

数据格式转换

2023年9月9日 20:05

取子集

- 1.索引
- 2.subset()
- 3.抽样sample()

删除添加固定行

负索引删除

数据集mtcars

mtcars[-1:-5,]删除1-5行的数据

cbind()合并列

rbind()合并行

rownames()取出行名

duplicated()判断是否有相同数据

翻转数据

t()转置

rev()翻转一行

transform()

Usage

transform(`_data`, ...)

Arguments

_data

The object to be transformed

排序

sort()从大到小排序

order()从大到小排序后给出索引值

数据中心化

数据集中的各项数据减去数据集的均值

scale()

scale(x, center = TRUE)

数据标准化

指在数据中心化后再除以数据集的标准差

scale()

scale(x, center = TRUE, scale = TRUE)

对数据进行指定函数操作

apply()返回对向量，矩阵或数据框进行函数处理后的数据

apply(X, MARGIN, FUN=)#x:数据，MARGIN=1为行，2为列,FUN使用的函数

sapply(x,FUN)对数据集x，分别用FUN函数处理后输出，FUN函数是处理向量的函数

lapply()类似apply，对列表操作

tapply()对因子进行分析

针对数据框变形问题

使用reshape2包

`melt(data, id.vars=)` 融合数据, 将多列数据变为一列
`id.vars='变量名'` 改变以哪个作为分类变量
`dcast(md, formula~variable, fun.aggregate=function)`
`md`: 融合后的数据集
`formula` 以什么为分类
`function`: 使用什么函数来处理

可使用tidyr和dplyr来平替reshape2包

可操作数据格式为行为不同次, 列为不同变量, 行需要有行名

`gather(数据, 操作位置)` 把宽数据变为长数据

`spread()` 把长数据转换为宽数据

`separate(数据, 列名, into=赋予新的列名, sep=分隔符)` 将类似'a.b.c'的数据分成'a,b,c'三列

`unite(数据, col=组合后的列名, 操作列的名字, sep=连接符)`

dplyr



data-transformation

对列进行操作

`select()`



`select(data, ...)` Extract columns as a table.
`mtcars > select(mpg, wt)`

`mutate()`

对行进行操作

`arrange()`

ARRANGE CASES



`arrange(data, ..., .by_group = FALSE)` Order rows by values of a column or columns (low to high), use with **`desc()` to order from high to low.**
`mtcars > arrange(mpg)`
`mtcars > arrange(desc(mpg))`

? 避免不同包中函数歧义可'包名::函数名'来调用

magrittr包中的管道符

★ 管道符

`%>%`

快捷键`ctrl+shift+M`

输出结果传递作为输入

例子

```
x <- c(1:15)
```

```
round(sqrt(exp(log(x))), 2)
```

可简化为

```
library(magrittr) ## import the magrittr package
```

```
x <- c(1:15)
```

```
x %>%
```

```
  log() %>%
```

```
  exp() %>%
```

```
  sqrt() %>%
```

```
  round(2)
```

爆破符

`%%$%`

引用数据框中的变量不再需要`attach`

例子:

```
library(magrittr)
```

```
mtcars %$%
```

```
  cor(dis, mpg)
```