Package ‘tidycomm’

September 22, 2019

Title  Data Modification and Analysis for Communication Research
Version  0.1.0
Description  Provides convenience functions for common data
modification and analysis tasks in communication research. This
includes functions for univariate and bivariate data analysis, index
generation and reliability computation, and intercoder reliability
tests. All functions follow the style and syntax of the tidyverse, and
are construed to perform their computations on multiple variables at
once. Functions for univariate and bivariate data analysis comprise
summary statistics for continuous and categorical variables, as well
as several tests of bivariate association including effect sizes.
Functions for data modification comprise index generation and
automated reliability analysis of index variables. Functions for
intercoder reliability comprise tests of several intercoder
reliability estimates, including simple and mean pairwise percent
agreement, Krippendorff’s Alpha (Krippendorff 2004, ISBN:
9780761915454), and various Kappa coefficients (Brennan & Prediger

License  GPL-3
URL  https://github.com/joon-e/tidycomm
BugReports  https://github.com/joon-e/tidycomm/issues
Depends  R (>= 2.10)
Imports  broom, dplyr, forcats, glue, magrittr, MBESS, purrr, rlang,
stringr, tibble, tidyr
Suggests  covr, knitr, rmarkdown, testthat (>= 2.1.0), tidyselect
VignetteBuilder  knitr
Encoding  UTF-8
LazyData  true
RoxygenNote  6.1.1
NeedsCompilation  no
**Description**

Add a rowwise mean or sum index of specific variables to the dataset.

**Usage**

```r
add_index(data, name, ..., type = "mean", na.rm = TRUE)
```

**Arguments**

- `data`: a tibble
- `name`: Name of the index column to compute.
- `...`: Variables used for the index.
- `type`: Type of index to compute. Either "mean" (default) or "sum".
- `na.rm`: a logical value indicating whether NA values should be stripped before the computation proceeds. Defaults to TRUE.

**Value**

- a tibble
## correlate

**Compute correlation coefficients**

### Description

Computes correlation coefficients for all combinations of the specified variables. If no variables are specified, all numeric (integer or double) variables are used.

### Usage

```r
correlate(data, ..., method = "pearson")
```

### Arguments

- **data**
  
  A tibble

- **...**
  
  Variables to compute correlations for (column names). Leave empty to compute for all numeric variables in data.

- **method**
  
  A character string indicating which correlation coefficient is to be computed. One of "pearson" (default), "kendall", or "spearman"

### Value

A tibble

### Examples

```r
WoJ %>% correlate(ethics_1, ethics_2, ethics_3)
WoJ %>% correlate()
```
Description

Computes contingency table for one independent (column) variable and one or more dependent (row) variables.

Usage

```r
crosstab(data, col_var, ..., add_total = FALSE, percentages = FALSE, chi_square = FALSE)
```

Arguments

- `data`: a tibble
- `col_var`: Independent (column) variable.
- `...`: Dependent (row) variables.
- `add_total`: Logical indicating whether a 'Total' column should be computed. Defaults to FALSE.
- `percentages`: Logical indicating whether to output column-wise percentages instead of absolute values. Defaults to FALSE.
- `chi_square`: Logical indicating whether a Chi-square test should be computed. Test results will be reported via message(). Defaults to FALSE.

Value

- a tibble

See Also

Other categorical: `tab_frequencies`

Examples

```r
WoJ %>% crosstab(reach, employment)
WoJ %>% crosstab(reach, employment, add_total = TRUE, percentages = TRUE, chi_square = TRUE)
```
**describe**

*Describe variables*

**Description**

Describe variables by several measures of central tendency and variability. If no variables are specified, all numeric (integer or double) variables are described.

**Usage**

```r
describe(data, ..., na.rm = TRUE)
```

**Arguments**

- `data` a tibble
- `...` Variables to describe (column names). Leave empty to describe all numeric variables in data.
- `na.rm` a logical value indicating whether NA values should be stripped before the computation proceeds. Defaults to TRUE.

**Value**

a tibble

**Examples**

```r
iris %>% describe()
mtcars %>% describe(mpg, am, cyl)
```

---

**fbposts**

*Facebook posts reliability test*

**Description**

45 political facebook posts coded by 6 coders for an intercoder reliability test, focused on populist messages.

**Usage**

```r
fbposts
```
get_reliability

Format

A data frame with 270 rows and 7 variables

- **post_id**: Numeric id of the coded Facebook post
- **coder_id**: Numeric id of the coder
- **type**: Type of Facebook post, one of "link", "photo", "status", or "video"
- **n_pictures**: Amount of pictures attached to the post, ranges from 0 to 6
- **pop_elite**: Populism indicator: Does the Facebook post attack elites?, 0 = "no attacks on elites", 1 = "attacks political actors", 2 = "attacks public administration actors", 3 = "attacks economical actors", 4 = "attacks media actors/journalists", 9 = "attacks other elites"
- **pop_people**: Populism indicator: Does the Facebook refer to 'the people'? 0 = "does not refer to 'the people'"", 1 = "refers to 'the people'"
- **pop_othering**: Populism indicator: Does the Facebook attack 'others'? 0 = "no attacks on 'others'"", 1 = "attacks other cultures", 2 = "attacks other political stances", 3 = "attacks other 'others'"

---

**get_reliability**

Get reliability estimates of index variables

---

Description

Get reliability estimates of index variables created with `add_index`.

Usage

```r
get_reliability(data, ..., type = "alpha", interval.type = NULL, bootstrap.samples = NULL, conf.level = NULL, progress = FALSE)
```

Arguments

- **data**: a tibble
- **...**: Index variables created with `add_index`. Leave empty to get reliability estimates for all index variables.
- **type**: Type of reliability estimate. See `ci.reliability`
- **interval.type**: Type of reliability estimate confidence interval. See `ci.reliability`
- **bootstrap.samples**: Number of bootstrap samples for CI calculation. See `ci.reliability`
- **conf.level**: Confidence level for estimate CI. See `ci.reliability`
- **progress**: Show progress for reliability estimate computation. Useful if using computationally intense computations (e. g., many bootstrapping samples) and many index variables.
tab_frequencies

Value

a tibble

See Also

add_index() to create index variables

Examples

WoJ %>%
  add_index(ethical_flexibility, ethics_1, ethics_2, ethics_3, ethics_4) %>%
  get_reliability()

---

Description

Tabulates frequencies for one or more categorical variable, including relative, and cumulative frequencies.

Usage

tab_frequencies(data, ...)

Arguments

data a tibble

... Variables to tabulate

Value

a tibble

See Also

Other categorical: crosstab

Examples

WoJ %>% tab_frequencies(employment)
WoJ %>% tab_frequencies(employment, country)
test_icr  

**Perform an intercoder reliability test**

**Description**

Performs an intercoder reliability test by computing various intercoder reliability estimates for the included variables.

**Usage**

`test_icr(data, unit_var, coder_var, ..., levels = NULL, na.omit = FALSE, agreement = TRUE, holsti = TRUE, kripp_alpha = TRUE, cohens_kappa = FALSE, fleiss_kappa = FALSE, brennan_prediger = FALSE)`

**Arguments**

- `data` a tibble
- `unit_var` Variable with unit identifiers
- `coder_var` Variable with coder identifiers
- `...` Variables to compute intercoder reliability estimates for. Leave empty to compute for all variables (excluding `unit_var` and `coder_var`) in `data`.
- `levels` Optional named vector with levels of test variables
- `na.omit` Logical indicating whether NA values should be stripped before computation. Defaults to `FALSE`.
- `agreement` Logical indicating whether simple percent agreement should be computed. Defaults to `TRUE`.
- `holsti` Logical indicating whether Holsti’s reliability estimate (mean pairwise agreement) should be computed. Defaults to `TRUE`.
- `kripp_alpha` Logical indicating whether Krippendorff’s Alpha should be computed. Defaults to `TRUE`.
- `cohens_kappa` Logical indicating whether Cohen’s Kappa should be computed. Defaults to `FALSE`.
- `fleiss_kappa` Logical indicating whether Fleiss’ Kappa should be computed. Defaults to `FALSE`.
- `brennan_prediger` Logical indicating whether Brennan & Prediger’s Kappa should be computed (extension to 3+ coders as proposed by von Eye (2006)). Defaults to `FALSE`.

**Value**

a tibble
References


Examples

```r
fbposts %>% test_icr(post_id, coder_id, pop_elite, pop_othering)
fbposts %>% test_icr(post_id, coder_id, levels = c(n_pictures = "ordinal"), fleiss_kappa = TRUE)
```

---

to_correlation_matrix  Create correlation matrix

Description

Turns the tibble exported from `correlate` into a correlation matrix.

Usage

```r
to_correlation_matrix(data)
```

Arguments

- `data` a tibble returned from `correlate`

Value

a tibble

Examples

```r
WoJ %>% correlate() %>% to_correlation_matrix()
```
t_test  

Compute t-tests

Description

Computes t-tests for one group variable and specified test variables. If no variables are specified, all numeric (integer or double) variables are used.

Usage

t_test(data, group_var, ..., var.equal = TRUE, paired = FALSE, pooled_sd = TRUE, levels = NULL, case_var = NULL)

Arguments

data  
a tibble

group_var  
group variable (column name)

...  
test variables (column names). Leave empty to compute t-tests for all numeric variables in data.

var.equal  
a logical variable indicating whether to treat the two variances as being equal. If TRUE then the pooled variance is used to estimate the variance otherwise the Welch (or Satterthwaite) approximation to the degrees of freedom is used. Defaults to TRUE.

paired  
a logical indicating whether you want a paired t-test. Defaults to FALSE.

pooled_sd  
a logical indicating whether to use the pooled standard deviation in the calculation of Cohen’s d. Defaults to TRUE.

levels  
optional: a vector of length two specifying the two levels of the group variable.

case_var  
optional: case-identifying variable (column name). If you set paired = TRUE, specifying a case variable will ensure that data are properly sorted for a dependent t-test.

Value

a tibble

Examples

WoJ %>% t_test(temp_contract, autonomy_selection, autonomy_emphasis)
WoJ %>% t_test(temp_contract)
WoJ %>% t_test(employment, autonomy_selection, autonomy_emphasis, levels = c("Full-time", "Freelancer"))
unianova

Compute one-way ANOVAs

Description

Computes one-way ANOVAs for one group variable and specified test variables. If no variables are specified, all numeric (integer or double) variables are used.

Usage

unianova(data, group_var, ..., descriptives = FALSE, post_hoc = FALSE)

Arguments

data a tibble
group_var group variable (column name)
... test variables (column names). Leave empty to compute ANOVAs for all numeric variables in data.
descriptives a logical indicating whether descriptive statistics (mean & standard deviation) for all group levels should be added to the returned tibble. Defaults to FALSE.
post_hoc a logical indicating whether post-hoc tests (Tukey’s HSD) should be computed. Results of the post-hoc test will be added in a list column of result tibbles.

Value

a tibble

Examples

WoJ %>% unianova(employment, autonomy_selection, autonomy_emphasis)
WoJ %>% unianova(employment)
WoJ %>% unianova(employment, descriptives = TRUE, post_hoc = TRUE)

WoJ Worlds of Journalism sample data

Description

A subset of data from the Worlds of Journalism 2012-16 study containing survey data of 1,200 journalists from five European countries.

Usage

WoJ
Format

A data frame with 1200 rows and 15 variables:

country  Country of residence
reach    Reach of medium
employment Current employment situation
temp_contract Type of contract (if current employment situation is either full-time or part-time
autonomy_selection Autonomy in news story selection, scale from 1 (no freedom at all) to 5 (complete freedom)
autonomy_emphasis Autonomy in news story emphasis, scale from 1 (no freedom at all) to 5 (complete freedom)
ethics_1 Agreement with statement "Journalists should always adhere to codes of professional ethics, regardless of situation and context", scale from 1 (strongly disagree) to 5 (strongly agree) (reverse-coded!)
ethics_2 Agreement with statement "What is ethical in journalism depends on the specific situation.", scale from 1 (strongly disagree) to 5 (strongly agree)
ethics_3 Agreement with statement "What is ethical in journalism is a matter of personal judgment.", scale from 1 (strongly disagree) to 5 (strongly agree)
ethics_4 Agreement with statement "It is acceptable to set aside moral standards if extraordinary circumstances require it.", scale from 1 (strongly disagree) to 5 (strongly agree)
work_experience Work experience as a journalist in years
trust_parliament Trust placed in parliament, scale from 1 (no trust at all) to 5 (complete trust)
trust_government Trust placed in government, scale from 1 (no trust at all) to 5 (complete trust)
trust_parties Trust placed in parties, scale from 1 (no trust at all) to 5 (complete trust)
trust_politicians Trust placed in politicians in general, scale from 1 (no trust at all) to 5 (complete trust)

Source

Index

*Topic datasets
  fbposts, 5
  WoJ, 11

add_index, 2, 6
add_index(), 7

ci.reliability, 6
correlate, 3, 9
crosstabs, 4, 7

describe, 5

fbposts, 5

gt.get.reliability, 6
gt.get.reliability(), 3

t_test, 10
tab.frequencies, 4, 7
test_icr, 8
tibble, 2–11

to_correlation_matrix, 9

unianova, 11

WoJ, 11