

# How can I calculate descriptive statistics in R, with an example?

Authored by  
**stats writer**

June 28, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I calculate descriptive statistics in R, with an example?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=155979>

R is a statistical programming language that allows users to easily calculate descriptive statistics for a given dataset. Descriptive statistics are numerical measures that summarize and describe the main characteristics of a dataset. These statistics provide valuable insights into the distribution, central tendency, and variability of the data. To calculate descriptive statistics in R, users can utilize built-in functions such as mean, median, mode, standard deviation, and range. These functions can be applied to a specific dataset or a subset of the data, providing a comprehensive understanding of the data's characteristics. For example, to calculate the mean of a dataset in R, users can use the function "mean(dataset)," where "dataset" represents the name of the dataset. This will output the average value of the dataset. By using these functions, users can easily and efficiently analyze their data and make informed decisions based on the results.

## Calculate Descriptive Statistics in R (With Example)

**Descriptive statistics are values that describe a dataset.**

**They help us gain an understanding of where of the dataset is located along with how the values are in the dataset.**

**There are two functions we can use to calculate descriptive statistics in R:**

**Method 1: Use summary() Function**

```
summary(my_data)
```

**The summary() function calculates the following values for each variable in a data frame in R:**

**Minimum 1st Quartile Median Mean 3rd Quartile Maximum**

## **Method 2: Use sapply() Function**

```
sapply(my_data, sd, na.rm=TRUE)
```

The `sapply()` function can be used to calculate descriptive statistics other than the ones calculated by the `summary()` function for each variable in a data frame.

For example, the `sapply()` function above calculates the standard deviation of each variable in a data frame.

The following example shows how to use both of these functions to calculate descriptive statistics for variables in a data frame in R.

### **Example: Calculating Descriptive Statistics in R**

Suppose we have the following data frame in R that contains three variables:

```
#create data frame
```

```
df <- data.frame(x=c(1, 4, 4, 5, 6, 7, 10, 12),  
y=c(2, 2, 3, 3, 4, 5, 11, 11),
```

```
z=c(8, 9, 9, 9, 10, 13, 15, 17))
```

```
#view data frame
```

```
df
```

```
x y z
```

```
1 1 2 8
```

```
2 4 2 9
```

```
3 4 3 9
```

```
4 5 3 9
```

```
5 6 4 10
```

```
6 7 5 13
```

```
7 10 11 15
```

```
8 12 11 17
```

We can use the `summary()` function to calculate a variety of descriptive statistics for each variable:

```
#calculate descriptive statistics for each variable
```

```
summary(df)
```

```
x y z
```

```
Min. : 1.000 Min. : 2.000 Min. : 8.00
```

```
1st Qu.: 4.000 1st Qu.: 2.750 1st Qu.: 9.00
```

```
Median : 5.500 Median : 3.500 Median : 9.50
```

**Mean : 6.125 Mean : 5.125 Mean :11.25**  
**3rd Qu.: 7.750 3rd Qu.: 6.500 3rd Qu.:13.50**  
**Max. :12.000 Max. :11.000 Max. :17.00**

**#calculate descriptive statistics for 'x' and 'z' only**  
**summary(df)**

**x z**

**Min. : 1.000 Min. : 8.00**  
**1st Qu.: 4.000 1st Qu.: 9.00**  
**Median : 5.500 Median : 9.50**  
**Mean : 6.125 Mean :11.25**  
**3rd Qu.: 7.750 3rd Qu.:13.50**  
**Max. :12.000 Max. :17.00**

**We can also use the `sapply()` function to calculate specific descriptive statistics for each variable.**

**For example, the following code shows how to calculate the standard deviation of each variable:**

**#calculate standard deviation for each variable**  
**sapply(df, sd, na.rm=TRUE)**

**x y z**

**3.522884 3.758324 3.327376**

**We can also use a function() within sapply() to calculate descriptive statistics.**

**For example, the following code shows how to calculate for each variable:**

```
#calculate range for each variable  
sapply(df, function(df) max(df, na.rm=TRUE)-min(df,  
na.rm=TRUE))
```

```
x y z  
11 9 9
```

**Lastly, we can create a complex function that calculates some descriptive statistic and then use this function with the sapply() function.**

**For example, the following code shows how to calculate of each variable in the data frame:**

```
#define function that calculates mode  
find_mode <- function(x) {  
u <- unique(x)
```

```
tab <- tabulate(match(x, u))
u
}

#calculate mode for each variable
apply(df, find_mode)
```

```
$x
```

```
4
```

```
$y
```

```
2 3 11
```

```
$z
```

```
9
```

From the output we can see:

The mode of variable x is 4. The mode of variable y is 2, 3, and 11 (since each of these values occurred most frequently) The mode of variable z is 9.

By using the `summary()` and `apply()` functions, we can calculate any descriptive statistics that we'd like for each variable in a data frame.

## Additional Resources

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

ARABPSYCHOLOGY.COM