Package: GEmetrics (via r-universe)

November 2, 2024

Title Best Linear Unbiased Prediction of Genotype-by-Environment Metrics

Version 1.0.0

Description Provides functions to calculate the best linear unbiased prediction of genotype-by-environment metrics: ecovalence, environmental variance, Finlay and Wilkinson regression and Lin and Binns superiority measure, based on a multi-environment genomic prediction model.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.3.1

Imports stats, BGLR

Depends R (>= 3.6.0)

NeedsCompilation no

Author Simon Rio [aut, cre] (<https://orcid.org/0000-0001-7014-8789>), Humberto Fanelli Carvalho [aut] (<https://orcid.org/0000-0003-0745-7583>), Julio Isidro y Sanchez [aut] (<https://orcid.org/0000-0002-9044-3221>)

Maintainer Simon Rio <simon.rio@cirad.fr>

Contents

BGLR_format *Format the phenotypic response matrix for BGLR*

Description

This function formats the phenotypic data as well as the kinship matrix for BGLR

Usage

```
BGLR_format(Pheno, K)
```

Arguments

Pheno   a data frame with three columns: "Y" for phenotypic values, "Genotype" for genotype names and "Environment" for environment names. All genotypes names must be included in the set or row/column names of the "K" matrix. The number of environments (J) must be at least two

K   a square kinship or genomic relationship matrix for N genotypes whose row and column names include those of the "Genotype" column of the "Pheno" matrix

Value

a list of two elements: a "BGLR_pheno" phenotypic response matrix with J columns to be used in BGLR and the corresponding "BGLR_K" kinship matrix.

Examples

```
## Set seed for reproductibility
set.seed(123)

## Load "wheat" dataset from BGLR
data("wheat",package = "BGLR")

## Generate a design data frame for all genotypes in 5 environments
Design <- expand.grid(Genotype=rownames(wheat.A),Environment=paste0("Env",1:5))

## Set sparseness by discarding 75% of the combinations
Design <- Design[-sample(nrow(Design),round(nrow(Design)*3/4)),]

## Simulate phenotypic data with default parameter values
DataSim <- Simulate_MET_data(Design=Design,K=wheat.A)

## Generate the phenotypic response matrix for BGLR and the corresponding kinship matrix
BGLR_data <- BGLR_format(Pheno=DataSim$Pheno,K=wheat.A)
head(BGLR_data$BGLR_pheno)
```

EnvBV_blup	*Compute the best linear unbiased prediction and the conditional variance matrix of environment-specific breeding values*

Description

This function calculates the best linear unbiased prediction and the conditional variance matrix of environment-specific breeding values

Usage

```
EnvBV_blup(Pheno, K, Omega_G, Omega_E)
```

Arguments

Pheno	a data frame with three columns: "Y" for phenotypic values, "Genotype" for genotype names and "Environment" for environment names. All genotypes names must be included in the set or row and column names of the "K" matrix. The number of environments must be at least two
K	a square kinship or genomic relationship matrix for N genotypes whose row and column names include those of the "Genotype" column of the "Pheno" matrix
Omega_G	a square matrix with genetic covariances between J environments whose row and column names correspond to those of the "Environment" column of the "Pheno" matrix
Omega_E	a square matrix with error covariances between J environments whose row and column names correspond to those of the "Environment" column of the "Pheno" matrix

Value

a list of two elements: a "G_hat" matrix of best linear unbiased prediction of environment-specific breeding values for the N genotypes (as rows) in J environments (as columns), and a square conditional variance matrix "P" of environment-specific breeding values of dimension NJxNJ where each row block of size N correspond to an environment and the rows of each block correspond to genotypes

Examples

```
## Set seed for reproductibility
set.seed(123)

## Load "wheat" dataset from BGLR
data("wheat",package = "BGLR")

## Generate a design data frame for all genotypes in 5 environments
Design <- expand.grid(Genotype=rownames(wheat.A),Environment=paste0("Env",1:5))
```

```
## Set sparseness by discarding 80% of the combinations
Design <- Design[-sample(nrow(Design),round(nrow(Design)*4/5)),]

## Simulate phenotypic data with default parameter values
DataSim <- Simulate_MET_data(Design=Design,K=wheat.A)


## Calculate the blup and the conditional variance matrix using simulated variance components
## this step can take several seconds
## note that variance can also be estimated (e.g. using BGLR)
BlupEnvBV <- EnvBV_blup(Pheno=DataSim$Pheno,K=wheat.A,Omega_G=DataSim$Omega_G,
                        Omega_E=DataSim$Omega_E)

## Display results
head(BlupEnvBV$G_hat)
BlupEnvBV$P[1:5,1:5]
```

GEmetrics_blup   *Best linear unbiased prediction of genotype-by-environment (GE) metrics*

Description

This function calculates the best linear unbiased prediction of the following GE metrics: ecovalence, environmental variance, Finlay and Wilkinson regression and Lin and Binns superiority measure. Ignoring the P matrix resumes to ignoring the condition variance term in the calculation

Usage

```
GEmetrics_blup(G_hat, metric, P = NULL)
```

Arguments

G_hat	a matrix of best linear unbiased prediction of environment-specific breeding values for N genotypes (as rows) in J environments (as columns)
metric	a character string indicating what GE metric to consider: "Ecovalence", "EnvironmentalVar", "FinlayWilkRegression", "LinBinns"
P	(optional) a square conditional variance matrix of environment-specific breeding values of dimension NJxNJ, where each row block of size N correspond to an environment and the rows of each block correspond to genotypes

Value

a vector of size N with best linear unbiased prediction of the GE metric

Examples

```
## Set seed for reproductibility
set.seed(123)

## Load "wheat" dataset from BGLR
data("wheat",package = "BGLR")

## Generate a design data frame for all genotypes in 5 environments
Design <- expand.grid(Genotype=rownames(wheat.A),Environment=paste0("Env",1:5))

## Set sparseness by discarding 80% of the combinations
Design <- Design[-sample(nrow(Design),round(nrow(Design)*4/5)),]

## Simulate phenotypic data with default parameter values
DataSim <- Simulate_MET_data(Design=Design,K=wheat.A)


## Calculate the blup and the conditional variance matrix using simulated variance components
## this step can take several seconds
## note that variance can also be estimated (e.g. using BGLR)
BlupEnvBV <- EnvBV_blup(Pheno=DataSim$Pheno,K=wheat.A,Omega_G=DataSim$Omega_G,
                        Omega_E=DataSim$Omega_E)


## Compute GE metric estimates
GEmetrics_hat <- GEmetrics_blup(G_hat=BlupEnvBV$G_hat,metric="Ecovalence",P=BlupEnvBV$P)
head(GEmetrics_hat)
```

Simulate_MET_data *Simulate multi-environment trials data*

Description

This function calculates the best linear unbiased prediction and the conditional variance matrix of environment-specific breeding values

Usage

```
Simulate_MET_data(Design, K, h2 = 0.5, rho = 0.5, sd_mu = 1)
```

Arguments

Design	a data frame with two columns: "Genotype" for genotype names and "Environment" for environment names. All genotypes names must be included in the set of row/column names of the "K" matrix. The number of environments (J) must be at least two
K	a square kinship or genomic relationship matrix for N genotypes whose row/column names include those of the "Genotype" column of the "Design" matrix

h2	heritability (numeric value between 0 and 1 excluded) of observations in each environment: either a scalar to set a common heritability for all environments, or a vector of heritabilities associated with each environment of size J. The default value is a heritability of 0.5 for all environments
rho	genetic correlation (numeric value between -1 and 1 excluded) between environment pairs: either a scalar to set a common genetic correlation between all environment pairs, or a square correlation matrix of dimension JxJ. The default value is a genetic correlation of 0.5 between all environment pairs
sd_mu	standard deviation (positive numeric value) of the Gaussian distribution in which environment means are drawn. The default value is 1

Value

a list of two elements: a "Pheno" data frame consisting of the "Design" data frame to which a "Y" column containing simulated phenotypic values has been added, and a "EnvBV" matrix of dimension NxJ containing the simulated environment-specific breeding values

Examples

```
## Set seed for reproductibility
set.seed(123)

## Load "wheat" dataset from BGLR
data("wheat",package = "BGLR")

## Generate a design data frame for all genotypes in 5 environments
Design <- expand.grid(Genotype=rownames(wheat.A),Environment=paste0("Env",1:5))

## Set sparseness by discarding 75% of the combinations
Design <- Design[-sample(nrow(Design),round(nrow(Design)*3/4)),]

## Simulate phenotypic data with default parameter values
DataSim <- Simulate_MET_data(Design=Design,K=wheat.A)

## Simulated phenotypes
head(DataSim$Pheno)

## Simulated environment-specific breeding values
head(DataSim$EnvBV)

## Genetic covariance matrix between environments
DataSim$Omega_G

## Error covariance matrix between environments
DataSim$Omega_E
```

Index