

How can I use dplyr to calculate the sum across multiple columns in a data frame?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I use dplyr to calculate the sum across multiple columns in a data frame?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=155698>

Dplyr is a popular R package that provides a set of tools for data manipulation and transformation. One of its useful functions is the ability to calculate the sum across multiple columns in a data frame. This can be achieved by using the "summarise" function, which allows you to specify the columns to be summed and the resulting column name. Additionally, dplyr also has the "mutate" function which can be used to add a new column with the sum of the specified columns. By utilizing these functions, you can efficiently and accurately calculate the sum across multiple columns in a data frame, making data analysis and manipulation more convenient.

Sum Across Multiple Columns Using dplyr

You can use the following methods to sum values across multiple columns of a data frame using dplyr:

Method 1: Sum Across All Columns

```
df %>%  
mutate(sum = rowSums(., na.rm=TRUE))
```

Method 2: Sum Across All Numeric Columns

```
df %>%  
mutate(sum = rowSums(across(where(is.numeric)),  
na.rm=TRUE))
```

Method 3: Sum Across Specific Columns

```
df %>%  
mutate(sum = rowSums(across(c(col1, col2))))
```

The following examples show how to each method with the following data frame that contains information about points scored by various basketball players during different games:

```
#create data frame
```

```
df <- data.frame(game1=c(22, 25, 29, 13, 22, 30),  
game2=c(12, 10, 6, 6, 8, 11),  
game3=c(NA, 15, 15, 18, 22, 13))
```

```
#view data frame
```

```
df
```

```
game1 game2 game3
```

```
1 22 12 NA
```

```
2 25 10 15
```

```
3 29 6 15
```

```
4 13 6 18
```

```
5 22 8 22
```

```
6 30 11 13
```

Example 1: Sum Across All Columns

The following code shows how to calculate the sum of values across all columns in the data frame:

```
library(dplyr)
```

```
#sum values across all columns
```

```
df %>%
```

```
mutate(total_points = rowSums(., na.rm=TRUE))
```

	game1	game2	game3	total_points
1	22	12	NA	34
2	25	10	15	50
3	29	6	15	50
4	13	6	18	37
5	22	8	22	52
6	30	11	13	54

Example 2: Sum Across All Numeric Columns

The following code shows how to calculate the sum of values across all numeric columns in the data frame:

```
library(dplyr) #sum values across all numeric columns
```

```
df %>%
```

```
mutate(total_points =
```

```
rowSums(across(where(is.numeric)), na.rm=TRUE))
```

	game1	game2	game3	total_points
1	22	12	NA	34
2	25	10	15	50

```
3 29 6 15 50
4 13 6 18 37
5 22 8 22 52
6 30 11 13 54
```

Example 3: Sum Across Specific Columns

The following code shows how to calculate the sum of values across the game1 and game2 columns only:

```
library(dplyr)

#sum values across game1 and game2 only
df %>%
  mutate(first2_sum = rowSums(across(c(game1,
game2))))
```

```
game1 game2 game3 first2_sum
1 22 12 NA 34
2 25 10 15 35
3 29 6 15 35
4 13 6 18 19
5 22 8 22 30
6 30 11 13 41
```

Additional Resources

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

ARABPSYCHOLOGY.COM