Package 'DataSum'

January 20, 2025

Type Package

Title Comprehensive Data Summarization for Statistical Analysis

Version 0.1.0

Description Summarizing data frames by calculating various statistical measures, including measures of central tendency, dispersion, skewness(), kurtosis(), and normality tests. The package leverages the 'moments' package for calculating statistical moments and related measures, the 'dplyr' package for data manipulation, and the 'nortest' package for normality testing. 'DataSum' includes functions such as getmode() for finding the mode(s) of a data vector, shapiro_normality_test() for performing the Shapiro-Wilk test (Shapiro & Wilk 1965 <doi:10.1093/biomet/52.3-4.591>) (or the Anderson-Darling test when the data length is outside the valid range for the Shapiro-Wilk test) (Stephens 1974 <doi:10.1080/01621459.1974.10480196>), Datum() for generating a comprehensive summary of a data vector with various statistics (including data type, sample size, mean, mode, median, variance, standard deviation, maximum, minimum, range, skewness(), kurtosis(), and normality test result) (Joanes & Gill 1998 <doi:10.1111/1467-9884.00122>), and DataSumm() for applying the Datum() function to each column of a data frame. Emphasizing the importance of normality testing, the package provides robust tools to validate whether data follows a normal distribution, a fundamental assumption in many statistical analyses and models.

License GPL-3

Encoding UTF-8

RoxygenNote 7.3.2

Imports moments, dplyr, nortest, stats

URL https://github.com/Uzairkhan11w/DataSum

BugReports https://github.com/Uzairkhan11w/DataSum/issues

NeedsCompilation no

Author Immad Ahmad Shah [aut],
Uzair Javid Khan [aut, cre],
Sukhdev Mishra [aut]

Maintainer Uzair Javid Khan <uzairkhan11w@gmail.com>

Repository CRAN

Date/Publication 2024-08-28 08:50:06 UTC

Contents

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   *Summarize a Single Vector*

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   *Get Mode of a Numeric Vector*

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

Perform Normality Test

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