

How can standard deviation be calculated using dplyr in R, and what are some examples of its usage?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can standard deviation be calculated using dplyr in R, and what are some examples of its usage?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=154486>

The standard deviation is a statistical measure that indicates the spread or variability of a set of data values from the mean. In R, the dplyr package can be used to calculate the standard deviation of a dataset. This can be done by using the summarize function, which allows for the calculation of various summary statistics, including the standard deviation.

For example, if we have a dataset containing the heights of individuals, we can use dplyr to calculate the standard deviation of the height variable. This can be achieved by using the following code:

```
dataset %>% summarize(sd = sd(height))
```

This will return the standard deviation of the height variable as a new column named "sd". Furthermore, dplyr allows for the calculation of standard deviation for specific groups within a dataset. For instance, if we want to calculate the standard deviation of height for males and females separately, we can use the group_by function in combination with summarize:

```
dataset %>% group_by(gender) %>% summarize(sd = sd(height))
```

This will return the standard deviation of height for each gender category. Additionally, dplyr also allows for the calculation of standard deviation for multiple variables at once by using the select function.

In summary, the dplyr package in R provides a convenient and efficient way to calculate standard deviation for various scenarios, such as calculating it for the entire dataset or for specific groups within the dataset. It is a useful tool for analyzing and understanding the variability of data values in a given dataset.

Calculate Standard Deviation Using dplyr (With Examples)

You can use the following methods to calculate the standard deviation of values in a data frame in :

Method 1: Calculate Standard Deviation of One Variable

```
library(dplyr)
```

```
df %>%
```

```
summarise(sd_var1 = sd(var1, na.rm=TRUE))
```

Method 2: Calculate Standard Deviation of Multiple Variables

```
library(dplyr)
```

```
df %>%
```

```
summarise(sd_var1 = sd(var1, na.rm=TRUE),  
sd_var2 = sd(var2, na.rm=TRUE))
```

Method 3: Calculate Standard Deviation of Multiple Variables, Grouped by Another Variable

```
library(dplyr)
```

```
df %>%
```

```
group_by(var3) %>%
```

```
summarise(sd_var1 = sd(var1, na.rm=TRUE),  
sd_var2 = sd(var2, na.rm=TRUE))
```

This tutorial explains how to use each method in practice with the following data frame in R:

```
#create data frame
```

```
df <- data.frame(team=c('A', 'A', 'A', 'A', 'B', 'B', 'B', 'B'),  
points=c(12, 15, 18, 22, 14, 17, 29, 35),  
assists=c(4, 4, 3, 6, 7, 8, 3, 10))
```

```
#view data frame
```

```
df
```

```
team points assists
```

```
1 A 12 4
```

```
2 A 15 4
```

```
3 A 18 3
```

```
4 A 22 6
```

```
5 B 14 7
```

```
6 B 17 8
```

```
7 B 29 3
```

```
8 B 35 10
```

Example 1: Calculate Standard Deviation of One Variable

The following code shows how to calculate the standard deviation of the points variable:

```
library(dplyr)
```

```
#calculate standard deviation of points variable
```

```
df %>%
```

```
summarise(sd_points = sd(points, na.rm=TRUE))
```

```
sd_points
```

```
1 7.995534
```

From the output we can see that the standard deviation of values for the points variable is 7.995534.

Example 2: Calculate Standard Deviation of Multiple Variables

The following code shows how to calculate the standard deviation of the points and the assists variables:

```
library(dplyr)
```

```
#calculate standard deviation of points and assists variables
```

```
df %>%
```

```
summarise(sd_points = sd(points, na.rm=TRUE),
```

```
sd_assists = sd(assists, na.rm=TRUE))
```

```
sd_points sd_assists
```

```
1 7.995534 2.559994
```

The output displays the standard deviation for both the points and assists variables.

Example 3: Calculate Standard Deviation of Multiple Variables, Grouped by Another Variable

The following code shows how to calculate the standard deviation of the points and the assists variables:

```
library(dplyr)

#calculate standard deviation of points and assists
variables
df %>%
group_by(team) %>%
summarise(sd_points = sd(points, na.rm=TRUE),
sd_assists = sd(assists, na.rm=TRUE))

# A tibble: 2 x 3
team sd_points sd_assists

1 A 4.27 1.26
2 B 9.91 2.94
```

The output displays the standard deviation for both the

points and assists variables for team A and team B.

Note: You can include a list of several variables in the `group_by()` function if you would like to group by multiple variables.

ARABPSYCHOLOGY.COM