短学期作业二

汪利军 *3140105707*

July 5, 2017

Contents

1 Problem 5.1

```
c(1,4)*c(2,3)
```

```
## [1]  2 12
```

2 Problem 5.4

```
Gender = c('M', 'F')
Weight = c(80, 60)
Eyes = c('Blue', 'Green')
Height = c(180, 160)
X = data.frame(Gender, Weight)
Y = data.frame(Eyes, Height)
rownames(X) = c('Jack', 'Julia')
rownames(Y) = c('Jack', 'Julia')
## merge
merge(X, Y, by = "row.names")
```

```
##   Row.names Gender Weight  Eyes Height
## 1      Jack      M     80  Blue    180
## 2     Julia      F     60 Green    160
```

3 Problem 5.11

```
weight = c(79, 90, 87, 63, 90, 71, 58, 80, 91, 89)
height = c(163, 163, 198, 164, 168, 178, 191, 194, 185, 176)
```

```
gender = c('M', 'F', 'M', 'F', 'F', 'F', 'M', 'F', 'F', 'M')
x = data.frame(weight, height, gender)
```

outputs the weights and heights of all women

```
x[x$gender == 'F', c(1,2)]
```

```
##   weight height
## 2     90    163
## 4     63    164
## 5     90    168
## 6     71    178
## 8     80    194
## 9     91    185
```

4 Problem 5.12

```
(1:3)[any(c(T, F, T))]
```

```
## [1] 1 2 3
```

```
(1:3)[all(c(T, F, T))]
```

```
## integer(0)
```

5 Problem 5.13

```
c(T, T, F) | c(F, T, F)
```

```
## [1]  TRUE  TRUE FALSE
```

```r
c(T, T, F) || c(F, T, F)
```

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

6 Problem 5.15

```r
paste(c("a", "b"), c("c", "d"), collapse = "", sep = "")
```

```
## [1] "acbd"
```

7 Chapter 5 Worksheet B

读取数据

```r
library(XLConnect)
```

```
## Loading required package: XLConnectJars
```

```
## XLConnect 0.2-13 by Mirai Solutions GmbH [aut],
##   Martin Studer [cre],
##   The Apache Software Foundation [ctb, cph] (Apache POI),
##   Graph Builder [ctb, cph] (Curvesapi Java library)
```

```
## http://www.mirai-solutions.com ,
## http://miraisolutions.wordpress.com
```

```r
connect <- loadWorkbook('Infarction.xls')
data <- readWorksheet(connect, 1)
```

7.1 Problem 5.1

注意到表格中缺失值为 “.”，为方便后续处理，先将 “.” 赋值为 NA。值得说明的是，为了让 data 能在 sapply 中进行更新，采用了 “<<-” 运算符，而非 “<-”，这样 data 数据集中的 “.” 重新赋值为了 NA

```
res = sapply(1:nrow(data), function(i) data[i, data[i, ] == '.'] <<- NA)
```

则包含缺失值的行为

```
which(!complete.cases(data))
```

```
##  [1]   2   5  10  20  29  33  38  49  62  71 153 195 200 202 344 346 362
## [18] 426
```

7.2 Problem 5.2

```
row.missing.count = sapply(1:nrow(data), function(i) sum(is.na(data[i,])))
## 超过一个缺失值的行
which(row.missing.count>1)
```

```
##  [1]   2  10  29  38  49  71 153 195 200 346 362 426
```

7.3 Problem 5.3

包含缺失值的列名为

```
colnames(data)[which(sapply(1:ncol(data), function(i) sum(is.na(data[i]))) > 0)]
```

```
## [1] "WEIGHT" "BMI"    "ATCD"
```

7.4 Problem 5.4

通过 complete.cases() 进行删掉含缺失值的行

```
data.rm.missing = data[complete.cases(data),]
```