

How do you calculate the five number summary in R, and can you provide examples?

Authored by
stats writer

May 4, 2024

RECOMMENDED CITATION

stats writer (2024). *How do you calculate the five number summary in R, and can you provide examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=142653>

The five number summary is a statistical technique used to summarize a set of data. It consists of five key values: the minimum value, the first quartile, the median, the third quartile, and the maximum value. These values provide a quick and concise overview of the distribution of data. In R, the five number summary can be calculated using the "summary()" function. This function takes in a dataset as its argument and returns the five number summary for each numerical column in the dataset. For example, if we have a dataset called "grades" with a column named "test_scores", we can calculate the five number summary by using the code "summary(grades\$test_scores)". This will give us the minimum value, first quartile, median, third quartile, and maximum value for the test scores in the "grades" dataset.

Calculate Five Number Summary in R (With Examples)

A five number summary is a way to summarize a dataset using the following five values:

The minimum The first quartile The median The third quartile The maximum

The five number summary is useful because it provides a concise summary of the distribution of the data in the following ways:

It tells us where the middle value is located, using the median. It tells us how spread out the data is, using the first and third quartiles. It tells us the range of the data, using the minimum and the maximum.

The easiest way to calculate a five number summary of a dataset in R is to use the `fivenum()` function from base

R:

fivenum(data)

The following example shows how to use this syntax in practice.

Example 1: Five Number Summary of Vector

The following code shows how to calculate the five number summary of a numeric vector in R:

```
#define numeric vector
```

```
data <- c(4, 6, 6, 7, 8, 9, 12, 13, 14, 15, 15, 18, 22)
```

```
#calculate five number summary of data
```

```
fivenum(data)
```

```
4 7 12 15 22
```

From the output we can see:

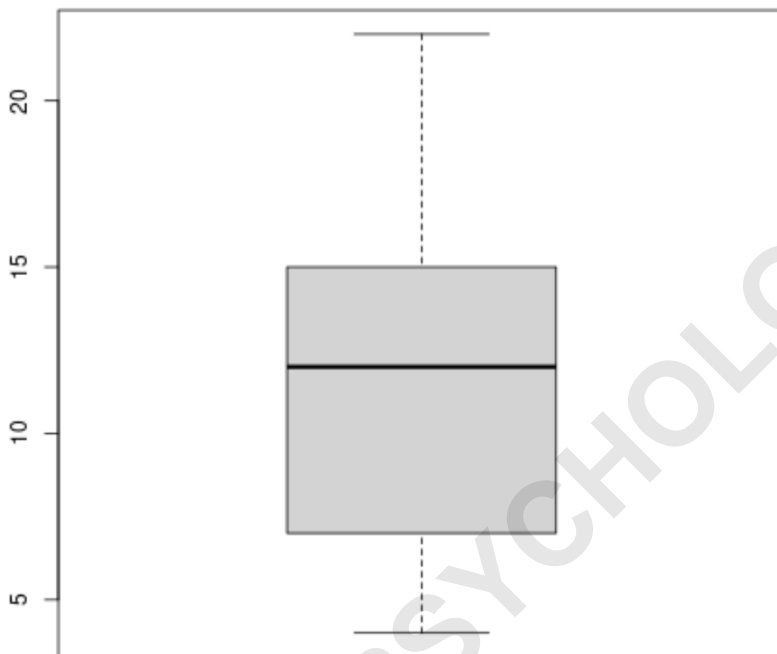
The minimum: 4The first quartile: 7The median: 12The third quartile: 15The maximum: 22

We can quickly visualize the five number summary by

creating a boxplot:

`boxplot(data)`

`4 7 12 15 22`



Here's how to interpret the boxplot:

The line at the bottom of the plot represents the minimum value (4). The line at the bottom of the box represents the first quartile (7). The line in the middle of the box represents the median (12). The line at the top of the box represents the third quartile (15). The line at the top of the plot represents the maximum value (22).

Example 2: Five Number Summary of Column in Data Frame

The following code shows how to calculate the five number summary of a specific column in a data frame:

```
#create data frame
df <- data.frame(team=c('A', 'B', 'C', 'D', 'E', 'F', 'G', 'H'),
points=c(99, 90, 86, 88, 95, 87, 85, 89),
assists=c(33, 28, 31, 39, 34, 30, 29, 25),
rebounds=c(30, 28, 24, 24, 28, 30, 31, 35))

#calculate five number summary of points column
fivenum(df$points)

85.0 86.5 88.5 92.5 99.0
```

Example 3: Five Number Summary of Multiple Columns

The following code shows how to use the `sapply()` function to calculate the five number summary of several columns in a data frame at once:

```
#create data frame
df <- data.frame(team=c('A', 'B', 'C', 'D', 'E', 'F', 'G', 'H'),
points=c(99, 90, 86, 88, 95, 87, 85, 89),
assists=c(33, 28, 31, 39, 34, 30, 29, 25),
```

```
rebounds=c(30, 28, 24, 24, 28, 30, 31, 35))
```

```
#calculate five number summary of points, assists, and  
rebounds column  
sapply(df, fivenum)
```

```
points assists rebounds
```

```
85.0 25.0 24.0
```

```
86.5 28.5 26.0
```

```
88.5 30.5 29.0
```

```
92.5 33.5 30.5
```

```
99.0 39.0 35.0
```

ARABPSYCHOLOGY.COM