Package 'NRejections'

January 20, 2025

Type Package

Title Metrics for Multiple Testing with Correlated Outcomes

Version 1.2.0

Author Maya B. Mathur, Tyler J. VanderWeele

Maintainer Maya B. Mathur <mmathur@stanford.edu>

Description Implements methods in Mathur and VanderWeele (in preparation) to characterize global evidence strength across W correlated ordinary least squares (OLS) hypothesis tests. Specifically, uses resampling to estimate a null interval for the total number of rejections in, for example, 95% of samples generated with no associations (the global null), the excess hits (the difference between the observed number of rejections and the upper limit of the null interval), and a test of the global null based on the number of rejections.

LazyData true

License GPL-2

Imports stats, doParallel, matrixcalc, StepwiseTest, foreach, mvtnorm

RoxygenNote 7.1.1

Suggests testthat

NeedsCompilation no

Repository CRAN

Date/Publication 2020-07-09 13:50:02 UTC

Contents

`adj_minP` *Adjust p-values using minP*

Description

Returns minP-adjusted p-values (single-step). See Westfall & Young (1993), pg. 48.

Usage

```
adj_minP(p, p.bt)
```

Arguments

`p`	Original dataset p-values (W-vector)
`p.bt`	Bootstrapped p-values (a W X B matrix)

References

Westfall, P. H., & Young, S. S. (1993). Resampling-based multiple testing: Examples and methods for p-value adjustment. Taylor & Francis Group.

Examples

```
# observed p-values for 3 tests
 pvals = c(0.00233103655078803, 0.470366742594242, 0.00290278216035089
)

# bootstrapped p-values for 5 resamples
p.bt = t( structure(c(0.308528665936264, 0.517319402377912, 0.686518314693482,
            0.637306248855186, 0.106805510862352, 0.116705315041494, 0.0732076817175753,
             0.770308936364482, 0.384405349738909, 0.0434358213611965, 0.41497067850141,
              0.513471489744384, 0.571213377144122, 0.628054979652722, 0.490196884985226
), .Dim = c(5L, 3L)) )

# adjust the p-values
adj_minP( p = pvals, p.bt = p.bt )
```

adj_Wstep *Return Wstep-adjusted p-values*

Description

Returns p-values adjusted based on Westfall & Young (1993)'s step-down algorithm (see pg. 66-67).

Usage

```
adj_Wstep(p, p.bt)
```

Arguments

p	Original dataset p-values (W-vector)
p.bt	Bootstrapped p-values (an W X B matrix)

References

Westfall, P. H., & Young, S. S. (1993). Resampling-based multiple testing: Examples and methods for p-value adjustment. Taylor & Francis Group.

Examples

```
# observed p-values for 3 tests
 pvals = c(0.00233103655078803, 0.470366742594242, 0.00290278216035089
)

# bootstrapped p-values for 5 resamples
p.bt = t( structure(c(0.308528665936264, 0.517319402377912, 0.686518314693482,
            0.637306248855186, 0.106805510862352, 0.116705315041494, 0.0732076817175753,
             0.770308936364482, 0.384405349738909, 0.0434358213611965, 0.41497067850141,
              0.513471489744384, 0.571213377144122, 0.628054979652722, 0.490196884985226
), .Dim = c(5L, 3L)) )

# adjust the p-values
adj_Wstep( p = pvals, p.bt = p.bt )
```

cell_corr *Cell correlation for simulating data*

Description

The user does not need to call this function. This internal function is called by make_corr_mat and populates a single cell. Assumes X1 is the covariate of interest.

Usage

```
cell_corr(vname.1, vname.2, rho.XX, rho.YY, rho.XY, nY, prop.corr = 1)
```

Arguments

vname.1	Quoted name of first variable
vname.2	Quoted name of second variable
rho.XX	Correlation between pairs of Xs
rho.YY	Correlation between all pairs of Ys
rho.XY	rho.XY Correlation between pairs of X-Y (of non-null ones)
nY	Number of outcomes
prop.corr	Proportion of X-Y pairs that are non-null (non-nulls will be first .prop.corr * .nY pairs)

---

corr_tests *Global evidence strength across correlated tests*

Description

This is the main wrapper function for the user to call. For an arbitrary number of outcome variables, regresses the outcome on an exposure of interest (X) and adjusted covariates (C). Returns the results of the original sample (statistics and inference corresponding to X for each model, along with the observed number of rejections), a 100*(1 - alpha.fam) percent null interval for the number of rejections in samples generated under the global null, the excess hits (the difference between the observed number of rejections and the upper null interval limit), and results of a test of the global null hypothesis at alpha.fam of the global null. The global test can be conducted based on the number of rejections or based on various FWER-control methods (see References).

Usage

```
corr_tests(
  d,
  X,
  C = NA,
  Ys,
  B = 2000,
  cores,
  alpha = 0.05,
  alpha.fam = 0.05,
  method = "nreject"
)
```

Arguments

d	Dataframe
X	Single quoted name of covariate of interest
C	Vector of quoted covariate names
Ys	Vector of quoted outcome names
B	Number of resamples to generate
cores	Number of cores to use for parallelization. Defaults to number available.
alpha	Alpha level for individual hypothesis tests
alpha.fam	Alpha level for global test and null interval
method	Which methods to report (ours, Westfall's two methods, Bonferroni, Holm, Romano)

Value

samp.res is a list containing the number of observed rejections (rej), the coefficient estimates of interest for each outcome model (bhats), their t-values (tvals), their uncorrected p-values at level alpha (pvals), and an N X W matrix of residuals for each model (resid).

nrej.bt contains the number of rejections in each bootstrap resample.

tvals.bt is a W X B matrix containing t-values for the resamples.

pvals.bt is a W X B matrix containing p-values for the resamples.

null.int contains the lower and upper limits of a 100*(1 - alpha.fam) percent null interval.

excess.hits is the difference between the observed rejections and the upper limit of the null interval.

global.test is a dataframe containing global test results for each user-specified method, including an indicator for whether the test rejects the global null at alpha.fam (reject), the p-value of the global test where possible (reject), and the critical value of the global test based on the number of rejections (crit).

References

Mathur, M.B., & VanderWeele, T.J. (in preparation). New metrics for multiple testing with correlated outcomes.

Romano, J. P., & Wolf, M. (2007). Control of generalized error rates in multiple testing. The Annals of Statistics, 1378-1408.

Westfall, P. H., & Young, S. S. (1993). Resampling-based multiple testing: Examples and methods for p-value adjustment. Taylor & Francis Group.

Examples

```
##### Example 1 #####
data(rock)

res = corr_tests( d = rock,
                  X = c("area"),
```

```
                  C = NA,
                  Ys = c("perm", "peri", "shape"),
                  method = "nreject" )

# mean rejections in resamples
# should be close to 0.05 * 3 = 0.15
mean( as.numeric(res$nrej.bt) )


##### Example 1 #####
cor = make_corr_mat( nX = 10,
nY = 20,
rho.XX = 0.10,
rho.YY = 0.5,
rho.XY = 0.1,
prop.corr = .4 )

d = sim_data( n = 300, cor = cor )

# X1 is the covariate of interest, and all other X variables are adjusted
all.covars = names(d)[ grep( "X", names(d) ) ]
C = all.covars[ !all.covars == "X1" ]

# may take 10 min to run
res = corr_tests( d,
                  X = "X1",
                  C = C,
                  Ys = names(d)[ grep( "Y", names(d) ) ],
                  method = "nreject" )

# look at the main results
res$null.int
res$excess.hits
res$global.test
```

`dataset_result` *Fit all models for a single dataset*

Description

The user does not need to call this function. For a single dataset, fits separate OLS models for W outcomes with or without centering the test statistics to enforce the global null.

Usage

```
dataset_result(
  d,
  X,
  C = NA,
```

```
  Ys,
  alpha = 0.05,
  center.stats = TRUE,
  bhat.orig = NA
)
```

Arguments

d	Dataframe
X	Single quoted name of covariate of interest
C	Vector of quoted covariate names
Ys	W-vector of quoted outcome names
alpha	Alpha level for individual tests
center.stats	Should test statistics be centered by original-sample estimates to enforce global null?
bhat.orig	Estimated coefficients for covariate of interest in original sample (W-vector). Can be left NA for non-centered stats.

Value

Returns a list containing the number of observed rejections (rej), the coefficient estimates of interest for each outcome model (bhats), their t-values (tvals), their uncorrected p-values at level alpha (pvals), and a matrix of residuals from each model (resid). The latter is used for residual resampling under the global null.

Examples

```
samp.res = dataset_result( X = "complaints",
           C = c("privileges", "learning"),
           Ys = c("rating", "raises"),
           d = attitude,
           center.stats = FALSE,
           bhat.orig = NA,  # bhat.orig is a single value now for just the correct Y
           alpha = 0.05 )
```

---

fit_model *Fit OLS model for a single outcome*

Description

The user does not need to call this function. Fits OLS model for a single outcome with or without centering the test statistics to enforce the global null.

Usage

```
fit_model(
  X,
  C = NA,
  Y,
  Ys,
  d,
  center.stats = FALSE,
  bhat.orig = NA,
  alpha = 0.05
)
```

Arguments

X	Single quoted name of covariate of interest
C	Vector of quoted covariate names
Y	Quoted name of single outcome for which model should be fit
Ys	Vector of all quoted outcome names
d	Dataframe
center.stats	Should test statistics be centered by original-sample estimates to enforce global null?
bhat.orig	Estimated coefficients for covariate of interest in original sample (W-vector). Can be left NA for non-centered stats.
alpha	Alpha level for individual tests

Examples

```
data(attitude)
fit_model( X = "complaints",
           C = c("privileges", "learning"),
           Y = "rating",
           Ys = c("rating", "raises"),
           d = attitude,
           center.stats = FALSE,
           bhat.orig = NA,
           alpha = 0.05 )
```

---

fix_input *Fix bad user input*

Description

The user does not need to call this function. Warns about and fixes bad user input: missing data on analysis variables or datasets containing extraneous variables.

Usage

```
fix_input(X, C, Ys, d)
```

Arguments

X	Single quoted name of covariate of interest
C	Vector of quoted covariate names
Ys	Vector of quoted outcome names
d	Dataframe

---

get_crit *Return ordered critical values for Wstep*

Description

The user does not need to call this function. This is an internal function for use by `adj_minP` and `adj_Wstep`.

Usage

```
get_crit(p.dat, col.p)
```

Arguments

p.dat	p-values from dataset (W-vector)
col.p	Column of resampled p-values (for the single p-value for which we're

---

make_corr_mat *Makes correlation matrix to simulate data*

Description

Simulates a dataset with a specified number of standard MVN covariates and outcomes with a specified correlation structure. If the function returns an error stating that the correlation matrix is not positive definite, try reducing the correlation magnitudes.

Usage

```
make_corr_mat(nX, nY, rho.XX, rho.YY, rho.XY, prop.corr = 1)
```

Arguments

nX	Number of covariates, including the one of interest
nY	Number of outcomes
rho.XX	Correlation between all pairs of Xs
rho.YY	Correlation between all pairs of Ys
rho.XY	Correlation between pairs of X-Y that are not null (see below)
prop.corr	Proportion of X-Y pairs that are non-null (non-nulls will be first prop.corr * nY pairs)

Examples

```
make_corr_mat( nX = 1,
nY = 4,
rho.XX = 0,
rho.YY = 0.25,
rho.XY = 0,
prop.corr = 0.8 )
```

---

resample_resid *Resample residuals for OLS*

Description

Implements the residual resampling OLS algorithm described in Mathur & VanderWeele (in preparation). Specifically, the design matrix is fixed while the resampled outcomes are set equal to the original fitted values plus a vector of residuals sampled with replacement.

Usage

```
resample_resid(
  d,
  X,
  C = NA,
  Ys,
  alpha,
  resid,
  bhat.orig,
  B = 2000,
  cores = NULL
)
```

Arguments

d	Dataframe
X	Single quoted name of covariate of interest
C	Vector of quoted covariate names
Ys	Vector of quoted outcome names
alpha	Alpha level for individual tests
resid	Residuals from original sample (W X B matrix)
bhat.orig	Estimated coefficients for covariate of interest in original sample (W-vector)
B	Number of resamples to generate
cores	Number of cores available for parallelization

Value

Returns a list containing the number of rejections in each resample, a matrix of p-values in the resamples, and a matrix of t-statistics in the resamples.

References

Mathur, M.B., & VanderWeele, T.J. (in preparation). New metrics for multiple testing with correlated outcomes.

Examples

```
samp.res = dataset_result( X = "complaints",
                C = c("privileges", "learning"),
                Ys = c("rating", "raises"),
                d = attitude,
                center.stats = FALSE,
               bhat.orig = NA,  # bhat.orig is a single value now for just the correct Y
                alpha = 0.05 )

resamps = resample_resid(  X = "complaints",
                  C = c("privileges", "learning"),
                  Ys = c("rating", "raises"),
                  d = attitude,
                  alpha = 0.05,
                  resid = samp.res$resid,
                  bhat.orig = samp.res$b,
                  B=20,
                  cores = 2)
```

`sim_data` *Simulate MVN data*

Description

Simulates one dataset with standard MVN correlated covariates and outcomes.

Usage

```
sim_data(n, cor)
```

Arguments

`n`	Number of rows to simulate
`cor`	Correlation matrix (e.g., from `make_corr_mat`)

Examples

```
cor = make_corr_mat( nX = 5,
nY = 2,
rho.XX = -0.06,
rho.YY = 0.1,
rho.XY = -0.1,
prop.corr = 8/40 )

d = sim_data( n = 50, cor = cor )
```

Index