“What is the easiest way to create summary tables in R?

Authored by

stats writer

May 2, 2024

The quickest and simplest method of generating summary tables in R is by using the built-in functions such as "table()" or "summary()" on a dataset. These functions allow for easy creation of summaries, including counts, means, and other statistical measures, for variables within a dataset. Additionally, using the "dplyr" package can also provide a straightforward approach to creating summary tables by utilizing functions such as "group_by()" and "summarize()" to group and summarize data. Overall, utilizing these functions and packages can greatly simplify the process of creating summary tables in R.

The Easiest Way to Create Summary Tables in R

The easiest way to create summary tables in R is to use the describe() and describeBy() functions from the psych library.

library(psych)#create summary table
describe(df)

#create summary table, grouped by a specific variable
describeBy(df, group=df$var_name)

The following examples show how to use these functions in practice.

Example 1: Create Basic Summary Table

Suppose we have the following data frame in R:

#create data frame
df <- data.frame(team=c('A', 'A', 'B', 'B', 'C', 'C', 'C'),

points=c(15, 22, 29, 41, 30, 11, 19),

rebounds=c(7, 8, 6, 6, 7, 9, 13),

steals=c(1, 1, 2, 3, 5, 7, 5))

#view data frame

df

team points rebounds steals

1 A 15 7 1

2 A 22 8 1

3 B 29 6 2

4 B 41 6 3

5 C 30 7 5

6 C 11 9 7

7 C 19 13 5

We can use the describe() function to create a summary table for each variable in the data frame:

library(psych) #create summary table

describe(df)

vars n mean sd median trimmed mad min max range skew kurtosis

team* 1 7 2.14 0.90 2 2.14 1.48 1 3 2 -0.22 -1.90

points 2 7 23.86 10.24 22 23.86 10.38 11 41 30 0.33 -1.41
rebounds 3 7 8.00 2.45 7 8.00 1.48 6 13 7 1.05 -0.38
steals 4 7 3.43 2.30 3 3.43 2.97 1 7 6 0.25 -1.73
se
team* 0.34
points 3.87
rebounds 0.93
steals 0.87

Here's how to interpret each value in the output:

vars: column numbern: Number of valid casesmean: The mean valuemedian: The median valuetrimmed: The trimmed mean (default trims 10% of observations from each end)mad: The median absolute deviation (from the median)min: The minimum valuemax: The maximum valuerange: The range of values (max - min)skew: The skewnesskurtosis: The kurtosisse: The standard error

It's important to note that any variable with an asterisk (*) symbol next to it is a categorical or logical variable that has been converted to a numerical variable with values that represent the numerical ordering of the values.

In our example, the variable 'team' has been converted to a numerical variable so we shouldn't interpret the summary statistics for it literally.

Also note that you can use the argument fast=TRUE to only calculate the most common summary statistics:

```
#create smaller summary table
describe(df, fast=TRUE)
```

```
vars n mean sd min max range se
team 1 7 NaN NA Inf -Inf -Inf NA
points 2 7 23.86 10.24 11 41 30 3.87
rebounds 3 7 8.00 2.45 6 13 7 0.93
steals 4 7 3.43 2.30 1 7 6 0.87
```

We can also choose to only compute the summary statistics for certain variables in the data frame:

```
#create summary table for just 'points' and 'rebounds' columns
describe(df, fast=TRUE)
```

```
vars n mean sd min max range se
points 1 7 23.86 10.24 11 41 30 3.87
```

rebounds 2 7 8.00 2.45 6 13 7 0.93

Example 2: Create Summary Table, Grouped by Specific Variable

```
#create summary table, grouped by 'team' variable
describeBy(df, group=df$team, fast=TRUE)

Descriptive statistics by group
group: A
vars n mean sd min max range se
team 1 2 NaN NA Inf -Inf -Inf NA
points 2 2 18.5 4.95 15 22 7 3.5
rebounds 3 2 7.5 0.71 7 8 1 0.5
steals 4 2 1.0 0.00 1 1 0 0.0
------------------------------------------------------------
group: B
vars n mean sd min max range se
team 1 2 NaN NA Inf -Inf -Inf NA
points 2 2 35.0 8.49 29 41 12 6.0
rebounds 3 2 6.0 0.00 6 6 0 0.0
steals 4 2 2.5 0.71 2 3 1 0.5
------------------------------------------------------------
group: C
vars n mean sd min max range se
team 1 3 NaN NA Inf -Inf -Inf NA
```

points 2 3 20.00 9.54 11 30 19 5.51

rebounds 3 3 9.67 3.06 7 13 6 1.76

steals 4 3 5.67 1.15 5 7 2 0.67

The output shows the summary statistics for each of the three teams in the data frame.

How to Calculate Five Number Summary in R