

# CIND 123 - Data Analytics: Basic Methods

Rui Zhang

Assignment 2 (10%)

Rui Zhang

DHA 500736315

---

## Instructions

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. Review this website for more details on using R Markdown <http://rmarkdown.rstudio.com>.

Use RStudio for this assignment. Complete the assignment by inserting your R code wherever you see the string “#INSERT YOUR ANSWER HERE”.

When you click the **Knit** button, a document (PDF, Word, or HTML format) will be generated that includes both the assignment content as well as the output of any embedded R code chunks.

Submit **both** the rmd and generated output files. Failing to submit both files will be subject to mark deduction.

## Sample Question and Solution

Use `seq()` to create the vector  $(1, 2, 3, \dots, 20)$ .

```
seq(1,20)
```

```
## [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
```

---

## Question 1

The Titanic Passenger Survival DataSet provides information on the fate of passengers on the fatal maiden voyage of the ocean liner “Titanic.” The dataset is available from the Department of Biostatistics at the Vanderbilt University School of Medicine (<http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/titanic3.csv>) in several formats. Store the Titanic DataSet `titanic_train` using the following commands.

```
library(titanic)
titanicDataset <- read.csv(file = "http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/titanic3.csv")
summary(titanicDataset)
```

```
##      pclass      survived      name      sex
## Min.   :1.000   Min.   :0.000   Length:1309   Length:1309
## 1st Qu.:2.000   1st Qu.:0.000   Class :character   Class :character
## Median :3.000   Median :0.000   Mode  :character   Mode  :character
## Mean   :2.295   Mean   :0.382
## 3rd Qu.:3.000   3rd Qu.:1.000
## Max.   :3.000   Max.   :1.000
##
##      age      sibsp      parch      ticket
## Min.   : 0.17   Min.   :0.0000   Min.   :0.000   Length:1309
## 1st Qu.:21.00   1st Qu.:0.0000   1st Qu.:0.000   Class :character
## Median :28.00   Median :0.0000   Median :0.000   Mode  :character
## Mean   :29.88   Mean   :0.4989   Mean   :0.385
## 3rd Qu.:39.00   3rd Qu.:1.0000   3rd Qu.:0.000
## Max.   :80.00   Max.   :8.0000   Max.   :9.000
## NA's    :263
##      fare      cabin      embarked      boat
## Min.   : 0.000   Length:1309   Length:1309   Length:1309
## 1st Qu.: 7.896   Class :character   Class :character   Class :character
## Median :14.454   Mode  :character   Mode  :character   Mode  :character
## Mean   :33.295
## 3rd Qu.:31.275
## Max.   :512.329
## NA's    :1
##      body      home.dest
## Min.   : 1.0   Length:1309
## 1st Qu.:72.0   Class :character
## Median :155.0   Mode  :character
## Mean   :160.8
## 3rd Qu.:256.0
## Max.   :328.0
## NA's    :1188
```

- a) Extract the columns `sex`, `age`, `cabin` and `survived` into a new data frame of the name ‘`titanicSubset`’.

```
#INSERT YOUR ANSWER HERE
```

- b) Use the `aggregate()` function to display the total number of survivors grouped by `sex`

```
#INSERT YOUR ANSWER HERE
```

- c) Use the `count()` function in `dplyr` package to display the total number of passengers within each Ticket Class `Pclass`.

*#INSERT YOUR ANSWER HERE*

- d) Answer the following graphically:

1. What was the survival rates for females and males?
2. What was the age distribution on the Titanic?

*#INSERT YOUR ANSWER HERE*

- e) Use the `for` loop and `if` control statements to list the children's names, aged 14 or under, on the Titanic.

*#INSERT YOUR ANSWER HERE*

---

## Question 2

In an experiment of rolling 10 dice simultaneously. Use the binomial distribution to calculate the followings:

- a) The probability of getting six 6's

*#INSERT YOUR ANSWER HERE*

- b) The probability of getting six, seven, or eight 3's

*#INSERT YOUR ANSWER HERE*

- c) The probability of getting six even numbers

*#INSERT YOUR ANSWER HERE*

---

### Question 3

In a shipment of 20 engines, history shows that the probability of any one engine proving unsatisfactory is 0.1

- a) Use the Binomial approximation to calculate the probability that at least three engines are defective?

*#INSERT YOUR ANSWER HERE*

- b) Use the Poisson approximation to calculate the probability that at least three engines are defective?

*#INSERT YOUR ANSWER HERE*

- c) Compare the results of parts a and b, then illustrate graphically on how well the Poisson probability distribution approximates the Binomial probability distribution.

*#INSERT YOUR ANSWER HERE*

---

## Question 4

Write a script in R to compute the following probabilities of a normal random variable with mean 16 and variance 9

a) lies between 14.4 and 20.3 (inclusive)

*#INSERT YOUR ANSWER HERE*

b) is greater than 21.8

*#INSERT YOUR ANSWER HERE*

c) is less or equal to 10.5

*#INSERT YOUR ANSWER HERE*

d) is less than 13 or greater than 19

*#INSERT YOUR ANSWER HERE*

---

END of Assignment #2.