

# Package ‘FDRestimation’

January 20, 2025

**Version** 1.0.1

**Date** 2022-03-21

**Title** Estimate, Plot, and Summarize False Discovery Rates

**Author** Megan Murray [aut, cre],  
Jeffrey Blume [aut]

**Maintainer** Megan Murray <megan.c.hollister@vanderbilt.edu>

**Depends** R (>= 3.4.0)

**Imports** stats, utils, graphics, Rdpack

**Description** The user can directly compute and display false discovery rates from inputted p-values or z-scores under a variety of assumptions. `p.fdr()` computes FDRs, adjusted p-values and decision reject vectors from inputted p-values or z-values. `get.pi0()` estimates the proportion of data that are truly null. `plot.p.fdr()` plots the FDRs, adjusted p-values, and the raw p-values points against their rejection threshold lines.

**License** MIT + file LICENSE

**URL** <doi:10.12688/f1000research.52999.2>

**LazyData** false

**RoxygenNote** 7.1.2

**Encoding** UTF-8

**RdMacros** Rdpack

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2022-04-01 16:30:02 UTC

## Contents

|  |    |
|--|----|
| <code>get.pi0</code> . . . . .             | 2  |
| <code>p.fdr</code> . . . . .               | 4  |
| <code>plot.p.fdr</code> . . . . .          | 7  |
| <code>print.summary.p.fdr</code> . . . . . | 9  |
| <code>summary.p.fdr</code> . . . . .       | 10 |

|              |           |
|--------------|-----------|
| <b>Index</b> | <b>13</b> |
|--------------|-----------|

get.pi0

*pi0 Estimation***Description**

This function estimates the null proportion of data or pi0 value.

**Usage**

```
get.pi0(
  pvalues,
  set.pi0 = 1,
  zvalues = "two.sided",
  estim.method = "last.hist",
  threshold = 0.05,
  default.odds = 1,
  hist.breaks = "scott",
  na.rm = TRUE
)
```

**Arguments**

|              |   |
|--------------|---|
| pvalues      | A numeric vector of raw p-values.   |
| set.pi0      | A numeric value to specify a known or assumed pi0 value in the interval $[0, 1]$ . Defaults to 1. Which means the assumption is that all inputted raw p-values come from the null distribution. |
| zvalues      | A numeric vector of z-values to be used in pi0 estimation or a string with options "two.sided", "greater" or "less". Defaults to "two.sided".   |
| estim.method | A string used to determine which method is used to estimate the pi0 value. Defaults to "last.hist".   |
| threshold    | A numeric value in the interval $[0, 1]$ used in a multiple comparison hypothesis tests to determine significance from the null. Defaults to 0.05.  |
| default.odds | A numeric value determining the ratio of $\pi_1/\pi_0$ used in the computation of lower bound FDR. Defaults to 1.   |
| hist.breaks  | A numeric or string variable representing how many breaks in the pi0 estimation histogram methods. Defaults to "scott".   |
| na.rm        | A Boolean TRUE or FALSE value indicating whether NA's should be removed from the inputted raw p-value vector before further computation. Defaults to TRUE.                                      |

**Details**

We run into errors or warnings when pvalues, zvalues, threshold or default.odds are not inputted correctly.

**Value**

An estimated null proportion:

pi0                    A numeric value representing the proportion of the given data that come from the null distribution. A value in the interval  $[0, 1]$ .

**References**

- Romain Francois (2014). *bibtex: bibtex parser*. R package version 0.4.0.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <https://www.R-project.org/>.
- Storey JD, Tibshirani R (2003). “Statistical significance for genomewide studies.” *Proceedings of the National Academy of Sciences*, **100**(16), 9440–9445.
- Meinshausen N, Rice J, others (2006). “Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses.” *The Annals of Statistics*, **34**(1), 373–393.
- Jiang H, Doerge RW (2008). “Estimating the proportion of true null hypotheses for multiple comparisons.” *Cancer informatics*, **6**, 117693510800600001.
- Nettleton D, Hwang JG, Caldo RA, Wise RP (2006). “Estimating the number of true null hypotheses from a histogram of p values.” *Journal of agricultural, biological, and environmental statistics*, **11**(3), 337.
- Pounds S, Morris SW (2003). “Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values.” *Bioinformatics*, **19**(10), 1236–1242.
- Murray MH, Blume JD (2020). “False Discovery Rate Computation: Illustrations and Modifications.” 2010.04680.

**See Also**

[plot.p.fdr](#), [p.fdr](#), [summary.p.fdr](#)

**Examples**

```
# Example 1
pi0 = 0.8
pi1 = 1-pi0
n = 10000
n.0 = ceiling(n*pi0)
n.1 = n-n.0

sim.data = c(rnorm(n.1,3,1),rnorm(n.0,0,1))
sim.data.p = 2*pnorm(-abs(sim.data))

get.pi0(sim.data.p, estim.method = "last.hist")
get.pi0(sim.data.p, estim.method = "storey")
get.pi0(sim.data.p, estim.method = "set.pi0")
```

p.fdr

*FDR Computation***Description**

This function computes FDRs and Method Adjusted p-values.

**Usage**

```
p.fdr(
  pvalues = NA,
  zvalues = "two.sided",
  threshold = 0.05,
  adjust.method = "BH",
  BY.corr = "positive",
  just.fdr = FALSE,
  default.odds = 1,
  estim.method = "set.pi0",
  set.pi0 = 1,
  hist.breaks = "scott",
  ties.method = "random",
  sort.results = FALSE,
  na.rm = TRUE
)
```

**Arguments**

|               |   |
|---------------|---|
| pvalues       | A numeric vector of raw p-values.   |
| zvalues       | A numeric vector of z-values to be used in pi0 estimation or a string with options "two.sided", "greater" or "less". Defaults to "two.sided".   |
| threshold     | A numeric value in the interval $[0, 1]$ used in a multiple comparison hypothesis tests to determine significance from the null. Defaults to 0.05.  |
| adjust.method | A string used to identify the p-value and false discovery rate adjustment method. Defaults to BH. Options are BH, BY, codeBon, Holm, Hoch, and Sidak.   |
| BY.corr       | A string of either "positive" or "negative" to determine which correlation is used in the BY method. Defaults to positive.  |
| just.fdr      | A Boolean TRUE or FALSE value which output only the FDR vector instead of the list output. Defaults to FALSE.   |
| default.odds  | A numeric value determining the ratio of $\pi_1/\pi_0$ used in the computation of one FDR. Defaults to 1.   |
| estim.method  | A string used to determine which method is used to estimate the null proportion or pi0 value. Defaults to set.pi0.  |
| set.pi0       | A numeric value to specify a known or assumed pi0 value in the interval $[0, 1]$ . Defaults to 1. Which means the assumption is that all inputted raw p-values come from the null distribution. |

|              |  |
|--------------|--|
| hist.breaks  | A numeric or string variable representing how many breaks are used in the $\pi_0$ estimation histogram methods. Defaults to "scott".                       |
| ties.method  | A string a character string specifying how ties are treated. Options are "first", "last", "average", "min", "max", or "random". Defaults to "random".      |
| sort.results | A Boolean TRUE or FALSE value which sorts the output in either increasing or non-increasing order dependent on the FDR vector. Defaults to FALSE.          |
| na.rm        | A Boolean TRUE or FALSE value indicating whether NA's should be removed from the inputted raw p-value vector before further computation. Defaults to TRUE. |

### Details

We run into errors or warnings when pvalues, zvalues, threshold, set.pi0, BY.corr, or default.odds are not inputted correctly.

### Value

A list containing the following components:

|                   |  |
|-------------------|--|
| fdrs              | A numeric vector of method adjusted FDRs.  |
| Results Matrix    | A numeric matrix of method adjusted FDRs, method adjusted p-values, and raw p-values.                                      |
| Reject Vector     | A vector containing Reject.H0 and/or FTR.H0 based off of the threshold value and hypothesis test on the adjusted p-values. |
| $\pi_0$           | A numeric value for the $\pi_0$ value used in the computations.  |
| threshold         | A numeric value for the threshold value used in the hypothesis tests.  |
| Adjustment Method | The string with the method name used in computation(needed for the plot.fdr function).                                     |

### References

- Romain Francois (2014). *bibtex: bibtex parser*. R package version 0.4.0.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <https://www.R-project.org/>.
- Efron B (2013). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge University Press. ISBN 9780511761362.
- Benjamini Y, Hochberg Y (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society*, **57**(1), 289–300.
- Shaffer JP (1995). "Multiple Hypothesis Testing." *Annual review of psychology*, **46**(1), 561–584.
- Storey JD, Tibshirani R (2003). "Statistical significance for genomewide studies." *Proceedings of the National Academy of Sciences*, **100**(16), 9440–9445.
- Benjamini Y, Yekutieli D (2001). "The control of the false discovery rate in multiple testing under dependency." *Annals of statistics*, 1165–1188.

- Meinshausen N, Rice J, others (2006). “Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses.” *The Annals of Statistics*, **34**(1), 373–393.
- Jiang H, Doerge RW (2008). “Estimating the proportion of true null hypotheses for multiple comparisons.” *Cancer informatics*, **6**, 117693510800600001.
- Nettleton D, Hwang JG, Caldo RA, Wise RP (2006). “Estimating the number of true null hypotheses from a histogram of p values.” *Journal of agricultural, biological, and environmental statistics*, **11**(3), 337.
- Pounds S, Morris SW (2003). “Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values.” *Bioinformatics*, **19**(10), 1236–1242.
- Holm S (1979). “A simple sequentially rejective multiple test procedure.” *Scandinavian journal of statistics*, 65–70.
- Bonferroni C (1936). “Teoria statistica delle classi e calcolo delle probabilita.” *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, **8**, 3–62.
- Hochberg Y (1988). “A sharper Bonferroni procedure for multiple tests of significance.” *Biometrika*, **75**(4), 800–802.
- Šidák Z (1967). “Rectangular confidence regions for the means of multivariate normal distributions.” *Journal of the American Statistical Association*, **62**(318), 626–633.
- Murray MH, Blume JD (2020). “False Discovery Rate Computation: Illustrations and Modifications.” 2010.04680.

### See Also

[plot.p.fdr](#), [summary.p.fdr](#), [get.pi0](#)

### Examples

```
# Example 1
pi0 = 0.8
pi1 = 1-pi0
n = 10000
n.0 = ceiling(n*pi0)
n.1 = n-n.0

sim.data = c(rnorm(n.1,3,1),rnorm(n.0,0,1))
sim.data.p = 2*pnorm(-abs(sim.data))

fdr.output = p.fdr(pvalues=sim.data.p, adjust.method="BH")

fdr.output$fdrs
fdr.output$pi0

# Example 2

sim.data.p = output = c(runif(800),runif(200, min=0, max=0.01))
fdr.output = p.fdr(pvalues=sim.data.p, adjust.method="Holm", sort.results = TRUE)

fdr.output$`Results Matrix`
```

---

`plot.p.fdr`*FDR plotting*

---

**Description**

This function creates a plot using a `x` (`p.fdr.object`).

**Usage**

```
## S3 method for class 'p.fdr'
plot(
  x,
  raw.pvalues = TRUE,
  adj.pvalues = TRUE,
  sig.line = TRUE,
  adj.sig.line = TRUE,
  threshold = NA,
  x.axis = "Rank",
  xlim = NA,
  ylim = c(0, 1),
  zvalues = "two.sided",
  legend.where = NA,
  legend.on = TRUE,
  main = NA,
  pch.adj.p = 17,
  pch.raw.p = 20,
  pch.adj.fdr = 20,
  col = c("dodgerblue", "firebrick2", "black"),
  ...
)
```

**Arguments**

|                           |   |
|---------------------------|---|
| <code>x</code>            | A <code>p.fdr</code> object that contains the list of output.   |
| <code>raw.pvalues</code>  | A Boolean TRUE or FALSE value to indicate whether or not to plot the raw p-value points. Defaults to TRUE.                          |
| <code>adj.pvalues</code>  | A Boolean TRUE or FALSE value to indicate whether or not to plot the adjusted p-value points. Defaults to TRUE.                     |
| <code>sig.line</code>     | A Boolean TRUE or FALSE value to indicate whether or not to plot the raw p-value significance line. Defaults to TRUE.               |
| <code>adj.sig.line</code> | A Boolean TRUE or FALSE value to indicate whether or not to plot the adjusted significance threshold. Defaults to TRUE.             |
| <code>threshold</code>    | A numeric value to determine the threshold at which we plot significance. Defaults to value used in the <code>p.fdr.object</code> . |

|              |  |
|--------------|--|
| x.axis       | A string variable to indicate what to plot on the x-axis. Can either be "Rank" or "Zvalues". Defaults to "Rank".   |
| xlim         | A numeric interval for x-axis limits.  |
| ylim         | A numeric interval for y-axis limits. Defaults to c(0,1).  |
| zvalues      | A numeric vector of z-values to be used in pi0 estimation or a string with options "two.sided", "greater" or "less". Defaults to "two.sided".  |
| legend.where | A string "bottomright", "bottomleft", "topleft", "topright". Defaults to "topleft" if x.axis="Rank" and "topright" if x.axis="Zvalues".  |
| legend.on    | A Boolean TRUE or FALSE value to indicate whether or not to print the legend.  |
| main         | A string variable for the title of the plot.   |
| pch.adj.p    | A plotting "character", or symbol to use for the adjusted p-value points. This can either be a single character or an integer code for one of a set of graphics symbols. Defaults to 17.   |
| pch.raw.p    | A plotting "character", or symbol to use for the raw p-value points. This can either be a single character or an integer code for one of a set of graphics symbols. Defaults to 20.  |
| pch.adj.fdr  | A plotting "character", or symbol to use for the adjusted FDR points. This can either be a single character or an integer code for one of a set of graphics symbols. Defaults to 20.   |
| col          | A vector of colors for the points and lines in the plot. If the input has 1 value all points and lines will be that same color. If the input has length of 3 then col.adj.fdr will be the first value, col.adj.p will be the second, and col.raw.p is the third. Defaults to c("dodgerblue", "firebrick2", "black"). |
| ...          | Graphical parameters. Any argument that can be passed to image.plot and to base plot, such as axes=FALSE, main='title', ylab='latitude'  |

## Details

We run into errors or warnings when zvalues or col are inputted incorrectly.

## References

- Romain Francois (2014). *bibtex: bibtex parser*. R package version 0.4.0.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <https://www.R-project.org/>.
- Benjamini Y, Hochberg Y (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society*, **57**(1), 289–300.
- Benjamini Y, Yekutieli D (2001). "The control of the false discovery rate in multiple testing under dependency." *Annals of statistics*, 1165–1188.
- Holm S (1979). "A simple sequentially rejective multiple test procedure." *Scandinavian journal of statistics*, 65–70.
- Hochberg Y (1988). "A sharper Bonferroni procedure for multiple tests of significance." *Biometrika*, **75**(4), 800–802.



Šidák Z (1967). “Rectangular confidence regions for the means of multivariate normal distributions.” *Journal of the American Statistical Association*, **62**(318), 626–633.

Bonferroni C (1936). “Teoria statistica delle classi e calcolo delle probabilita.” *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, **8**, 3–62.

Murray MH, Blume JD (2020). “False Discovery Rate Computation: Illustrations and Modifications.” 2010.04680.

### See Also

[summary.p.fdr](#), [p.fdr](#), [get.pi0](#)

### Examples

```
# Example 1

sim.data.p = c(runif(80),runif(20, min=0, max=0.01))
fdr.output = p.fdr(pvalues=sim.data.p)

plot(fdr.output)
plot(fdr.output, x.axis="Zvalues")
```

---

`print.summary.p.fdr` *Print the summary of p.fdr.object*

---

### Description

This function prints the summary a p.fdr.object.

### Usage

```
## S3 method for class 'summary.p.fdr'
print(x, digits = 3, ...)
```

### Arguments

|                     |  |
|---------------------|--|
| <code>x</code>      | A list of output from the summary.p.fdr function.                                      |
| <code>digits</code> | A numeric value for the number of desired digits in the summary output. Defaults to 3. |
| <code>...</code>    | Further arguments passed to or from other methods.                                     |

### Details

We run into errors or warnings when

## References

Romain Francois (2014). *bibtex: bibtex parser*. R package version 0.4.0.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <https://www.R-project.org/>.

Murray MH, Blume JD (2020). "False Discovery Rate Computation: Illustrations and Modifications." 2010.04680.

## See Also

[plot.p.fdr](#), [p.fdr](#), [get.pi0](#)

## Examples

```
# Example 1
pi0 = 0.8
pi1 = 1-pi0
n = 10
n.0 = ceiling(n*pi0)
n.1 = n-n.0

sim.data = c(rnorm(n.1,5,1),rnorm(n.0,0,1))
sim.data.p = 2*pnorm(-abs(sim.data))

fdr.output = p.fdr(pvalues=sim.data.p, adjust.method="BH")

summary(fdr.output)
```

---

summary.p.fdr

*Summary of p.fdr.object*

---

## Description

This function summarizes a p.fdr object.

## Usage

```
## S3 method for class 'p.fdr'
summary(object, digits = 5, ...)
```

## Arguments

|        |  |
|--------|--|
| object | A list of output from the p.fdr function.  |
| digits | A numeric value for the number of desired digits in the summary output. Defaults to 3. |
| ...    | Additional arguments affecting the summary produced.                                   |

**Details**

We run into errors or warnings when

**Value**

A list containing the following components:

|                           |   |
|---------------------------|---|
| Range                     | The range on the false discovery rates.   |
| Significant Findings      | The number of significant findings. Found using the adjusted p-values and the given threshold. This is also the number of times we decide to reject the null hypothesis that the data is generated from a standard normal distribution. |
| Inconclusive Findings     | The number of inconclusive findings. Found using the adjusted p-values and the given threshold. This is also the number of times we fail to reject the null hypothesis that the data is generated from a standard normal distribution.  |
| Assumed/Estimated $\pi_0$ | the assumed or estimated $\pi_0$ value depending on how the p.fdr function was run.   |
| Number of Tests           | The total number of multiple comparison tests completed.  |
| Adjustment Method         | The adjustment method used in the p.fdr function.   |

**References**

- Romain Francois (2014). *bibtex: bibtex parser*. R package version 0.4.0.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <https://www.R-project.org/>.
- Murray MH, Blume JD (2020). "False Discovery Rate Computation: Illustrations and Modifications." 2010.04680.

**See Also**

[plot.p.fdr](#), [p.fdr](#), [get.pi0](#)

**Examples**

```
# Example 1
pi0 = 0.8
pi1 = 1-pi0
n = 10
n.0 = ceiling(n*pi0)
n.1 = n-n.0

sim.data = c(rnorm(n.1,5,1),rnorm(n.0,0,1))
sim.data.p = 2*pnorm(-abs(sim.data))

fdr.output = p.fdr(pvalues=sim.data.p, adjust.method="BH")
```

```
summary(fdr.output)
```

# Index

- \* **FDR adjusted p-values null proportion**

- get.pi0, 2

- p.fdr, 4

- \* **FDR**

- get.pi0, 2

- p.fdr, 4

- plot.p.fdr, 7

- print.summary.p.fdr, 9

- summary.p.fdr, 10

- \* **p-values**

- get.pi0, 2

- p.fdr, 4

- plot.p.fdr, 7

- print.summary.p.fdr, 9

- \* **plot FDR adjusted p-values**

- plot.p.fdr, 7

- \* **plot**

- plot.p.fdr, 7

- \* **summary FDR adjusted p-values null proportion**

- print.summary.p.fdr, 9

- \* **summary FDR adjusted p-values**

- summary.p.fdr, 10

- \* **summary**

- print.summary.p.fdr, 9

- summary.p.fdr, 10

get.pi0, 2, 6, 9–11

p.fdr, 3, 4, 9–11

plot.p.fdr, 3, 6, 7, 10, 11

print.summary.p.fdr, 9

summary.p.fdr, 3, 6, 9, 10