

論增

獲得幫助並與社區討論課程材料。

論壇 所有主題帖

Search

Q

此主題僅對十二月 19 - 一月 22 班次的學生可視。

← All Course Discussions

Ø

Quiz Related: Different outputs while splitting the data. Help me with reasoning.

Suyash Dewangan Week 3 · 4天前

So, this is the case for Iris data and we are supposed to find the mean of Sepal.Length for three species.

First attempt:

f <- levels(iris[,5])

sapply(split(iris[,1],f),mean)

setosa versicolor virginica

5.842 5.822 5.866

Second attempt:

sapply(split(iris\$Sepal.Length,factor(iris\$Species)),mean)

setosa versicolor virginica

5.006 5.936 6.588

Why is it different when I use factor levels from levels()?

最早 最熱門 最新

涂

涂玉臻 Mentor·17小時前

Hi Suyash,

This is a great finding and a good question which is worth being noted. I took a few moments to figure.

Let's jump to the conclusion first: **only the second result is correct**. You can be assured by calling

The problem in your first method is that the length of f is only 3, and that length is shorter than iris\$Sepal.Length. According to the help document of split function, "**f is recycled as necessary** and if the length of x is not a multiple of the length of f a warning is printed," which means when you split iris\$Sepal.Length with f, R regards it as:

```
Sepal.Length Species
1
2
  1
             5.1 setosa
3
  2
             4.9 versicolor
4
  3
             4.7 virginica
5
  4
             4.6 setosa
6
  5
             5.0 versicolor
             5.4 virginica
```

In fact, however, these first six rows above are all setosa. Check it with:

- 1 head(iris)
- 2 # Also compare the difference of these two following lines:
- 3 split(iris\$Sepal.Length, iris\$Species)
- 4 split(iris\$Sepal.Length, f)

Cheers,

Yu-Zhen

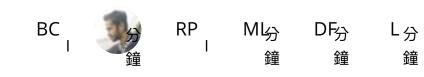
⚠ 1 · Hide 1 Reply

Thank you very much, Yu-Zhen!! Greatly appreciate your time and explanation.:) ₾ 0 贊 涂 回复 回复 1 > 涂 回复 回复

描述

歡迎來到課程討論論壇! 提出疑問題、討論話題和查找分享您目標的同學。 瀏覽下面熱門主題帖或 邊欄中的其他論壇。





分鐘







了解成為社區助教的更多信息