $\alpha$  to  $k$  if `use.alpha = TRUE`.

The data may also be a matrix. In that case y would be entered as `matrix(c(y1, n1-y1, y2, n2-y2), 2, 2, byrow = TRUE)`

**Value**

An object of class `rrsi` with the following fields:

estimate	matrix of point and interval estimates - see details
estimator	either "PF" or "RR"
y	data vector
rnd	how many digits to round the display
k	likelihood ratio criterion
alpha	complement of confidence level

**Note**

Level tested: High.

**Author(s)**

David Siev <david.siev@aphis.usda.gov>

## References

- Royall R. *Statistical Evidence: A Likelihood Paradigm*. Chapman & Hall, Boca Raton, 1997. Section 7.6
- Ralston ML, Jennrich RI, 1978. DUD, A Derivative-Free Algorithm for Nonlinear Least Squares. *Technometrics* 20:7-14.

## Examples

```
RRlsi(c(4,24,12,28))

# 1/8 likelihood support interval for PF

# corresponds to 95.858% confidence
# (under certain assumptions)

# PF
#   PF   LL   UL
# 0.6111 0.0168 0.8859
```

---

RRmh	<i>Mantel-Haenszel method, CI for common RR over strata or clusters with sparse data.</i>
------	---

---

## Description

Estimates confidence intervals for the risk ratio or prevented fraction from clustered or stratified data, using a Mantel-Haenszel estimator for sparse data.

## Usage

```
RRmh(formula = NULL, data = NULL, compare = c("b", "a"),
      Y, alpha = 0.05, pf = TRUE, rnd = 3)
```

## Arguments

formula	Formula of the form $\text{cbind}(y, n) \sim x + \text{cluster}(w)$ , where $y$ is the number positive, $n$ is the group size, $x$ is a factor with two levels of treatment, and $w$ is a factor indicating the clusters.
data	data.frame containing variables for formula
compare	Text vector stating the factor levels: <code>compare[1]</code> is the control or reference group to which <code>compare[2]</code> is compared
Y	Matrix of data, $K \times 4$ . Each row is a stratum or cluster. The columns are $y_2, n_2, y_1, n_1$ , where the $y$ 's are the number of positive in each group, and the $n$ is the total in each group. If data entered by formula and dataframe, $Y$ is generated automatically.
pf	Estimate <i>RR</i> or its complement <i>PF</i> ?
alpha	Complement of the confidence level.
rnd	Number of digits for rounding. Affects display only, not estimates.

## Details

Based on the Mantel-Haenszel (1959) procedure for sparse data developed by Greenland and Robins (1985). The confidence limits are based on asymptotic normality of the log(risk ratio). Agresti and Hartzel (2000) favor this procedure for small, sparse data sets, but they warn that it is less efficient than maximum likelihood for large data sets.

## Value

An object of class `rr1` with the following fields.

<code>estimate</code>	vector of point and interval estimates: point estimate, lower confidence limit, upper confidence limit
<code>estimator</code>	either "PF" or "RR"
<code>y</code>	y matrix of the data
<code>compare</code>	groups compared
<code>rnd</code>	how many digits to round the display
<code>alpha</code>	complement of confidence level

## Note

If either all y1's or all y2's are zero, a division by zero may occur, and a NaN returned for some values.

Level tested: Low.

See vignette *Examples for Stratified Designs* for more examples (enter ?PF).

Call to this function may be one of two formats: (1) specify data and formula or (2) as a matrix Y

```
RRmh(formula, data, compare = c('b', 'a'), pf = TRUE, alpha = 0.05, rnd = 3)
```

```
RRmh(Y, pf = TRUE, alpha = 0.05, rnd = 3)
```

## Author(s)

Christopher Tong <Christopher.H.Tong@aphis.usda.gov>

## References

Mantel N, Haenszel W, 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 22: 719-748.

Greenland S, Robins JM, 1985. Estimation of a common effect parameter from sparse follow-up data. *Biometrics* 41: 55-68. Errata, 45: 1323-1324.

Agresti A, Hartzel J, 2000. Strategies for comparing treatments on a binary response with multi-centre data. *Statistics in Medicine* 19: 1115-1139.

Lachin JM, 2000. *Biostatistical Methods: The Assessment of Relative Risks* (Wiley, New York), Sec. 4.3.1.

**See Also**[rr1](#)**Examples**

```
## Table 1 from Gart (1985)
## as data frame
RRmh(cbind(y,n) ~ tx + cluster(clus), Table6 , pf = FALSE)

# RR estimates

# RR
# 95% interval estimates
#
# RR LL UL
# 2.67 1.37 5.23
#

## or as matrix
RRmh(Y = table6, pf = FALSE)
```

---

`rrmp`*Data class rrmp*

---

**Description**

Data class rrmp

**Fields**

estimate vector with point and interval estimate  
 estimator either "PF" or "IDR"  
 y data vector  
 rnd how many digits to round display  
 alpha complement of c.i.  
 xtable 2 x 2 data matrix  
 compare text vector, same as input  
 frecvec data arrayed a vector of length 4  
 multvec data.frame showing the multinomial representation of the data

**Author(s)**

Marie Vendettuoli &lt;marie.c.vendettuoli@aphis.usda.gov&gt;

**See Also**[RRmpWald](#)



RRmpWald

Wald confidence intervals for RR from matched pairs

**Description**

Estimates confidence intervals for the risk ratio or prevented fraction from matched pairs.

**Usage**

```
RRmpWald(formula = NULL, data = NULL,
  compare = c("con", "vac"), affected = 1, x,
  alpha = 0.05, pf = TRUE, tdist = TRUE, df = NULL,
  rnd = 3)
```

**Arguments**

formula	Formula of the form $y \sim x + \text{cluster}(w)$ , where $y$ is the indicator for an individual's positive response, $x$ is a factor with two levels of treatment, and $w$ identifies the pairs.
data	data.frame containing variables in formula
compare	Text vector stating the factor levels: compare[1] is the control or reference group to which compare[2] is compared
affected	Indicator for positive response
x	Alternative data input. Instead of formula and data frame, data may be input as vector or 2x2 table
alpha	Complement of the confidence level
pf	Estimate <i>RR</i> or its complement <i>PF</i> ?
tdist	Use t distribution?
df	Degrees of freedom. When NULL, the function will default to $df = N - 2$ , where $N$ is the total number of pairs.
rnd	Number of digits for rounding. Affects display only, not estimates.

**Details**

Estimates confidence intervals for the risk ratio or prevented fraction from matched pairs. The response is the tetranomial vector [11, 12, 21, 22], where the first index is the row and the the second index is the column when displayed as a 2x2 table. Wald type confidence intervals are found by applying the delta method to the multinomial variance. This method fails when there are no responders in one of the treatment groups.

Alternative forms of data entry are illustrated by the output, say  $Y$ , where  $c(Y\$xtable) = Y\$freqvec = Y\$multvec\$Freq$ .

If  $RR = 0$  ( $PF = 1$ ), the function will return degenerate interval.

**Value**

A `rrmp` object with the following fields:

<code>estimate</code>	vector of point and interval estimates - see details
<code>estimator</code>	either "PF" or "RR"
<code>compare</code>	text vector, same as input
<code>alpha</code>	complement of confidence level
<code>rnd</code>	how many digits to round the display
<code>xtable</code>	data arrayed in 2x2 matrix
<code>freqvec</code>	data arrayed a 4-vector
<code>multvec</code>	data frame showing the multinomial representation of the data

**Note**

Level tested: Low.

Experimental functions for estimating profile likelihood intervals are in the `CVBmisc` package.

Call to this function may be one of two formats: (1) specify data and formula or (2) as a vector or 2 x 2 table `x`

```
RRmpWald(formula, data, compare = c('con', 'vac'), affected = 1, alpha = 0.05, pf = TRUE, tdist = TRUE, df = NULL, rnd = 3)
```

```
RRmpWald(x, compare = c('con', 'vac'), affected = 1, alpha = 0.05, pf = TRUE, tdist = TRUE, df = NULL, rnd = 3)
```

**Author(s)**

David Siev <david.siev@aphis.usda.gov>

**Examples**

```
RRmpWald(pos ~ tx + cluster(cage), New, compare = c('con', 'vac'))
```

```
# PF
# 95% interval estimates
#
# PF    LL    UL
# 0.550 0.183 0.752
```

```
RRmpWald(x = c(2, 9, 1, 6))
```

```
# PF
# 95% interval estimates
#
# PF    LL    UL
# 0.727 0.124 0.915
```

---

RRor	<i>RR estimate from logistic regression.</i>
------	--

---

### Description

Model based interval estimate of the risk ratio or prevented fraction from a logistic regression model.

### Usage

```
RRor(fit = NULL, beta.hat = NULL, var.beta.hat = NULL,
     degf = NULL, which = c(1, 2), pf = T, norm = F,
     alpha = 0.05, rnd = 3)
```

### Arguments

fit	A <code>glm</code> object.
beta.hat	Parameters estimates from a logistic regression with no intercept.
var.beta.hat	Variance-covariance matrix from a logistic regression with no intercept.
degf	Degrees of freedom.
which	Numeric vector indicating which parameters to compare, so that $RR = compare[2]/compare[1]$
pf	Estimate $RR$ or its complement $PF$ ?
norm	Estimate confidence interval using quantiles of Gaussian rather than t distribution quantiles?
alpha	Complement of the confidence level.
rnd	Number of digits for rounding. Affects display only, not estimates.

### Details

Estimates confidence intervals using the delta method on parameters from a generalized linear model with logit link.

$RR = \mu_2/\mu_1$ , where  $\mu_i$  are the estimated probabilities from the model.

### Value

A `rror` object with the following fields.

estimate	vector with point and interval estimate
estimator	either $PF$ or $RR$
mu	matrix with rows giving probability estimates for each of the groups
rnd	how many digits to round the display
alpha	complement of confidence level
norm	logical indicating Gaussian or t interval
degf	degrees of freedom

**Note**

Level tested: Moderate. (Extensive in S+, not as much in R. They have different glm() functions.)

Call to this function may be one of two formats: (1) specify fit or (2) beta.hat, var.beta.hat, degf

```
RRor(fit, degf = NULL, pf = TRUE, alpha = 0.05, which = c(1, 2), norm = TRUE, rnd = 3)
```

```
RRor(beta.hat, var.beta.hat, degf, pf = TRUE, alpha = 0.05, which = c(1, 2), norm = TRUE, rnd = 3)
```

**Author(s)**

David Siev <david.siev@aphis.usda.gov>

**References**

No references yet.

**See Also**

[rror](#), [phiWt](#), [tauWt](#). See the package vignette for more examples.

**Examples**

```
bird.fit <- glm(cbind(y, n - y) ~ tx - 1, binomial, bird)
RRor(tauWt(bird.fit))

# 95% t intervals on 4 df
#
# PF
#   PF    LL    UL
# 0.500 -0.583 0.842
#
#   mu.hat    LL    UL
# txcon 0.733 0.943 0.3121
# txvac 0.367 0.752 0.0997
#
# See the package vignette for more examples
```

---

rror

*Data class rror*

---

**Description**

Data class rror

**Fields**

estimate vector with point and interval estimate  
 estimator either "PF" or "IDR"  
 y data vector  
 rnd how many digits to round display  
 alpha complement of c.i.  
 norm logical indicating Gaussian or t interval  
 degf degrees of freedom  
 mu matrix with rows giving probability estimates for each of the groups

**Author(s)**

Marie Vendettuoli <marie.c.vendettuoli@aphis.usda.gov>

**See Also**

[RRor](#)

---

 RRotsst

---

*RR exact CI, OTSST method.*


---

**Description**

Estimates confidence interval for the risk ratio or prevented fraction, exact method based on the score statistic (inverts one two-sided test).

**Usage**

```
RRotsst(y, alpha = 0.05, pf = TRUE, stepstart = 0.1,
        iter.max = 36, converge = 1e-06, rnd = 3,
        trace.it = FALSE, nuisance.points = 120, gamma = 1e-06)
```

**Arguments**

y	Data vector c(y1, n1, y2, n2) where y are the positives, n are the total, and group 1 is compared to group 2.
alpha	Complement of the confidence level.
pf	Estimate <i>RR</i> or its complement <i>PF</i> ?
trace.it	Verbose tracking of the iterations?
iter.max	Maximum number of iterations
converge	Convergence criterion
rnd	Number of digits for rounding. Affects display only, not estimates.
stepstart	starting interval for step search

nuisance.points	number of points over which to evaluate nuisance parameter
gamma	parameter for Berger-Boos correction (restricts range of nuisance parameter evaluation)

### Details

Estimates confidence intervals based on the score statistic that are 'exact' in the sense of accounting for discreteness. The score statistic is used to select tail area tables, and the binomial probability is estimated over the tail area by taking the maximum over the nuisance parameter. Algorithm is a simple step search.

The data may also be a matrix. In that case `y` would be entered as `matrix(c(y1, n1 - y1, y2, n2 - y2), 2, 2, byrow =`

### Value

An object of class `rr1` with the following fields:

estimate	vector with point and interval estimate
estimator	either "PF" or "RR"
y	data vector
rnd	how many digits to round the display
alpha	complement of confidence level

### Note

Level tested: Moderate.

### Author(s)

David Siev <david.siev@aphis.usda.gov>

### References

- Koopman PAR, 1984. Confidence intervals for the ratio of two binomial proportions. *Biometrics* 40:513-517.
- Agresti A, Min Y, 2001. On small-sample confidence intervals for parameters in discrete distribution. *Biometrics* 57: 963-971.
- Berger RL, Boos DD, 1994. P values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association* 89:214-220.

### See Also

[RRtosst](#), [rr1](#).

**Examples**

```
## Not run: RRotsst(c(4, 24, 12, 28), rnd = 3)

# PF
# 95% interval estimates

#   PF   LL   UL
# 0.6111 0.0148 0.8519
## End(Not run)
```

---

RRsc	<i>RR score based asymptotic CI.</i>
------	--------------------------------------

---

**Description**

Estimates confidence intervals for the risk ratio or prevented fraction based on the score statistic.

**Usage**

```
RRsc(y = NULL, alpha = 0.05, pf = TRUE, trace.it = FALSE,
      iter.max = 18, converge = 1e-06, rnd = 3)
```

**Arguments**

<code>y</code>	Data vector <code>c(y1, n1, y2, n2)</code> where <code>y</code> are the positives, <code>n</code> are the total, and group 1 is compared to group 2.
<code>alpha</code>	Complement of the confidence level.
<code>pf</code>	Estimate <i>RR</i> or its complement <i>PF</i> ?
<code>trace.it</code>	Verbose tracking of the iterations?
<code>iter.max</code>	Maximum number of iterations
<code>converge</code>	Convergence criterion
<code>rnd</code>	Number of digits for rounding. Affects display only, not estimates.

**Details**

Estimates are returned for three estimators based on the score statistic. The score method was introduced by Koopman (1984). Gart and Nam's modification (1988) includes a skewness correction. The method of Miettinen and Nurminen (1985) is a version made slightly more conservative than Koopman's by including a factor of  $(N-1)/N$ . The starting estimate for the DUD algorithm is obtained by the modified Katz method (log method with 0.5 added to each cell). Both forms of the Katz estimate may be retrieved from the returned object using `RRsc()$estimate`.

The data may also be a matrix. In that case `y` would be entered as `matrix(c(y1, n1-y1, y2, n2-y2), 2, 2, byrow = TRUE)`.

**Value**

A `rrsc` object with the following fields.

<code>estimate</code>	matrix of point and interval estimates - see details
<code>estimator</code>	either "PF" or "RR"
<code>y</code>	data vector
<code>rnd</code>	how many digits to round the display
<code>alpha</code>	complement of confidence level

**Note**

Level tested: High.

**Author(s)**

David Siev <david.siev@aphis.usda.gov>

**References**

Gart JJ, Nam J, 1988. Approximate interval estimation of the ratio of binomial parameters: a review and corrections for skewness. *Biometrics* 44:323-338.

Koopman PAR, 1984. Confidence intervals for the ratio of two binomial proportions. *Biometrics* 40:513-517.

Miettinen O, Nurminen M, 1985. Comparative analysis of two rates. *Statistics in Medicine* 4:213-226.

Ralston ML, Jennrich RI, 1978. DUD, A Derivative-Free Algorithm for Nonlinear Least Squares. *Technometrics* 20:7-14.

**See Also**

[rrsc](#)

**Examples**

```
RRsc(c(4, 24, 12, 28))

# PF
# 95% interval estimates

# PF    LL    UL
# MN method  0.611 0.0251 0.857
# score method 0.611 0.0328 0.855
# skew corr   0.611 0.0380 0.876
```



---

rrsc	<i>Data class rrsc</i>
------	------------------------

---

**Description**

Data class rrsc

**Fields**

estimate vector with point and interval estimate  
 rnd how many digits to round display  
 alpha complement of c.i.  
 estimator either "PF" or "RR"  
 y data vector

**Author(s)**

Marie Vendettuoli <marie.c.vendettuoli@aphis.usda.gov>

**See Also**

[RRsc](#)

---

rrsi	<i>Data class rrsi</i>
------	------------------------

---

**Description**

Data class rrsi

**Fields**

y numeric data vector  
 k likelihood ratio criterion  
 rnd digits to round display  
 alpha complement of c.i.  
 estimate vector with point and interval estimate  
 estimator either "PF" or "IDR"

**Author(s)**

Marie Vendettuoli <marie.c.vendettuoli@aphis.usda.gov>

**See Also**

[IDRlsi](#), [RRlsi](#)

RRstr

*Gart-Nam method, CI for common RR over strata or clusters.***Description**

Estimates confidence intervals for the risk ratio or prevented fraction from clustered or stratified data.

**Usage**

```
RRstr(formula = NULL, data = NULL, compare = c("b", "a"),
      Y, alpha = 0.05, pf = TRUE, trace.it = FALSE,
      iter.max = 24, converge = 1e-06, rnd = 3,
      multiplier = 0.7, divider = 1.1)
```

**Arguments**

formula	Formula of the form <code>cbind(y, n) ~ x + cluster(w)</code> , where <code>y</code> is the number positive, <code>n</code> is the group size, <code>x</code> is a factor with two levels of treatment, and <code>w</code> is a factor indicating the clusters.
data	data.frame containing variables of formula
compare	Text vector stating the factor levels: <code>compare[1]</code> is the control or reference group to which <code>compare[2]</code> is compared
Y	Matrix of data. Each row is a stratum or cluster. The columns are <code>y2, n2, y1, n1</code> . If data entered by formula and dataframe, <code>Y</code> is generated automatically.
pf	Estimate <i>RR</i> or its complement <i>PF</i> ?
alpha	Size of the homogeneity test and complement of the confidence level.
trace.it	verbose tracking of the iterations?
iter.max	Maximum number of iterations
converge	Convergence criterion
rnd	Number of digits for rounding. Affects display only, not estimates.
multiplier	internal control parameter for algorithm
divider	internal control parameter for algorithm

**Details**

Uses the DUD algorithm to estimate confidence intervals by the method of Gart.

**Value**

A `rrstr` object with the following fields:

<code>estimate</code>	matrix of point and interval estimates - starting value, MLE, and skewness corrected
<code>hom</code>	list of homogeneity statistic, p-value, and degrees of freedom
<code>estimator</code>	either "PF" or "RR"
<code>y</code>	Y matrix of the data
<code>compare</code>	groups compared
<code>rnd</code>	how many digits to round the display
<code>alpha</code>	size of test; complement of confidence level

**Note**

See vignette *Examples for Stratified Designs* for more examples (enter ?PF).

Level tested: High.

Call to this function may be one of two formats: (1) specify data and formula or (2) as a matrix Y

```
RRstr(formula, data, compare = c('b','a'), pf = TRUE, alpha = 0.05, trace.it = FALSE, iter.max = 24, 1.1)
```

```
RRstr(Y, compare = c('b','a'), pf = TRUE, alpha = 0.05, trace.it = FALSE, iter.max = 24, converge = 1.1)
```

**Author(s)**

David Siev <david.siev@aphis.usda.gov>

**References**

- Gart JJ, 1985. Approximate tests and interval estimation of the common relative risk in the combination of  $2 \times 2$  tables. *Biometrika* 72:673-677.
- Gart JJ, Nam J, 1988. Approximate interval estimation of the ratio of binomial parameters: a review and corrections for skewness. *Biometrics* 44:323-338.
- Ralston ML, Jennrich RI, 1978. DUD, A Derivative-Free Algorithm for Nonlinear Least Squares. *Technometrics* 20:7-14.

**See Also**

[rrstr](#)

**Examples**

```
## Table 1 from Gart (1985)
## as data frame
RRstr(cbind(y,n) ~ tx + cluster(clus), Table6 , pf = FALSE)

# Test of homogeneity across clusters

# stat      0.954
# df         3
# p          0.812

# RR estimates

#          RR  LL  UL
# starting  2.66 1.37 5.18
# mle       2.65 1.39 5.03
# skew corr  2.65 1.31 5.08

## or as matrix
RRstr(Y = table6, pf = FALSE)
```

---

rrstr

*Data class rrstr*


---

**Description**

Data class rrstr

**Fields**

estimate vector with point and interval estimate  
 rnd how many digits to round display  
 alpha complement of c.i.  
 estimator either "PF" or "RR"  
 hom list of homogeneity statistic, p-value, and degrees of freedom  
 y matrix of data  
 compare groups compared

**Author(s)**

Marie Vendettuoli <marie.c.vendettuoli@aphis.usda.gov>

**See Also**

[RRstr](#)

---

RRtosst *RR exact CI, TOSST method.*

---

### Description

Estimates confidence interval for the risk ratio or prevented fraction; exact method based on the score statistic (inverts two one-sided tests).

### Usage

```
RRtosst(y, alpha = 0.05, pf = TRUE, stepstart = 0.1,
        iter.max = 36, converge = 1e-06, rnd = 3,
        trace.it = FALSE, nuisance.points = 120, gamma = 1e-06)
```

### Arguments

y	Data vector c(y1, n1, y2, n2) where y are the positives, n are the total, and group 1 is compared to group 2.
alpha	Complement of the confidence level.
pf	Estimate <i>RR</i> or its complement <i>PF</i> ?
trace.it	Verbose tracking of the iterations?
iter.max	Maximum number of iterations
converge	Convergence criterion
rnd	Number of digits for rounding. Affects display only, not estimates.
stepstart	starting interval for step search
nuisance.points	number of points over which to evaluate nuisance parameter
gamma	parameter for Berger-Boos correction (restricts range of nuisance parameter evaluation)

### Details

Estimates confidence intervals based on the score statistic that are 'exact' in the sense of accounting for discreteness. Inverts two one-sided score tests. The score statistic is used to select tail area tables, and the binomial probability is estimated over the tail area by taking the maximum over the nuisance parameter. Algorithm is a simple step search.

The data may also be a matrix. In that case y would be entered as `matrix(c(y1, n1-y1, y2, n2-y2), 2, 2, byrow = TRUE)`.

**Value**

A `rr1` object with the following fields.

<code>estimate</code>	vector with point and interval estimate
<code>estimator</code>	either "PF" or "RR"
<code>y</code>	data vector
<code>rnd</code>	how many digits to round the display
<code>alpha</code>	complement of confidence level

**Note**

Level tested: Moderate.

**Author(s)**

David Siev <david.siev@aphis.usda.gov>

**References**

- Koopman PAR, 1984. Confidence intervals for the ratio of two binomial proportions. *Biometrics* 40:513-517.
- Agresti A, Min Y, 2001. On small-sample confidence intervals for parameters in discrete distribution. *Biometrics* 57: 963-971.
- Berger RL, Boos DD, 1994. P values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association* 89:214-220.

**See Also**

[RRotsst](#), [rr1](#)

**Examples**

```
## Not run: RRtosst(c(4, 24, 12, 28))

# PF
# 95% interval estimates

#   PF   LL   UL
# 0.611 0.012 0.902
## End(Not run)
```

---

rsb	<i>Rao-Scott weights.</i>
-----	---------------------------

---

**Description**

Rao-Scott weights.

**Usage**

```
rsb(y, n, id = NULL)
```

**Arguments**

y	Number positive.
n	Total number.
id	Factor for estimating the weights by subset.

**Details**

Estimates the cluster design effect  $d_i$  as the variance inflation due to clustering by the method of Rao and Scott. `rsb` estimates the  $d_i$  for use by `rsbWt` or other functions.

**Value**

A list with the following elements.

w	vector of weights
d	vector of $d_i$ estimates

**Note**

Level tested: Moderate.

**Author(s)**

David Siev <david.siev@aphis.usda.gov>

**References**

Rao JNK, Scott AJ, 1992. A simple method for the analysis of clustered binary data. *Biometrics* 48:577-585.

**See Also**

[rsbWt](#). See the package vignette for more examples.

**Examples**

```
# Weil's rat data (Table 1 of Rao and Scott)
rsb(rat$y, rat$n, rat$group)$d
# control treated
# 1.232495 3.952861
```

---

rsbWt	<i>Rao-Scott weighting.</i>
-------	-----------------------------

---

**Description**

Rao-Scott weighting of clustered binomial observations.

**Usage**

```
rsbWt(fit = NULL, subset.factor = NULL, fit.only = T)
```

**Arguments**

fit	A <code>glm</code> object.
subset.factor	Factor for estimating the weights by subset.
fit.only	Return only the new fit? If FALSE, also returns the weights and phi estimates.

**Details**

Estimates the cluster design effect  $d_i$  as the variance inflation due to clustering by the method of Rao and Scott. Observations are then weighted by the inverse of the  $d_i$ .

**Value**

A list with the following elements.

fit	the new model fit, updated by the estimated weights
weights	vector of weights
d	vector of $d_i$ estimates

**Note**

Level tested: Moderate.

**Author(s)**

David Siev <david.siev@aphis.usda.gov>

**References**

Rao JNK, Scott AJ, 1992. A simple method for the analysis of clustered binary data. *Biometrics* 48:577-585.



**See Also**

[RRor](#), [rsb](#). See the package vignette for more examples.

**Examples**

```
birdm.fit <- glm(cbind(y,n-y)~tx-1,binomial,birdm)
RRor(rsbWt(birdm.fit))
#
# 95% t intervals on 4 df
#
# PF
#   PF   LL   UL
# 0.479 -1.061 0.868
#
#   mu.hat   LL   UL
# txcon 0.768 0.968 0.2659
# txvac 0.400 0.848 0.0737
#
# See the package vignette for more examples
```

---

Set1

*Set1 dataset*

---

**Description**

Set1 dataset

**Format**

a data.frame with 6 observation of the following 4 variables, no NAs

**y** number positive

**n** total number in group tx x clus

**tx** treatment 'vac' or 'con'

**clus** cluster ID

**References**

We need some references

---

set1	<i>set1 dataset</i>
------	---------------------

---

**Description**

set1 dataset

**Format**

a 3 x 4 matrix of data in [Set1](#)

**References**

we need some references!

---

Table6	<i>Table6 dataset</i>
--------	-----------------------

---

**Description**

Table6 dataset

**Format**

a data.frame with 8 observations of the following 4 variables, no NAs

**y** number positive

**n** total number in group tx x clus

**tx** treatment 'a' or 'b'

**clus** cluster ID

**References**

Table 1 from Gart (1985)

---

table6	<i>table6 dataset</i>
--------	-----------------------

---

**Description**

table6 dataset

**Format**

matrix for of data in [Table6](#)

---

tauWt	<i>Binomial dispersion: intra-cluster correlation parameter.</i>
-------	--

---

**Description**

MME estimates of binomial dispersion parameter tau (intra-cluster correlation).

**Usage**

```
tauWt(fit, subset.factor = NULL, fit.only = TRUE,
      iter.max = 12, converge = 1e-06, trace.it = FALSE)
```

**Arguments**

fit	A <code>glm</code> object.
subset.factor	Factor for estimating tau by subset.
fit.only	Return only the final fit? If FALSE, also returns the weights and tau estimates.
iter.max	Maximum number of iterations.
converge	Convergence criterion: difference between model degrees of freedom and Pearson's chi-square. Default 1e-6.
trace.it	Display print statements indicating progress

**Details**

Estimates binomial dispersion parameter  $\tau$  by the method of moments. Iteratively refits the model by the Williams procedure, weighting the observations by  $1/\phi_{ij}$ , where  $\phi_{ij} = 1 + \tau_j(n_{ij} - 1)$ ,  $j$  indexes the subsets, and  $i$  indexes the observations.

**Value**

A list with the following elements.

fit	the new model fit, updated by the estimated weights
weights	vector of weights
phi	vector of phi estimates

**Note**

Level tested: Moderate. (Extensive in S+, not as much in R. They have different `glm()` functions.)

**Author(s)**

David Siev <david.siev@aphis.usda.gov>

## References

- Williams DA, 1982. Extra-binomial variation in logistic linear models. *Applied Statistics* 31:144-148.
- Wedderburn RWM, 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61:439-447.

## See Also

[phiWt](#), [RRor](#), see package vignette for more examples

## Examples

```
birdm.fit <- glm(cbind(y,n-y)~tx-1, binomial, birdm)
RRor(tauWt(birdm.fit))

# 95% t intervals on 4 df
#
# PF
#   PF   LL   UL
# 0.489 -0.578 0.835
#
#   mu.hat   LL   UL
# txcon 0.737 0.944 0.320
# txvac 0.376 0.758 0.104
#
# See the package vignette for more examples
```

# Index

## \*Topic **datasets**

- bird, 3
- birdm, 4
- New, 8
- rat, 11
- Set1, 33
- set1, 34
- Table6, 34
- table6, 34

bird, 3

birdm, 4

getPFTable, 4

getPFTable, rr1-method (getPFTable), 4

getPFTable, rror-method (getPFTable), 4

getPFTable, rrsc-method (getPFTable), 4

getPFTable, rrsi-method (getPFTable), 4

getPFTable, rrstr-method (getPFTable), 4

glm, 9, 19, 32, 35

IDRlsi, 5, 8, 25

IDRsc, 6, 7, 12

New, 8

PF (PF-package), 2

pf, 9

PF-package, 2

phiWt, 9, 20, 36

print.rr1, 11

print.rrmp (print.rr1), 11

print.rror (print.rr1), 11

print.rrsc (print.rr1), 11

print.rrsi (print.rr1), 11

print.rrstr (print.rr1), 11

rat, 11

rr1, 7, 9, 12, 15, 16, 22, 30

RRlsi, 12, 25

RRmh, 14

rrmp, 16, 18

RRmpWald, 16, 17

RRor, 10, 19, 21, 33, 36

rror, 19, 20, 20

RRotsst, 12, 21, 30

RRsc, 23, 25

rrsc, 9, 24, 25

rrsi, 6, 9, 13, 25

RRstr, 26, 28

rrstr, 9, 27, 28

RRtosst, 12, 22, 29

rsb, 31, 33

rsbWt, 31, 32

Set1, 33, 34

set1, 34

Table6, 34, 34

table6, 34

tauWt, 10, 20, 35