How can the median value be calculated for rows in R?

Authored by

stats writer

June 27, 2024

The median value for rows in R can be calculated by using the median() function, which is a statistical measure that represents the middle value of a dataset. This function takes the values of a row and arranges them in ascending order, then identifies the middle value. If there is an even number of values, the median is calculated by taking the average of the two middle values. This process can be applied to each row in a dataset to determine the median value for each individual row. By using the median function, R provides a simple and efficient way to calculate the median value for rows in a dataset.

Calculate the Median Value of Rows in R

You can use the following methods to calculate the median value of rows in R:

Method 1: Calculate Median of Rows Using Base R

```
df$row_median = apply(df, 1, median, na.rm=TRUE)
```

Method 2: Calculate Median of Rows Using dplyr

```
library(dplyr)

df %>%
rowwise() %>%
mutate(row_median =
median(c_across(where(is.numeric)), na.rm=TRUE))
```

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

Example 1: Calculate Median of Rows Using Base R

Suppose we have the following data frame in R that shows the points scored by various basketball players during three different games:

```
#create data frame
df <- data.frame(game1=c(10, 12, 14, 15, 16, 18, 19),
game2=c(14, 19, 13, 8, 15, 15, 17),
game3=c(9, NA, 15, 25, 26, 30, 19))

#view data frame
df
```

```
game1 game2 game3
1 10 14 9
2 12 19 NA
3 14 13 15
4 15 8 25
5 16 15 26
6 18 15 30
7 19 17 19
```

We can use the apply() function from base R to create a new column that shows the median value of each row:

```
#calculate median of each row
df$row_median = apply(df, 1, median, na.rm=TRUE)

#view updated data frame
df
```

```
game1 game2 game3 row_median
1 10 14 9 10.0
2 12 19 NA 15.5
3 14 13 15 14.0
4 15 8 25 15.0
5 16 15 26 16.0
6 18 15 30 18.0
7 19 17 19 19.0
```

The new column called row_median contains the median value of each row in the data frame.

Example 2: Calculate Median of Rows Using dplyr

Suppose we have the following data frame in R that shows the points scored by various basketball players during three different games:

```
#create data frame
df <- data.frame(player=c('A', 'B', 'C', 'D', 'E', 'F', 'G'),
```

```
game1=c(10, 12, 14, 15, 16, 18, 19),
game2=c(14, 19, 13, 8, 15, 15, 17),
game3=c(9, NA, 15, 25, 26, 30, 19))

#view data frame
df

player game1 game2 game3
1 A 10 14 9
2 B 12 19 NA
3 C 14 13 15
4 D 15 8 25
5 E 16 15 26
6 F 18 15 30
7 G 19 17 19
```

We can use the mutate() function from the dplyr package to create a new column that shows the median value of each row for the numeric columns only:

```
library(dplyr)

#calculate median of rows for numeric columns only
df %>%
rowwise() %>%
```

```
mutate(row_median =
median(c_across(where(is.numeric)), na.rm=TRUE))

# A tibble: 7 x 5
# Rowwise:
player game1 game2 game3 row_median

1 A 10 14 9 10
2 B 12 19 NA 15.5
3 C 14 13 15 14
4 D 15 8 25 15
5 E 16 15 26 16
6 F 18 15 30 18
7 G 19 17 19 19
```