Assignment 1

Question 1

```
# question 1
apply(Titanic, c(1, 4), sum)
```

```
##       Survived
## Class   No Yes
##   1st  122 203
##   2nd  167 118
##   3rd  528 178
##   Crew 673 212
```

```
apply(Titanic, c(2, 4), sum)
```

```
##         Survived
## Sex        No Yes
##   Male   1364 367
##   Female  126 344
```

```
apply(Titanic, c(3, 4), sum)
```

```
##        Survived
## Age       No Yes
##   Child   52  57
##   Adult 1438 654
```

Question 2

```
# question 2
set.seed(1234)
k <- rnorm(1000)
sum(k < -1.65 | k > 1.65)/1000
```

```
## [1] 0.098
```

10 % of the generate numbers are outside the -1.645, 1.645 limit. This matches our expectation because 1.65 is 1.65 standard deviations from the mean and we know that 10% of values should be outside the 1.65 standard deviation range. ## Question 3

```
# question 3
breaks <- c(-99999,-1.645,1.645,99999)
x <- rnorm(1000)
t <- table(cut(x,breaks))
names(t) <- c('small', 'medium', 'large')
t
```

```
##  small medium  large
##     51    903     46
```

Question 4

```
# question 4
rnd <- round(runif(1000,0,1000))
sum(rnd %% 5 == 0)
```

```
## [1] 191
```

This matches out expectation. Randomly generated numbers should have 20% divisible by 5 because 20% of all numbers are divisible by 5.

Question 5

```
# question 5
f_day <- function(year) {
  return (paste(toString(year),'-01-01',sep=''))
}

l_day <- function(year) {
  return (paste(toString(year),'-12-31',sep=''))
}

all_days <- function(year){
  return(table(weekdays(seq(as.Date(f_day(year)), as.Date(l_day(year)), by="days"))))

}

for (year in c(2011:2020)) {
  print(year)
  print(all_days(year))
}
```

```
## [1] 2011
##
##    Friday    Monday  Saturday    Sunday  Thursday   Tuesday Wednesday
##        52        52        53        52        52        52        52
## [1] 2012
##
##    Friday    Monday  Saturday    Sunday  Thursday   Tuesday Wednesday
##        52        53        52        53        52        52        52
## [1] 2013
##
##    Friday    Monday  Saturday    Sunday  Thursday   Tuesday Wednesday
##        52        52        52        52        52        53        52
## [1] 2014
##
##    Friday    Monday  Saturday    Sunday  Thursday   Tuesday Wednesday
##        52        52        52        52        52        52        53
## [1] 2015
##
##    Friday    Monday  Saturday    Sunday  Thursday   Tuesday Wednesday
##        52        52        52        52        53        52        52
## [1] 2016
```

```
##
##    Friday    Monday  Saturday    Sunday  Thursday   Tuesday Wednesday
##        53        52        53        52        52        52        52
## [1] 2017
##
##    Friday    Monday  Saturday    Sunday  Thursday   Tuesday Wednesday
##        52        52        52        53        52        52        52
## [1] 2018
##
##    Friday    Monday  Saturday    Sunday  Thursday   Tuesday Wednesday
##        52        53        52        52        52        52        52
## [1] 2019
##
##    Friday    Monday  Saturday    Sunday  Thursday   Tuesday Wednesday
##        52        52        52        52        52        53        52
## [1] 2020
##
##    Friday    Monday  Saturday    Sunday  Thursday   Tuesday Wednesday
##        52        52        52        52        53        52        53
```

Question 6

```
dat <- read.delim('Groceries_dataset.csv', header=TRUE)

# number of rows and columns
dim(dat)
```

```
## [1] 38765     1
```

```
#summary of data
summary(dat)
```

```
##  Member_number.Date.itemDescription
##  Length:38765
##  Class :character
##  Mode  :character
```