

Package ‘rosetta’

February 8, 2019

Title Parallel Use of Statistical Packages in Teaching

Version 0.0.1

Description When teaching statistics, it can often be desirable to uncouple the content from specific software packages. To ease such efforts, the Rosetta Stats website (<<https://rosettastats.com>>) allows comparing analyses in different packages. This package is the companion to the Rosetta Stats website, aiming to provide functions that produce output that is similar to output from other statistical packages, thereby facilitating 'software-agnostic' teaching of statistics.

Maintainer Gjalt-Jorn Peters <gjalt-jorn@userfriendlyscience.com>

License GPL (>= 3)

URL <https://rosetta.openscience>

BugReports <https://github.com/psytext/rosetta>

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Depends R (>= 3.0.0)

Imports car (>= 3.0.2), ggplot2 (>= 2.2.1), gridExtra (>= 2.3), methods (>= 3.0.0), lme4 (>= 1.1.19), multcompView (>= 0.1-0), pander (>= 0.6.3), psych (>= 1.8.4), rio (>= 0.5.10), ufs (>= 0.1.0), userfriendlyscience (>= 0.7.1)

NeedsCompilation no

Author Gjalt-Jorn Peters [aut, cre],
Ron Pat-El [ctb],
Peter Verboon [ctb]

Repository CRAN

Date/Publication 2019-02-08 15:50:03 UTC

R topics documented:

crossTab	2
exportToSPSS	3
fanova	5
freq	7
oneway	9
posthocTGH	11

Index	13
--------------	-----------

crossTab	<i>Cross tables</i>
----------	---------------------

Description

This function produces a cross table, computes Chi Square, and computes the point estimate and confidence interval for Cramer's V.

Usage

```
crossTab(x, y = NULL, conf.level = 0.95, digits = 2,
         pValueDigits = 3, ...)
```

```
## S3 method for class 'crossTab'
print(x, digits = x$input$digits,
      pValueDigits = x$input$pValueDigits, ...)
```

```
## S3 method for class 'crossTab'
pander(x, digits = x$input$digits,
       pValueDigits = x$input$pValueDigits, ...)
```

Arguments

x	Either a crosstable to analyse, or one of two vectors to use to generate that crosstable. The vector should be a factor, i.e. a categorical variable identified as such by the 'factor' class).
y	If x is a crosstable, y can (and should) be empty. If x is a vector, y must also be a vector.
conf.level	Level of confidence for the confidence interval.
digits	Minimum number of digits after the decimal point to show in the result.
pValueDigits	Minimum number of digits after the decimal point to show in the Chi Square p value in the result.
...	Extra arguments to crossTab are passed on to [ufs::confIntV()].

Value

The results of [ufs::confIntV()], but also prints the cross table and the chi square test results.

Examples

```
crossTab(infert$education, infert$induced, samples=50);
```

exportToSPSS

Basic SPSS translation functions

Description

Basic functions to make working with R easier for SPSS users: `getData` and `getDat` provide an easy way to load SPSS datafiles, and `exportToSPSS` to write to a datafile and syntax file that SPSS can import; `filterBy` and `useAll` allow easy temporary filtering of rows from the dataframe; `mediaan` and `modus` compute the median and mode of ordinal or numeric data.

Usage

```
exportToSPSS(dat, savfile = NULL, datafile = NULL, codefile = NULL,
  fileEncoding = "UTF-8", newLinesInString = " |n| ")
```

```
filterBy(dat, expression, replaceOriginalDataframe = TRUE,
  envir = parent.frame())
```

```
getData(filename = NULL, file = NULL,
  errorMessage = "[defaultErrorMessage]", applyRtoLabels = TRUE,
  use.value.labels = FALSE, to.data.frame = TRUE,
  stringsAsFactors = FALSE, silent = FALSE, ...)
```

```
getDat(..., dfName = "dat", backup = TRUE)
```

```
mediaan(vector)
```

```
modus(vector)
```

```
useAll(dat, replaceFilteredDataframe = TRUE)
```

Arguments

<code>dat</code>	Dataframe to process: for <code>filterBy</code> , dataframe to filter rows from; for <code>useAll</code> , dataframe to restore ('unfilter').
<code>savfile</code>	The name of the SPSS format <code>.sav</code> file (alternative for writing a datafile and a codefile).
<code>datafile</code>	The name of the data file, a comma separated values file that can be read into SPSS by using the code file.
<code>codefile</code>	The name of the code file, the SPSS syntax file that can be used to import the data file.

<code>fileEncoding</code>	The encoding to use to write the files.
<code>newLinesInString</code>	A string to replace newlines with (SPSS has problems reading newlines).
<code>expression</code>	Logical expression determining which rows to keep and which to drop. Can be either a logical vector or a string which is then evaluated. If it's a string, it's evaluated using 'with' to evaluate the expression using the variable names.
<code>replaceOriginalDataframe</code>	Whether to also replace the original dataframe in the parent environment. Very messy, but for maximum compatibility with the 'SPSS way of doing things', by default, this is true. After all, people who care about the messiness/inappropriateness of this function wouldn't be using it in the first place :-)
<code>envir</code>	The environment where to create the 'backup' of the unfiltered dataframe, for when <code>useAll</code> is called and the filter is deactivated again.
<code>filename, file</code>	It is possible to specify a path and filename to load here. If not specified, the default R file selection dialogue is shown. <code>file</code> is still available for backward compatibility but will eventually be phased out.
<code>errorMessage</code>	The error message that is shown if the file does not exist or does not have the right extension; "[defaultErrorMessage]" is replaced with a default error message (and can be included in longer messages).
<code>applyRioLabels</code>	Whether to apply the labels supplied by Rio. This will make variables that has value labels into factors.
<code>use.value.labels</code>	Only useful when reading from SPSS files: whether to read variables with value labels as factors (TRUE) or numeric vectors (FALSE).
<code>to.data.frame</code>	Only useful when reading from SPSS files: whether to return a dataframe or not.
<code>stringsAsFactors</code>	Whether to read strings as strings (FALSE) or factors (TRUE).
<code>silent</code>	Whether to suppress potentially useful information.
<code>...</code>	Additional options, passed on to the function used to import the data (which depends on the extension of the file).
<code>dfName</code>	The name of the dataframe to create in the parent environment.
<code>backup</code>	Whether to backup an object with name <code>dfName</code> , if one already exists in the parent environment.
<code>vector</code>	For median and modus, the vector for which to find the median or mode.
<code>replaceFilteredDataframe</code>	Whether to replace the filtered dataframe passed in the 'dat' argument (see <code>replaceOriginalDataframe</code>).

Value

`getData` returns the imported dataframe, with the filename from which it was read stored in the 'filename' attribute.

`getDat` is a simple wrapper for `getData()` which creates a dataframe in the parent environment, by default with the name 'dat'. Therefore, calling `getDat()` in the console will allow the user to select

a file, and the data from the file will then be read and be available as 'dat'. If an object with dfName (i.e. 'dat' by default) already exists, it will be backed up with a warning. `getDat()` therefore returns nothing.

`mediaan` returns the median, or, in the case of a factor where the median is in between two categories, both categories.

`modus` returns the mode.

Note

`getData()` currently can't read from LibreOffice or OpenOffice files. There doesn't seem to be a platform-independent package that allows this. Non-CRAN package `ROpenOffice` from Omega-Hat should be able to do the trick, but fails to install (manual download and installation using <http://www.omegahat.org> produces "ERROR: dependency 'Rcompression' is not available for package 'ROpenOffice'" - and manual download and installation of `RCompression` produces "Please define LIB_ZLIB; ERROR: configuration failed for package 'Rcompression'"). If you have any suggestions, please let me know!

Examples

```
## Not run:
### Open a dialogue to read an SPSS file
getData();

## End(Not run)

### Get a median and a mode
mediaan(c(1,2,2,3,4,4,5,6,6,6,7));
modus(c(1,2,2,3,4,4,5,6,6,6,7));

### Create an example dataframe
(exampleDat <- data.frame(x=rep(8, 8), y=rep(c(0,1), each=4)));
### Filter it, replacing the original dataframe
(filterBy(exampleDat, "y=0"));
### Restore the old dataframe
(useAll(exampleDat));
```

Description

This function is meant as a userfriendly wrapper to approximate the way analysis of variance is done in SPSS.

Usage

```
fanova(data, y, between = NULL, covar = NULL, plot = FALSE,
       levene = FALSE, digits = 2, contrast = NULL)
```

Arguments

<code>data</code>	The dataset containing the variables to analyse.
<code>y</code>	The dependent variable. For oneway anova, factorial anova, or ancova, this is the name of a variable in dataframe <code>data</code> . For repeated measures anova, this is a vector with the names of all variable names in dataframe <code>data</code> , e.g. <code>c('t0_value', 't1_value', 't2_value')</code> .
<code>between</code>	A vector with the variables name(s) of the between subjects factor(s).
<code>covar</code>	A vector with the variables name(s) of the covariate(s).
<code>plot</code>	Whether to produce a plot. Note that a plot is only produced for oneway and twoway anova and oneway repeated measures designs: if covariates or more than two between-subjects factors are specified, not plot is produced. For twoway anova designs, the second predictor is plotted as moderator (and the first predictor is plotted on the x axis).
<code>levene</code>	Whether to show Levene's test for equality of variances (using car's leveneTest function but specifying mean as function to compute the center of each group).
<code>digits</code>	Number of digits (actually: decimals) to use when printing results. The p-value is printed with one extra digit.
<code>contrast</code>	This functionality has not been implemented yet.

Details

This wrapper uses [oneway](#) and [lm](#) and [lmer](#) in combination with car's [Anova](#) function to conduct the analysis of variance.

Value

Mainly, this function prints its results, but it also returns them in an object containing three lists:

<code>input</code>	The arguments specified when calling the function
<code>intermediate</code>	Intermediat objects and values
<code>output</code>	The results such as the plot.

Author(s)

Gjalt-Jorn Peters

Maintainer: Gjalt-Jorn Peters <gjalt-jorn@userfriendlyscience.com>

See Also

[regr](#) and [logRegr](#) for similar functions for linear and logistic regression and [oneway](#), [lm](#), [lmer](#) and [Anova](#) for the functions used behind the scenes.

Examples

```

### Oneway anova with a plot
fanova(dat=mtcars, y='mpg', between='cyl', plot=TRUE);

### Factorial anova
fanova(dat=mtcars, y='mpg', between=c('vs', 'am'), plot=TRUE);

### Ancova
fanova(dat=mtcars, y='mpg', between=c('vs', 'am'), covar='hp');

### Don't run these examples to not take too much time during testing
### for CRAN
## Not run:
### Repeated measures anova; first generate datafile
dat <- mtcars[, c('am', 'drat', 'wt')];
names(dat) <- c('factor', 't0_dependentVar', 't1_dependentVar');
dat$factor <- factor(dat$factor);

### Then do the repeated measures anova
fanova(dat, y=c('t0_dependentVar', 't1_dependentVar'),
       between='factor', plot=TRUE);

## End(Not run)

```

freq

Frequency tables

Description

Function to show frequencies in a manner similar to what SPSS' "FREQUENCIES" command does. Note that frequency is an alias for freq.

Usage

```

freq(vector, digits = 1, nsmall = 1, transposed = FALSE, round = 1,
     plot = FALSE, plotTheme = ggplot2::theme_bw())

## S3 method for class 'freq'
print(x, digits = x$input$digits,
     nsmall = x$input$nsmall, transposed = x$input$transposed, ...)

## S3 method for class 'freq'
pander(x, ...)

frequencies(..., digits = 1, nsmall = 1, transposed = FALSE,
           round = 1, plot = FALSE, plotTheme = ggplot2::theme_bw())

```

```
## S3 method for class 'frequencies'
print(x, ...)
```

```
## S3 method for class 'frequencies'
pander(x, prefix = "###", ...)
```

Arguments

vector	A vector of values to compute frequencies for.
digits	Minimum number of significant digits to show in result.
nsmall	Minimum number of digits after the decimal point to show in the result.
transposed	Whether to transpose the results when printing them (this can be useful for blind users).
round	Number of digits to round the results to (can be used in conjunction with digits to determine format of results).
plot	If true, a histogram is shown of the variable.
plotTheme	The ggplot2 theme to use.
x	The 'freq' or 'frequencies' object to print.
...	For 'frequencies', the variables of which to provide frequencies; for the 'print' methods, additional arguments are passed on to the 'print' function.
prefix	The prefix to use when printing 'frequencies', to easily prepend Markdown headers.

Value

An object with several elements, the most notable of which is:

dat A dataframe with the frequencies

For frequencies, these objects are in a list of their own.

Examples

```
### Create factor vector
ourFactor <- factor(mtcars$gear, levels=c(3,4,5),
                   labels=c("three", "four", "five"));
### Add some missing values
factorWithMissings <- ourFactor;
factorWithMissings[10] <- factorWithMissings[20] <- NA;

### Show frequencies
freq(ourFactor);
freq(factorWithMissings);

### ... Or for all of them at one
```



```
frequencies(ourFactor, factorWithMissings);
```

oneway	<i>oneway</i>
--------	---------------

Description

The oneway function wraps a number of analysis of variance functions into one convenient interface that is similar to the oneway anova command in SPSS.

Usage

```
oneway(y, x, posthoc = NULL, means = FALSE, fullDescribe = FALSE,
       levene = FALSE, plot = FALSE, digits = 2, omegasq = TRUE,
       etasq = TRUE, corrections = FALSE, pvalueDigits = 3, t = FALSE,
       conf.level = 0.95, posthocLetters = FALSE,
       posthocLetterAlpha = 0.05, overrideVarNames = NULL, silent = FALSE)
```

```
## S3 method for class 'oneway'
print(x, digits = x$input$digits,
      pvalueDigits = x$input$pvalueDigits, na.print = "", ...)
```

```
## S3 method for class 'oneway'
pander(x, digits = x$input$digits,
       pvalueDigits = x$input$pvalueDigits, headerStyle = "***",
       na.print = "", ...)
```

Arguments

y	y has to be a numeric vector.
x	x has to be vector that either is a factor or can be converted into one.
posthoc	Which post-hoc tests to conduct. Valid values are any correction methods in <code>p.adjust.methods</code> (at the time of writing of this document, "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"), as well as "tukey" and "games-howell".
means	Whether to show the means for the y variable in each of the groups determined by the x variable.
fullDescribe	If TRUE, not only the means are shown, but all statistics acquired through the 'describe' function in the 'psych' package are shown.
levene	Whether to show Levene's test for equality of variances (using car's leveneTest function but specifying mean as function to compute the center of each group).
plot	Whether to show a plot of the means of the y variable in each of the groups determined by the x variable.

<code>digits</code>	The number of digits to show in the output.
<code>omegasq</code>	Whether to show the omega squared effect size.
<code>etasq</code>	Whether to show the eta squared effect size (this is biased and generally advised against; omega squared is less biased).
<code>corrections</code>	Whether to show the corrections for unequal variances (Welch and Brown-Forsythe).
<code>pvalueDigits</code>	The number of digits to show for p-values; smaller p-values will be shown as <code><.001</code> or <code><.0001</code> etc.
<code>t</code>	Whether to transpose the dataframes with the means (if requested) and the anova results. This can be useful for blind people.
<code>conf.level</code>	Confidence level to use when computing the confidence interval for η^2 . Note that the function we use doubles the 'unconfidence' level to maintain consistency with the NHST value (see http://yatani.jp/HCIstats/ANOVA#RCODEOneWay , http://daniellakens.blogspot.nl/2014/06/calculating-confidence-intervals-for.html or Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. <i>Psychological methods</i> , 9(2), 164-82. doi:10.1037/1082-989X.9.2.164
<code>posthocLetters</code>	Whether to also compute and show the letters signifying differences between groups when conducting post hoc tests. This requires package <code>multcompView</code> to be installed.
<code>posthocLetterAlpha</code>	The alpha to use when determining whether groups have different means when using <code>posthocLetters</code> .
<code>overrideVarNames</code>	Can be used to override the variable names (most useful in functions).
<code>silent</code>	Whether to show warnings and other diagnostic information or remain silent.
<code>na.print</code>	How to print missing values.
<code>...</code>	Any additional arguments are passed to the 'print' or 'pander' function.
<code>headerStyle</code>	The header pre- and suffix to use when pandering the result (useful when working with Markdown).

Value

A list of three elements:

<code>input</code>	List with input arguments
<code>intermediate</code>	List of intermediate objects, such as the <code>aov</code> and <code>Anova</code> (from the <code>car</code> package) objects.
<code>output</code>	List with <code>etasq</code> , the effect size, and <code>dat</code> , a dataframe with the Oneway Anova results.

Note

By my knowledge the Brown-Forsythe correction was not yet available in R. I took this from the original paper (directed there by Field, 2014). Note that this is the corrected F value, not the Brown-Forsythe test for normality!

Author(s)

Gjalt-Jorn Peters

Maintainer: Gjalt-Jorn Peters <gjalt-jorn@userfriendlyscience.com>

References

Brown, M., & Forsythe, A. (1974). *The small sample behavior of some statistics which test the equality of several means*. *Technometrics*, 16(1), 129-132. <https://doi.org/10.2307/1267501>

Field, A. (2014) *Discovering statistics using SPSS* (4th ed.). London: Sage.

Steiger, J. H. (2004). *Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis*. *Psychological methods*, 9(2), 164-82. doi:10.1037/1082-989X.9.2.164

Examples

```
### Do a oneway Anova
oneway(y=ChickWeight$weight, x=ChickWeight$Diet);

### Also order means and transpose the results
oneway(y=ChickWeight$weight, x=ChickWeight$Diet, means=TRUE, t=TRUE);
```

 posthocTGH

posthocTGH

Description

This function is used by the 'oneway' function for oneway analysis of variance in case a user requests post-hoc tests using the Tukey or Games-Howell methods.

Usage

```
posthocTGH(y, x, method = c("games-howell", "tukey"),
  conf.level = 0.95, digits = 2, p.adjust = "none",
  formatPvalue = TRUE)
```

```
## S3 method for class 'posthocTGH'
print(x, digits = x$input$digits, ...)
```

Arguments

y	y has to be a numeric vector.
x	x has to be vector that either is a factor or can be converted into one.
method	Which post-hoc tests to conduct. Valid values are "tukey" and "games-howell".
conf.level	Confidence level of the confidence intervals.

digits	The number of digits to show in the output.
p.adjust	Any valid <code>p.adjust</code> method.
formatPvalue	Whether to format the p values according to APA standards (i.e. replace all values lower than .001 with '<.001'). This only applies to the printing of the object, not to the way the p values are stored in the object.
...	Any additional arguments are passed on to the 'print' function.

Value

A list of three elements:

input	List with input arguments
intermediate	List of intermediate objects.
output	List with two objects 'tukey' and 'games.howell', containing the outcomes for the respective post-hoc tests.

Note

This function is based on a file that was once hosted at http://www.psych.yorku.ca/cribbie/6130/games_howell.R, but has been removed since. It was then adjusted for implementation in the `userfriendlyscience` package. Jeffrey Baggett needed the confidence intervals, and so emailed them, after which his updated function was used. In the meantime, it appears Aaron Schlegel (<https://rpubs.com/aaronsc32>) independently developed a version with confidence intervals and posted it on RPubS at <https://rpubs.com/aaronsc32/games-howell-test>.

Also, for some reason, `p.adjust` can be used to specify additional correction of *p* values. I'm not sure why I implemented this, but I'm not entirely sure it was a mistake either. Therefore, in `userfriendlyscience` version 0.6-2, the default of this setting changed from "holm" to "none" (also see <https://stats.stackexchange.com/questions/83941/games-howell-post-hoc-test-in-r>).

Author(s)

Gjalt-Jorn Peters (Open University of the Netherlands) & Jeff Bagget (University of Wisconsin - La Crosse)

Maintainer: Gjalt-Jorn Peters <gjalt-jorn@userfriendlyscience.com>

Examples

```
### Compute post-hoc statistics using the tukey method
posthocTGH(y=ChickWeight$weight, x=ChickWeight$Diet, method="tukey");
### Compute post-hoc statistics using the games-howell method
posthocTGH(y=ChickWeight$weight, x=ChickWeight$Diet);
```

Index

- *Topic **bivar**
 - crossTab, 2
- *Topic **file**
 - exportToSPSS, 3
- *Topic **hplot**
 - fanova, 5
- *Topic **htest**
 - fanova, 5
- *Topic **univar**
 - exportToSPSS, 3
 - freq, 7
- *Topic **utilities**
 - exportToSPSS, 3
 - oneway, 9
 - posthocTGH, 11

Anova, 6

basicSPSStranslation (exportToSPSS), 3

confIntV (crossTab), 2

cramersV (crossTab), 2

crossTab, 2

exportToSPSS, 3

fanova, 5

filterBy (exportToSPSS), 3

freq, 7

frequencies (freq), 7

Frequency (freq), 7

getDat (exportToSPSS), 3

getData (exportToSPSS), 3

leveneTest, 6, 9

lm, 6

lmer, 6

logRegr, 6

mean, 6, 9

mediaan (exportToSPSS), 3

modus (exportToSPSS), 3

oneway, 6, 9

p.adjust, 12

pander.crossTab (crossTab), 2

pander.freq (freq), 7

pander.frequencies (freq), 7

pander.oneway (oneway), 9

posthocTGH, 11

print.crossTab (crossTab), 2

print.freq (freq), 7

print.frequencies (freq), 7

print.oneway (oneway), 9

print.posthocTGH (posthocTGH), 11

regr, 6

useAll (exportToSPSS), 3

userfriendlyscience, 12