

How can I create a categorical variable from a continuous variable in R?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I create a categorical variable from a continuous variable in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=164851>

Creating a categorical variable from a continuous variable in R involves converting a continuous variable, which contains numerical data, into a categorical variable, which contains distinct categories or groups. This can be achieved by assigning numerical ranges or labels to the data points in the continuous variable. This process allows for easier analysis and interpretation of the data in terms of categories rather than numerical values. In R, there are various methods and functions that can be used to create a categorical variable from a continuous variable, such as the `cut()` and `factor()` functions. By following these steps, a categorical variable can be easily created from a continuous variable in R.

Create Categorical Variable from Continuous in R

You can use the `cut()` function in R to create a categorical variable from a continuous one.

This function uses the following basic syntax:

```
df$cat_variable <- cut(df$continuous_variable,  
breaks=c(5, 10, 15, 20, 25),  
labels=c('A', 'B', 'C', 'D'))
```

Note that `breaks` specifies the values to split the continuous variable on and `labels` specifies the label to give to the values of the new categorical variable.

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

Example: Create Categorical Variable from Continuous in R

Suppose we have the following data frame in R:

```
#create data frame
```

```
df <- data.frame(team=c('A', 'B', 'C', 'D', 'E', 'F', 'G', 'H'),  
points=c(78, 82, 86, 94, 99, 104, 109, 110))
```

```
#view data frame
```

```
df
```

```
team points
```

```
1 A 78
```

```
2 B 82
```

```
3 C 86
```

```
4 D 94
```

```
5 E 99
```

```
6 F 104
```

```
7 G 109
```

```
8 H 110
```

Currently points is a continuous variable.

We can use the cut() function to cut it into a categorical variable:

```
#add new column that cuts 'points' into categories
```

```
df$cat <- cut(df$points,  
breaks=c(70, 80, 90, 100, 110),  
labels=c('Bad', 'OK', 'Good', 'Great'))
```

```
#view updated data frame
```

```
df
```

```
team points cat
```

```
1 A 78 Bad
```

```
2 B 82 OK
```

```
3 C 86 OK
```

```
4 D 94 Good
```

```
5 E 99 Good
```

```
6 F 104 Great
```

```
7 G 109 Great
```

```
8 H 110 Great
```

We created a new categorical variable called `cat` that classifies each team in the data frame as `Bad`, `OK`, `Good`, or `Great` based on their points.

We can use the `class()` function to check the class of this new variable:

```
#check class of 'cat' column
```

```
class(df$cat)
```

```
"factor"
```

We can see that the cat variable is a factor.

We can also use the table() function to count the occurrences of each category in the cat variable:

```
#count occurrences of each category in 'cat' variable
```

```
table(df$cat)
```

```
Bad OK Good Great
```

```
1 2 2 3
```

```
#add new column that cuts 'points' into categories
```

```
df$cat <- cut(df$points, breaks=c(70, 80, 90, 100, 110))
```

```
#view updated data frame
```

```
df
```

```
team points cat
```

```
1 A 78 (70,80]
```

```
2 B 82 (80,90]
```

```
3 C 86 (80,90]
```

```
4 D 94 (90,100]
```

5 E 99 (90,100]

6 F 104 (100,110]

7 G 109 (100,110]

8 H 110 (100,110]

In some cases, you may actually prefer this to using custom labels.

Additional Resources

ARABPSYCHOLOGY.COM