Package 'st4gi'

April 25, 2016

Type Package

Title Statistical tools for genetic improvement

Version 0.7

Date 2016-25-04

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 (>= 3.0.0)

Imports lme4, shiny

License MIT + file LICENSE

LazyData true

Suggests testthat

RoxygenNote 5.0.1.9000

NeedsCompilation no

R topics documented:

`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"), size = c(1, 1))
```

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.
`size`	Relative size for symbols and labels.

Details

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

Value

It returns the genotype, environment and interaction means, the interaction effects matrix, 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 and the data are balanced.

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, biplot2 by default
ammi("y", "geno", "env", "rep", met8x12)

# Run AMMI for trait y, biplot1
ammi("y", "geno", "env", "rep", met8x12, biplot = 1)
```

---

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, nr = NULL, rdf = NULL, rms = NULL,
  method = "AMMI", f = 0.5, biplot = 2, biplot1 = "effects",
  title = NULL, xlab = NULL, color = c("darkorange", "black", "gray"),
  size = c(1, 1))
```

Arguments

int.mean	GxE means matrix, genotypes in rows, environments in columns.
trait	Name of the trait.
nr	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.
size	Relative size for symbols and labels.

Details

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

Value

It returns the genotype, environment and interaction means, the interaction effects matrix, 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 and `nr`, `rms` and `rdf` are specified.

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")

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

---

aovmet *ANOVA for MET with a RCBD*

Description

Fit an analysis of variance model for a multi environment trial (MET) with a RCBD in each environment.

Usage

```
aovmet(trait, geno, env, rep, data, maxp = 0.1)
```

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, default is 10%.

Details

If data is unbalanced, missing values are estimated up to an specified maximum proportion, 10% by default. Genotypes and environments are considered as fixed factors while the blocks are considered as random and nested into the environments.

Value

It returns the ANOVA table.

Author(s)

Raul Eyzaguirre.

Examples

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

# Run ANOVA for MET
aovmet("y", "geno", "env", "rep", met8x12)
```

cdt *Compute derived traits*

Description

Compute derived traits for a given fieldbook.

Usage

```
cdt(fb, plot.size = NULL)
```

Arguments

fb	The name of the fieldbook data frame.
plot.size	Plot size in square meters.

Details

The data frame must use the labels (lower or upper case) listed in function checknames. See ?checknames for details.

Value

It returns a data frame with the original and derived traits.

Author(s)

Raul Eyzaguirre.

Examples

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

# Compute derived traits
cdt(pjpz09, 4.5)
```

checkdata01 *Check data for a RCBD*

Description

This function checks the frequencies of genotypes in a RCBD.

Usage

```
checkdata01(trait, treat, rep, data)
```

Arguments

trait	The trait to analyze.
treat	The treatments.
rep	The replications.
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 treatment without data, and if the design is balanced.

Value

Three control values (c1, c2, and c3), the number of missing values nmis, the proportion of missing values (pmis), the number of treatments (nt), and the number of replications (nr).

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, rep, data)
```

Arguments

trait	The trait to analyze.
geno	The genotypes.
env	The environments.
rep	The replications.
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

Three control values (c1, c2, and c3), the number of missing values nmis, the proportion of missing values (pmis), the number of genotypes (ng), the number of environments (ne), and the number of replications (nr).

Author(s)

Raul Eyzaguirre.

`checknames` *Check fieldbook traits names*

Description

Check that fieldbook traits names correspond with the names defined in the document "PROCEDURES FOR THE EVALUATION AND ANALYSIS OF SWEETPOTATO TRIALS".

Usage

```
checknames(fb)
```

Arguments

`fb`	The name of the fieldbook data frame.

Details

The data frame must use the following labels (lower or upper case):

- L : Locations (LOC is also valid)
- Y : Years
- S : Seasons
- G : Genotypes (GENO is also valid)
- NAME : Names for genotypes
- E : Environments (ENV is also valid)
- R : Replications (REP is also valid)
- NOPS : Number of plants sowed
- 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 using CIP color charts
- RFCS : Root secondary flesh color using CIP color charts
- SCOL : Storage root skin color (1-9)
- FCOL : Storage root flesh color (1-9)
- RFCP : Storage root primary flesh color (1-9)
- RFCS : Storage root secondary flesh color (1-9)
- 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
- DM : Storage root dry matter content (%)
- DMRY : Dry matter root yield
- DMVF : Fresh weight vines for dry matter assessment
- DMVD : Dry weight of DMVF samples
- DMV : Vines dry matter content (%)
- DMFY : Dry matter foliage 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 (mg/100 g dry weight)
- ZN : Zinc (mg/100 g dry weight)

- CA : Calcium (mg/100 g dry weight)
- MG : Magnesium (mg/100 g dry weight)
- BC : Beta-carotene (mg/100 g dry weight)
- BC.CC : Beta-carotene with color charts
- TC : Total carotenoids (mg/100 g dry weight)
- 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 data frame with all traits names in upper case, and a list of the traits with names not included in the list shown above.

Author(s)

Raul Eyzaguirre.

Examples

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

# Check the trait names
checknames(pjpz09)
```

docomp *Do computations over some factors*

Description

Do computations for several traits for some specific factors.

Usage

```
docomp(do, traits, factors, addcol = NULL, data)
```

Arguments

do	The computation to perform. Implemented options are count, and standard functions like mean, median, min, max, sd, var, sum, etc.
traits	List of traits.
factors	List of factors.
addcol	Additional columns to keep.
data	The name of the data frame containing the data.

Details

This function do a specific computation 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 computations.

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 and then for each genotype
# across the two locations.
traits <- c("rytha", "bc", "dm", "star", "nocr")
factors <- c("geno", "loc")
output1 <- docomp("mean", traits, factors, data = spg)
docomp("mean", traits, "geno", data = output1)

# Compute maxima for all the traits across the two replications
# for each genotype and location.
docomp("max", traits, factors, data = spg)
```

elston *Elston Index*

Description

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

Usage

```
elston(traits, geno, env = NULL, rep = NULL, data, means = "single",
  model = "gxe", lb = 1)
```

Arguments

traits	List of traits.
geno	The genotypes.
env	The environments.
rep	The replications.
data	The name of the data frame containing the data.
means	The genotypic means to compute the index, "single" or "fitted". The default is "single". See details for more information.
model	Type of model to fit means if means = "fitted", "gxe" for a model with gxe interaction or "g+e" for a model without interaction. The default is "gxe". See details for more information.
lb	Lower bound. 1 for $k = min(x)$ and 2 for $k = (n \times min(x) - max(x))/(n-1)$

Details

The Elston index is a weight free index. It is assumed that all the traits are in the same direction where the highest the value the better the genotype. To include any trait with an opposite direction it must be transformed by multiplication by -1 before.

If means = "fitted" and model = "gxe" then the arguments env and rep must be specified. If means = "fitted" and model = "g+e" then only the argument env must be specified. If means = "single" and env and rep are specified, then single arithmetic means are computed over the replications for each genotype at each environment and then for each genotype over environments. In any other case single arithmetic means are computed over all the observations for each genotype.

Value

It returns a data frame with the genotypic means for each trait, the Elston index, and the rank for each genotype according to the index.

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", data = spg)
```

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

Description

This data set has data for root yield in tons per hectare (`rytha`), foliage yield in tons per hectare (`fytha`), and dry matter (`dm`) for an experiment with 13 genotypes (`geno`), 12 environments (`env`), and 2 replications (`rep`) in each environment following a RCBD.

Usage

```
megaclones
```

Format

A data frame with 6 columns and 312 rows.

Source

International Potato Center, sweetpotato experimental data.

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

Description

This data set has the yields per plot (`y`) for an experiment with 8 genotypes (`geno`), 12 environments (`env`), and 3 replications (`rep`) in each evironment following a RCBD.

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 with a dotplot*

Description

Function to plot means and confidence limits.

Usage

```
msdplot(trait, groups, data, conf = 0.95, nmax = 10, dotplot = "TRUE",
  sort.means = "none", main = NULL, xlab = "groups", ylab = "",
  colors = c("orange", "orange", "black"), pch = 4, lwd = 2, x.las = 1,
  jf = 0.1, dist = 0.1)
```

Arguments

trait	The trait to plot.
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.
nmax	Maximum number of points for the compulsory dotplot, default is 10.
dotplot	Logical. If TRUE, a dotplot is shown. If FALSE it will only suppress the dots if the number of data points is larger than nmax.
sort.means	Sort for means. Options are "none", "increasing", and "decreasing", "none" by default.
main	Main title.
xlab	Title for x axis.
ylab	Title for y axis.
colors	Color for mean symbols, confidence interval lines, and data points.
pch	Ploting character for means.
lwd	Width for ploting characters for means.
x.las	x axes labels orientation.
jf	Jitter factor for dots.
dist	Horizontal distance between the means and the dots.

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 represented by horizontal lines, a vertical line representing a confidence limit or a number of standard deviations, and alternatively the individual data points.

Author(s)

Raul Eyzaguirre

Examples

```
# Simulate some data
mydata <- data.frame(y = rnorm(50, sample(40:60, 5), sample(5:10, 5)),
                     g = rep(1:5, 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, treat, rep, data, maxp = 0.1, tol = 1e-06)
```

Arguments

`trait`	The trait to estimate missing values.
`treat`	The treatments.
`rep`	The replications.
`data`	The name of the data frame.
`maxp`	Maximum allowed proportion of missing values to estimate, defaults to 10%.
`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 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 = 0.1, tol = 1e-06)
```

Arguments

`trait`	The trait to estimate missing values.
`geno`	The genotypes.
`env`	The environments.
`rep`	The replications.
`data`	The name of the data frame.
`maxp`	Maximum allowed proportion of missing values to estimate, default is 10%.
`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 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 = NULL, data, means = "single",
  model = "gxe", dgg = NULL, units = "sdu", sf = 0.1)
```

Arguments

`traits`	List of traits.
`geno`	The genotypes.
`env`	The environments.
`rep`	The replications. Must be defined if `model = "gxe"`.
`data`	The name of the data frame containing the data.
`means`	The genotypic means to compute the index, `"single"` or `"fitted"`. The default is `"single"`. See details for more information.
`model`	Type of model, `"gxe"` for a model with gxe interaction or `"g+e"` for a model without interaction. The default is `"gxe"`. See details for more information.
`dgg`	Desired genetic gains. The default is one standard deviation for each trait.
`units`	Units for dgg, `"actual"` or `"sdu"`. See details for more information.
`sf`	Selected fraction. The default is 0.1.

Details

The Pesek-Baker is an index where relative economic weights have been replaced by desired gains.

By default a model with components for genotypes, environments, genotypes by environments interaction and replications nested into environments is fitted (`model = "gxe"`). If `model = "g+e"` then a model with components for genotypes and environments is fitted, and in this case the gxe variance includes the gxe plus the error variance. Response to selection is only computed when `model = "gxe"`.

If `means = "fitted"` then the model specified in `model` is used to fit the means of the genotypes. Otherwise single arithmetic means are computed over the replications for each genotype at each environment and then for each genotype over environments.

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".

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, and
- $Pesek.Baker.Index, a data frame with the genotypic means for each trait, the Pesek-Baker index, and the rank for each genotype according to the index.

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 number of plants sowed (nops), number of plants harvested (noph), number of commercial roots (nocr), number of non commercial roots (nonc), commercial root weight (crw), non commercial root weght (ncrw), total root weight (trw), and vine weight (vw) for an experiment with 102 genotypes (geno) and 2 replications (rep) with a RCBD.

Usage

```
pjpz09
```

Format

A data frame with 10 columns and 204 rows.

Source

International Potato Center, sweetpotato experimental data.

rcbd *ANOVA for a RCBD*

Description

Fit an analysis of variance model for a RCBD.

Usage

```
rcbd(trait, treat, rep, data, maxp = 0.1)
```

Arguments

trait	The trait to analyze.
treat	The treatments.
rep	The replications.
data	The name of the data frame containing the data.
maxp	Maximum allowed proportion of missing values to estimate, default is 10%.

Details

If data is unbalanced, missing values are estimated up to an specified maximum proportion, 10% by default.

Value

It returns ANOVA table.

Author(s)

Raul Eyzaguirre.

Examples

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

# Get a copy with some missing values for trw
temp <- pjpz09
temp[c(10, 20, 30), "trw"] <- NA

# Run ANOVA for trw
rcbd("trw", "geno", "rep", temp)
```

---

rsa *Regression Stability Analysis*

Description

Function to run the regression stability analysis (Yates and Cochran, 1938, Finlay and Wilkinson, 1963).

Usage

```
rsa(trait, geno, env, rep, data, maxp = 0.1)
```

Arguments

trait	The trait to analyze.
geno	The genotypes.
env	The environments.
rep	The replications.
data	The name of the data frame containing the data.
maxp	Maximum allowed proportion of missing values to estimate, default is 10%.

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, 10% by default. For the ANOVA table, genotypes and environments are considered as fixed factors while the blocks are considered as random and nested into the environments. 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. In both cases the X values are centered on zero, so the intercepts of the models correspond to the means of the genotypes or environments.

Value

It returns the regression stability analysis decomposition of the GxE interaction for genotypes and environments (Heterogeneity among regressions and deviation from regression), the coefficient of variation, and the following regression stability measures for genotypes and environments:

- a the intercept.
- b the slope.
- se the standard error for the slope.
- MSe the mean square error.
- MSentry the variance of the genotype means across environments and the environment means across genotypes.
- MSinter the variance of the genotype interaction effects across environments and the environment interaction effects across genotypes.

Author(s)

Raul Eyzaguirre.

References

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.

`rts4` *Response to selection with several locations in two steps (two years)*

Description

It computes the response to selection for each step in a two steps selection with several locations for a given number of genotypes at step 1, the number of locations, replications and selected genotypes at step 1, the number of locations, replications and selected genotypes at step 2, and the genotypic, genotypic by location, genotypic by year, genotypic by location by year, and error variances.

Usage

```
rts4()
```

Details

It uses package `shiny` for the web layout. Type `rts4()` in the R console to run the app.

Value

It returns the response to selection at step 1 and 2.

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(fb, plot.size, f = 3, width = 240, file = TRUE)
```

Arguments

fb	The name of the fieldbook data frame.
plot.size	Plot size in square meters.
f	Factor for extreme values detection. See details.
width	Number of columns for the output.
file	Logigal, if TRUE the output goes to a file.

Details

The data frame must use the labels (lower or upper case) listed in function checknames. See ?checknames for details. Extreme values are detected using the interquartile range. The rule is to detect any value out of the interval $[Q_1 - f \times IQR; Q_3 + f \times IQR]$. By default f = 3.

Value

If file = TRUE it returns a file with name checks.txt with a list of all rows with some kind of inconsistency and all rows with outliers. If file = FALSE the output is shown in the R console.

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 an experiment with 8 genotypes (geno), 2 locations (loc), and 2 replications (rep) in each location following a RCBD.

Usage

```
spg
```

Format

A data frame with 8 columns and 32 rows.

Source

International Potato Center, sweetpotato experimental data.

suma *Compute sum of two traits*

Description

Compute the sum of two traits. Missing values do not propagate.

Usage

```
suma(a, b)
```

Arguments

a	Name of trait 1 to sum
b	Name of trait 2 to sum.

Details

Missing values do not propagate. If NA is present for both traits then NA is applied to the sum.

Value

It returns the sum of the two traits.

Author(s)

Raul Eyzaguirre.

Examples

```
# The data
head(pjpz09)

# Compute total biomass as the sum of trw and vw
suma(pjpz09$trw, pjpz09$vw)
```

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, maxp = 0.1, conf = 0.95, title = NULL,
  color = c("darkorange", "black", "gray"), size = c(1, 1))
```

Arguments

trait	The trait to analyze.
geno	The genotypes.
env	The environments.
rep	The replications.
data	The name of the data frame containing the data.
maxp	Maximum allowed proportion of missing values to estimate, default is 10%.
conf	Probability for the Tai limits.
title	Main title for plot.
color	Color for symbols, labels and lines.
size	Relative size for symbols and labels.

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