assinment1

Yuhao Wang(UNI: yw3204)

9/15/2018

Part 1

i

```r
titanic <- read.table("Titanic.txt", header = TRUE, as.is = TRUE)
```

ii

```r
nrow(titanic)
```

```
## [1] 891
```

```r
ncol(titanic)
```

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

iii

```r
titanic$Survivor.Word <- "survived"
ind = titanic$Survived == 0
titanic[ind, ]$Survivor.Word <- "died"
head(titanic$Survivor.Word)
```

```
## [1] "died"     "survived" "survived" "survived" "died"     "died"
```

```r
head(titanic$Survived)
```

```
## [1] 0 1 1 1 0 0
```

Part 2

i

```r
apply(titanic[, c("Survived", "Age", "Fare")], 2, mean)
```

```
##   Survived        Age       Fare
##  0.3838384         NA 32.2042080
```

The mean of Survived tells us the percentage of people survived. Age has a mean value of NA because some value in age are NA.

ii

```
ind1 = titanic$Sex == "female"
p1 = sum(titanic[ind1, ]$Survived == 1) / nrow(titanic)
round(p1, 2)
```

[1] 0.26

The proportion is 0.26.

iii

```
ind2 = titanic$Survived == 1
p2 = sum(titanic[ind2, ]$Sex == "female") / nrow(titanic[ind2, ])
round(p2, 2)
```

[1] 0.68

The proportion is 0.68.

iv

```
classes <- sort(unique(titanic$Pclass))
Pclass.Survival <- vector("numeric", length = 3)
names(Pclass.Survival) <- classes

for (i in 1:3) {
    ind_temp = titanic$Pclass == i
    p_temp = sum(titanic[ind_temp, ]$Survived == 1) / nrow(titanic[ind_temp, ])
    Pclass.Survival[i] = round(p_temp, 2)
}
Pclass.Survival
```

1   2   3
0.63 0.47 0.24

v

```
Pclass.Survival2 <- vector("numeric", length = 3)
names(Pclass.Survival2) <- classes
Pclass.Survival2 <- tapply(titanic$Survived, titanic$Pclass, mean)
round(Pclass.Survival2, 2)
```

1   2   3
0.63 0.47 0.24

vi

Yes, the more advanced the Pcalss, the higher the survival rate.