How can the “complete.cases” function be used in R? Provide examples.

Authored by

stats writer

May 1, 2024

The "complete.cases" function in R is a useful tool for identifying and handling missing values in a dataset. This function allows users to determine which rows or observations in a dataset contain complete or partial missing values, and then manipulate the data accordingly. It can be used in a variety of scenarios, such as data cleaning, imputation, or analysis.

For example, if a dataset contains missing values denoted as NA, the "complete.cases" function can be used to identify all rows with complete data by returning a logical vector of TRUE and FALSE values. This can then be used to subset the dataset and remove the incomplete observations.

Another use case for the "complete.cases" function is in imputation, where missing values are replaced with estimated values. The function can be used to identify the incomplete rows, and then the mean or median of the complete cases can be used to fill in the missing values.

In summary, the "complete.cases" function is a helpful tool for data management and analysis in R, allowing users to efficiently identify and handle missing values in a dataset.

Use complete.cases in R (With Examples)

You can use the complete.cases() function in R to remove missing values in a vector, matrix, or data frame.

This function uses the following basic syntax:

#remove missing values from vector
x <- x

#remove rows with missing values in any column of data frame
df <- df

#remove rows with NA in specific columns of data frame
df <- df),]

The following examples show how to use this function in practice.

Example 1: Remove Missing Values from Vector

The following code shows how to remove all NA values from a vector:

#define vector
x <- c(1, 24, NA, 6, NA, 9)

#remove NA values from vector
x <- x

x

1 24 6 9

Example 2: Remove Rows with NA in Any Column of Data Frame

The following code shows how to remove rows with NA values in any column of a data frame :

```
#define data frame
df <- data.frame(x=c(1, 24, NA, 6, NA, 9),
y=c(NA, 3, 4, 8, NA, 12),
z=c(NA, 7, 5, 15, 7, 14))

#view data frame
df

x y z
1 1 NA NA
2 24 3 7
3 NA 4 5
4 6 8 15
5 NA NA 7
6 9 12 14

#remove rows with NA value in any column data frame
df <- df

#view data frame
df

x y z
2 24 3 7
4 6 8 15
```

6 9 12 14

Example 3: Remove Rows with NA in Specific Columns of Data Frame

The following code shows how to remove rows with NA values in specific columns of a data frame :

```
#define data frame
df <- data.frame(x=c(1, 24, NA, 6, NA, 9),
y=c(NA, 3, 4, 8, NA, 12),
z=c(NA, 7, 5, 15, 7, 14))

#view data frame
df

x y z
1 1 NA NA
2 24 3 7
3 NA 4 5
4 6 8 15
5 NA NA 7
6 9 12 14#remove rows with NA value in y or z column
df <- df), ]

#view data frame
```

df

x y z

2 24 3 7

3 NA 4 5

4 6 8 15

6 9 12 14