How can the apply() function be used on specific columns in R?

Authored by

stats writer

June 27, 2024

The apply() function in R allows for the application of a specific function to a dataset or a specific subset of columns. This function can be used to perform a variety of tasks, such as calculating summary statistics, applying custom functions, or manipulating data within the specified columns. By specifying the columns as an argument, the apply() function can be used to efficiently perform operations on only those columns, rather than the entire dataset. This allows for more targeted and precise data manipulation and analysis in R programming.

R: Use apply() Function on Specific Columns

Often you may want to use the apply() function to apply a function to specific columns in a data frame in R.

However, the apply() function first forces all columns in a data frame to have the same object type before applying a function, which can sometimes have unintended consequences.

A better choice is the lapply() function, which uses the following basic syntax:

```
df <- lapply(df, my_function)
```

This particular example applies the function my_function to only col1 and col2 in the data frame.

The following example shows how to use this syntax in practice.

Example: Apply Function to Specific Columns of Data Frame

Suppose we have the following data frame in R:

```
#create data frame
df <- data.frame(team=c('A', 'A', 'A', 'A', 'B', 'B', 'B', 'B'),
points=c(19, 22, 15, NA, 14, 25, 25, 25),
rebounds=c(10, 6, 3, 7, 11, 13, 9, 12),
assists=c(4, 4, 3, 6, 7, 5, 10, 8))

#view data frame
df

team points rebounds assists
1 A 19 10 4
2 A 22 6 4
3 A 15 3 3
4 A NA 7 6
5 B 14 11 7
6 B 25 13 5
7 B 25 9 10
8 B 25 12 8
```

Now suppose we define the following function that multiplies values by 2 and then adds 1:

```
#define function
my_function <- function(x) x*2 + 1
```

We can use the following lapply() function to apply this function only to the points and rebounds columns in the data frame:

```
#apply function to specific columns
df <- lapply(df, my_function)

#view updated data frame
df

team points rebounds assists
1 A 39 21 4
2 A 45 13 4
3 A 31 7 3
4 A NA 15 6
5 B 29 23 7
6 B 51 27 5
7 B 51 19 10
8 B 51 25 8
```

From the output we can see that we multiplied each value in the points and rebounds columns by 2 and

then added 1.

Also notice that the team and assists columns remained unchanged.

Additional Resources

The following tutorials explain how to perform other common tasks in R: