How can I replace NA values with the median in R?

Authored by

stats writer

June 27, 2024

"Replacing missing or NA values with the median in R can be accomplished using the 'na.aggregate' function from the 'zoo' package. This function will automatically calculate the median and replace all NA values with that value. Alternatively, the 'na.rm' argument can be used with the 'median' function to calculate the median without including NA values. Both methods provide a simple and efficient way to handle missing data in R."

Replace NA with Median in R

You can use the following methods to replace NA values with the median using functions from the dplyr and tidyr packages in R:

Method 1: Replace NA values with Median in One Column

df %>% mutate(across(col1, ~replace_na(., median(., na.rm=TRUE))))

Method 2: Replace NA values with Median in Several Columns

df %>% mutate(across(c(col1, col2), ~replace_na(., median(., na.rm=TRUE))))

Method 3: Replace NA values with Median in All Numeric Columns

```
df %>% mutate(across(where(is.numeric), ~replace_na(., median(., na.rm=TRUE))))
```

The following examples show how to use each method in practice with the following data frame:

```
#create data frame
df <- data.frame(player=c('A', 'B', 'C', 'D', 'E'),
points=c(17, 13, NA, 9, 25),
rebounds=c(3, 4, NA, NA, 8),
blocks=c(1, 1, 2, 4, NA))

#view data frame
df
```

```
player points rebounds blocks
1 A 17 3 1
2 B 13 4 1
3 C NA NA 2
4 D 9 NA 4
5 E 25 8 NA
```

Example 1: Replace NA Values with Median in One Column

The following code shows how to replace the NA values

in the points column with the median value of the points column:

```
library(dplyr)
library(tidyr)

#replace NA values in points column with median of points column
df <- df %>% mutate(across(points, ~replace_na(., median(., na.rm=TRUE))))
#view updated data frame
df

player points rebounds blocks
1 A 17 3 1
2 B 13 4 1
3 C 15 NA 2
4 D 9 NA 4
5 E 25 8 NA
```

The median value in the points column was 15, so the one NA value in the points column was replaced with 15.

All other columns remained unchanged.

Example 2: Replace NA Values with Median in Several Columns

The following code shows how to replace the NA values in the points and blocks columns with their respective column medians:

```
library(dplyr)
library(tidyr)

#replace NA values in points and blocks columns with their respective medians
df <- df %>% mutate(across(c(points, blocks), ~replace_na(., median(., na.rm=TRUE))))

#view updated data frame
df

player points rebounds blocks
1 A 17 3 1.0
2 B 13 4 1.0
3 C 15 NA 2.0
4 D 9 NA 4.0
5 E 25 8 1.5
```

Example 3: Replace NA Values with Median in All Numeric

Columns

The following code shows how to replace the NA values in every numeric columns with their respective median value:

```
library(dplyr)
library(tidyr)

#replace NA values in all numeric columns with their respective medians
df <- df %>% mutate(across(where(is.numeric), ~replace_na(., median(., na.rm=TRUE))))

#view updated data frame
df

player points rebounds blocks
1 A 17 3 1.0
2 B 13 4 1.0
3 C 15 4 2.0
4 D 9 4 4.0
5 E 25 8 1.5
```

Notice that the NA values in all numeric columns have

been replaced with their respective column medians.

The one column that was not numeric (player) has remained unchanged.

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