Package 'SASOM'

October 16, 2024

Type Package

Title SASOM: Score-Based Association Tests for Somatic Mutations and Multinomial Outcomes

Version 0.1.0

Author Meiling Liu, Yang Liu, and Qianchuan He

Maintainer Meiling Liu <mliu@fredhutch.org>

Description Test for association between a set of somatic mutations and multinomial outcomes by using score statistics.

Depends R (>= 3.5.0), CompQuadForm, nnet

License LGPL (>= 2.0)

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

NeedsCompilation no

R topics documented:

`DAPC` *Data-Adaptive P-value Combination (DAPC) Method*

Description

The `DAPC` function combines two independent p-values using the Data-Adaptive P-value Combination (DAPC) method. This method is designed to provide an adaptive approach for combining p-values from independent statistical tests, which may offer better performance compared to traditional methods like Fisher's or Tippet's, especially under certain data conditions.

Usage

```
DAPC(x)
```

Arguments

x	A numeric vector of length 2, containing two independent p-values to be combined. These p-values are assumed to come from separate statistical tests.

Details

The DAPC method is a flexible p-value combination technique that adapts to the distribution of the input p-values. The method uses a data-adaptive weighting scheme to improve power and control the type I error rate, making it suitable for genomic studies where associations between multiple variants and outcomes are tested.

This function is used within the SASOM framework to combine p-values from different tests (e.g., fixed and random effects) in a robust manner.

Value

The function returns a single numeric value:

p The combined p-value calculated using the DAPC method. This p-value represents the overall significance of the two independent tests.

Author(s)

Meiling Liu, Yang Liu, and Qianchuan He

References

Liu, M., Liu, Y., Wu, M.C., Hsu, L. and He, Q., 2021. A method for subtype analysis with somatic mutations. Bioinformatics, 37(1), pp.50-56.

See Also

SASOM for performing the association test between somatic mutations and multinomial outcomes.

Examples

```
# Example of combining two independent p-values
p_values <- c(0.1, 0.3)
combined_p <- DAPC(p_values)

# Print the combined p-value
print(combined_p)
```

SASOM *Score-Based Association Test for Multinomial Outcomes*

Description

The SASOM function implements a score-based test for evaluating the association between somatic mutations and multinomial outcomes.

Usage

```
SASOM(y, X, G, W, p.comb)
```

Arguments

y	A matrix or a factor vector representing the outcome. If a matrix is provided, each row corresponds to a subject, and each column corresponds to a subtype, with the sum of each row equal to 1, indicating the subject's assignment to one of the subtypes. If a factor vector is provided, each element corresponds to the subtype assignment of each subject.
X	A covariate matrix, where each row corresponds to a subject and each column to a covariate. Covariates could include clinical or demographic variables.
G	A matrix of genetic variants, where each row represents a subject, and each column represents a variant. This matrix captures the variant information for each subject under study.
W	A variant annotation matrix where each row represents a genomic feature (such as a gene or pathway), and each column represents a variant. The number of columns in `W` must match the number of rows in `G`. This matrix provides additional annotation information for variants, such as their functional category.
p.comb	A character string specifying the method to combine p-values from fixed and random effects. Must be one of `"all"`, `"DAPC"`, `"fisher"`, or `"tippet"`.

"all" All combination methods (DAPC, Fisher's, and Tippet's) are used.

"DAPC" P-values are combined using the DAPC (Data-Adaptive P-value Combination) method.

"fisher" P-values are combined using Fisher's method.

"tippet" P-values are combined using Tippet's method.

Details

The `SASOM` function employs score-based statistics to test for associations between somatic mutations (e.g., genetic variants) and multinomial outcomes, such as disease subtypes. Both fixed effects (which assume the variant has a consistent effect across all subjects) and random effects (which allow the variant effect to vary) are tested. The function offers flexible p-value combination methods for robust conclusions.

The package accommodates a variety of settings in association studies, particularly when testing multiple outcomes (subtypes) is required, and when the outcome may be influenced by both covariates and genomic features. By incorporating multiple combination methods, `SASOM` provides a versatile framework for hypothesis testing in genomic research.

The function is particularly suited for studies where somatic mutations or variants in genomic regions are hypothesized to impact categorical outcomes, such as different cancer subtypes or stages of a disease.

Value

The function returns a list containing the following components:

pval.theta	P-value for testing the fixed variant effect. This evaluates whether the somatic mutation has a consistent association with the outcome across all subjects.
pval.tau	P-value for testing the random variant effect, which allows the variant effect to vary across subjects.
p.fisher	The overall p-value combining `pval.theta` and `pval.tau` using Fisher's method. This is used for global association testing.

`pval.tippet` The overall p-value combining `pval.theta` and `pval.tau` using Tippet's method, providing a more conservative approach.

`pval.dapc` The overall p-value combining `pval.theta` and `pval.tau` using the Data-Adaptive P-value Combination (DAPC) method, which can offer better performance under certain conditions.

Author(s)

Meiling Liu (<mliu@fredhutch.org>), Yang Liu, and Qianchuan He

References

For a comprehensive discussion on score-based association testing and p-value combination methods, please refer to:

Liu, M., Liu, Y., Wu, M.C., Hsu, L. and He, Q., 2021. A method for subtype analysis with somatic mutations. Bioinformatics, 37(1), pp.50-56.

Examples

```
data("SASOM.data") # Load the sample dataset
attach(SASOM.data) # Attach the data to the environment
result <- SASOM(y, X, G, W, p.comb = "all") # Run the SASOM test
print(result) # View the results
detach(SASOM.data) # Detach the dataset
```

SASOM.data *Example Dataset for SASOM Analysis*

Description

`SASOM.data` is a synthetic dataset created to illustrate the usage of the `SASOM` function, which tests for associations between somatic mutations and multinomial outcomes. The dataset includes simulated genetic, covariate, and outcome data for a population of subjects.

Usage

```
data(SASOM.data)
```

Format

A list containing the following components:

`y` A numeric matrix representing the multinomial outcomes for the subjects. Each row corresponds to a subject, and each column corresponds to a subtype. The rows sum to 1, indicating that each subject belongs to exactly one subtype. There are three subtypes in total.

`X` A numeric matrix of covariates, including an intercept term (e.g., clinical or demographic covariates). Each row corresponds to a subject, and each column corresponds to a covariate.

`G` A numeric matrix of genetic variant data. Each row represents a subject, and each column represents a genetic variant. The matrix contains information on the presence or absence of somatic mutations for each subject.

`W` A numeric matrix of variant annotations. Each row corresponds to a genomic feature (e.g., genes or pathways), and each column corresponds to a variant. This matrix provides additional biological or functional information about the variants in `G`, linking each variant to relevant genomic features.

Details

This dataset is intended to demonstrate the application of the `SASOM` package for testing the association between somatic mutations and multinomial outcomes, such as disease subtypes.

Source

Simulated data for illustrative purposes.

References

Liu, M., Liu, Y., Wu, M.C., Hsu, L. and He, Q., 2021. A method for subtype analysis with somatic mutations. Bioinformatics, 37(1), pp.50-56.

Examples

```
data("SASOM.data") # Load the example data
str(SASOM.data) # Display the structure of the dataset
## Example usage of the SASOM function with the dataset
out <- SASOM(SASOM.data$y, SASOM.data$X, SASOM.data$G, SASOM.data$W, p.comb = "all")
```

Index