Package 'st4gi'

April 27, 2015

Type Package

Title Statistical tools for genetic improvement

Version 0.2.1

Date 2015-01-23

Author Raul Eyzaguirre

Maintainer Raul Eyzaguirre <R.EYZAGUIRRE@CGIAR.ORG>

Description Statistical tools for the analysis of experimental data for crop genetic improvement

Depends R (>= 3.1.2)

Imports lme4, shiny

License MIT + file LICENSE

LazyData true

Suggests testthat

NeedsCompilation no

R topics documented:

ammi 2
ammigxe 3
checkdata01 5
checkdata02 5
domeans 6
elston 7
met8x12 8
msdplot 8
mveb 9
mvemet 10
pesekbaker 11
pjpz09 13
rsa 13
rts1 14
rts2 15
rts3 16
spconsis 16
spg 19
tai 19

Index **21**

`ammi` *AMMI or GGE with data at plot level*

Description

This function runs AMMI (Gollob, H. R., 1968) or GGE biplot (Yan , W. et al., 2000) with data at plot level.

Usage

```
ammi(trait, geno, env, rep, data, method = "AMMI", f = 0.5, biplot = 2,
  biplot1 = "effects", title = NULL, xlab = NULL,
  color = c("darkorange", "black", "gray"), ...)
```

Arguments

`trait`	The trait to analyze.
`geno`	The genotypes.
`env`	The environments.
`rep`	The replications or blocks. A RCBD is assumed.
`data`	The name of the data frame containing the data.
`method`	AMMI or GGE.
`f`	Scaling factor, defaults to 0.5.
`biplot`	Choose 1 for the trait-PC1 biplot and 2 for the PC1-PC2 biplot.
`biplot1`	Choose "effects" or "means" for biplot1.
`title`	Main title for biplot1 or biplot2.
`xlab`	Xlab for biplot1.
`color`	Color for lines, symbols and/or labels for environments, genotypes and axes.
`...`	Additional graphic parameters.

Details

Significance of PCs are evaluated only with `method = "AMMI"` and if the data are balanced.

Value

It returns the first and second PC values for genotypes and environments, a table with the contribution of each PC, a dispersion plot of means or effects against the first PC, or a dispersion plot of PC1 against PC2. Significance of PCs are included in the contributions table only if method is set to AMMI.

Author(s)

Raul Eyzaguirre

References

Gollob, H. R. (1968). A Statistical Model which combines Features of Factor Analytic and Analysis of Variance Techniques, Psychometrika, Vol 33(1): 73-114.

Yan, W. et al. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot, Crop Sci., Vol 40: 597-605.

See Also

svd

Examples

```
# The data
head(met8x12)
str(met8x12)

# Run AMMI for trait y, biplot1 by default
ammi("y", "geno", "env", "rep", met8x12)

# Run AMMI for trait y, biplot2
ammi("y", "geno", "env", "rep", met8x12, biplot = 2)
```

---

ammigxe *AMMI or GGE with data from an interaction means matrix*

Description

This function runs AMMI (Gollob, H. R., 1968) or GGE biplot (Yan , W. et al., 2000) with data from an interaction means matrix.

Usage

```
ammigxe(int.mean, trait = NULL, rep.num = NULL, rdf = NULL, rms = NULL,
  method = "AMMI", f = 0.5, biplot = 2, biplot1 = "effects",
  title = NULL, xlab = NULL, color = c("darkorange", "black", "gray"),
  ...)
```

Arguments

int.mean	GxE means matrix, genotypes in rows, environments in columns.
trait	Name of the trait.
rep.num	Number of replications.
rdf	Residual degrees of freedom.
rms	Residual mean square.
method	AMMI or GGE.
f	Scaling factor, defaults to 0.5.
biplot	1 for the trait-PC1 biplot and 2 for the PC1-PC2 biplot.
biplot1	Choose "effects" or "means" for biplot1.
title	Main title for biplot1 or biplot2.

xlab	Xlab for biplot1.
color	Color for lines, symbols and/or labels for environments, genotypes and axes.
...	Additional graphic parameters.

Details

Significance of PCs are evaluated only with `method = "AMMI"` and if `rep.num`, `rms` and `rdf` are specified.

Value

It returns the first and second PC values for genotypes and environments, a table with the contribution of each PC, a dispersion plot of means or effects against the first PC, or a dispersion plot of PC1 against PC2. Significance of PCs are included in the contributions table only if `rep.num`, `rms` and `rdf` are specified and method is set to AMMI.

Author(s)

Raul Eyzaguirre

References

Gollob, H. R. (1968). A Statistical Model which combines Features of Factor Analytic and Analysis of Variance Techniques, Psychometrika, Vol 33(1): 73-114.

Yan, W. et al. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot, Crop Sci., Vol 40: 597-605.

See Also

`svd`

Examples

```
# The data
head(met8x12)
str(met8x12)

# Compute GxE means
int.mean <- tapply(met8x12$y, list(met8x12$geno, met8x12$env), mean, na.rm = TRUE)

# Run AMMI with GxE means matrix, biplot2
ammigxe(int.mean, trait = "y", biplot = 2)

# Run GGE with GxE means matrix, biplot2
ammigxe(int.mean, trait = "y", method = "GGE", biplot = 2)
```

checkdata01 *Check data for a RCBD*

Description

This function checks the frequencies of genotypes in a RCBD.

Usage

```
checkdata01(trait, geno, data)
```

Arguments

trait	The trait to analyze.
geno	The genotypes.
data	The name of the data frame.

Details

This function checks if there is more than one replication in a RCBD, if there is any genotype without data, and if the design is balanced.

Value

c1, c2, c3, three control values.

Author(s)

Raul Eyzaguirre.

checkdata02 *Check data for a MET in a RCBD*

Description

This function checks the frequencies of genotypes in each environment in a RCBD.

Usage

```
checkdata02(trait, geno, env, data)
```

Arguments

trait	The trait to analyze
geno	The genotypes
env	The environments
data	The name of the data frame

Details

This function checks if there is more than one replication in a RCBD in several environments, if there is any genotype without data for some specific environments, and if the design is balanced.

Value

c1, c2, c3, three control values

Author(s)

Raul Eyzaguirre

`domeans` *Compute means over some factors*

Description

Compute means for several traits for some specific factors.

Usage

```
domeans(traits, factors, addcol = NULL, data)
```

Arguments

`traits`	List of traits to compute means.
`factors`	List of factors.
`addcol`	Additional columns to keep.
`data`	The name of the data frame containing the data.

Details

This function computes means for all the `traits` for each level's combination of the `factors`. Additional columns can be kept if specified in `addcol`.

Value

It returns a data frame with the means.

Author(s)

Raul Eyzaguirre

Examples

```
# The data
head(spg)
str(spg)

# Compute means for all the traits across the two replications
# for each genotype and location.
traits <- c('rytha', 'bc', 'dm', 'star', 'nocr')
factors <- c('geno', 'loc')
domeans(traits, factors, data=spg)

# Save the output in a data.frame with name 'output1'
# and compute means for each genotype across the two locations.
output1 <- domeans(traits, factors, data=spg)
domeans(traits, 'geno', data=output1)
```

`elston` *Elston Index*

Description

Function to compute the Elston index (Elston, R. C., 1963).

Usage

```
elston(traits, geno, data, lb = 1)
```

Arguments

`traits`	List of traits.
`geno`	The genotypes.
`data`	The name of the data frame containing the data.
`lb`	Lower bound. 1 for k = min(x) and 2 for k = (n*min(x) - max(x))/(n-1)

Details

The Elston index is a weight free index.

Value

It returns the Elston index value and the Elston index value sorted in descending order.

Author(s)

Raul Eyzaguirre

References

Elston, R. C. (1963). A weight-free index for the purpose of ranking or selection with respect to several traits at a time. Biometrics. 19(1): 85-97.

Examples

```
# The data
head(spg)
str(spg)

# Run Elston index with all the traits
elston(c("rytha", "bc", "dm", "star", "nocr"), "geno", spg)
```

`met8x12` *Yields for a multi-environment trial (MET)*

Description

This data set gives the yields per plot for 8 genotypes in 12 environments with a RCBD with 3 blocks in each environment.

Usage

```
met8x12
```

Format

A data frame with 4 columns and 288 rows.

Source

International Potato Center, sweetpotato experimental data.

`msdplot` *Plot means and standard deviations*

Description

Function to plot means and confidence limits.

Usage

```
msdplot(trait, groups, data, conf = 0.95, sort.means = "none",
  main.title = NULL, x.title = "groups", y.title = "", col = "black",
  bg = "darkorange", col.lines = "black")
```

Arguments

trait	The trait to analyze.
groups	The grouping factor.
data	The name of the data frame containing the data.
conf	Probability for the confidence limits or number of standard deviations.
sort.means	Sort for means, "none" by default.
main.title	Main title.
x.title	Title for x axis.
y.title	Title for y axis.
col	Line color for circles.
bg	Background color for circles.
col.lines	Line color for confidenced interval lines.

Details

An alternative to the controversial dynamite plots. If conf is set to a value greater than or equal to 1, then it is interpreted as number of standard deviations.

Value

It returns a plot with the means and confidence limits for each group.

Author(s)

Raul Eyzaguirre

Examples

```
# Simulate some data
mydata <- data.frame(y = rnorm(100, sample(80:120, 10), sample(10:20, 10)),
                     g = rep(1:10, 10))

# Draw the plot
msdplot("y", "g", mydata)
```

---

mveb *Estimation of missing values for a RCBD*

Description

Function to estimate missing values for a Randomized Complete Block Design (RCBD) by the least squares method.

Usage

```
mveb(trait, geno, rep, data, maxp = 0.05, tol = 1e-06)
```

Arguments

trait	The trait to estimate missing values.
geno	The genotypes.
rep	The replications or blocks. A RCBD is assumed.
data	The name of the data frame.
maxp	Maximum allowed proportion of missing values to estimate, defaults to 5%.
tol	Tolerance for the convergence of the iterative estimation process.

Details

A data.frame with data for a RCBD with at least two replications and at least one datum for each treatment must be loaded. Experimental data with only one replication, any treatment without data, or more missing values than specified in maxp will generate an error message.

Value

It returns a data frame with name new.data and the number est.num and proportion est.prop of estimated missing values. The new.data data frame contains the experimental layout and columns trait and trait.est with the original data and the original data plus the estimated values.

Author(s)

Raul Eyzaguirre.

Examples

```
# The data
head(met8x12)
str(met8x12)

# Choose one environment
temp <- subset(met8x12, env=="TM80N")

# Missing value in the first row
head(temp)

# Estimate the missing value
mveb("y", "geno", "rep", temp)
```

---

mvemet *Estimation of missing values for a MET in a RCBD*

Description

Function to estimate missing values for a Multi Environment Trial (MET) with a Randomized Complete Block Design (RCBD) by the least squares method.

Usage

```
mvemet(trait, geno, env, rep, data, maxp = 0.05, tol = 1e-06)
```

Arguments

trait	The trait to estimate missing values.
geno	The genotypes.
env	The environments.
rep	The replications or blocks. A RCBD is assumed.
data	The name of the data frame.
maxp	Maximum allowed proportion of missing values to estimate, defaults to 5%.
tol	Tolerance for the convergence of the iterative estimation process.

Details

A data.frame with data for a MET in a RCBD with at least two replications and at least one datum for each treatment must be loaded. Experimental data with only one replication, any treatment without data, or more missing values than specified in maxp will generate an error message.

Value

It returns a data frame with name new.data and the number est.num and proportion est.prop of estimated missing values. The new.data data frame contains the experimental layout and columns trait and trait.est with the original data and the original data plus the estimated values.

Author(s)

Raul Eyzaguirre.

Examples

```
# The data
head(met8x12)
str(met8x12)

# Estimate the missing values
mvemet("y", "geno", "env", "rep", met8x12)
```

---

pesekbaker *Pesek-Baker Index*

Description

Function to compute the Pesek-Baker index (Pesek, J. and R.J. Baker., 1969).

Usage

```
pesekbaker(traits, geno, env, rep, data, dgg = NULL, units = "sdu",
  sf = 0.1)
```

Arguments

traits	List of traits.
geno	The genotypes.
env	The environments.
rep	The replications or blocks.
data	The name of the data frame containing the data.
dgg	Desired genetic gains, defaults to one standard deviation for each trait.
units	Units for dgg, "actual" or "sdu". See details for more information.
sf	Selected fraction, defaults to 0.1.

Details

The Pesek-Baker is an index where relative economic weights have been replaced by desired gains. If dgg is not specified, the standard deviations of the traits are used. It means that the desired genetic gains are equal to one standard deviation for each trait. dgg can be specified in actual units (units = "actual") or in standard deviations (units = "sdu"), defaults to "sdu". For example, if you have a trait which is expressed in kilograms and with a standard deviation of 5 kilograms, typing dgg = 2 means a desired genetic gain of 2 standard deviations that corresponds to 10 kilograms. If you type dgg = 2 and units = "actual" then this means a desired genetic gain of 2 kilograms. If dgg = NULL then the desired genetic gain will be one standard deviation, no matter if units is set as "actual" or "sdu". To compute the index the package lme4 is needed.

Value

It returns:

- $Desired.Genetic.Gains, the desired genetic gains in actual units,
- $Standard.Deviations, the estimated standard deviations,
- $Genetic.Variances, the estimated genetic variances,
- $Correlation.Matrix, the estimated correlation matrix,
- $Index.Coefficients, the index coefficients,
- $Response.to.Selection, the response to selection,
- $Std.Response.to.Selection, the standardized response to selection,
- $Pesek.Baker.Index, the Pesek-Baker index value, and
- $Sorted.Pesek.Baker.Index the Pesek-Baker index value sorted in descending order.

Author(s)

Raul Eyzaguirre

References

Pesek, J. and R.J. Baker.(1969). Desired improvement in relation to selection indices. Can. J. Plant. Sci. 9:803-804.

Examples

```
# The data
head(spg)
str(spg)

# Run Pesek-Baker index with all the traits
pesekbaker(c("rytha", "bc", "dm", "star", "nocr"), "geno", "loc", "rep", spg)

# Use different desired genetic gains for each trait,
# more weight on bc and dm, less on star and nocr.
pesekbaker(c("rytha", "bc", "dm", "star", "nocr"), "geno", "loc", "rep", spg,
           dgg = c(1, 1.5, 1.5, 0.8, 0.8))
```

---

pjpz09 *Data for a yield trial*

Description

This data set contains data for several traits for a one location experiment.

Usage

```
pjpz09
```

Format

A data frame with 10 columns and 204 rows.

Source

International Potato Center, sweetpotato experimental data.

---

rsa *Regression Stability Analysis*

Description

Function to run the regression stability analysis (Yates and Cochran, 1938, Finlay and Wilkinson, 1963). This implementation follows the formulas of Eberhart and Russell (1966).

Usage

```
rsa(trait, geno, env, rep, data, maxp = 0.05)
```

Arguments

trait	The trait to analyze.
geno	The genotypes.
env	The environments.
rep	The replications or blocks.
data	The name of the data frame containing the data.
maxp	Maximum allowed proportion of missing values to estimate, defaults to 5%.

Details

The regression stability analysis is evaluated with a balanced data set. If data is unbalanced, missing values are estimated up to an specified maximum proportion, 5% by default. To run a regression stability analysis you need a set of genotypes evaluated in a set of environments. At least 3 genotypes or environments are needed. In a regression stability analysis for genotypes grown at several environments, for each genotype a simple linear regression of individual yield (Y) on the mean yield of all genotypes for each environment (X) is fitted. In a similar way, for each environment, a simple linear regression of individual yield (Y) on the mean yield of all environments for each genotype (X) is fitted.

Value

It returns the regression stability analysis for genotypes and environments. It also returns the coefficient of variation.

Author(s)

Raul Eyzaguirre

References

Eberhart, S. A. and Russell, W. A. (1966). Stability Parameters for Comparing Varieties. Crop Sci. 6: 36-40.

Finlay, K. W., and Wilkinson, G. N. (1963). The Analysis of Adaption in a Plant-Breeding Programme. Aust. J. Agric. Res. 14: 742-754.

Yates, F., and Cochran, W. G. (1938). The Analysis of Group Experiments. J. Agric. Sci. 28: 556-580.

Examples

```
# The data
head(met8x12)
str(met8x12)

# Run regression stability analysis
rsa("y", "geno", "env", "rep", met8x12)
```

---

rts1 *Response to selection for a single experiment*

Description

It finds the optimum number of replications to get the maximum response to selection for a single experiment for a given plot capacity, number of selected genotypes, genotypic variance and error variance.

Usage

```
rts1()
```

Details

It uses package `shiny` for the web layout. Type `rts1()` in the R console to run the app.

Value

It returns a plot of response to seleccion versus number of replications and computes the optimum number of replications and the response to selection at this optimum value.

Author(s)

Raul Eyzaguirre.

`rts2` *Response to selection with several locations*

Description

It finds the optimum number of replications to get the maximum response to selection with several locations for a given plot capacity, number of locations, number of selected genotypes, genotypic variance, genotypic by location variance, and error variance.

Usage

```
rts2()
```

Details

It uses package `shiny` for the web layout. Type `rts2()` in the R console to run the app.

Value

It returns a plot of response to seleccion versus number of replications and computes the optimum number of replications and the response to selection at this optimum value.

Author(s)

Raul Eyzaguirre.

rts3 *Response to selection with several locations and years*

Description

It finds the optimum number of replications to get the maximum response to selection with several locations and years for a given plot capacity, number of locations, number of years, number of selected genotypes, genotypic variance, genotypic by location variance, genotypic by year variance, genotypic by location by year variance, and error variance.

Usage

```
rts3()
```

Details

It uses package `shiny` for the web layout. Type `rts3()` in the R console to run the app.

Value

It returns a plot of response to seleccion versus number of replications and computes the optimum number of replications and the response to selection at this optimum value.

Author(s)

Raul Eyzaguirre.

spconsis *Check consistency for sweetpotato experimental data*

Description

Set of rules to check for consistency of sweetpotato experimental data. Data labels must be defined as specified in the PROCEDURES FOR THE EVALUATION AND ANALYSIS OF SWEETPOTATO TRIALS document.

Usage

```
spconsis(data, plot.size, width = 240)
```

Arguments

`data`	The name of the data frame.
`plot.size`	Plot size in square meters.
`width`	Number of columns for the output file.

Details

The data frame must use the following labels:

- L : Locations
- Y : Year
- S : Season
- G : Genotypes
- NAME : Names for genotypes
- E : Environments
- R : Replications or blocks
- NOPS : Number of plants planted
- NOPE : Number of plants established
- VIR1 : Virus symptoms (1-9), first evaluation
- VIR2 : Virus symptoms (1-9), second evaluation
- VIR3 : Virus symptoms (1-9), third evaluation
- ALT1 : Alternaria symptoms (1-9), first evaluation
- ALT2 : Alternaria symptoms (1-9), second evaluation
- VV1 : Vine vigor (1-9), first evaluation
- VV2 : Vine vigor2 (1-9), second evaluation
- VW : Vine weight
- NOPH : Number of plants harvested
- NOPR : Number of plants with roots
- NOCR : Number of commercial roots
- NONC : Number of non commercial roots
- CRW : Commercial root weight
- NCRW : Non commercial root weight
- RFCP : Root primary flesh color
- RFCS : Root secondary flesh color
- SCOL : Storage root skin color
- FCOL : Storage root flesh color
- RS : Root size (1-9)
- RF : Root form (1-9)
- DAMR : Root defects (1-9)
- RSPR : Root sprouting (1-9)
- WED1 : Weevil damage (1-9), first evaluation
- WED2 : Weevil damage2 (1-9), second evaluation
- DMF : Fresh weight of roots for dry matter assessment
- DMD : Dry weight of DMF samples
- DMVF : Fresh weight vines for dry matter assessment
- DMVD : Dry weight of DMVF samples
- DM : Storage root dry matter content (

- DMFY : Dry matter foliage yield
- DMRY : Dry matter root yield
- FRAW1 : Root fiber (1-9), first determination
- SURAW1 : Root sugar (1-9), first determination
- STRAW1 : Root starch (1-9), first determination
- COOF1 : Cooked fiber (1-9), first evaluation
- COOSU1 : Cooked sugars (1-9), first evaluation
- COOST1 : Cooked starch (1-9), first evaluation
- COOT1 : Cooked taste (1-9), first evaluation
- COOAP1 : Cooked appearance (1-9), first evaluation
- FRAW2 : Root fiber (1-9), second determination
- SURAW2 : Root sugar (1-9), second determination
- STRAW2 : Root starch (1-9), second determination
- COOF2 : Cooked fiber (1-9), second evaluation
- COOSU2 : Cooked sugars (1-9), second evaluation
- COOST2 : Cooked starch (1-9), second evaluation
- COOT2 : Cooked taste (1-9), second evaluation
- COOAP2 : Cooked appearance (1-9), second evaluation
- PROT : Protein
- FE : Iron in dry weight
- ZN : Zinc in dry weight
- CA : Calcium in dry weight
- MG : Magnesium in dry weight
- BC : Beta-carotene in dry weight (NIRS)
- BC.CC : Beta-carotene with color charts
- TC : Total carotenoids in dry weight (NIRS)
- STAR : Starch
- FRUC : Fructose
- GLUC : Glucose
- SUCR : Sucrose
- MALT : Maltose
- TRW : Total root weight
- CYTHA : Commercial root yield t/ ha
- RYTHA : Total root yield t/ha
- ACRW : Average commercial root weight = CRW/NOCR
- NRPP : Number of roots per plant
- YPP : Yield per plant Kg
- CI : Percent marketable roots (commercial index)
- HI : Harvest index
- SHI : Harvest sowing index (survival)
- BIOM : Biomass yield
- FYTHA : Foliage total yield t/ha
- RFR : Root foliage ratio

Value

It returns a file with name checks.txt with a list of all rows with some kind of inconsistency and all rows with outliers.

Author(s)

Raul Eyzaguirre.

Examples

```
# The data
 head(pjpz09)
 str(pjpz09)

 # Check the data
 spconsis(pjpz09, 4.5)
```

---

spg *Some traits for a multi-environment trial (MET)*

Description

This data set has data for root yield in tons per hectare (rytha), beta-carotene (bc), dry matter (dm), starch (star) and number of commercial roots (nocr) for 8 genotypes in 2 locations with a RCBD with 2 blocks in each location.

Usage

```
spg
```

Format

A data frame with 8 columns and 32 rows.

Source

International Potato Center, sweetpotato experimental data.

---

tai *Tai's stability analysis*

Description

This function runs Tai's stability analysis (Tai, G. C. C., 1971). It assumes a RCBD with fixed effects for genotypes and random effects for environments.

Usage

```
tai(trait, geno, env, rep, data, conf = 0.95, title = NULL,
  color = c("darkorange", "black", "gray"), ...)
```

Arguments

trait	The trait to analyze.
geno	The genotypes.
env	The environments.
rep	The replications or blocks. A RCBD is assumed.
data	The name of the data frame containing the data.
conf	Probability for the Tai limits.
title	Main title for plot.
color	Color for symbols, labels and lines.
...	Additional graphic parameters.

Details

The limits for alpha and lambda are computed using the mean squares from an ANOVA table for a RCBD with blocks nested into environments. If the data set is unbalanced, a warning is produced.

Value

It returns the Tai graph for stability analysis and the values of alpha and lambda for each genotype.

Author(s)

Raul Eyzaguirre

References

Tai, G. C. C. (1971). Genotypic Stability Analysis and Its Application to Potato Regional Trials, Crop Science, Vol 11.

Examples

```
# The data
head(met8x12)
str(met8x12)

# Run Tai for trait y
tai("y", "geno", "env", "rep", met8x12)
```

Index

ammi, 2
ammigxe, 3

checkdata01, 5
checkdata02, 5

domeans, 6

elston, 7

met8x12, 8
msdplot, 8
mveb, 9
mvemet, 10

pesekbaker, 11
pjpz09, 13

rsa, 13
rts1, 14
rts2, 15
rts3, 16

spconsis, 16
spg, 19

tai, 19