

Package ‘iPRSue’

September 5, 2025

Title Individual Polygenic Risk Score Uncertainty Estimation

Version 1.0.0

Description Provides tools for estimating uncertainty in individual polygenic risk scores (PRSs) using both sampling-based and analytical methods, as well as the Best Linear Unbiased Estimator (BLUE). These methods quantify variability in PRS estimates for both binary and quantitative traits. See Henderson (1975) <[doi:10.2307/2529430](https://doi.org/10.2307/2529430)> for more details.

URL <https://github.com/DoviniJ/iPRSue>

License GPL-3

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Depends R (>= 2.10)

Imports data.table,
bigstatsr,
logistf,
stats,
utils

LazyData true

R topics documented:

BLUE_estimates_BT	2
BLUE_estimates_QT	3
GWAS_BT	5
GWAS_QT	6
iPRSue_estimates_BT	7
iPRSue_estimates_QT	8
Index	11

BLUE_estimates_BT	<i>BLUE_estimates_BT function</i>
-------------------	-----------------------------------

Description

Estimates individual-level polygenic risk scores (PRS) with uncertainty using a frequentist approach for binary traits. This implementation applies Firth's bias-reduced logistic regression on the discovery sample, computes the coefficient covariance matrix, and uses the delta method to derive PRS variance and confidence intervals.

Usage

```
BLUE_estimates_BT(
  discovery_pheno,
  discovery_geno_mat,
  target_pheno,
  target_geno_mat,
  significance_level = 0.05,
  max_iterations = 100
)
```

Arguments

discovery_pheno	Character. Path to the phenotype file for the discovery dataset. Assumes no header and that the binary trait is in the third column.
discovery_geno_mat	Character. Path to the genotype matrix file for the discovery dataset. Assumes no header.
target_pheno	Character. Path to the phenotype file for the target dataset. Assumes no header and individual IDs in the second column.
target_geno_mat	Character. Path to the genotype matrix file for the target dataset. Assumes no header.
significance_level	Numeric. Significance level for confidence intervals (e.g., 0.05 for 95% CI). Default is 0.05.
max_iterations	Integer. Maximum number of iterations allowed in Firth logistic regression. Default is 100.

Details

The function fits a Firth logistic regression model using the `logistf` package to reduce small-sample bias in the discovery set. It extracts SNP effect estimates and their covariance matrix, and propagates this uncertainty through to the individual-level PRS in the target dataset via the delta method. Confidence intervals are derived assuming normality.

Missing or non-estimable coefficients and variances are set to zero.

Value

A data frame with the following columns:

IID Individual identifier (from the target phenotype file).

PRS Estimated polygenic risk score for each individual.

Variance Estimated variance of the PRS.

Lower_Limit Lower bound of the confidence interval.

Upper_Limit Upper bound of the confidence interval.

Examples

```
bpd <- system.file("Bpd_0_1.txt", package = "iPRSue", mustWork = TRUE)
bpt <- system.file("Bpt.txt", package = "iPRSue", mustWork = TRUE)
gd <- system.file("Gd.txt", package = "iPRSue", mustWork = TRUE)
gt <- system.file("Gt.txt", package = "iPRSue", mustWork = TRUE)

results <- BLUE_estimates_BT(
  discovery_pheno = bpd,
  discovery_geno_mat = gd,
  target_pheno = bpt,
  target_geno_mat = gt,
  significance_level = 0.05,
  max_iterations = 100
)
head(results)
```

BLUE_estimates_QT	<i>BLUE_estimates_QT function</i>
-------------------	-----------------------------------

Description

Estimates individual-level polygenic risk scores (PRS) with uncertainty using a frequentist approach for quantitative traits. This implementation fits a multiple linear regression model in the discovery dataset, computes the coefficient covariance matrix, and applies the delta method to propagate uncertainty to the target dataset.

Usage

```
BLUE_estimates_QT(
  discovery_pheno,
  discovery_geno_mat,
  target_pheno,
  target_geno_mat,
  significance_level = 0.05
)
```

Arguments

discovery_pheno	Character. Path to the phenotype file for the discovery dataset. Assumes no header and that the quantitative trait is in the third column.
discovery_geno_mat	Character. Path to the genotype matrix file for the discovery dataset. Assumes no header.
target_pheno	Character. Path to the phenotype file for the target dataset. Assumes no header and individual IDs in the second column.
target_geno_mat	Character. Path to the genotype matrix file for the target dataset. Assumes no header.
significance_level	Numeric. Significance level for confidence intervals (e.g., 0.05 for 95% CI). Default is 0.05.

Details

The function fits a multiple linear regression model (lm) using the discovery data. The estimated SNP effects and their covariance matrix are used to compute PRS and associated uncertainty for each individual in the target dataset. Confidence intervals are constructed using the normal approximation.

Missing or non-estimable coefficients and variances are set to zero.

Value

A data frame with the following columns:

IID Individual identifier (from the target phenotype file).

PRS Estimated polygenic risk score for each individual.

Variance Estimated variance of the PRS.

Lower_Limit Lower bound of the confidence interval.

Upper_Limit Upper bound of the confidence interval.

Examples

```

qpd <- system.file("Qpd.txt", package = "iPRSue", mustWork = TRUE)
qpt <- system.file("Qpt.txt", package = "iPRSue", mustWork = TRUE)
gd  <- system.file("Gd.txt",  package = "iPRSue", mustWork = TRUE)
gt  <- system.file("Gt.txt",  package = "iPRSue", mustWork = TRUE)

results <- BLUE_estimates_QT(
  discovery_pheno = qpd,
  discovery_geno_mat = gd,
  target_pheno = qpt,
  target_geno_mat = gt,
  significance_level = 0.05
)
head(results)

```

GWAS_BT

*GWAS_BT function***Description**

Performs genome-wide association study (GWAS) using plink2 logistic model and outputs the GWAS summary statistics with additive SNP effects (beta) and standard errors (se)

Usage

```
GWAS_BT(plink_path, b_file, discovery_pheno, discovery_cov, thread = 20)
```

Arguments

plink_path	Path to the PLINK executable application
b_file	Prefix of the binary files, where all .fam, .bed and .bim files have a common prefix
discovery_pheno	Name (with file extension) of the phenotype file containing family ID, individual ID and phenotype of the discovery dataset as columns, without heading
discovery_cov	Name (with file extension) of the covariate file containing family ID, individual ID, and covariate(s) of the discovery dataset as columns, without heading. If no covariate is used, have a constant column (e.g. vector of 1s)
thread	Number of threads used (default = 20)

Details

The function uses logistic regression to regress the binary phenotype (1/2 coding for controls/cases) on each SNP separately using plink 2. Then the estimated effects and standard errors are adjusted for standardization. It is optional to employ covariates in the model. If no covariate is used, create your covariate file with a constant in the 3rd column (e.g. vector of 1s).

Value

A data frame with two columns:

beta Estimated effect size (log odds) for each SNP.

se Standard error of the estimated effect size.

Examples

```
## Not run:
results <- GWAS_BT(
  plink_path = "./plink2",
  b_file = "./binary_file_prefix",
  discovery_pheno = "./discovery_phenotype_file",
  discovery_cov = "./discovery_covariate_file",
  thread = 48
)
head(gwas_results)

## End(Not run)
```

GWAS_QT

*GWAS_QT function***Description**

Performs genome-wide association study (GWAS) using plink2 linear model and outputs the GWAS summary statistics with additive SNP effects (beta) and standard errors (se)

Usage

```
GWAS_QT(plink_path, b_file, discovery_pheno, discovery_cov, thread = 20)
```

Arguments

plink_path	Path to the PLINK executable application
b_file	Prefix of the binary files, where all .fam, .bed and .bim files have a common prefix
discovery_pheno	Name (with file extension) of the phenotype file containing family ID, individual ID and phenotype of the discovery dataset as columns, without heading
discovery_cov	Name (with file extension) of the covariate file containing family ID, individual ID, and covariate(s) of the discovery dataset as columns, without heading. If no covariate is used, have a constant column (e.g. vector of 1s)
thread	Number of threads used (default = 20)

Details

The function uses linear regression to regress the quantitative phenotype on each SNP separately using plink 2. Then the estimated effects and standard errors are adjusted for standardization. The phenotype is standardized prior to analysis. It is optional to employ covariates in the model. If no covariate is used, create your covariate file with a constant in the 3rd column (e.g. vector of 1s).

Value

A data frame with two columns:

beta Estimated effect size from linear regression.

se Standard error of the effect size estimate.

Examples

```
## Not run:
results <- GWAS_QT(
  plink_path = "./plink2",
  b_file = "./binary_file_prefix",
  discovery_pheno = "./discovery_phenotype_file",
  discovery_cov = "./discovery_covariate_file",
  thread = 48
)
head(results)

## End(Not run)
```

iPRSue_estimates_BT *iPRSue_estimates_BT function*

Description

Computes individual-level polygenic risk scores (PRS) with uncertainty estimates using a simulation-based approach for binary traits. This implementation follows the iPRSue framework, simulating multiple PRSs by sampling from the GWAS effect size distribution and deriving individual-level confidence intervals.

Usage

```
iPRSue_estimates_BT(
  gwas,
  target_pheno,
  target_geno_mat,
  no_of_PRSs = 500,
  significance_level = 0.05,
  seed = NULL
)
```

Arguments

gwas	A data frame with GWAS summary statistics for binary traits. Must contain beta and se columns representing estimated SNP effect sizes and their standard errors.
target_pheno	Character. Path to the target phenotype file. Assumes no header and individual IDs in the second column.
target_geno_mat	Character. Path to the genotype matrix of target individuals. No header is expected; columns correspond to SNPs.
no_of_PRSs	Integer. Number of simulations used to construct PRS uncertainty intervals. Default is 500.
significance_level	Numeric. Significance level for confidence intervals (e.g., 0.05 gives 95% CI). Default is 0.05.
seed	Integer or NULL. Random seed for reproducibility. If NULL, results may vary across runs. Default is NULL.

Details

For each SNP, the function simulates no_of_PRSs effect sizes from a normal distribution defined by its GWAS beta and SE. These sampled betas are multiplied by the genotype matrix to generate PRS replicates for each individual. Confidence intervals are then calculated using the specified significance level.

This function is designed for binary traits and should be used with GWAS summary statistics obtained from logistic regression.

Value

A data frame containing the following columns:

IID Individual identifier (from target phenotype file).

PRS Mean of simulated PRSs for each individual.

Variance Variance across simulated PRSs.

Lower_Limit Lower bound of the confidence interval.

Upper_Limit Upper bound of the confidence interval.

Examples

```
## Not run:
# Step 1: Run GWAS on binary trait
results <- GWAS_BT(
  plink_path = "./plink2",
  b_file = "./binary_file_prefix",
  discovery_pheno = "./discovery_phenotype_file",
  discovery_cov = "./discovery_covariate_file",
  thread = 48
)

# Step 2: Estimate individual PRS with uncertainty
bpt <- system.file("Bpt.txt", package = "iPRSue", mustWork = TRUE)
gt <- system.file("Gt.txt", package = "iPRSue", mustWork = TRUE)

prs_estimates <- iPRSue_estimates_BT(
  gwas = results,
  target_pheno = bpt,
  target_geno_mat = gt,
  no_of_PRSs = 500,
  significance_level = 0.05,
  seed = 123
)
head(prs_estimates)

## End(Not run)
```

iPRSue_estimates_QT *iPRSue_estimates_QT function*

Description

Computes individual-level polygenic risk scores (PRS) with uncertainty estimates using a simulation-based approach for quantitative traits. This implementation follows the iPRSue framework, simulating multiple PRSs by sampling from the GWAS effect size distribution and deriving individual-level confidence intervals.

Usage

```
iPRSue_estimates_QT(
  gwas,
  target_pheno,
  target_geno_mat,
  no_of_PRSs = 500,
  significance_level = 0.05,
  seed = NULL
)
```

Arguments

gwas	A data frame with GWAS summary statistics for a quantitative trait. Must contain beta and se columns representing estimated SNP effect sizes and their standard errors.
target_pheno	Character. Path to the target phenotype file. Assumes no header and individual IDs in the second column.
target_geno_mat	Character. Path to the genotype matrix of target individuals. No header is expected; columns correspond to SNPs.
no_of_PRSs	Integer. Number of simulations used to construct PRS uncertainty intervals. Default is 500.
significance_level	Numeric. Significance level for confidence intervals (e.g., 0.05 gives 95% CI). Default is 0.05.
seed	Integer or NULL. Random seed for reproducibility. If NULL, results may vary across runs. Default is NULL.

Details

For each SNP, the function simulates no_of_PRSs effect sizes from a normal distribution defined by its GWAS beta and SE. These sampled betas are multiplied by the genotype matrix to generate PRS replicates for each individual. Confidence intervals are then calculated using the specified significance level.

This function is designed for quantitative traits and should be used with GWAS summary statistics obtained from linear regression.

Value

A data frame containing the following columns:

IID Individual identifier (from target phenotype file).

PRS Mean of simulated PRSs for each individual.

Variance Variance across simulated PRSs.

Lower_Limit Lower bound of the confidence interval.

Upper_Limit Upper bound of the confidence interval.

Examples

```
## Not run:
# Step 1: Run GWAS on quantitative trait
results <- GWAS_QT(
  plink_path = "./plink2",
  b_file = "./binary_file_prefix",
  discovery_pheno = "./discovery_phenotype_file",
  discovery_cov = "./discovery_covariate_file",
  thread = 48
)

# Step 2: Estimate individual PRS with uncertainty
qpt <- system.file("Qpt.txt", package = "iPRSue", mustWork = TRUE)
gt <- system.file("Gt.txt", package = "iPRSue", mustWork = TRUE)

prs_estimates <- iPRSue_estimates_QT(
  gwas = results,
  target_pheno = qpt,
  target_geno_mat = gt,
  no_of_PRSs = 500,
  significance_level = 0.05,
  seed = 123
)
head(prs_estimates)

## End(Not run)
```

Index

BLUE_estimates_BT, [2](#)
BLUE_estimates_QT, [3](#)

GWAS_BT, [5](#)
GWAS_QT, [6](#)

iPRSue_estimates_BT, [7](#)
iPRSue_estimates_QT, [8](#)