Package 'drawsample'

March 1, 2021

Type Package

Title Draw Samples with the Desired Properties from a Data Set

Version 1.0.0

Language en-US

Maintainer Kubra Atalay Kabasakal <katalay@hacettepe.edu.tr>

Description A tool to sample data with the desired properties. Samples can be
drawn by purposive sampling with determining distributional
conditions, such as deviation from normality (skewness and kurtosis),
and sample size in quantitative research studies.
For purposive sampling, a researcher has something in mind and participants that
fit the purpose of the study are included (Etikan, Musa, & Alkassim, 2015)
<doi:10.11648/j.ajtas.20160501.11>. Purposive sampling can be useful for answering
many research questions (Klar & Leeper, 2019)
<doi:10.1002/9781119083771.ch21>.

License GPL-2

Encoding UTF-8

Imports dplyr, lattice, tibble, psych, magrittr, moments, readxl,
shiny, shinycssloaders, shinydashboard, xlsx, utils, purrr,
rlang

Suggests rmarkdown, knitr, testthat (>= 3.0.0)

LazyData true

RoxygenNote 7.1.1

URL https://github.com/atalay-k/drawsample

Depends R (>= 2.10)

BugReports https://github.com/atalay-k/drawsample/issues

Config/testthat/edition 3

NeedsCompilation no

Author Kubra Atalay Kabasakal [aut, cre]
(<https://orcid.org/0000-0002-3580-5568>),
Huseyin Yıldız [ctb] (<https://orcid.org/0000-0003-2387-263X>)

R topics documented:

---

drawsample-package   *Draw Samples with the Desired Properties from a Data Set*

Description

draw_sample, function takes a sample of the specified sample size,#' skewness, and kurtosis form a data set (dist)with or without resampling. Fleishman's power method (doi: 10.1007/BF02293811) was used for the desired skewness and kurtosis level. Therefore, the coefficient of skewness can be chosen between 0 and 3.6. Although the kurtosis coefficient varies for each skewness coefficient and varies from -1.2 and 20. If convenient kurtosis and skew values are not provided, no solutions can be found and an error is given.

Author(s)

Maintainer: Kubra Atalay Kabasakal <katalay@hacettepe.edu.tr> (ORCID)

Other contributors:

- Huseyin Yıldız <huseyinyildiz35@gmail.com> (ORCID) [contributor]

References

Fleishman AI (1978). A Method for Simulating Non-normal Distributions. *Psychometrika, 43, 521-532.* doi: 10.1007/BF02293811.

Atalay Kabasakal, K. & Gunduz, T . (2020). Drawing a Sample with Desired Properties from Population in R Package "drawsample".*Journal of Measurement and Evaluation in Education and Psychology,11*(4),405-429. doi: 10.21031/epod.790449

See Also

Useful links:

- https://github.com/atalay-k/drawsample
- Report bugs at https://github.com/atalay-k/drawsample/issues

constants_table	*Fleishman's Power Method Transformation Constants*

Description

This table includes Fleishman's Power Method Transformation constants.

Usage

```
constants_table
```

Format

A `data.frame` with 5 columns, which are

Skew the skewness value

Kurtosis the standardized kurtosis value

b Outcome that is based on `Skew,Kurtosis`

c Outcome that is based on `Skew,Kurtosis`

d Outcome that is based on `Skew,Kurtosis`

References

Fleishman AI (1978). A Method for Simulating Non-normal Distributions. *Psychometrika, 43, 521-532.* doi: 10.1007/BF02293811.

Fialkowski, A. C. (2018). SimMultiCorrData: Simulation of Correlated Data with Multiple Variable Types. R package version 0.2.2. Retrieved from https://cran.r-project.org/web/packages/SimMultiCorrData/index.html

See Also

find_constants

draw_sample	*Draw Samples with the Desired Properties from a Data Set*

Description

A function to sample data with desired properties.

Usage

```
draw_sample(
  dist,
  n,
  skew,
  kurts,
  replacement = FALSE,
  output_name = c("sample", "default")
)
```

Arguments

dist	data frame:consists of id and scores with no missing
n	numeric: desired sample size
skew	numeric: the skewness value
kurts	numeric: the kurtosis value
replacement	logical:Sample with or without replacement? (default is FALSE).
output_name	character: a vector of two components. The first component is the name of the output file, user can change the second component.

Details

The execution of the function may take some time since it tries to obtain the specified value for skewness and kurtosis.

Value

This function returns a `list` including following:

- a matrix: Descriptive statistics of the given data, the reference vector and the sample.

- a data frame: The id's and scores of the sample

- graph: Histograms for the "data" and the "sample"

References

Fleishman AI (1978). A Method for Simulating Non-normal Distributions. *Psychometrika, 43, 521-532.* doi: 10.1007/BF02293811.

Fialkowski, A. C. (2018). SimMultiCorrData: Simulation of Correlated Data with Multiple #' Variable Types. R package version 0.2.2. Retrieved from https://cran.r-project.org/web/packages/SimMultiCorrData/index.html

Atalay Kabasakal, K. & Gunduz, T. (2020). Drawing a Sample with Desired Properties from Population in R Package "drawsample".*Journal of Measurement and Evaluation in Education and Psychology,11*(4),405-429. doi: 10.21031/epod.790449

Examples

```
# Example data provided with package
data(example_data)
## Not run:
# Draw a sample based on Score_1(from negatively skewed to normal)
# draw_sample(dist=example_data[,c(1,2)],n=200,skew = 0,kurts = 0,
# output_name = c("sample", "1"))
# Draw a sample based on Score_2 (from negatively skewed to positively skewed)
# draw_sample(dist=example_data[,c(1,3)],n=200,skew = 1,kurts = 1,
# output_name = c("sample", "2"))
# Draw a sample based on Score_2 (from negatively skewed to positively skewed
# with replacement)
# draw_sample(dist=example_data[,c(1,3)],n=200,skew = 0.5,kurts = 0.4,
# replacement=TRUE,output_name = c("sample", "3"))

## End(Not run)
```

---

draw_sample_n   *Draw Samples with the Desired Properties from a Data Set*

Description

A function to sample data with desired properties.

Usage

```
draw_sample_n(
  dist,
  n,
  skew,
  kurts,
  location = 0,
  delta_var = 0,
  output_name = c("sample", "default")
)
```

Arguments

dist	data frame:consists of id and scores with no missing
n	numeric: desired sample size
skew	numeric: the skewness value
kurts	numeric: the kurtosis value
location	numeric: the value for adjusting mean (default is 0).
delta_var	numeric: the value for adjusting variance (default is 0).
output_name	character: a vector of two components. The first component is the name of the output file, user can change the second component.

Details

The desired skewness and kurtosis values cannot be met while the function execution is faster. The attributes of kurtosis are in doubt. This is because the range of kurtosis is greater than the skewness. For `location` values can be entered to position the midpoint or mean of the distribution differently. For `delta_var` the value can be entered for how much will increase or decrease the variability of reference distribution. In other words, the reference distribution is generated as the standard normal distribution, unless the user changes the default values of the `location` and `delta_var` arguments.

Value

This function returns a `list` including following:

- a matrix: Descriptive statistics of the given data, the reference vector and the sample.
- a data frame: The id's and scores of the sample
- graph: Histograms for the "data" and the "sample"

References

Fleishman AI (1978). A Method for Simulating Non-normal Distributions. *Psychometrika, 43, 521-532*. doi: 10.1007/BF02293811.

Fialkowski, A. C. (2018). SimMultiCorrData: Simulation of Correlated Data with Multiple #' Variable Types. R package version 0.2.2. Retrieved from https://cran.r-project.org/web/packages/SimMultiCorrData/index.html

Examples

```
# Example data provided with package
data(example_data)
## Not run:
# Draw a sample based on Score_1
# draw_sample_n(dist=example_data[,c(1,2)],n=200,skew = 0,kurts = 0, location=0, delta_var=0,
# output_name = c("sample", "4"))
# Draw a sample based on Score_2 (location par)
# draw_sample_n(dist=example_data[,c(1,3)],n=200,skew = 1,kurts = 1,location=-0.5,delta_var=0,
# output_name = c("sample", "5"))
# Draw a sample based on Score_2 (delta_var par)
# draw_sample_n(dist=example_data[,c(1,3)],n=200,skew = 0.5,kurts = 0.4,location=0,delta_var=0.3,
# output_name = c("sample", "6"))

## End(Not run)
```

---

draw_sample_shiny   *Draw Samples with a Shiny Applications*

Description

Performing package functions with user friendly 'shiny' interface.

Usage

```
draw_sample_shiny()
```

Examples

```
# Example data provided with package
data(example_data)
## Not run:
# if(interactive()){
## Run this code for launching the 'shiny' application
#   draw_sample_shiny()
#   }
#
## End(Not run)
```

---

example_data   *Example Data*

Description

The example data set is made of 500 subjects ids and total scores from two different tests.

Usage

```
data(example_data)
```

Format

A data.frame with 3 columns, which are

ID students' id

Score_1 Scores of test 1

Score_2 Scores of test 2

Index