

Exercise 1

191820019 陈文杰

2020/9/25

R Markdown

#例1. RStudio开启时，就会默认启动一系列的数据集。其中一个数据集，名称是rivers，这是一个向量，它描述的是北美141条主要河流的长度。请用这一数据集，完成下述问题：

- 1、查看这一数据集
- 2、计算这一数据集的元素个数、平均数、中位数、标准差、方差、最大值、最小值。并将结果放置于一个名为rivers.Des的向量中。

解：

##[策略]

- 利用R自带统计函数length

##[过程|结果]

```
data(rivers)
Length <- length(rivers)
Mean <- mean(rivers)
Median <- median(rivers)
Sd <- sd(rivers)
Var <- var(rivers)
Max <- max(rivers)
Min <- min(rivers)
rviers.Des <- c(Length, Mean, Median, Sd, Var, Max, Min)
rviers.Des
```

```
## [1] 141.0000 591.1844 425.0000 493.8708 243908.4086 3710.0000
## [7] 135.0000
```

#例2.尝试使用R语言中的cor这一函数。使用women这一系统自带数据集，完成下述问题：

- 1、这一数据集基本结构如何。多少行？多少列？
- 2、请计算第一列的均值和方差。
- 3、请选择第一列的值超过60的行，形成一个新的数据集，名称为women.Height60
- 4、请使用R中cor这一函数，计算women这一数据集前2列的相关系数。
- 5、cor这一函数中的一个参数为method，它的取值包括，person，kendall，spearman。请简略写下他们之间的区别（不超过100个字），并分别用三种方法，重复问题4。
- 6、写一个名为Women.Cor的函数，这一函数具有两个参数，参数1能够接受一个data.frame，参数2能够接受一个vector，要求函数功能是根据传入的vector获取对应行，并计算这些行前2列值的相关系数（pearson方法）。例如，如果传递：参数1是women数据集，参数2是c(4:8)，返回的是women数据集4到8行数据，前2列的相关系数。

解：

##[策略]

- 利用dim函数查看数据框的形状
- 用df[]选取数据框子集
- 利用cor () 函数计算变量相关性

##[过程|结果]

```
women
```

```
##2.1
women[,1]
```

```
## [1] 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72
```

```
dim(women)
```

```
## [1] 15 2
```

```
##2.2
coll = women$height
height.mean = mean(coll)
height.var = var(coll)
##2.3
women.Height60 = women[which(women$height >60),]
#subset(women. women$height>60)
##2.4
cor(women[,1],women[,2])
```

```
## [1] 0.9954948
```

```
##2.5
help(cor)
```

```
## starting httpd help server ... done
```

```
#Person 皮尔逊相关(用于线性相关的连续型变量)
#两个连续变量间呈线性相关时，使用Pearson积差相关系数，不满足积差相关分析的适用条件时，使用Spearman秩相关系数来描述。
cor(women[,1],women[,2],method="pearson")
```

```
## [1] 0.9954948
```

```
#Sperman 斯皮尔曼相关(可用于rank variable)
#Spearman相关系数又称秩相关系数，是利用两变量的秩次大小作线性相关分析，对原始变量的分布不作要求，属于非参数统计方法，适用范围要广些。
cor(women[,1],women[,2],method="spearman")
```

```
## [1] 1
```

#kendall 单调相关性 (用于category variable)

#Kendall's tau-b等级相关系数: 用于反映分类变量相关性的指标, 适用于两个分类变量均为有序分类的情况。对相关的有序变量进行非参数相关检验; 取值范围在-1-1之间, 此检验适合于正方形表格;

```
cor(women[,1],women[,2],method="kendall")
```

```
## [1] 1
```

##2.6

```
Women.Cor <- function(df,vec){  
  df.row = df[vec,]  
  df.cor = cor(df.row[,1],df.row[,2])  
  return (df.cor)  
}  
Women.Cor(women,c(4:8))
```

```
## [1] 1
```