Package: kpmt (via r-universe)

August 31, 2024

Version 0.1.0

Date 2017-07-28

Title Known Population Median Test

Description Functions that implement the known population median test.

Author Matthew M Parks <map2085@med.cornell.edu>

Maintainer Matthew M Parks <map2085@med.cornell.edu>

Imports matrixStats, stats, methods, utils

License MIT + file LICENSE

RoxygenNote 6.0.1

NeedsCompilation no

Depends R (>= 2.10)

Date/Publication 2017-08-01 04:51:06 UTC

Repository https://map2085.r-universe.dev

RemoteUrl https://github.com/cran/kpmt

RemoteRef HEAD

RemoteSha 2aad85343bec38651e594481dec94229917c9e60

Contents

---

genefeat   *Transcript features of the protein coding genes of the human genome*

Description

Various quantitative features of the transcripts of the set of all 19,962 protein coding genes in the annotated human genome, according to Ensembl 84. Features include length of the Coding Sequence (CDS), lengths of the 5' and 3' UTRs, and the relative usage for each amino acid.

Usage

```
genefeat
```

Format

matrix.

Source

Ensembl

References

Yates, et. al. (2016) Nucleic Acids Res. 2016 44 Database issue:D710-6. "Ensembl 2016". (doi)

GO0007186   *The genes pertaining to GO term GO:0007186 (G-protein coupled receptor signaling pathway)*

Description

The names of all 868 protein coding genes annotated as members of Gene Ontology (GO) term GO:0007186 (G-protein coupled receptor signaling pathway), according to the Gene Ontology consortium.

Usage

```
GO0007186
```

Format

character vector.

Source

Gene Ontology

References

Ashburner et al. Gene ontology: tool for the unification of biology (2000) Nat Genet 25(1):25-9. (doi: 10.1038/75556) The Gene Ontology Consortium. Gene Ontology Consortium: going forward. (2015) Nucl Acids Res 43 Database issue (doi)

---

kpmt	*Known Population Median Test*

Description

Performs the known population median test.

Usage

```
kpmt(pop, obs = NULL, med = NULL, size = NULL, tail = "two-sided",
  verbose = FALSE)
```

Arguments

pop
: [data frame, matrix, vector] numeric values for the whole population. If a data frame or matrix is given, it should have format:
    - rownames = population member names (e.g. gene names)
    - colnames = features to test (e.g. relative codon usage, UTR length, MFE, etc.)

    For a data frame or matrix, the test will be performed on each column, separately. If a named vector is given, it should have format:
    - names = population member names.
    - values = numeric values of the feature.

obs
: [character vector or named list of character vectors] a character vector of population member names, or a named list of character vectors of population member names.
    1. If obs is a list, then each list element name should correspond to a feature name of pop.
    2. If obs is a vector, then it is considered to be the same sample for each population feature.
    3. if size = NULL, then obs is considered to be a sample of population member names
    4. if size is non-NULL, then obs is considered to be the observed median values per column of pop. (length obs must be equal to number of features in pop)

med
: [number or vector] pre-computed minimal medians of pop.

size
: [integer] size of the set which generated the observed median. If obs is a sample, i.e. contains population member names, then size must be NULL.

tail
: ["two-sided", "lower", "upper"] if NULL, then the minimum of lower and upper will be reported.

verbose
: display extra messages for tracking execution.

Value

data frame with columns:

- "name" a column from pop
- "median.sample" min median of the sample
- "median.all" min median of the whole population
- "median.background" min median of the non-sampled members.
- "logp" log of p.value if sample median is different from all
- "p.value" p.value if sample median is different from all
- "FDR" only if > 30 features, i.e columns of pop

each row is a different population feature, i.e. column of pop.

Examples

```
data(genefeat)
data(GO0007186)
res  <-  kpmt( pop = genefeat , obs = GO0007186 )
```

---

logStirling   *Log Stirling*

Description

Computes the log of the Stirling approximation of n!.

Usage

```
logStirling(n)
```

Arguments

n   integer or vector of integers.

Value

Stirling approximation of log(n!). If n <= 14, then computes log(n!) directly, i.e. no Stirling approximation.

---

minMedian   *Minimal median*

Description

Computes the minimal median of a vector or matrix.

Usage

```
minMedian(x)
```

Arguments

x   a vector or matrix of real numbers. if a matrix, then the minimal median will be
                        computed for each column.

Value

the minimal median of x.

Examples

```
minMedian(1:6) # returns 3
```

---

validate_accuracy   *Validate computational accuracy of Known Population Median Test*
                        algorithm

Description

The implementation of the exact analytical solution of the Known Population Median Test involves
approximations using the Stirling series and is therefore suspect for computational error. This
function creates a population and empirically computes p-values via resampling. It then compares
these empirical p-values to those calculated by the Known Population Median Test, and returns the
error in log(p). WARNING! Takes a long time.

Usage

```
validate_accuracy(N = 50, n = 10, nrep = 1e+08)
```

Arguments

N   population size
n   sample size
nrep   number of resampling samples for empirically estimating p-values.

Value

data frame with log(p) error and other information.

Index