

Package ‘acro’

August 19, 2024

Title A Tool for Automating the Statistical Disclosure Control of Research Outputs

Version 0.1.2

Maintainer Jim Smith <James.Smith@uwe.ac.uk>

Description Assists researchers and output checkers by distinguishing between research output that is safe to publish, output that requires further analysis, and output that cannot be published because of substantial disclosure risk. A paper about the tool was presented at the UN-ECE Expert Meeting on Statistical Data Confidentiality 2023; see <<https://uwe-repository.worktribe.com/output/11060964>>.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

SystemRequirements Python (>= 3.8)

Imports reticulate, admiraldev, png

Depends R (>= 2.10)

LazyData true

Suggests spelling, testthat (>= 3.0.0)

Config/testthat/edition 3

Language en-US

URL <https://github.com/AI-SDC/ACRO-R>

BugReports <https://github.com/AI-SDC/ACRO-R/issues>

NeedsCompilation no

Author Jim Smith [cre, ctb] (<<https://orcid.org/0000-0001-7908-1859>>),
Maha Albashir [aut, ctb],
Richard John Preen [ctb] (<<https://orcid.org/0000-0003-3351-8132>>)

Repository CRAN

Date/Publication 2024-08-19 18:40:11 UTC

Contents

acro_add_comments	2
acro_add_exception	3
acro_crosstab	3
acro_custom_output	4
acro_finalise	4
acro_glm	5
acro_hist	5
acro_init	6
acro_lm	6
acro_pivot_table	7
acro_print_outputs	8
acro_remove_output	8
acro_rename_output	9
acro_surv_func	9
acro_table	10
create_virtualenv	10
install_acro	11
lung	11
nursery_data	12

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

acro_add_comments	<i>Add comments to outputs</i>
-------------------	--------------------------------

Description

Add comments to outputs

Usage

```
acro_add_comments(name, comment)
```

Arguments

name	The name of the output.
comment	The comment.

Value

No return value, called for side effects

acro_add_exception *Adds an exception request to an output.*

Description

Adds an exception request to an output.

Usage

```
acro_add_exception(name, reason)
```

Arguments

name	The name of the output.
reason	The comment.

Value

No return value, called for side effects

acro_crosstab *Compute a simple cross tabulation of two (or more) factors.*

Description

Compute a simple cross tabulation of two (or more) factors.

Usage

```
acro_crosstab(index, columns, values = NULL, aggfunc = NULL)
```

Arguments

index	Values to group by in the rows.
columns	Values to group by in the columns.
values	Array of values to aggregate according to the factors. Requires aggfunc be specified.
aggfunc	If specified, requires values be specified as well.

Value

Cross tabulation of the data

acro_custom_output	<i>Adds an unsupported output to the results dictionary</i>
--------------------	---

Description

Adds an unsupported output to the results dictionary

Usage

```
acro_custom_output(filename, comment = NULL)
```

Arguments

filename	The name of the file that will be added to the list of the outputs.
comment	An optional comment.

Value

No return value, called for side effects

acro_finalise	<i>Creates a results file for checking.</i>
---------------	---

Description

Creates a results file for checking.

Usage

```
acro_finalise(path, ext)
```

Arguments

path	Name of a folder to save outputs.
ext	Extension of the results file. Valid extensions are json or xlsx.

Value

No return value, called for side effects

acro_glm	<i>Fits Logit or Probit model.</i>
----------	------------------------------------

Description

Fits Logit or Probit model.

Usage

```
acro_glm(formula, data, family)
```

Arguments

formula	The formula specifying the model.
data	The data for the model.
family	Decide whether to fit a logit or probit model.

Value

Regression Results Wrapper

acro_hist	<i>Histogram</i>
-----------	------------------

Description

Histogram

Usage

```
acro_hist(  
  data,  
  column,  
  breaks = 10,  
  freq = TRUE,  
  col = NULL,  
  filename = "histogram.png"  
)
```

Arguments

data	The object holding the data.
column	The column that will be used to plot the histogram.
breaks	Number of histogram bins to be used.
freq	If False, the result will contain the number of samples in each bin. If True, the result is the value of the probability density function at the bin.
col	The color of the plot.
filename	The name of the file where the plot will be saved.

Value

The histogram.

acro_init *Initialise an ACRO object*

Description

Initialise an ACRO object

Usage

```
acro_init(suppress = FALSE)
```

Arguments

suppress	Whether to automatically apply suppression.
----------	---

Value

No return value, called for side effects

acro_lm *Fits Ordinary Least Squares Regression*

Description

Fits Ordinary Least Squares Regression

Usage

```
acro_lm(formula, data)
```

Arguments

formula	The formula specifying the model.
data	The data for the model.

Value

Regression Results Wrapper.

acro_pivot_table	<i>Pivot table</i>
------------------	--------------------

Description

Pivot table

Usage

```
acro_pivot_table(
  data,
  values = NULL,
  index = NULL,
  columns = NULL,
  aggfunc = "mean"
)
```

Arguments

data	The data to operate on.
values	Column to aggregate, optional.
index	If an array is passed, it must be the same length as the data. The list can contain any of the other types (except list). Keys to group by on the pivot table index. If an array is passed, it is being used as the same manner as column values.
columns	If an array is passed, it must be the same length as the data. The list can contain any of the other types (except list). Keys to group by on the pivot table column. If an array is passed, it is being used as the same manner as column values.
aggfunc	If list of strings passed, the resulting pivot table will have hierarchical columns whose top level are the function names

Value

Cross tabulation of the data.

acro_print_outputs *Prints the current results dictionary.*

Description

Prints the current results dictionary.

Usage

acro_print_outputs()

Value

No return value, called for side effects

acro_remove_output *Remove outputs*

Description

Remove outputs

Usage

acro_remove_output(name)

Arguments

name Key specifying which output to remove, e.g., 'output_0'.

Value

No return value, called for side effects

acro_rename_output *Rename outputs*

Description

Rename outputs

Usage

```
acro_rename_output(old, new)
```

Arguments

old	The old name of the output.
new	The new name of the output.

Value

No return value, called for side effects

acro_surv_func *Survival analysis*

Description

Survival analysis

Usage

```
acro_surv_func(time, status, output, filename = "kaplan-meier.png")
```

Arguments

time	An array of times (censoring times or event times).
status	Status at the event time.
output	A string determine the type of output. Available options are table or plot.
filename	The name of the file where the plot will be saved.

Value

The survival table or plot.

acro_table	<i>Compute a simple cross tabulation of two (or more) factors.</i>
------------	--

Description

Compute a simple cross tabulation of two (or more) factors.

Usage

```
acro_table(index, columns, dnn = NULL, deparse.level = 0, ...)
```

Arguments

index	Values to group by in the rows.
columns	Values to group by in the columns.
dnn	The names to be given to the dimensions in the result
deparse.level	Controls how the default dnn is constructed.
...	Any other parameters.

Value

Cross tabulation of the data

create_virtualenv	<i>Create a python virtual environment</i>
-------------------	--

Description

Create a python virtual environment

Usage

```
create_virtualenv(...)
```

Arguments

...	Any other parameters.
-----	-----------------------

Value

No return value, called for side effects

install_acro	<i>Install acro</i>
--------------	---------------------

Description

Install acro

Usage

```
install_acro(envname = "r-acro", ...)
```

Arguments

envname	the name of the Python virtual environment
...	Any other parameters.

Value

No return value, called for side effects

lung	<i>Lung Cancer Survival Data</i>
------	----------------------------------

Description

The lung dataset contains information about lung cancer survival.

Usage

lung

Format

A data frame with columns:

inst institutional identification

time Survival time in months.

status Survival status (1 = death, 0 = censored).

age Age of the patient at the start of the study.

sex Gender of the patient.

ph.ecog Performance status (Eastern Cooperative Oncology Group).

ph.karno 'Karnofsky' performance status.

pat.karno 'Karnofsky' performance status as assessed by the patient.

meal.cal Daily caloric intake at the start of the study.

wt.loss Weight loss in the last six months.

Examples

```
data(lung)
```

```
nursery_data
```

```
Nursery Database
```

Description

This dataset is originated from a hierarchical decision model created to evaluate applications for nursery schools.

Usage

```
nursery_data
```

Format

A data frame with columns: A data frame with 12960 rows and 9 columns:

parents Parents' occupation

has_nurs Child's nursery

form Form of the family

children Number of children

housing Housing conditions

finance Financial standing of the family

social Social conditions

health Health conditions

recommend The ranking of applications for nursery schools

Source

<https://www.openml.org/search?type=data&status=active&id=26&sort=runs>

Examples

```
data(nursery_data)
```

Index

* datasets

lung, [11](#)

nursery_data, [12](#)

acro_add_comments, [2](#)

acro_add_exception, [3](#)

acro_crosstab, [3](#)

acro_custom_output, [4](#)

acro_finalise, [4](#)

acro_glm, [5](#)

acro_hist, [5](#)

acro_init, [6](#)

acro_lm, [6](#)

acro_pivot_table, [7](#)

acro_print_outputs, [8](#)

acro_remove_output, [8](#)

acro_rename_output, [9](#)

acro_surv_func, [9](#)

acro_table, [10](#)

create_virtualenv, [10](#)

install_acro, [11](#)

lung, [11](#)

nursery_data, [12](#)