

Package ‘FDRsampsiz

January 15, 2016

Type Package

Title Compute Sample Size that Meets Requirements for Average Power and FDR

Version 1.0

Author Stan Pounds <stanley.pounds@stjude.org>

Maintainer Stan Pounds <stanley.pounds@stjude.org>

Depends R (>= 2.15.1)

Imports stats

Date 2016-01-06

Description Defines a collection of functions to compute average power and sample size for studies that use the false discovery rate as the final measure of statistical significance.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2016-01-15 10:28:09

R topics documented:

FDRsampsiz-package	2
afdr	3
alpha.fdr	4
alpha.power	5
average.power	6
fdr.power	7
fdr.sampsiz	8
n.fdr	9
pi.star	11
power.cox	11
power.hart	12
power.li	13
power.onesampt	14
power.oneway	15

power.ranksum	16
power.signtest	17
power.tcorr	18
power.twosampt	18

Index	20
--------------	-----------

FDRsampsizе-package	<i>An R package to Perform Power and Sample Size Calculations for Microarray Studies</i>
---------------------	--

Description

A general approach to performing power and sample size calculations for microarray studies has been developed in the literature. However, the software associated with those articles implements the approach only for studies that will perform the t-test or one-way ANOVA to compare gene expression across two or more groups. Here, we describe a set of R routines that implement the general method for power and sample size calculations for a wider variety of statistical tests. These routines accept the name of a function that computes the power for the statistical test of interest and thus have the flexibility to perform calculations for virtually any statistical test with a known power formula.

Details

Package:	FDRsampsizе
Type:	Package
Version:	1.0
Date:	2016-01-06
License:	GPL(>=2)

Author(s)

Stan Pounds <stanley.pounds@stjude.org>

References

A Onar-Thomas, S Pounds. FDRsampsizе: An R package to Perform Generalized Power and Sample Size Calculations for Planning Studies that use the False Discovery Rate to Measure Significance, Manuscript 2016.

Pounds, Stan, and Cheng Cheng. "Sample size determination for the false discovery rate." *Bioinformatics* 21.23 (2005): 4263-4271.

Jung, Sin-Ho. "Sample size for FDR-control in microarray data analysis." *Bioinformatics* 21.14 (2005): 3097-3104.

`afdr`*Compute the anticipated FDR*

Description

Compute the anticipated FDR for given sample size, p-value threshold, and effect size.

Usage

```
afdr (n, alpha, pow.func, eff.size, lam = 0.95, eps = 1e-04,  
      ...)
```

Arguments

<code>n</code>	sample size (scalar)
<code>alpha</code>	p-value cut-off (scalar)
<code>pow.func</code>	an R function that computes statistical power
<code>eff.size</code>	effect size vector
<code>lam</code>	p-value at which to evaluate ensemble PDF (for pi.star)
<code>eps</code>	epsilon for numerical differentiation
<code>...</code>	additional arguments for the functions

Details

The aFDR is defined by Pounds and Cheng (2005) as the anticipated false discovery rate incurred by performing all tests with p-value threshold alpha given the same size effect size and power function.

Value

the aFDR

References

Pounds, Stan, and Cheng Cheng. "Sample size determination for the false discovery rate." *Bioinformatics* 21.23 (2005): 4263-4271.

Jung, Sin-Ho. "Sample size for FDR-control in microarray data analysis." *Bioinformatics* 21.14 (2005): 3097-3104.

Examples

```
afdr(n=50,alpha=0.01,pow.func=power.twowampt,eff.size=rep(c(1,0),c(100,900)))
```

alpha.fdr	<i>Find the fixed p-value threshold that controls the FDR at a specified level</i>
-----------	--

Description

Find the p-value threshold that satisfies an FDR requirement (if such a threshold exists)

Usage

```
alpha.fdr (fdr, n, pow.func, eff.size, null.effect, lam = 0.95,
          eps = 1e-04, tol = 1e-04, ...)
```

Arguments

fdr	Desired FDR, scalar
n	sample size
pow.func	an R function to compute statistical power
eff.size	effect size vector
null.effect	value of effect size that corresponds to the null hypothesis
lam	the lambda parameter in computing the π_0 (proportion of tests with a true null) estimate of Storey (2002)
eps	epsilon for numerical differentiation
tol	tolerance for bisection solution to alpha
...	additional arguments for the functions

Value

a list with the following components:

fdr	the FDR at that alpha
alpha	the determined alpha
OK	indicates if the requirement is met

References

A Onar-Thomas, S Pounds "FDRsampsiz: An R package to Perform Generalized Power and Sample Size Calculations for Planning Studies that use the False Discovery Rate to Measure Significance", Manuscript 2015.

Pounds, Stan, and Cheng Cheng. "Sample size determination for the false discovery rate." *Bioinformatics* 21.23 (2005): 4263-4271.

Jung, Sin-Ho. "Sample size for FDR-control in microarray data analysis." *Bioinformatics* 21.14 (2005): 3097-3104.

Examples

```
alpha.fdr(fdr=0.1,n=50,pow.func=power.twosampt,
          eff.size=rep(0:1,c(900,100)),null.effect=0)
```

alpha.power

Find the p-value threshold that gives a specified average power

Description

Find p-value cut-off that yields desired average power given n and effect size

Usage

```
alpha.power(ave.pow, n, pow.func, eff.size, null.effect, tol = 1e-06,
            ...)
```

Arguments

ave.pow	desired average power (scalar)
n	sample size
pow.func	name of R function to compute statistical power
eff.size	effect size vector
null.effect	value of effect size that corresponds to null hypothesis
tol	tolerance for bisection solution to alpha
...	additional arguments for the functions

Value

a list with the following components:

alpha	desired value of alpha
ave.pow	average power at that alpha

References

A Onar-Thomas, S Pounds. "FDRsampsiz: An R package to Perform Generalized Power and Sample Size Calculations for Planning Studies that use the False Discovery Rate to Measure Significance", Manuscript 2015. Pounds, Stan, and Cheng Cheng. "Sample size determination for the false discovery rate." *Bioinformatics* 21.23 (2005): 4263-4271.

Jung, Sin-Ho. "Sample size for FDR-control in microarray data analysis." *Bioinformatics* 21.14 (2005): 3097-3104.

Examples

```
alpha.power(ave.pow=0.8,n=50,pow.func=power.twosampt,
            eff.size=rep(0:1,c(900,100)),null.effect=0)
```

average.power	<i>Compute average power for a given sample size</i>
---------------	--

Description

Compute average power for given sample size, effect size, and p-value threshold

Usage

```
average.power (n, alpha, pow.func, eff.size, null.effect, ...)
```

Arguments

n	sample size
alpha	p-value cut off (scalar)
pow.func	an R function to compute statistical power
eff.size	effect size vector
null.effect	value of effect size that corresponds to null hypothesis
...	additional arguments for the functions

Value

average power (scalar)

References

Pounds, Stan, and Cheng Cheng. "Sample size determination for the false discovery rate." *Bioinformatics* 21.23 (2005): 4263-4271. Gadbury GL, et al. (2004) Power and sample size estimation in high dimensional biology. *Statistical Methods in Medical Research* 13(4):325-38. Jung, Sin-Ho. "Sample size for FDR-control in microarray data analysis." *Bioinformatics* 21.14 (2005): 3097-3104.

Examples

```
average.power(n=50,alpha=0.01,pow.func=power.tvosampt,  
             eff.size=rep(0:1,c(900,100)),null.effect=0)
```

fdr.power

Compute the average power at a specific FDR control level

Description

Compute the average power at a specific level of FDR control for a given effect size and sample size

Usage

```
fdr.power (fdr, n, pow.func, eff.size, null.effect, lam = 0.95,
           eps = 1e-04, tol = 1e-04, ...)
```

Arguments

fdr	Desired FDR, scalar
n	sample size
pow.func	name of R function to compute statistical power
eff.size	effect size vector; will be provided as the third argument of pow.func
null.effect	value of effect size that corresponds to null hypothesis
lam	name of R function to compute statistical power
eps	epsilon for numerical differentiation
tol	tolerance for bisection solution to alpha
...	additional arguments for the functions

Value

average power (scalar) of the tests with a false null hypothesis

References

A Onar-Thomas, S Pounds "FDRsampsiz: An R package to Perform Generalized Power and Sample Size Calculations for Planning Studies that use the False Discovery Rate to Measure Significance", Manuscript 2016.

Gadbury GL, et al. (2004) Power and sample size estimation in high dimensional biology. *Statistical Methods in Medical Research* 13(4):325-38.

Pounds S and Cheng C (2005) Sample size determination for the false discovery rate. *Bioinformatics* 21(23): 4263-71.

Examples

```
fdr.power(fdr=0.10, n=50, pow.func=power.twosamt,
          eff.size=rep(0:1, c(900, 100)), null.effect=0)
```

fdr.sampsize	<i>Determine sample size required to achieve a desired average power while controlling the FDR at a specified level.</i>
--------------	--

Description

Determines the sample size needed to achieve a desired average power while controlling the FDR at a specified level.

Usage

```
fdr.sampsize (fdr, ave.pow, pow.func, eff.size, null.effect, max.n = 500,
             min.n = 5, tol = 1e-05, eps = 1e-05, lam = 0.95, ...)
## S3 method for class 'fdr.sampsize'
print(x,...)
```

Arguments

fdr	Desired FDR (scalar numeric)
ave.pow	Desired average power (scalar numeric)
pow.func	Character string name of function to compute power; must accept n, alpha, and eff.size as its first three arguments. Other optional arguments may also be provided.
eff.size	Numeric vector of effect sizes; internally, this will be provided as the third argument of pow.func
null.effect	Scalar value of the effect size under the null hypothesis. This may be 0 or 1 for tests that respectively use differences or ratios for comparisons.
max.n	Maximum n to consider
min.n	Minimum n to consider
tol	Tolerance for bisection calculations
eps	Epsilon for numerical differentiation
lam	Lambda for computing anticipated pi0 estimate, see Storey 2002.
x	result of the fdr.sampsize function
...	additional arguments for pow.func

Details

This function checks the technical conditions regarding whether the desired FDR can be achieved by min.n or max.n before calling n.fdr. Thus, for most applications, fdr.sampsize should be used instead of n.fdr.

Value

fdr.sampsiz returns an object of class 'FDRsampsiz', which is a list with the following components:

OK	indicates if the requirement is met
n	the computed sample size
alpha	the p-value threshold that gives the desired FDR
fdr.hat	anticipated value of the FDR estimate given n and effect size
act.fdr	actual expected FDR given n and effect size
ave.pow	average power
act.pi	actual value of pi0, the proportion of tests with a true null hypothesis.
pi.hat	expected value of the pi0 estimate
eff.size	input effect size vector

References

A Onar-Thomas, S Pounds. "FDRsampsiz: An R package to Perform Generalized Power and Sample Size Calculations for Planning Studies that use the False Discovery Rate to Measure Significance", Manuscript 2015.

Pounds, Stan, and Cheng Cheng. "Sample size determination for the false discovery rate." *Bioinformatics* 21.23 (2005): 4263-4271.

Jung, Sin-Ho. "Sample size for FDR-control in microarray data analysis." *Bioinformatics* 21.14 (2005): 3097-3104.

Examples

```
power.twosampt          # show the power.cox function
res=fdr.sampsiz(fdr=0.1,
                ave.pow=0.8,
                pow.func=power.twosampt,
                eff.size=rep(c(1,0),c(10,990)),
                null.effect=0)
res
```

n.fdr *Find the sample size that meets desired FDR and power criteria*

Description

Find smallest sample size that meets requirements for average power and FDR

Usage

```
n.fdr (ave.pow, fdr, pow.func, eff.size, null.effect, lam = 0.95,
      eps = 1e-04, n0 = 5, n1 = 500, tol = 1e-06, ...)
```

Arguments

ave.pow	required average power (scalar)
fdr	required FDR (scalar)
pow.func	name of R function that computes statistical power
eff.size	effect size vector
null.effect	Value of effect size that indicates null
lam	p-value at which to evaluate ensemble PDF
eps	epsilon for numerical differentiation
n0	smallest sample size to be considered for bisection
n1	maximum sample size to be considered for bisection
tol	tolerance for solving for alpha in iterations
...	additional arguments for the functions

Details

This performs the sample size calculation without checking whether the minimum or maximum sample size satisfy the desired requirements. The `fdr.sampsize` function checks these conditions and then calls `n.fdr`. Thus, many users will may prefer to use the `sampsize.fdr` function instead of `n.fdr`.

Value

a list with the following components:

n	a sample size estimate
alpha	the p-value cut-off
fdr.hat	an approximation to the expected value of the FDR estimate given n
ave.pow	the average power
fdr.act	the actual FDR given n
pi.hat	expected value of the pi.hat estimator given n
act.pi	actual pi0

References

A Onar-Thomas, S Pounds. "FDRsampsiz: An R package to Perform Generalized Power and Sample Size Calculations for Planning Studies that use the False Discovery Rate to Measure Significance", Manuscript 2015.

Pounds, Stan, and Cheng Cheng. "Sample size determination for the false discovery rate." *Bioinformatics* 21.23 (2005): 4263-4271.

Jung, Sin-Ho. "Sample size for FDR-control in microarray data analysis." *Bioinformatics* 21.14 (2005): 3097-3104.

pi.star *Compute the anticipated null proportion estimate*

Description

Compute an approximation of the expected value of the null proportion estimate given the sample size and effect size.

Usage

```
pi.star (n, pow.func, eff.size, lam = 0.95, eps = 1e-04, ...)
```

Arguments

n	sample size
pow.func	an R function to compute statistical power
eff.size	effect size vector
lam	p-value at which to numerically evaluate p-value pdf (scalar)
eps	epsilon for numerical differentiation
...	additional arguments for the functions

Value

scalar value for approximated $E(\hat{\pi})$

References

#> Pounds, Stan, and Cheng Cheng. "Sample size determination for the false discovery rate." *Bioinformatics* 21.23 (2005): 4263-4271.

power.cox *Compute the power of a single-predictor Cox regression model*

Description

Use the formula of Hseih and Lavori (2000) to compute the power of a single-predictor Cox model.

Usage

```
power.cox (n, alpha, logHR, v)
```

Arguments

n	number of events (scalar)
alpha	p-value threshold (scalar)
logHR	log hazard ratio (vector)
v	variance of predictor variable (vector)

Value

vector of power estimates for two-sided test

References

Hsieh, FY and Lavori, Philip W (2000) Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials* 21(6):552-560.

Examples

```
power.cox          # show the power.cox function
res=fdr.sampsize(fdr=0.1,
                 ave.pow=0.8,
                 pow.func=power.cox,
                 eff.size=rep(c(log(2),0),c(100,900)),
                 null.effect=0,
                 v=1)
res
```

power.hart

Compute Power for RNA-seq Experiments Assuming Negative Binomial Distribution.

Description

Use the formula of Hart et al (2013) to compute power for comparing RNA-seq expression across two groups assuming a negative binomial distribution.

Usage

```
power.hart (n, alpha, log.fc, mu, sig)
```

Arguments

n	per-group sample size (scalar)
alpha	p-value threshold (scalar)
log.fc	log fold-change (vector), usual null hypothesis is log.fc=0
mu	read depth per gene (vector, same length as log.fc)
sig	coefficient of variation (CV) per gene (vector, same length as log.fc)

Details

This function is based on equation (1) of Hart et al (2013). It assumes a negative binomial model for RNA-seq read counts and equal sample size per group.

Value

vector of power estimates for the set of two-sided tests

References

SN Hart, TM Therneau, Y Zhang, GA Poland, and J-P Kocher (2013). Calculating Sample Size Estimates for RNA Sequencing Data. *Journal of Computational Biology* 20: 970-978.

Examples

```
power.hart      # show the power function
n.hart=2*(qnorm(0.975)+qnorm(0.9))^2*(1/20+0.6^2)/(log(2)^2) # Equation 6 of Hart et al
power.hart(n.hart,0.05,log(2),20,0.6)                    # Recapitulate 90% power
res=fdr.sampsize(fdr=0.1,
                 ave.pow=0.8,
                 pow.func=power.hart,
                 eff.size=rep(c(log(2),0),c(100,900)),
                 null.effect=0,mu=5,sig=1)
res
```

power.li	<i>Compute Power for RNA-Seq Experiments Assuming Poisson Distribution</i>
----------	--

Description

Use the formula of Li et al (2013) to compute power for comparing RNA-seq expression across two groups assuming the Poisson distribution.

Usage

```
power.li (n, alpha, rho, mu0, w = 1, type = "w")
```

Arguments

n	per-group sample size
alpha	p-value threshold
rho	fold-change, usual null hypothesis is that rho=1
mu0	average count in control group
w	ratio of total number of
type	type of test: "w" for Wald, "s" for score, "lw" for log-transformed Wald, "ls" for log-transformed score.

Details

This function computes the power for each of a series of two-sided tests defined by the input parameters. The power is based on the sample size formulas in equations 10-13 of Li et al (2013). Also, note that the null.effect is set to 1 in the examples because the usual null hypothesis is that the fold-change = 1.

Value

vector of power estimates for two-sided tests

References

C-I Li, P-F Su, Y Guo, and Y Shyr (2013). Sample size calculation for differential expression analysis of RNA-seq data under Poisson distribution. *Int J Comput Biol Drug Des* 6(4). doi:10.1504/IJCBDD.2013.056830

Examples

```
power.li      # show the power function
power.li(88,0.05,1.25,5,0.5,"w") # recapitulate 80% power in Table 1 of Li et al (2013)
res=fdr.sampsize(fdr=0.1,
                 ave.pow=0.8,
                 pow.func=power.li,
                 eff.size=rep(c(1.5,1),c(100,900)),
                 null.effect=1,
                 mu0=5,w=1,type="w")

res
```

power.onesampt	<i>Compute power of the one-sample t-test</i>
----------------	---

Description

Estimate power of the one-sample t-test; Uses classical power formula for one-sample t-test

Usage

```
power.onesampt (n, alpha, delta, sigma = 1)
```

Arguments

n	number of events (scalar)
alpha	p-value threshold (scalar)
delta	difference of actual mean from null mean (vector)
sigma	standard deviation (vector or scalar, default=1)

Value

vector of power estimates for two-sided test

Examples

```

power.onesampt      # show the power function
res=fdr.sampsize(fdr=0.1,
                 ave.pow=0.8,
                 pow.func=power.onesampt,
                 eff.size=rep(c(2,0),c(100,900)),
                 null.effect=0,
                 sigma=1)

res

```

power.oneway

Compute power of one-way ANOVA

Description

Compute power of one-way ANOVA; Uses classical power formula for ANOVA; Assumes equal variance and sample size

Usage

```
power.oneway (n, alpha, theta, k = 2)
```

Arguments

n	per-group sample size (scalar)
alpha	p-value threshold (scalar)
theta	sum of ((group mean - overall mean)/stdev)^2 across all groups for each hypothesis test (vector)
k	the number of groups to be compared, default k=2

Details

For many applications, the null effect is zero for the parameter theta described above.

Value

vector of power estimates for test of equal means

Examples

```

power.oneway      # show the power function
res=fdr.sampsize(fdr=0.1,
                 ave.pow=0.8,
                 pow.func=power.oneway,
                 eff.size=rep(c(2,0),c(100,900)),
                 null.effect=0,
                 k=3)

res

```

power.ranksum *Compute power of the rank-sum test*

Description

Compute power of rank-sum test; Uses formula of Noether (JASA 1987)

Usage

```
power.ranksum (n, alpha, p)
```

Arguments

n	sample size (scalar)
alpha	p-value threshold (scalar)
p	Pr (Y>X), as in Noether (JASA 1987)

Details

In most applications, the null effect size will be designated by $p = 0.5$, which indicates that Thus, in the example below, the argument null.effect=0.5 is specified in the call to fdr.sampsize.

Value

vector of power estimates for two-sided tests

References

Noether, Gottfried E (1987) Sample size determination for some common nonparametric tests. Journal of the American Statistical Association, 82:645-647.

Examples

```
power.ranksum                    # show the power function
res=fdr.sampsize(fdr=0.1,
                 ave.pow=0.8,
                 pow.func=power.ranksum,
                 eff.size=rep(c(0.8,0.5),c(100,900)),
                 null.effect=0.5)
res
```

power.signtest	<i>Compute power of the sign test</i>
----------------	---------------------------------------

Description

Use the Noether (1987) formula to compute the power of the sign test

Usage

```
power.signtest (n, alpha, p)
```

Arguments

n	sample size (scalar)
alpha	p-value threshold (scalar)
p	Pr (Y>X), as in Noether (JASA 1987)

Details

In most applications, the null effect size will be designated by $p = 0.5$ instead of $p = 0$. Thus, in the call to `fdr.sampsize`, we specify `null.effect=0.5` in the example below.

Value

vector of power estimates for two-sided tests

References

Noether, Gottfried E (1987) Sample size determination for some common nonparametric tests. *Journal of the American Statistical Association*, 82:645-647.

Examples

```
power.signtest      # show the power function
res=fdr.sampsize(fdr=0.1,
                 ave.pow=0.8,
                 pow.func=power.signtest,
                 eff.size=rep(c(0.8,0.5),c(100,900)),
                 null.effect=0.5)
res
```

power.tcorr *Compute Power of the t-test for non-zero correlation*

Description

Estimate power of t-test for non-zero correlation; Uses classical power formula for t-test

Usage

power.tcorr (n, alpha, rho)

Arguments

n	sample size (scalar)
alpha	p-value threshold (scalar)
rho	population correlation coefficient (vector)

Details

For many applications, the null.effect is rho=0.

Value

vector of power estimates for two-sided tests

Examples

```
power.tcorr      # show the power function
res=fdr.sampsize(fdr=0.1,
                 ave.pow=0.8,
                 pow.func=power.tcorr,
                 eff.size=rep(c(0.3,0),c(100,900)),
                 null.effect=0)

res
```

power.twosampt *Compute power of the two-samples t-test*

Description

Estimate power of the two-samples t-test; Uses classical power formula for two-sample t-test; Assumes equal variance and sample size

Usage

power.twosampt (n, alpha, delta, sigma = 1)

Arguments

n	per-group sample size (scalar)
alpha	p-value threshold (scalar)
delta	difference between population means (vector)
sigma	standard deviation (vector or scalar)

Details

For many applications, the null.effect is zero difference of means.

Value

vector of power estimates for two-sided test

Examples

```
power.twosampt      # show the power function
res=fdr.sampsize(fdr=0.1,
                 ave.pow=0.8,
                 pow.func=power.twosampt,
                 eff.size=rep(c(2,0),c(100,900)),
                 null.effect=0,
                 sigma=1)
res
```

Index

[afdr](#), [3](#)
[alpha.fdr](#), [4](#)
[alpha.power](#), [5](#)
[average.power](#), [6](#)

[fdr.power](#), [7](#)
[fdr.sampsize](#), [8](#)
[FDRsampsiz](#) ([FDRsampsiz](#)-package), [2](#)
[FDRsampsiz](#)-package, [2](#)

[n.fdr](#), [9](#)

[pi.star](#), [11](#)
[power.cox](#), [11](#)
[power.hart](#), [12](#)
[power.li](#), [13](#)
[power.onesampt](#), [14](#)
[power.oneway](#), [15](#)
[power.ranksum](#), [16](#)
[power.signtest](#), [17](#)
[power.tcorr](#), [18](#)
[power.twosampt](#), [18](#)
[print.fdr.sampsize](#) ([fdr.sampsize](#)), [8](#)