What is the error in using colMeans(x, na.rm = TRUE) when ‘x’ is not numeric?

Authored by

stats writer

June 25, 2024

The error in using the function colMeans(x, na.rm = TRUE) when the input 'x' is not numeric is that the function is designed to calculate the mean of numerical data. When the input is not numeric, the function will not be able to perform the necessary calculations and will result in an error. It is important to ensure that the input for this function is numeric in order to obtain accurate results.

Fix: Error in colMeans(x, na.rm = TRUE) : 'x' must be numeric

One error message you may encounter when using R is:

Error in colMeans(x, na.rm = TRUE) : 'x' must be numeric

This error usually occurs when you attempt to use the prcomp() function to perform , yet one or more of the columns in the data frame you're using is not numeric.

There are two ways to get around this error:

Method 1: Convert Non-Numeric Columns to Numeric

Method 2: Remove Non-Numeric Columns from Data Frame

The following examples show how to use each method in practice.

How to Reproduce the Error

Suppose we attempt to perform principal components analysis on the following data frame that contains a character column:

```
#create data frame
df <- data.frame(team=c('A', 'A', 'C', 'B', 'C', 'B', 'B', 'C', 'A'),
points=c(12, 8, 26, 25, 38, 30, 24, 24, 15),
rebounds=c(10, 4, 5, 5, 4, 3, 8, 18, 22))

#view data frame
df
```

```
team points rebounds
1 A 12 10
2 A 8 4
3 C 26 5
4 B 25 5
5 C 38 4
6 B 30 3
7 B 24 8
8 C 24 18
9 A 15 22
```

#attempt to calculate principal components
prcomp(df)

Error in colMeans(x, na.rm = TRUE) : 'x' must be numeric

The team column is a character column, which causes an error when we attempt to use the prcomp() function.

Method 1: Convert Non-Numeric Columns to Numeric

One way to avoid the error is to convert the team column to a numeric column before using the prcomp() function:

#convert character column to numeric
df$team <- as.numeric(as.factor(df$team))

#view updated data frame
df

team points rebounds
1 1 12 10
2 1 8 4
3 3 26 5
4 2 25 5

5 3 38 4

6 2 30 3

7 2 24 8

8 3 24 18

9 1 15 22

#calculate principal components

prcomp(df)

Standard deviations (1, .., p=3):

9.8252704 6.0990235 0.4880538

Rotation (n x k) = (3 x 3):

PC1 PC2 PC3

team -0.06810285 0.04199272 0.99679417

points -0.91850806 0.38741460 -0.07907512

rebounds 0.38949319 0.92094872 -0.01218661

This time we don't receive any error because each column in the data frame is numeric.

Method 2: Remove Non-Numeric Columns from Data Frame

Another way to avoid the error is to simply remove any non-numeric columns from the data frame before using the prcomp() function:

#remove non-numeric columns from data frame

df_new <- df

#view new data frame

df_new

points rebounds

1 12 10

2 8 4

3 26 5

4 25 5

5 38 4

6 30 3

7 24 8

8 24 18

9 15 22

#calculate principal components

prcomp(df_new)

Standard deviations (1, .., p=2):

9.802541 6.093638

Rotation (n x k) = (2 x 2):

PC1 PC2

points 0.9199431 0.3920519
rebounds -0.3920519 0.9199431

Once again, we we don't receive any error because each column in the data frame is numeric.

Note: In most cases, the first method is the preferred solution because it allows you to use all of the data rather than removing some of the columns.

The following tutorials explain how to fix other common errors in R: