

# GR5206 Homework 1

Hanao Li

## Part 1: Importing Data into R

```
# i.)  
# Import the dataset into R and called titanic  
titanic <- read.table('Titanic.txt', header = TRUE, as.is = TRUE)  
  
# ii.)  
# check the number of rows and columns titanic have  
nrow(titanic)
```

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

```
ncol(titanic)
```

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

So there are 891 rows and 12 columns.

```
# iii.)  
# Create a new variable called Survived.Word from the variable 'Survived'  
titanic['Survived.Word'] <- NA  
titanic$Survived.Word[titanic$Survived == 1] <- 'survived'  
titanic$Survived.Word[titanic$Survived == 0] <- 'died'
```

## Part 2: Exploring the Data in R

```
# i.)  
lists <- c('Survived', 'Age', 'Fare')  
small <- titanic[, names(titanic) %in% lists]  
apply(small, 2, mean)
```

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

From the results, the mean of Survived tells us that the survival rate for the Titanic passengers is about 0.38

```
summary(small$Age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      0.42   20.12   28.00   29.70   38.00   80.00   177
```

The mean of Age has a value of NA and this is because there are 177 missing values for the 'Age' variable

```
# ii.)
# Compute the proportion of female passengers who survived the titanic disaster
round(sum(titanic$Sex == 'female' & titanic$Survived == 1) / sum(titanic$Sex == 'female'), 2)
```

```
## [1] 0.74
```

The proportion is 0.74

```
# iii.)
# Compute the proportion of female passengers out of the survivors
round(sum(titanic$Sex == 'female' & titanic$Survived == 1) / sum(titanic$Survived == 1), 2)
```

```
## [1] 0.68
```

The proportion is 0.68

```
# iv.)
# Compute the survival rate for each class using for loop
classes <- sort(unique(titanic$Pclass))
Pclass.Survival <- vector('numeric', 3)
names(Pclass.Survival) <- classes

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

```
##      1      2      3
## 0.63 0.47 0.24
```

The survival rate is 0.63, 0.47 and 0.24 for Class 1, Class 2 and Class 3, respectively.

```
# v.)
# Compute the survival rate for each class using tapply
Pclass.Survival2 <- vector('numeric', 3)
names(Pclass.Survival2) <- classes

Pclass.Survival2 <- round(tapply(titanic$Survived, titanic$Pclass, function(x){sum(x) / length(x)}), 2)
Pclass.Survival2
```

```
##      1      2      3
## 0.63 0.47 0.24
```

The survival rate is 0.63, 0.47 and 0.24 for Class 1, Class 2 and Class 3, respectively.

```
# vi.)
# Summary
```

From the results we had, we could tell that there is a relationship between survival rate and class. The better the Class, the higher the survival rate. Class 1 had the highest survival rate while Class 3 had the lowest survival rate.