

How to Center Data in R (With Examples)

Authored by
stats writer

December 11, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Center Data in R (With Examples)*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=107156>

Centering data in R is a data pre-processing technique that shifts the mean of the data to zero. This process can be done using the `scale()` function, which subtracts the mean from each value in the dataset. This is useful when comparing data between different datasets, as it allows you to compare the relative values of the data, rather than the absolute value. Centering data can also be used when performing regression analysis, to minimize the influence of outliers. In summary, centering data in R is a simple but useful data pre-processing technique.

To **center** a dataset means to subtract the mean value from each individual observation in the dataset.

For example, suppose we have the following dataset:

Raw Data
4
6
9
13
14
17
18
19
19
21

It turns out that the mean value is 14. Thus, to center this dataset we would subtract 14 from each individual observation:

Raw Data		Centered Data
4	$4 - 14 =$	-10
6	$6 - 14 =$	-8
9	$9 - 14 =$	-5
13	$13 - 14 =$	-1
14	$14 - 14 =$	0
17	$17 - 14 =$	3
18	$18 - 14 =$	4
19	$19 - 14 =$	5
19	$19 - 14 =$	5
21	$21 - 14 =$	7

Note that the mean value of the centered dataset is zero.

This tutorial provides several examples of how to center data in R.

Example 1: Center the Values of a Vector

The following code shows how to use the function from base R to center the values in a vector:

```
#create vector
```

```
data <- c(4, 6, 9, 13, 14, 17, 18, 19, 19, 21)
```

```
#subtract the mean value from each observation in the vector
```

```
scale(data, scale=FALSE)
```

```
-10
-8
-5
-1
0
3
4
5
5
7
```

```
attr("scaled:center")
14
```

The resulting values are the centered values of the dataset. The `scale()` function also tells us that the mean value of the dataset is 14.

Note that the **`scale()`** function, by default, subtracts the mean from each individual observation and then divides by the standard deviation.

By specifying **`scale=FALSE`**, we tell R not to divide by the standard deviation.

Example 2: Center the Columns in a Data Frame

The following code shows how to use the function and the function from base R to center the values of each column of a data frame:

```
#create data frame
df <- data.frame(x = c(1, 4, 5, 6, 6, 8, 9),
y = c(7, 7, 8, 8, 8, 9, 12),
z = c(3, 3, 4, 4, 6, 7, 7))

#center each column in the data frame
df_new <- sapply(df, function(x) scale(x, scale=FALSE))

#display data frame
df_new

x y z
-4.5714286 -1.4285714 -1.8571429
-1.5714286 -1.4285714 -1.8571429
```

```
-0.5714286 -0.4285714 -0.8571429  
0.4285714 -0.4285714 -0.8571429  
0.4285714 -0.4285714 1.1428571  
2.4285714 0.5714286 2.1428571  
3.4285714 3.5714286 2.1428571
```

We can verify that the mean of each column in the new data frame is equal to zero by using the **colMeans()** function:

```
colMeans(df_new)
```

```
x y z  
2.537653e-16 -2.537653e-16 3.806479e-16
```

The values are shown in scientific notation, but each value is essentially equal to zero.