Package: DescribeDF (via r-universe)

August 31, 2024

Type Package

Title Description of a Data Frame

Version 0.2.1

Author Sandip Garai [aut, cre], Ranjit Kumar Paul [aut]

Maintainer Sandip Garai <sandipnicksandy@gmail.com>

Description Helps to describe a data frame in hand. Has been developed during PhD work of the maintainer. More information may be obtained from Garai and Paul (2023) <doi:10.1016/j.iswa.2023.200202>.

License GPL-3

Encoding UTF-8

Imports dplyr, e1071, psych, fNonlinear, tseries

NeedsCompilation no

Date/Publication 2023-05-09 07:30:12 UTC

Repository https://sandipgarai.r-universe.dev

RemoteUrl https://github.com/cran/DescribeDF

RemoteRef HEAD

RemoteSha fb30e3b466967ee6930fda27d31231ef2852ba04

Contents

df_descstat *Descriptive Statistics of A Data Frame*

Description

Gives descriptive statistics for a data frame (df) excluding the first column. First column mentions different statistics. Other columns are the Statistics values of the particular column (series). "p_value <= 0.01: ***; p_value <= 0.05: **; p_value <= 0.1: *".

Usage

```
df_descstat(df)
```

Arguments

df	Data Frame with first column as serial number or date

Value

- desc_df - A table contains 10 descriptive statistics rowwise

References

- Garai, S., & Paul, R. K. (2023). Development of MCS based-ensemble models using CEEMDAN decomposition and machine intelligence. Intelligent Systems with Applications, 18, 200202.
- Garai, S., Paul, R. K., Rakshit, D., Yeasin, M., Paul, A. K., Roy, H. S., Barman, S. & Manjunatha, B. (2023). An MRA Based MLR Model for Forecasting Indian Annual Rainfall Using Large Scale Climate Indices. International Journal of Environment and Climate Change, 13(5), 137-150.

Examples

```
# create a vector of dates
dates <- seq(as.Date("2021-01-01"), as.Date("2021-01-05"), by = "day")

# create vectors of random numeric data for columns A through E
A <- runif(5, 0, 1)
B <- runif(5, 0, 1)
C <- runif(5, 0, 1)
D <- runif(5, 0, 1)
E <- runif(5, 0, 1)

# combine the vectors into a data frame
df <- data.frame(Date = dates, A = A, B = B, C = C, D = D, E = E)

# print the data frame
print(df)
```

```
# Data Description
df_descstat(df)
```

`df_nonlinearity` *Non linearity test of A Data Frame*

Description

Gives non linearity test result for a df excluding the first column (contains serial number or date). This will give a list of data frames. Data frames are named as the names of columns of the data frame. First column mentions different statistics (eps). Other columns are the Statistics values of the particular dimension. "p_value <= 0.01: ***; p_value <= 0.05: **; p_value <= 0.1: *".

Usage

```
df_nonlinearity(df)
```

Arguments

`df`	Data Frame with first column as serial number or date

Value

- result_list - List of data frames named as the column names of provided data frame. Each df is such that first column mentions different statistics and other columns are the Statistics values of the particular dimension.

References

- Garai, S., & Paul, R. K. (2023). Development of MCS based-ensemble models using CEEMDAN decomposition and machine intelligence. Intelligent Systems with Applications, 18, 200202.
- Garai, S., Paul, R. K., Rakshit, D., Yeasin, M., Paul, A. K., Roy, H. S., Barman, S. & Manjunatha, B. (2023). An MRA Based MLR Model for Forecasting Indian Annual Rainfall Using Large Scale Climate Indices. International Journal of Environment and Climate Change, 13(5), 137-150.

Examples

```
# Create a sequence of numbers from 1 to 100
serial <- 1:100

# Create six vectors of random numbers, one for each column
col1 <- rnorm(100)
col2 <- rnorm(100)
col3 <- rnorm(100)
col4 <- rnorm(100)
```

```
col5 <- rnorm(100)
col6 <- rnorm(100)

# Combine the vectors into a data frame
df <- data.frame(serial, col1, col2, col3, col4, col5, col6)
df_nonlinearity(df)
```

---

df_stationarity *Descriptive Statistics of A Data Frame*

Description

Gives a list of three data frames: 'ADF', 'PP', 'KPSS'. This will also indicate whether the data is stationary or not according to the null hypothesis of the corresponding tests. The data frame must contain serial number or date or anything in the 1st column.This function will exclude the 1st column of the data frame and will perform tests on other columns. "p_value <= 0.01: ***; p_value <= 0.05: **; p_value <= 0.1: *".

Usage

```
df_stationarity(df)
```

Arguments

df	Data Frame with first column as serial number or date

Value

- test_results - List of three data frames: 'ADF', 'PP', 'KPSS'

References

- Garai, S., & Paul, R. K. (2023). Development of MCS based-ensemble models using CEEMDAN decomposition and machine intelligence. Intelligent Systems with Applications, 18, 200202.
- Garai, S., Paul, R. K., Rakshit, D., Yeasin, M., Paul, A. K., Roy, H. S., Barman, S. & Manjunatha, B. (2023). An MRA Based MLR Model for Forecasting Indian Annual Rainfall Using Large Scale Climate Indices. International Journal of Environment and Climate Change, 13(5), 137-150.

Examples

```
# create a vector of dates
dates <- seq(as.Date("2021-01-01"), as.Date("2021-01-05"), by = "day")

# create vectors of random numeric data for columns A through E
A <- runif(5, 0, 1)
B <- runif(5, 0, 1)
```

```
C <- runif(5, 0, 1)
D <- runif(5, 0, 1)
E <- runif(5, 0, 1)

# combine the vectors into a data frame
df <- data.frame(Date = dates, A = A, B = B, C = C, D = D, E = E)

# print the data frame
print(df)

# stationarity results
df_stationarity(df)
```

Index