

Package ‘descstatsr’

October 12, 2018

Type Package

Title Descriptive Univariate Statistics

Version 0.1.0

Author Harish Kumar

Maintainer Harish Kumar <kadamatiharish@gmail.com>

Description It generates summary statistics on the input dataset using different descriptive univariate statistical measures on entire data or at a group level. Though there are other packages which does similar job but each of these are deficient in one form or other, in the measures generated, in treating numeric, character and date variables alike, no functionality to view these measures on a group level or the way the output is represented. Given the foremost role of the descriptive statistics in any of the exploratory data analysis or solution development, there is a need for a more constructive, structured and refined version over these packages. This is the idea behind the package and it brings together all the required descriptive measures to give an initial understanding of the data quality, distribution in a faster,easier and elaborative way.The function brings an additional capability to be able to generate these statistical measures on the entire dataset or at a group level. It calculates measures of central tendency (mean, median), distribution (count, proportion), dispersion (min, max, quantile, standard deviation, variance) and shape (skewness, kurtosis). Addition to these measures, it provides information on the data type, count on no. of rows, unique entries and percentage of missing entries. More importantly the measures are generated based on the data types as required by them,rather than applying numerical measures on character and data variables and vice versa. Output as a dataframe object gives a very neat representation, which often is useful when working with a large number of columns. It can easily be exported as csv and analyzed further or presented as a summary report for the data.

License GPL-2

Encoding UTF-8

LazyData true**Imports** datasets,zoo,moments**Suggests** knitr**RoxygenNote** 6.1.0**NeedsCompilation** no**Repository** CRAN**Date/Publication** 2018-10-12 13:10:03 UTC

R topics documented:

desc_stats 2

Index 4

desc_stats *Descriptive Univariate Statistics*

Description

The function summarizes the input data using different descriptive univariate statistical measures on grouped or ungrouped level.

Usage

```
desc_stats(dataset, show_levels = 5, decimal_points = 2,
           group_variable = NULL, miss_val = NULL)
```

Arguments

dataset	A data.frame object, an input dataset for which descriptive statistics needs to be calculated
show_levels	An integer value. It controls how many top character/factor levels with their proportions needs to be displayed in descending order of their proportions, by default it is set to 5.
decimal_points	An integer value. It controls no of decimal points to which numeric data needs to be rounded off, by default it is set to 2.
group_variable	A character vector. Specify the character or factor variable/variables on whose unique group levels the data should be split and univariate statistics needs to be generated.
miss_val	A character vector. Specify different strings which needs to be considered as missing values.

Details

The functions calculates following measures on the input data:

Measures of Central Tendency: Mean, Median

Measures of Distribution: Count, Proportion

Measures of Dispersion: Min, Max, Quantile, Standard Deviation, Variance

Measures of shape: Skewness, Kurtosis

Addition to these measures, the function provides information on the data type, count on no. of rows, unique entries and percentage of missing entries

All the above statistics can be generated for the entire data or at a group level. The variables/variables specified to group_variable parameter splits the data into groups based on the unique levels of the variable/variables specified and calculates descriptive statistics on each of these levels. .

Value

A data.frame object with descriptive univariate statistics listed for numerical,categorical and date variables at group level, if specified, else for entire data.

Examples

```
desc_stats(iris,show_levels=2,decimal_points=2,group_variable=c("Species"),miss_val=c("unknown"))
desc_stats(iris,show_levels=2,decimal_points=2,group_variable=c("Species"))
desc_stats(iris,show_levels=2,decimal_points=2)
desc_stats(iris,show_levels=2)
desc_stats(iris)
```

Index

desc_stats, 2