How can I calculate the standard deviation of columns in R?

Authored by

stats writer

May 5, 2024

In order to calculate the standard deviation of columns in R, you can use the "sd" function. This function takes in a vector or a data frame column as its argument and returns the standard deviation of the values in that column. It is important to note that the data in the column must be numeric for this function to work properly. Additionally, the "na.rm" argument can be used to exclude any missing values from the calculation. Overall, using the "sd" function in R allows for a quick and efficient way to calculate the standard deviation of columns in a dataset.

Calculate Standard Deviation of Columns in R

You can use the following basic syntax to calculate the standard deviation of columns in R:

```
#calculate standard deviation of one column
sd(df$col1)

#calculate standard deviation of all columns
sapply(df, sd)

#calculate standard deviation of specific columns
sapply(df, sd)
```

The following examples show how to use this syntax in practice with the following data frame:

```
#create data frame
df <- data.frame(team=c('A', 'B', 'C', 'D', 'E'),
points=c(99, 91, 86, 88, 95),
assists=c(33, 28, 31, 39, 34),
```

rebounds=c(30, 28, 24, 24, 28))

#view data frame
df

team points assists rebounds
1 A 99 33 30
2 B 91 28 28
3 C 86 31 24
4 D 88 39 24
5 E 95 34 28

Example 1: Standard Deviation of One Column

The following code shows how to calculate the standard deviation of one column in the data frame:

#calculate standard deviation of 'points' column
sd(df$points)

5.263079

The standard deviation of values in the 'points' column is 5.263079.

Example 2: Standard Deviation of All Columns

The following code shows how to calculate the standard deviation of every column in the data frame:

```
#calculate standard deviation of all columns in data frame
sapply(df, sd)

team points assists rebounds
NA 5.263079 4.062019 2.683282
Warning message:
In var(if (is.vector(x) || is.factor(x)) x else as.double(x), na.rm = na.rm) :
NAs introduced by coercion
```

Since the 'team' column is a character variable, R returns NA and gives us a warning.

However, it successfully computes the standard deviation of the other three numeric columns.

Example 3: Standard Deviation of Specific Columns

The following code shows how to calculate the standard deviation of specific columns in the data

frame:

#calculate standard deviation of 'points' and 'rebounds' columns
sapply(df, sd)

points rebounds
5.263079 2.683282

Note that we could use column index values to select columns as well:

#calculate standard deviation of 'points' and 'rebounds' columns
sapply(df, sd)

points rebounds
5.263079 2.683282

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