

Package: DataSum (via r-universe)

August 29, 2024

Type Package

Title The DataSumm function takes a data frame as input and applies the Datum function to each column, returning a data frame with the summary statistics for each column

Version 0.1.0

Author Mr. Uzair Javid Khan and Dr. Immad Ahmad Shah

Maintainer Mr. Uzair Javid Khan <uzairkhan11@gmail.com>

Description The moments package provides functions for calculating various statistical moments and related measures, such as skewness and kurtosis. The dplyr package is used for data manipulation, and the nortest package is used for normality testing. The find_mode function takes a data vector as input and returns the mode(s) of the data. The shapiro_normality_test function performs a Shapiro-Wilk normality test on the input data, and returns ``Normal" if the data is normally distributed (p-value > 0.05), and ``Not Normal" otherwise. If the data length is outside the valid range for the Shapiro-Wilk test (3 to 5000), it performs an Anderson-Darling normality test instead. The Datum function takes a data vector as input and returns a data frame with various summary statistics, including data type, sample size, mean, mode, median, variance, standard deviation, maximum, minimum, range, skewness, kurtosis, and normality test result. If the data is numeric, it calculates the statistics accordingly. If the data is character or factor, it provides the mode and marks the other statistics as not applicable (NA). The DataSumm function takes a data frame as input and applies the Datum function to each column, returning a data frame with the summary statistics for each column. Measures of Central Tendency Mean: The average of the values, calculated by summing all the values and dividing by the number of values. Median: The middle value when the data is arranged in order. If there are an even number of values, the median is the average of the two middle values. Mode: The value that appears most frequently in the data set. Measures of Dispersion Range: The difference between the largest and

smallest values in the data set. Variance: A measure of how spread out the values are from the mean, calculated as the average squared deviation from the mean. Standard Deviation: The square root of the variance, providing a measure of the average amount each value deviates from the mean. Other Measures Skewness: A measure of the asymmetry of the probability distribution of a random variable around its mean. Positive skewness indicates a distribution with an asymmetric tail extending towards more positive values. Kurtosis: A measure of the "peakedness" of the probability distribution of a random variable. Normality: A test to determine if the data follows a normal (Gaussian) distribution, such as the Shapiro-Wilk test.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

Imports moments, dplyr, nortest, stats

Repository <https://uzairkhan11w.r-universe.dev>

RemoteUrl <https://github.com/uzairkhan11w/datasum>

RemoteRef HEAD

RemoteSha 153606be389b45d8ae7325469214e9076c43f8b0

Contents

DataSumm	2
Datum	3
getmode	4
shapiro_normality_test	4
Index	5

DataSumm

Summarize an Entire Data Frame

Description

This function summarizes each column of a data frame by calculating various statistics.

Usage

DataSumm(data)

Arguments

data A data frame.

Value

A data frame with summary statistics for each column.

Examples

```
DataSumm(iris)
```

Datum	<i>Summarize a Single Vector</i>
-------	----------------------------------

Description

This function summarizes a single vector by calculating various statistics.

Usage

```
Datum(data)
```

Arguments

data A numeric, character, or factor vector.

Value

A data frame with summary statistics.

Examples

```
Datum(rnorm(100))
```

getmode	<i>Get Mode of a Numeric Vector</i>
---------	-------------------------------------

Description

This function calculates the mode of a numeric vector.

Usage

```
getmode(data)
```

Arguments

data	A numeric vector.
------	-------------------

Value

The mode of the numeric vector.

Examples

```
getmode(c(1, 2, 2, 3, 4))
```

shapiro_normality_test	<i>Perform Normality Test</i>
------------------------	-------------------------------

Description

This function performs the Shapiro-Wilk test if the sample size is between 3 and 5000. Otherwise, it performs the Anderson-Darling test.

Usage

```
shapiro_normality_test(data)
```

Arguments

data	A numeric vector.
------	-------------------

Value

A character string indicating whether the data is "Normal" or "Not Normal".

Examples

```
shapiro_normality_test(rnorm(100))
```

Index

DataSumm, [2](#)

Datum, [3](#)

getmode, [4](#)

shapiro_normality_test, [4](#)