R documentation

of 'ds.skewness.Rd'

July 24, 2020

`ds.skewness` *Calculates the skewness of a server-side numeric variable*

Description

This function calculates the skewness of a numeric variable that is stored on the server-side (Opal server).

Usage

```
ds.skewness(x = NULL, method = 1, type = "both", datasources = NULL)
```

Arguments

`x`	a character string specifying the name of a numeric variable.
`method`	an integer value between 1 and 3 selecting one of the algorithms for computing skewness. For more information see **Details**. The default value is set to 1.
`type`	a character string which represents the type of analysis to carry out. `type` can be set as: `'combine'`, `'split'` or `'both'`. For more information see **Details**. The default value is set to `'both'`.
`datasources`	a list of `DSConnection-class` objects obtained after login. If the `datasources` argument is not specified the default set of connections will be used: see `datashield.connections_d`

Details

This function is similar to the function `skewness` in R package `e1071`.

The function calculates the skewness of an input variable `x` with three different methods:

(1) If `method` is set to 1 the following formula is used $skewness = \frac{\sum_{i=1}^{N}(x_i-\bar{(x)})^3/N}{(\sum_{i=1}^{N}((x_i-\bar{(x)})^2)/N)^(3/2)}$, where $\bar{x}$ is the mean of x and N is the number of observations.

(2) If `method` is set to 2 the following formula is used $skewness = \frac{\sum_{i=1}^{N}(x_i-\bar{(x)})^3/N}{(\sum_{i=1}^{N}((x_i-\bar{(x)})^2)/N)^(3/2)} * \frac{\sqrt{(N(N-1)}}{n-2}$.

(3) If `method` is set to 3 the following formula is used $skewness = \frac{\sum_{i=1}^{N}(x_i-\bar{(x)})^3/N}{(\sum_{i=1}^{N}((x_i-\bar{(x)})^2)/N)^(3/2)} * (\frac{N-1}{N})^(3/2)$.

The type argument can be set as follows:
(1) If type is set to 'combine', 'combined', 'combines' or 'c', the global skewness is returned.
(2) If type is set to 'split', 'splits' or 's', the skewness is returned separately for each study.
(3) If type is set to 'both' or 'b', both sets of outputs are produced.

If x contains any missing value, the function removes those before the calculation of the skewness.

Server functions called: skewnessDS1 and skewnessDS2

Value

ds.skewness returns a matrix showing the skewness of the input numeric variable, the number of valid observations and the validity message.

Author(s)

Demetris Avraam, for DataSHIELD Development Team

Examples

```
## Not run:
  ## Version 6, for version 5 see the Wiki

  # connecting to the Opal servers

  require('DSI')
  require('DSOpal')
  require('dsBaseClient')

  builder <- DSI::newDSLoginBuilder()
  builder$append(server = "study1",
                 url = "http://192.168.56.100:8080/",
                 user = "administrator", password = "datashield_test&",
                 table = "CNSIM.CNSIM1", driver = "OpalDriver")
  builder$append(server = "study2",
                 url = "http://192.168.56.100:8080/",
                 user = "administrator", password = "datashield_test&",
                 table = "CNSIM.CNSIM2", driver = "OpalDriver")
  builder$append(server = "study3",
                 url = "http://192.168.56.100:8080/",
                 user = "administrator", password = "datashield_test&",
                 table = "CNSIM.CNSIM3", driver = "OpalDriver")
  logindata <- builder$build()

  connections <- DSI::datashield.login(logins = logindata, assign = TRUE, symbol = "D")

 #Calculate the skewness of LAB_TSC numeric variable for each study separately and combined

  ds.skewness(x = "D$LAB_TSC",
              method = 1,
              type = "both",
             datasources = connections)

  # Clear the Datashield R sessions and logout
  DSI::datashield.logout(connections)
```

```
## End(Not run)
```

Index