How can I calculate a moving average by group in R?

Authored by

stats writer

June 27, 2024

To calculate a moving average by group in R, you can use the "group_by" function to group the data by a specific variable. Then, use the "mutate" function to create a new column that calculates the moving average for each group. This can be done by using the "rollmean" or "rollapply" functions from the "zoo" package. Finally, you can use the "ungroup" function to remove the grouping and view the complete data set with the added moving average column.

Calculate a Moving Average by Group in R

In time series analysis, a moving average represents the average value of a certain number of previous periods.

You can use the following basic syntax to calculate a moving average by group in R:

```
library(dplyr)library(zoo)

#calculate moving average by group
df %>%
group_by(variable1)
mutate(moving_avg = rollmean(variable2, k=3, fill=NA, align='right'))
```

This particular example calculates a 3-period moving average of variable2, group by variable1.

This code utilizes the group_by() function from the

dplyr package and the rollmean() function from the zoo package.

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

Example: Calculate Moving Average by Group in R

Suppose we have the following data frame in R that shows the sales of some product during consecutive days at two different stores:

```
#create data frame
df <- data.frame(store=rep(c('A', 'B'), each=7),
sales=c(4, 4, 3, 5, 6, 5, 7, 4, 8, 7, 2, 5, 4, 6))

#view data frame
df

store sales
1 A 4
2 A 4
3 A 3
4 A 5
5 A 6
6 A 5
```

7 A 7

8 B 4

9 B 8

10 B 7

11 B 2

12 B 5

13 B 4

14 B 6

We can use the following syntax to create a new column called moving_avg3 that displays the 3-day moving average value of sales for each store:

```
library(dplyr)
library(zoo)

#calculate 3-day moving average of sales, grouped by store
df %>%
group_by(store) %>%
mutate(moving_avg3 = rollmean(sales, k=3, fill=NA, align='right'))

# A tibble: 14 x 3
# Groups: store
```

store sales moving_avg3

1 A 4 NA

2 A 4 NA

3 A 3 3.67

4 A 5 4

5 A 6 4.67

6 A 5 5.33

7 A 7 6

8 B 4 NA

9 B 8 NA

10 B 7 6.33

11 B 2 5.67

12 B 5 4.67

13 B 4 3.67

14 B 6 5

Note: The value for k in the rollmean() function controls the number of previous periods used to calculate the moving average.

The moving_avg3 column shows the moving average value of sales for the previous 3 periods.

For example, the first 3-day moving average of sales for

store A is calculated as:

3-Day Moving Average = (4 + 4 + 3) / 3 = 3.67

The next 3-day moving average of sales for store A is calculated as:

3-Day Moving Average = (4 + 3 + 5) / 3 = 4

Note that the first two values for the moving average for each store are NA because there weren't enough previous periods to use for the moving average.

How to Plot Multiple Columns in R
How to Average Across Columns in R
How to Calculate the Mean by Group in R