Package 'heritability'

August 24, 2023

Type Package

Title Marker-Based Estimation of Heritability Using Individual Plant or Plot Data

Version 1.4

Date 2023-08-21

Author Willem Kruijer, with a contribution from Ian White (the internal function pin). Contains data collected by Padraic Flood and Rik Kooke.

Maintainer Willem Kruijer <willem.kruijer@wur.nl>

Depends R (>= 4.0), MASS (>= 3.1.20)

Suggests knitr, rmarkdown

Description

Implements marker-based estimation of heritability when observations on genetically identical replicates are available. These can be either observations on individual plants or plot-level data in a field trial. Heritability can then be estimated using a mixed model for the individual plant or plot data. For comparison, also mixed-model based estimation using genotypic means and estimation of repeatability with ANOVA are implemented. For illustration the package contains several datasets for the model species Arabidopsis thaliana.

License GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2023-08-24 07:00:02 UTC

R topics documented:

`heritability-package` *Marker-Based Estimation of Heritability Using Individual Plant or Plot Data.*

Description

The package implements marker-based estimation of heritability when observations on genetically identical replicates are available. These can be either observations on individual plants (e.g. in a growth chamber) or plot-level data in a field trial. The function `marker_h2` estimates heritability using a mixed model for the individual plant or plot data, as proposed in Kruijer *et al.* For comparison, also mixed-model based estimation using genotypic means (`marker_h2_means`) and estimation of repeatability with ANOVA (`repeatability`) are implemented. For illustration the package contains several datasets for the model species *Arabidopsis thaliana*.

Author(s)

Willem Kruijer Maintainer: Willem Kruijer <willlem.kruijer@wur.nl>

References

Kruijer, W. *et al.* (2015) Marker-based estimation of heritability in immortal populations. *Genetics*, Vol. 199(2), p. 1-20.

Examples

```
# A) marker-based estimation of heritability, given individual plant-data
# and a marker-based relatedness matrix:
data(LDV)
data(K_atwell)
# This may take up to 30 sec.
#out1 <- marker_h2(data.vector=LDV$LDV,geno.vector=LDV$genotype,
#                  covariates=LDV[,4:8],K=K_atwell)
#
# B) marker-based estimation of heritability, given genotypic means
# and a marker-based relatedness matrix:
data(means_LDV)
data(R_matrix_LDV)
data(K_atwell)
out2 <- marker_h2_means(data.vector=means_LDV$LDV,geno.vector=means_LDV$genotype,
                        K=K_atwell,Dm=R_matrix_LDV)
#
# C) estimation of repeatability using ANOVA:
data(LDV)
out3 <- repeatability(data.vector=LDV$LDV,geno.vector= LDV$genotype,
                      covariates.frame=as.data.frame(LDV[,3]))
```

BT_LW_H *Bolting time and leaf width for the Arabidopsis hapmap population.*

Description

Bolting time and leaf width for the Arabidopsis hapmap population

Usage

```
data(BT_LW_H)
```

Format

A data frame with phenotypic observations on bolting time and leaf width:

`genotype` a factor, the levels being the accession or ecotype identifiers

`BT` Bolting time, in number of days

`LW` Leaf width

`replicate` The replicate (or block) each plant is contained in (factor with levels `1` to `3`)

`rep1` numeric encoding of the factor replicate: equals 1 if the plant is in replicate 1 and 0 otherwise

`rep2` numeric encoding of the factor replicate: equals 1 if the plant is in replicate 2 and 0 otherwise

Author(s)

Willem Kruijer <willlem.kruijer@wur.nl>; experiments conducted by Rik Kooke <rik.kooke@gmail.com>

References

- Kruijer, W. *et al.* (2015) Marker-based estimation of heritability in immortal populations. *Genetics*, Vol. 199(2), p. 1-20.

See Also

For the corresponding genetic relatedness matrix, see `K_hapmap`.

Examples

```
data(BT_LW_H)
```

floweringTime *Flowering time data taken from Atwell* et al. *(2010).*

Description

Two data-frames containing individual plant data on flowering time under different conditions: LDV (Flowering time under long days and vernalization) and LD (Flowering time under long days, without vernalization).

Usage

```
data(LD); data(LDV)
```

Format

Data-frames with flowering time observations, genotype and design information:

genotype a factor, the levels being the accession or ecotype identifiers

LD Flowering time under long days, in number of days

LDV Flowering time under long days and vernalization, in number of days

replicate The replicate (or block) each plant is contained in (factor with levels 1 to 6)

rep1 numeric encoding of the factor replicate: equals 1 if the plant is in replicate 1 and 0 otherwise

rep2 numeric encoding of the factor replicate: equals 1 if the plant is in replicate 2 and 0 otherwise

rep3 numeric encoding of the factor replicate: equals 1 if the plant is in replicate 3 and 0 otherwise

rep4 numeric encoding of the factor replicate: equals 1 if the plant is in replicate 4 and 0 otherwise

rep5 numeric encoding of the factor replicate: equals 1 if the plant is in replicate 5 and 0 otherwise

Details

All plants that had not flowered by the end of the experiment were given a phenotypic value of 200. Only accessions for which SNP-data are available are included here: 167 accessions in case of LD and 168 accessions in case of LDV.

References

- Atwell, S., Y. S. Huang, B. J. Vilhjalmsson, G. Willems, M. Horton, *et al.* (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627-631.
- Kruijer, W. *et al.* (2015) Marker-based estimation of heritability in immortal populations. *Genetics*, Vol. 199(2), p. 1-20.

See Also

For the corresponding genetic relatedness matrix, see K_atwell.

Examples

```
data(LD); data(LDV)
```

`K_arabidopsis` *Marker-based relatedness matrices for 3 populations of Arabidopsis thaliana.*

Description

Marker-based relatedness matrices based on the SNP-data from Horton *et al.* (2012). Three matrices are provided: (a) `K_atwell`, for the 199 accessions studied in Atwell *et al.* (2010). (b) `K_hapmap`, for a subset of 350 accessions taken from the Arabidopsis hapmap (Li *et al.*, 2010). (c) `K_swedish`, for 304 Swedish accessions. All of these are part of the world-wide regmap of 1307 accessions, described in Horton *et al.* (2012).

Usage

```
data(K_atwell); data(K_hapmap); data(K_swedish)
```

Format

Matrices whose row- and column names are the ecotype or seed-stock IDs of the accessions.

Details

The matrices were computed using equation (2.2) in Astle and Balding (2009); see also Goddard *et al.* (2009). The `heritability`-package does not contain functions to construct relatedness matrices from genotypic data, but such functions can be found in many other software packages. For example, GCTA (Yang *et al.*, 2011), LDAK (Speed *et al.*, 2012), Fast-LMM (Lippert, 2011) and GEMMA (Zhou and Stephens, 2012).

References

- W. Astle and D.J. Balding (2009) Population Structure and Cryptic Relatedness in Genetic Association Studies. *Statistical Science*, Vol. 24, No. 4, 451-471.
- Atwell, S., Y. S. Huang, B. J. Vilhjalmsson, G. Willems, M. Horton, *et al.* (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* 465: 627-631.
- Goddard, M.E., Naomi R. Wray, Klara Verbyla and Peter M. Visscher (2009) Estimating Effects and Making Predictions from Genome-Wide Marker Data. *Statistical Science*, Vol. 24, No. 4, 517-529.
- Horton, M. W., A. M. Hancock, Y. S. Huang, C. Toomajian, S. Atwell, *et al.* (2012) Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. *Nature Genetics* 44: 212-216.
- Li, Y., Y. Huang, J. Bergelson, M. Nordborg, and J. O. Borevitz (2010) Association mapping of local climate-sensitive quantitative trait loci in arabidopsis thaliana. *PNAS* vol. 107, number 49.

- Lippert, C., J. Listgarten, Y. Liu, C.M. Kadie, R.I. Davidson, *et al.* (2011) FaST linear mixed models for genome-wide association studies. *Naure methods* 8: 833-835.
- Speed, D., G. Hemani, M. R. Johnson, and D.J. Balding (2012) Improved heritability estimation from genome-wide snps. *the American journal of human genetics* 91: 1011-1021.
- Yang, J., S.H. Lee, M.E. Goddard, and P.M. Visscher (2011) GCTA: a tool for genomewide complex trait analysis. *the American journal of human genetics* 88: 76-82.
- Zhou, X., and M. Stephens, (2012) Genome-wide efficient mixed-model analysis for association studies. *Nature genetics* 44: 821-824.

See Also

For phenotypic data for the population described in Atwell *et al.* (2010), see LD and LDV. For phenotypic data for the hapmap, see BT_LW_H and LA_H. For phenotypic data for the Swedish regmap, see LA_S.

Examples

```
data(K_atwell)
data(K_hapmap)
data(K_swedish)
```

`leafArea` Arabidopsis *leaf area data for the hapmap and Swedish regmap population.*

Description

Arabidopsis leaf area data for the hapmap and Swedish regmap population.

Usage

```
data(LA_H); data(LA_S)
```

Format

Data frame with leaf area observations:

`genotype` a factor, the levels being the accession identifiers

`LA13_H` Leaf area 13 days after sowing, in numbers of pixels (hapmap)

`LA13_S` Leaf area 13 days after sowing, in numbers of pixels (Swedish regmap)

`replicate` The replicate (or block) each plant is contained in (factor with levels 1 to 4)

`rep1` numeric encoding of the factor replicate: equals 1 if the plant is in replicate 1 and 0 otherwise

`rep2` numeric encoding of the factor replicate: equals 1 if the plant is in replicate 2 and 0 otherwise

`rep3` numeric encoding of the factor replicate: equals 1 if the plant is in replicate 3 and 0 otherwise

`x` The within image x-coordinate of the plant. A factor with levels 1 2 3

`y` The within image y-coordinate of the plant. A factor with levels `1 2 3 4`

`x1` numeric encoding of the factor `x`: equals 1 if the plant is in position 1 and 0 otherwise

`x2` numeric encoding of the factor `x`: equals 1 if the plant is in position 2 and 0 otherwise

`y1` numeric encoding of the factor `y`: equals 1 if the plant is in position 1 and 0 otherwise

`y2` numeric encoding of the factor `y`: equals 1 if the plant is in position 2 and 0 otherwise

`y3` numeric encoding of the factor `y`: equals 1 if the plant is in position 3 and 0 otherwise

Author(s)

Willem Kruijer <willlem.kruijer@wur.nl>; experiments conducted by Padraic Flood <flood@mpipz.mpg.de>

References

- Kruijer, W. *et al.* (2015) Marker-based estimation of heritability in immortal populations. *Genetics*, Vol. 199(2), p. 1-20.

See Also

For the corresponding genetic relatedness matrices, see `K_hapmap` and `K_swedish`.

Examples

```
data(LA_H); data(LA_S)
```

`marker_h2` *Compute a marker-based estimate of heritability, given phenotypic observations at individual plant or plot level.*

Description

Given a genetic relatedness matrix and phenotypic observations at individual plant or plot level, this function computes REML-estimates of the genetic and residual variance and their standard errors, using the AI-algorithm (Gilmour *et al.* 1995). Based on this, heritability estimates and confidence intervals are given (the estimator h_r^2 in Kruijer *et al.*).

Usage

```
marker_h2(data.vector, geno.vector, covariates = NULL, K, alpha = 0.05,
          eps = 1e-06, max.iter = 100, fix.h2 = FALSE, h2 = 0.5)
```

Arguments

data.vector	A vector of phenotypic observations. Needs to be of type numeric. May contain missing values.
geno.vector	A vector of genotype labels, either a factor or character. This vector should correspond to data.vector, and hence needs to be of the same length.
covariates	A data-frame or matrix with optional covariates, the rows corresponding to the phenotypic observations in data.vector and geno.vector. May contain missing values. Factors are not allowed, and need to be encoded by columns of type numeric or integer. The data-frame or matrix should not contain an intercept, which is included by default.
K	A genetic relatedness or kinship matrix, typically marker-based. Must have row- and column-names corresponding to the levels of geno.vector
alpha	Confidence level, for the 1-alpha confidence intervals.
eps	Numerical precision, used as convergence criterion in the AI-algorithm.
max.iter	Maximal number of iterations in the AI-algorithm.
fix.h2	Compute the log-likelihood and inverse AI-matrix for a fixed heritability value. Default is FALSE.
h2	When fix.h2 is TRUE, the value of the heritability. Must be of type numeric, between 0 and 1.

Details

- Given phenotypic observations Y_{ij} for genotypes $i = 1, ..., n$ and replicates $j = 1, ..., n_i$, the mixed model $Y_{ij} = \mu + G_i + E_{ij}$ is assumed. The vector of additive genetic effects $(G_1, ..., G_n)'$ follows a multivariate normal distribution with mean zero and covariance $\sigma^2_A K$, where σ^2_A is the additive genetic variance, and K is a genetic relatedness matrix derived from a dense set of markers. The errors E_{ij} are independent and normally distributed with variance σ^2_E. Under certain assumptions (see Speed *et al.* 2012) the marker- or chip-heritability $h^2 = \sigma^2_A/(\sigma^2_A + \sigma^2_E)$ equals the narrow-sense heritability.
- It is assumed that the genetic relatedness matrix K is scaled such that $trace(PKP) = n - 1$, where P is the projection matrix $I_n - 1_n 1_n'/n$, for the identity matrix I_n and 1_n being a column vector of ones. If this is not the case, K is automatically scaled prior to fitting the mixed model.
- The model can optionally include a term $X_{ij}\beta$, where X_{ij} is the row vector with observations on k extra covariates and the vector β contains their effects. In this case the argument covariates should be the (N x k) matrix or data-frame with rows X_{ij} (N being the total number of observations). Observations where either Y_{ij} or any of the covariates is missing are discarded.
- Confidence intervals for heritability are constructed using the delta-method and the inverse AI-matrix. The delta-method can be applied either directly to the function $(\sigma^2_A, \sigma^2_E) -> \sigma^2_A/(\sigma^2_A + \sigma^2_E)$ or to the function $(\sigma^2_A, \sigma^2_E) -> log(\sigma^2_A/\sigma^2_E)$. In the latter case, a confidence interval for $log(\sigma^2_A/\sigma^2_E)$ is obtained, which is back-transformed to a confidence interval for heritability. This approach (proposed in Kruijer *et al.*) has the advantage that intervals are always contained in the unit interval.
- The AI-algorithm is run for max.iter iterations. If by then there is no convergence a warning is printed and the current estimates are returned.

Value

A list with the following components:

- va: REML-estimate of the (additive) genetic variance.
- ve: REML-estimate of the residual variance.
- h2: Plug-in estimate of heritability: $va/(va + ve)$.
- conf.int1: 1-alpha confidence interval for heritability.
- conf.int2: 1-alpha confidence interval for heritability, obtained by application of the delta method on a logarithmic scale.
- inv.ai: The inverse of the average information (AI) matrix.
- loglik: The log-likelihood.

Author(s)

Willem Kruijer.

References

- Gilmour *et al.* Gilmour, A.R., R. Thompson and B.R. Cullis (1995) Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. *Biometrics*, volume 51, number 4, 1440-1450.
- Kruijer, W. *et al.* (2015) Marker-based estimation of heritability in immortal populations. *Genetics*, Vol. 199(2), p. 1-20.
- Speed, D., G. Hemani, M. R. Johnson, and D.J. Balding (2012) Improved heritability estimation from genome-wide snps. *the American journal of human genetics* 91: 1011-1021.

See Also

For marker-based estimation of heritability using genotypic means, see `marker_h2_means`.

Examples

```
data(LD)
data(K_atwell)
# Heritability estimation for all observations:
#out <- marker_h2(data.vector=LD$LD,geno.vector=LD$genotype,
#                 covariates=LD[,4:8],K=K_atwell)
# Heritability estimation for a randomly chosen subset of 20 accessions:
set.seed(123)
sub.set <- which(LD$genotype %in% sample(levels(LD$genotype),20))
out <- marker_h2(data.vector=LD$LD[sub.set],geno.vector=LD$genotype[sub.set],
                 covariates=LD[sub.set,4:8],K=K_atwell)
```

`marker_h2_means` *Compute a marker-based estimate of heritability, given genotypic means.*

Description

Given a genetic relatedness matrix and genotypic means, this function computes REML-estimates of the genetic and residual variance and their standard errors, using the AI-algorithm (Gilmour *et al.* 1995). Based on this, heritability estimates and confidence intervals are given (the estimator h^2_m in Kruijer *et al.*).

Usage

```
marker_h2_means(data.vector, geno.vector, K, Dm=NULL, alpha = 0.05, eps = 1e-06,
       max.iter = 100, fix.h2 = FALSE, h2 = 0.5, grid.size=99)
```

Arguments

`data.vector`	A vector of phenotypic observations, typically genotypic means. Needs to be of type numeric. May contain missing values.
`geno.vector`	A vector of genotype labels, either a factor or character. This vector should correspond to `data.vector`, and hence needs to be of the same length.
`K`	A genetic relatedness or kinship matrix, typically marker-based. Must have row- and column-names corresponding to the levels of `geno.vector`
`Dm`	Covariance of the genotypic means contained in data.vector; see details. Should be of class matrix, with row- and column-names corresponding to the levels of `geno.vector`
`alpha`	Confidence level, for the 1-alpha confidence intervals.
`eps`	Numerical precision, used as convergence criterion in the AI-algorithm.
`max.iter`	Maximal number of iterations in the AI-algorithm.
`fix.h2`	Compute the log-likelihood and inverse AI-matrix for a fixed heritability value. Default is `FALSE`.
`h2`	When `fix.h2` is `TRUE`, the value of the heritability. Must be of type numeric, between 0 and 1.
`grid.size`	If the AI-algorithm has not converged after `max.iter` iterations, the likelihood is computed on the grid of heritability values 1/(`grid.size`+1),...,`grid.size`/(`grid.size`+1); see details.

Details

- Given phenotypic observations Y_i for genotypes $i = 1, ..., n$, the mixed model $Y_i = \mu + G_i + E_i$ is assumed. Typically, the Y_i are genotypic means or BLUEs obtained from fitting a linear (mixed) model to the raw data, containing several plants or plots for each genotype. The vector of additive genetic effects $(G_1, ..., G_n)'$ follows a multivariate normal distribution with

mean zero and covariance $\sigma^2_A K$, where σ^2_A is the additive genetic variance, and K is a genetic relatedness matrix derived from a dense set of markers. The vector of errors $(E_1, ..., E_n)'$ follows a multivariate normal distribution with mean zero and covariance $\sigma^2_E D_m$, where D_m is the covariance of the means obtained from the initial analysis. In case of a completely randomized design with r_i replicates for genotypes $i = 1, ..., n$, D_m is diagonal with elements $1/r_i$. Under certain assumptions (see Speed *et al.* 2012) the marker- or chip-heritability $h^2 = \sigma^2_A/(\sigma^2_A + \sigma^2_E)$ equals the narrow-sense heritability.

- As in the marker_h2 function, it is assumed that the genetic relatedness matrix K is scaled such that $trace(PKP) = n-1$, where P is the projection matrix $I_n - 1_n 1_n'/n$, for the identity matrix I_n and 1_n being a column vector of ones. If this is not the case, K is automatically scaled prior to fitting the mixed model.
- No covariates can be included, as it is assumed that these are available at plant- or plot level, and accounted for in the genotypic means.
- The resulting heritability estatimes are less accurate than those obtained from individual plant or plot data, and the likelihood can be monotone in $h^2 = \sigma^2_A/(\sigma^2_A + \sigma^2_E)$. If the AI-algorithm has not converged after `max.iter` iterations, the likelihood is computed on the grid of heritability values 1/(`grid.size`+1),...,`grid.size`/(`grid.size`+1)
- As in the marker_h2 function, confidence intervals for heritability are constructed using the delta-method and the inverse AI-matrix. The delta-method can be applied either directly to the function $(\sigma^2_A, \sigma^2_E) -> \sigma^2_A/(\sigma^2_A + \sigma^2_E)$ or to the function $(\sigma^2_A, \sigma^2_E) -> log(\sigma^2_A/\sigma^2_E)$. In the latter case, a confidence interval for $log(\sigma^2_A/\sigma^2_E)$ is obtained, which is back-transformed to a confidence interval for heritability. This approach (proposed in Kruijer *et al.*) has the advantage that intervals are always contained in the unit interval.

Value

A list with the following components:

- va: REML-estimate of the (additive) genetic variance.
- ve: REML-estimate of the residual variance.
- h2: Plug-in estimate of heritability: $va/(va + ve)$.
- conf.int1: 1-alpha confidence interval for heritability.
- conf.int2: 1-alpha confidence interval for heritability, obtained by application of the delta method on a logarithmic scale.
- inv.ai: The inverse of the average information (AI) matrix.
- loglik: The log-likelihood.
- loglik.vector: Empty numeric vector if the AI-algorthm converged within $max.iter$ iterations. Otherwise it contains the log-likelihood on a grid.

Author(s)

Willem Kruijer.

References

- Gilmour *et al.* Gilmour, A.R., R. Thompson and B.R. Cullis (1995) Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. *Biometrics*, volume 51, number 4, 1440-1450.
- Kruijer, W. *et al.* (2015) Marker-based estimation of heritability in immortal populations. *Genetics*, Vol. 199(2), p. 1-20.
- Speed, D., G. Hemani, M. R. Johnson, and D.J. Balding (2012) Improved heritability estimation from genome-wide snps. *the American journal of human genetics* 91: 1011-1021.

See Also

For marker-based estimation of heritability using individual plant or plot data, see `marker_h2`.

Examples

```
data(means_LDV)
data(R_matrix_LDV)
data(K_atwell)
out <- marker_h2_means(data.vector=means_LDV$LDV,geno.vector=means_LDV$genotype,
                       K=K_atwell,Dm=R_matrix_LDV)
# Takes about a minute:
#data(means_LD)
#data(R_matrix_LD)
#out <- marker_h2_means(data.vector=means_LD$LD,geno.vector=means_LD$genotype,
#                       K=K_atwell,Dm=R_matrix_LD)
# The likelihood is monotone increasing:
#plot(x=(1:99)/100,y=out$loglik.vector,type="l",ylab="log-likelihood",lwd=2,
#     main='',xlab='h2',cex.lab=2,cex.axis=2.5)
```

`means_floweringTime` *Flowering time from Atwell* et al. *(2010): accession means.*

Description

Accession means for the flowering time data contained in `LD` and `LDV`.

Usage

```
data(means_LD); data(means_LDV)
```

Format

Data-frames with flowering time means:

`genotype` a factor, the levels being the accession or ecotype identifiers

`LD` Flowering time under long days, in number of days

`LDV` Flowering time under long days and vernalization, in number of days

Details

Following Kruijer *et al.* (appendix A) these means were defined as the least-squares estimate for the factor accession, in a linear model containing both accession and replicate effects. Consequently there are differences compared to Atwell *et al.* (2010), where just the arithmetic averages are considered.

References

- Atwell, S., Y. S. Huang, B. J. Vilhjalmsson, G. Willems, M. Horton, *et al.* (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627-631.
- Kruijer, W. *et al.* (2015) Marker-based estimation of heritability in immortal populations. *Genetics*, Vol. 199(2), p. 1-20.

See Also

Together with the covariance matrices contained in `R_matrix_LD` and `R_matrix_LDV`, the means contained in `means_LD` and `means_LDV` can be used to estimate heritability, using the function `marker_h2_means`. For the corresponding genetic relatedness matrix, see `K_atwell`. For the individual plant data, see `floweringTime`.

Examples

```
data(means_LD)
data(means_LDV)
```

`repeatability` *ANOVA-based estimates of repeatability*

Description

Given a population where each genotype is phenotyped for a number of genetically identical replicates (either individual plants or plots in a field trial), the repeatability or intra-class correlation can be estimated by $V_g/(V_g+V_e)$, where $V_g = (MS(G) - MS(E))/r$ and $V_e = MS(E)$. In these expressions, r is the number of replicates per genotype, and $MS(G)$ and $MS(E)$ are the mean sums of squares for genotype and residual error obtained from analysis of variance. In case $MS(G) < MS(E)$, V_g is set to zero. See Singh *et al.* (1993) or Lynch and Walsh (1998), p.563. When the genotypes have differing numbers of replicates, r is replaced by $\bar{r} = (n-1)^{-1}(R_1 - R_2/R_1)$, where $R_1 = \sum r_i$ and $R_2 = \sum r_i^2$. Under the assumption that all differences between genotypes are genetic, repeatability equals broad-sense heritability; otherwise it only provides an upper-bound for broad-sense heritability.

Usage

```
repeatability(data.vector, geno.vector, line.repeatability = FALSE,
              covariates.frame = data.frame())
```

Arguments

`data.vector` A vector of phenotypic observations. Needs to be of type numeric. May contain missing values.

`geno.vector` A vector of genotype labels, either a factor or character. This vector should correspond to `data.vector`, and hence needs to be of the same length.

`line.repeatability`
If `TRUE`, the line-repeatability or line-heritability $\sigma^2_G/(\sigma^2_G + \sigma^2_E/r)$ is estimated, otherwise (the default) the repeatability at plot- or plant level, which is $\sigma^2_G/(\sigma^2_G + \sigma^2_E)$.

`covariates.frame`
A data-frame with additional covariates, the rows corresponding to `geno.vector` and the phenotypic observations in `data.vector`. May contain missing values. Each column can be numeric or a factors.

Value

A list with the following components:

- repeatability: the estimated repeatability.
- gen.variance: the estimated genetic variance.
- res.variance: the estimated residual variance.
- line.repeatability: whether repeatability was estimated at the individual plant or plot level (the default), or at the level of genotypic means (in the latter case, `line.repeatability=TRUE`)
- average.number.of.replicates: The average number of replicates. See the description above.
- conf.int: Confidence interval for repeatability. See Singh *et al.* (1993) or Lynch and Walsh (1998)

Author(s)

Willem Kruijer <willem.kruijer@wur.nl>

References

- Kruijer, W. *et al.* (2015) Marker-based estimation of heritability in immortal populations. *Genetics*, Vol. 199(2), p. 1-20.
- Lynch, M., and B. Walsh (1998) Genetics and Analysis of Quantitative Traits. Sinauer Associates, 1st edition.
- Singh, M., S. Ceccarelli, and J. Hamblin (1993) Estimation of heritability from varietal trials data. *Theoretical and Applied Genetics* 86: 437-441.

Examples

```
repeatability(data.vector=rep(rnorm(26),each=5) + rnorm(5*26),
              geno.vector=rep(letters,each=5))
```

R_matrix *Covariance matrix of the accession means for flowering time.*

Description

Covariance matrices of the accession means for flowering time contained in means_LD and means_LDV, derived from the Atwell *et al.* (2010) data.

Usage

```
data(R_matrix_LDV);data(R_matrix_LD)
```

Format

Matrix whose row- and column names are the ecotype-IDs of the accessions contained in LD and LDV.

Details

The matrix was computed as in Kruijer *et al.*, Appendix A.

References

- Atwell, S., Y. S. Huang, B. J. Vilhjalmsson, G. Willems, M. Horton, *et al.* (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627-631.
- Kruijer, W. *et al.* (2015) Marker-based estimation of heritability in immortal populations. *Genetics*, Vol. 199(2), p. 1-20.

See Also

Together with the corresponding means contained in means_LD and means_LDV, these matrices can be used to estimate heritability, using the function marker_h2_means.

Examples

```
data(R_matrix_LD); data(R_matrix_LDV)
```

Index