

Package ‘pepa’

April 28, 2016

Type Package

Title Package for the Execution of Pre Cooked Analysis

Version 0.3.1

Date 2016-04-26

Author Raul Eyzaguirre

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

Description This packages creates automatic reports for different types of statistical methodologies.

Depends R (>= 3.0.0), st4gi

Imports agricolae, rmarkdown

License MIT + file LICENSE

Copyright International Potato Center (2015)

LazyData true

Suggests testthat

RoxygenNote 5.0.1.9000

NeedsCompilation no

R topics documented:

pty	2
pty.aovmet	2
pty.elston	3
pty.pesekbaker	4
pty.spconsis	5
repo.abd	6
repo.crd	7
repo.met	8
repo.rcbd	9
Index	11

pty	<i>Pepa tells you</i>
-----	-----------------------

Description

Explain an R object in plain English if she knows about it

Usage

```
pty(x, author = "International Potato Center")
```

Arguments

x	An R object.
author	Author.

Details

It uses a set of templates to explain R objects in plain English. It aims to produce automatic reports for some standard statistical procedures, most of them included in the `st4gi` package.

Value

It returns an explanation about the selected R object.

Author(s)

Raul Eyzaguirre.

Examples

```
# Pepa tells you something about a data frame:
pty(pj pz09)
```

pty.aovmet	<i>Pepa tells you about a MET with a RCBD</i>
------------	---

Description

Explain a fitted model for a multi environment trial (MET) with a RCBD in each environment in plain English.

Usage

```
pty.aovmet(trait, geno, env, rep, data, maxp = 0.1,
  author = "International Potato Center")
```

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%.
author	Author.

Details

It fits a linear model for a MET with a RCBD and explains the results. 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 an explanation about the MET with a RCBD fitted model.

Author(s)

Raul Eyzaguirre.

Examples

```
pty.aovmet("y", "geno", "env", "rep", met8x12)
```

pty.elston

Pepa tells you about the Elston index

Description

Explain the results of the Elston index in plain English.

Usage

```
pty.elston(traits, geno, env = NULL, rep = NULL, data, means = "single",
  model = "gxe", lb = 1, author = "International Potato Center")
```

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 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 $k = (n \times \min(x) - \max(x)) / (n - 1)$
author	Author.

Details

The Elston index is a weight free index.

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 an explanation about the Elston index.

Author(s)

Raul Eyzaguirre.

Examples

```
pty.elston(c("rytha", "bc", "dm", "star", "nocr"), "geno", data = spg)
```

pty.pesekbaker

Pepa tells you about the Pesek-Baker index

Description

Explain the results of the Pesek-Baker index in plain English.

Usage

```
pty.pesekbaker(traits, geno, env, rep = NULL, data, means = "single",
  model = "gxe", dgg = NULL, units = "sdu", sf = 0.1,
  author = "International Potato Center")
```

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.
author	Author.

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 an explanation about the Pesek-Baker index.

Author(s)

Raul Eyzaguirre.

Examples

```
pty.pesekbaker(c("rytha", "bc", "dm", "star", "nocr"), "geno", "loc", "rep", spg)
```

pty.spconsis

Pepa tells you about the consistency of your data

Description

It checks your data for inconsistencies.

Usage

```
pty.spconsis(data, plot.size, f = 3, author = "International Potato Center")
```

Arguments

data	The name of the data frame.
plot.size	Plot size in square meters.
f	Factor for extreme values detection. See details.
author	Author.

Details

The data frame must use the labels (lower or upper case) specified in the function `spconsis` of package `st4gi`. Type `?spconsis` to see the list. 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

It returns a list of all rows with some kind of inconsistency and all rows with outliers.

Author(s)

Raul Eyzaguirre.

Examples

```
pty.spconsis(pjz09, 4.5)
```

 repo.abd

Automatic report for an augmented block design (ABD)

Description

Produces an automatic report for selected traits in an experiment with an ABD.

Usage

```
repo.abd(traits, treat, rep, data, author = "International Potato Center")
```

Arguments

traits	The traits to analyze.
treat	The treatments.
rep	The replications.
data	The name of the data frame.
author	Author.

Details

It fits a linear model for an ABD and explains the results.

The Tukey HSD method for pairwise differences is applied. Assumptions of the model are evaluated with residual plots.

Value

It returns an explanation about the ABD fitted model.

Author(s)

Raul Eyzaguirre.

Examples

```
# A subset that looks like an ABD
temp <- pjpz09[c(1, 2, 9, 10, 13, 14, 27, 29, 31, 33, 35, 37, 40, 42, 44, 46, 48, 50, 203, 204), ]
repo.abd(c("trw", "vw", "crw"), "geno", "rep", temp)

# With some missing values
temp[c(1, 2, 3, 4), "trw"] <- NA
temp[c(1, 2, 3), "vw"] <- NA
temp[c(1, 10, 15), "crw"] <- NA
repo.abd(c("nocr", "trw", "vw", "crw"), "geno", "rep", temp)
```

 repo.crd

Automatic report for a Completely Randomized Design (CRD)

Description

Produces an automatic report for selected traits in an experiment with a CRD.

Usage

```
repo.crd(traits, treat, data, maxp = 0.1,
  author = "International Potato Center")
```

Arguments

traits	The traits to analyze.
treat	The treatments.
data	The name of the data frame.
maxp	Maximum allowed proportion of missing values to estimate, default is 10%.
author	Author.

Details

It fits a linear model for a CRD and explains the results.

Under the assumption of fixed effects an ANOVA table is computed. If the ANOVA results in a significant value then the Tukey HSD method for pairwise differences is applied. Assumptions of the model are evaluated with residual plots.

Under the assumption of random effects the model is estimated using REML and the variance components are shown.

Value

It returns an explanation about the CRD fitted model.

Author(s)

Raul Eyzaguirre.

Examples

```
repo.crd(c("trw", "vw", "crw"), "geno", pjpz09)

# With a small data set
temp <- pjpz09[1:18, ]
repo.crd(c("trw", "vw", "crw"), "geno", temp)
```

repo.met

Automatic report for a MET with a RCBD

Description

Produces an automatic report for selected traits in a multi environment trial (MET) with a RCBD in each environment.

Usage

```
repo.met(traits, geno, env, rep, data, maxp = 0.1,
  author = "International Potato Center")
```

Arguments

traits	The traits 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%.
author	Author.

Details

It fits a linear model for a MET with a RCBD for the selected trait. 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 an automatic report about the MET with a RCBD fitted model.

Author(s)

Raul Eyzaguirre.

Examples

```
repo.met(c("rytha", "fytha"), "geno", "env", "rep", megaclones)
```

repo.rcbd	<i>Automatic report for a Randomized Complete Block Design (RCBD)</i>
-----------	---

Description

Produces an automatic report for selected traits in an experiment with a RCBD.

Usage

```
repo.rcbd(traits, treat, rep, data, maxp = 0.1,
  author = "International Potato Center")
```

Arguments

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

Details

It fits a linear model for a RCBD and explains the results.

Under the assumption of fixed effects an ANOVA table is computed with missing values estimated up to a specified percentage (10% by default). If the ANOVA results in a significant value for treatments then the Tukey HSD method for pairwise differences is applied. Assumptions of the model are evaluated with residual plots.

Under the assumption of random effects the model is estimated using REML and the variance components are shown. Missing values are not estimated in this case.

Value

It returns an explanation about the RCBD fitted model.

Author(s)

Raul Eyzaguirre.

Examples

```
repo.rcbd(c("trw", "vw", "crw"), "geno", "rep", pjpz09)

# With a small data set
temp <- pjpz09[1:18, ]
repo.rcbd(c("trw", "vw", "crw"), "geno", "rep", temp)
```

Index

pty, [2](#)
pty.aovmet, [2](#)
pty.elston, [3](#)
pty.pesekbaker, [4](#)
pty.spconsis, [5](#)

repo.abd, [6](#)
repo.crd, [7](#)
repo.met, [8](#)
repo.rcbd, [9](#)