

How can the `tapply()` function be used in R?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can the `tapply()` function be used in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=154787>

The tapply() function in R is used to split and apply a function to subsets of a data frame or vector based on a specific variable or factor. This allows for easy calculation of summary statistics or application of custom functions to different subsets of data. The function takes in three arguments: the vector or data frame, the variable to use for splitting, and the function to apply. The result is a vector or data frame with the applied function's output for each subset. This function is especially useful for data analysis and manipulating large datasets, making it a valuable tool for data scientists and statisticians.

Use the tapply() Function in R (With Examples)

The tapply() function in R can be used to apply some function to a vector, grouped by another vector.

This function uses the following basic syntax:

```
tapply(X, INDEX, FUN, ..)
```

where:

X: A vector to apply a function to
INDEX: A vector to group by
FUN: The function to apply

The following examples show how to use this function 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'),  
position=c('G', 'G', 'F', 'F', 'G', 'G', 'F', 'F'),  
points=c(14, 19, 13, 8, 15, 15, 17, 19),
```

```
assists=c(4, 3, 3, 5, 9, 14, 15, 12))
```

```
#view data frame
```

```
df
```

```
team position points assists
```

```
1 A G 14 4
```

```
2 A G 19 3
```

```
3 A F 13 3
```

```
4 A F 8 5
```

```
5 B G 15 9
```

```
6 B G 15 14
```

```
7 B F 17 15
```

```
8 B F 19 12
```

Example 1: Apply Function to One Variable, Grouped by One Variable

The following code shows how to use the tapply() function to calculate the mean value of points, grouped by team:

```
#calculate mean of points, grouped by team
```

```
tapply(df$points, df$team, mean)
```

```
A B
```

13.5 16.5

From the output we can see:

The mean value of points for team A is 13.5. The mean value of points for team B is 16.5.

Note that you can also include additional arguments after the function, such as `na.rm`, to indicate that you wish to calculate the mean while ignoring NA values in the data frame:

```
#calculate mean of points, grouped by team  
tapply(df$points, df$team, mean, na.rm=TRUE)
```

**A B
13.5 16.5**

Example 2: Apply Function to One Variable, Grouped by Multiple Variables

The following code shows how to use the `tapply()` function to calculate the mean value of points, grouped by team and position:

```
#calculate mean of points, grouped by team and
```

position

**tapply(df\$points, list(df\$team, df\$position), mean,
na.rm=TRUE) F G**

A 10.5 16.5

B 18.0 15.0

The mean value of points for team A and position F is 10.5. The mean value of points for team A and position G is 16.5. The mean value of points for team B and position F is 18.0. The mean value of points for team B and position G is 15.0.

Note: In this example we grouped by two variables, but we can include as many variables as we'd like in the list() function to group by even more variables.