

# IST772– Problem Set 3

Sathish Kumar Rajendiran

Attribution statement: 1. I did this homework by myself, with help from the book and the professor.

```
set.seed(772)
```

## Chapter 3, Exercise 2

*For the remaining exercises in this set, we will use one of R’s built-in data sets, called the “trees” data set. According to the documentation for R, the trees data set contains information on the measurements of the girth, height and volume of timber in 31 felled black cherry trees. Use the summary(trees) command to reveal basic information about the trees data set. You will find that trees contains three different variables. Name the variables (1 pt). Use the dim(trees) command to show the dimensions of the trees data set. The second number in the output, 3, is the number of columns in the data set, in other words the number of variables. What is the first number (1 pt)? Report it and describe briefly what you think it signifies.*

### 1) Variables: Trees data set has 3 variables and 31 observations

- Variables \* [1] Girth - numeric | Tree diameter (rather than girth, actually) in inches \* [2] Height - numeric | Height in ft \* [3] Volume - numeric | Volume of timber in cubic ft

### 2) Dimensions: Trees data set has 3 variables/columns and 31 observations.

- Dimensions first number from the dim(trees) represent - number of observations (31) available in this dataset. This data set provides measurements of the diameter, height and volume of timber in 31 felled black cherry trees.

Please find more details from the R code below,

```
cat("\n***** Chapter 3, Exercise 2 - Begins *****\n")
```

```
##
```

```
## ***** Chapter 3, Exercise 2 - Begins *****
```

```

# ?trees # Diameter, Height and Volume for Black Cherry Trees

# This data set provides measurements of the diameter, height and volume of timber in 31
# felled black cherry trees.
# Note that the diameter (in inches) is erroneously labelled Girth in the data. It is measured at
# 4 ft 6 in above the ground.

cat("-----\n")

## -----

myTrees <- trees # copy "trees" into myTrees data set

cat("1) Column/Variable names from trees data set are:")

## 1) Column/Variable names from trees data set are:

cat("\n\n ", colnames(myTrees))

##
##      Girth Height Volume

cat("\n\nSummary of trees data set is:\n")

##
## Summary of trees data set is:

summary(myTrees) #summary of "myTrees" data set

##      Girth      Height      Volume
## Min.   : 8.30   Min.   :63   Min.   :10.20
## 1st Qu.:11.05   1st Qu.:72   1st Qu.:19.40
## Median :12.90   Median :76   Median :24.20
## Mean   :13.25   Mean   :76   Mean   :30.17
## 3rd Qu.:15.25   3rd Qu.:80   3rd Qu.:37.30
## Max.   :20.60   Max.   :87   Max.   :77.00

cat("-----\n\n")

## -----

cat("2) Number of Observations from trees data set is:", nrow(trees))

## 2) Number of Observations from trees data set is: 31

```

```

cat("\nDimensions of trees data set is:\n")

##
## Dimensions of trees data set is:

dim(myTrees) # Number of rows and columns; 31 observations (rows) and 3 columns/variables

## [1] 31  3

cat("\n-----\n")

##
## -----

cat("Structure of trees data set is:\n")

## Structure of trees data set is:

str(myTrees) #structure of "myTrees" data set / 31 observations of 3 variables Girth, Height and Volume

## 'data.frame':    31 obs. of  3 variables:
##  $ Girth : num  8.3 8.6 8.8 10.5 10.7 10.8 11 11 11.1 11.2 ...
##  $ Height: num  70 65 63 72 81 83 66 75 80 75 ...
##  $ Volume: num  10.3 10.3 10.2 16.4 18.8 19.7 15.6 18.2 22.6 19.9 ...

cat("\n***** Chapter 3, Exercise 2 - Ends *****\n")

##
## ***** Chapter 3, Exercise 2 - Ends *****

```

## Chapter 3, Exercise 3

When a data set contains more than one variable, R offers a subsetting operator, \$, to access each variable individually. (NB. the backslash is needed in the notebook file because a dollar sign by itself means to shift to math mode. In R code, you would just use the dollar sign, without the back slash.) For the exercises below, we are interested only in the contents of one of the variables in the data set, called Girth. We can access the Girth variable by itself, using the \$, with this expression: `trees$Girth`. Run the following commands, add a comment to each line saying what each command does, report the output, and briefly explain each piece of output (1 pt for summary, head, and mean; 1 pt for new variable, and 0.50 quantile):

### 1) Summary, Head and Mean:

- `Summary(trees$Girth)` - `summary()` function summarizes the Girth (Tree diameter in inches) variables spread of distribution by different quantiles range. Min, Max, Mean, 1st Quartile, 2nd Quartile i.e. (Median) 3rd Quartile and 4th Quartile (Max).
  - Min. 1st Qu. Median Mean 3rd Qu. Max.

– 8.30 11.05 12.90 13.25 15.25 20.60

- Head: `head()` function returns the first few observations from the overall population. In this case `head(trees$Girth)` returns top 6 records/observations of the variable Girth.
- Mean: `mean()` function returns average value from the population distribution. In this case, `mean(trees$Girth)` is 13.24839.

## 2) New Variable, Median (50% quantile):

- New Variable: `myTreeGirth <- trees$Girth` creates a new variable called `myTreeGirth` and assigns values of Girth variable from trees dataset.
- median: `median()` function returns the central point/ middle value of the population when sorted sequentially (smallest to largest). `myTreeGirth` has total 31 observations; hence, the median/middle value is 16th value. i.e. 12.9 median is also the 50th percentile value of the distribution.

Please find more details from the R code below,

```
cat("\n***** Chapter 3, Exercise 3 - Begins *****\n")

##
## ***** Chapter 3, Exercise 3 - Begins *****

cat("-----\n")

## -----

cat("\n1) Summary of 1st variable Girth is:\n")

##
## 1) Summary of 1st variable Girth is:

summary(trees$Girth) # Output summary of Girth variable to console

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      8.30   11.05   12.90   13.25   15.25   20.60

cat("-----\n")

## -----

cat("\n2) Top 6 or head of observations of 1st variable Girth is:\n")

##
## 2) Top 6 or head of observations of 1st variable Girth is:
```

```
head(trees$Girth)
```

```
## [1] 8.3 8.6 8.8 10.5 10.7 10.8
```

```
cat("-----\n")
```

```
## -----
```

```
cat("\n3) mean/average value of observations of 1st variable Girth is:\n")
```

```
##
```

```
## 3) mean/average value of observations of 1st variable Girth is:
```

```
mean(trees$Girth)
```

```
## [1] 13.24839
```

```
cat("-----\n")
```

```
## -----
```

```
cat("\n4) Assign values of 1st variable Girth to new vector(variable) myTreeGirth\n")
```

```
##
```

```
## 4) Assign values of 1st variable Girth to new vector(variable) myTreeGirth
```

```
myTreeGirth <- trees$Girth
```

```
cat("-----\n")
```

```
## -----
```

```
cat("\n5) 50th Quantile or Median of 1st variable Girth from the new variable myTreeGirth:\n")
```

```
##
```

```
## 5) 50th Quantile or Median of 1st variable Girth from the new variable myTreeGirth:
```

```
quantile(myTreeGirth,0.50)
```

```
## 50%
```

```
## 12.9
```

```
median(myTreeGirth)
```

```
## [1] 12.9
```

```
myTreeGirth[16] # 16th value/central/middle point value

## [1] 12.9

cat("\n***** Chapter 3, Exercise 3 - Ends *****\n")

##
## ***** Chapter 3, Exercise 3 - Ends *****
```

## Chapter 3, Exercise 4

*In the second to last command of the previous exercise, you created a copy of the girth data from the trees data set and put it in a new vector called myTreeGirth. You can continue to use this myTreeGirth variable for the rest of the exercises below. Create a histogram for that variable. Then write code that will display the 2.5% and 97.5% quantiles of the distribution for that variable (1 pt for histogram and quantiles). Write an interpretation of the variable, including descriptions of the mean, median (1 pt for mean and median), shape of the distribution (1 pt), and the 2.5% and 97.5% quantiles. Make sure to clearly describe what the 2.5% and 97.5% quantiles signify (1 pt).*

### 1) Histogram, Quantiles:

- histogram - function `hist(myTreeGirth)` - computes a histogram of new variable “myTreeGirth” to console as vertical bar graph with bar sizes representing the range of frequency distribution of observations.
- quantiles: `quantile(myTreeGirth,prob=0.025)` and `quantile(myTreeGirth,prob=0.975)` function returns the 2.5% and 97.5% quantiles of the distribution of the new variable “myTreeGirth”
  - 2.5% quantile is 8.525
  - 97.5% quantile is 18.65

### 2) Mean, Median (50% quantile):

- mean: `mean(myTreeGirth)` returns the mean or the average distribution value of the new variable “myTreeGirth” which is 13.24839
- median: `median()` function returns the central point/ middle value of the population when sorted sequentially (smallest to largest). myTreeGirth has total 31 observations; hence, the median/middle value is 16th value. i.e. 12.9 median is also the 50th percentile value of the distribution.

### 3) Shape:

- Shape: Shape of the histogram - represents an uniform frequency distribution of values across 31 observations. Shape has consistent high and low bars across. Values 10 to 12 ranges in high frequencies while the ~18 having the least frequency.

#### 4) Interpret Quantiles:

- 2.5% quantile: represents low end of the left tail of the density curve.
- 97.5% quantile represents the low end of right tail of the density curve. region between these two ranges suggests the possibility of 95% of the distribution of observations and is called the central region. In comparison with 50% quantile (median) - left hand side of the median has high frequency/densely packed values than the right side of the median. Values are sparsed more on this side.

Please find more details from the R code below,

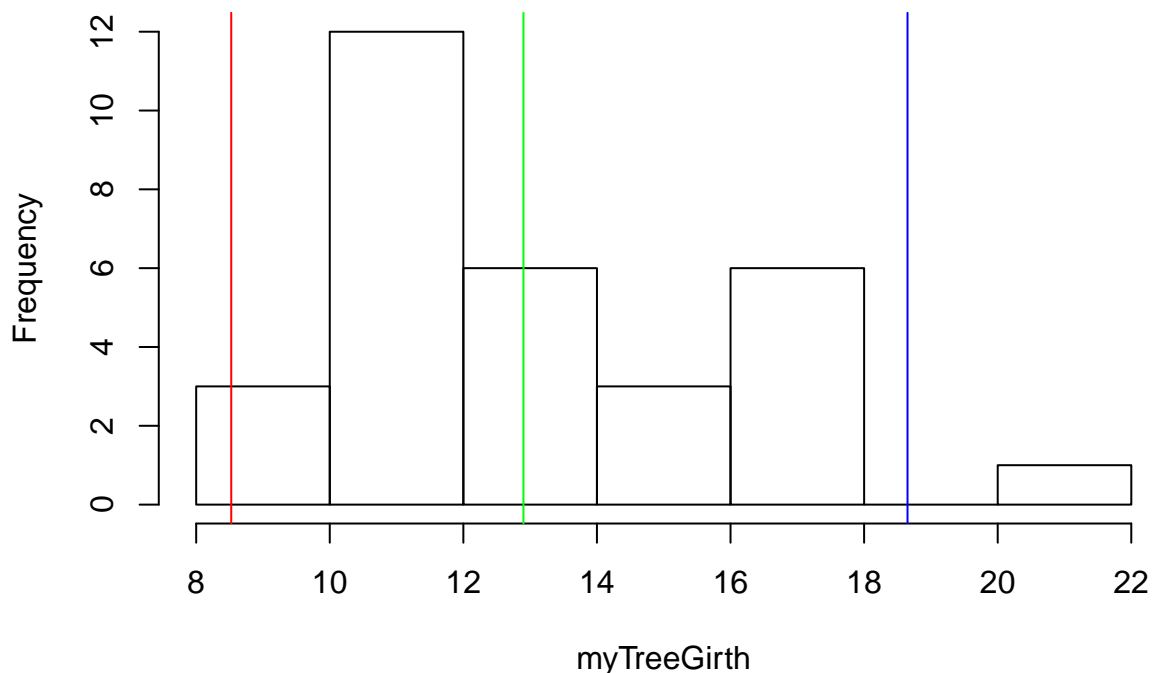
```
cat("\n***** Chapter 3, Exercise 4 - Begins *****\n")

##
## ***** Chapter 3, Exercise 4 - Begins *****

?hist

# cat("\n1) Histogram of 1st variable (vector) Girth is:\n")
hist(myTreeGirth) # Output histogram of Girth variable to console with 2.5% and 97.5% quantile lines
abline(v=quantile(myTreeGirth,prob=0.025),col="red")
abline(v=quantile(myTreeGirth,prob=0.975),col="blue")
abline(v=quantile(myTreeGirth,prob=0.50),col="green")
```

**Histogram of myTreeGirth**



```
cat("\n***** Chapter 3, Exercise 4 - Continues *****\n")
```

```
##
```

```
## ***** Chapter 3, Exercise 4 - Continues *****
```

```
cat("\n2) Quantiles:\n")
```

```
##
```

```
## 2) Quantiles:
```

```
rawQ2.5 <- quantile(myTreeGirth, prob=0.025) #2.5% of the quantile
rawQ97.5 <- quantile(myTreeGirth, prob=0.975) #97.5% of the quantile
rawQ50 <- quantile(myTreeGirth, prob=0.50) # 50% of the quantile (median)
cat(" 2.5% Quantile of the variable Girth is :",rawQ2.5,"\n 97.5% Quantile of the
    variable Girth is :",rawQ97.5)
```

```
## 2.5% Quantile of the variable Girth is : 8.525
```

```
## 97.5% Quantile of the
```

```
## variable Girth is : 18.65
```

```
cat("\n 50% Quantile of the variable Girth is :",rawQ50)
```

```
##
```

```
## 50% Quantile of the variable Girth is : 12.9
```

```
cat("\n-----\n")
```

```
##
```

```
## -----
```

```
cat("\n3) Mean and Median:\n")
```

```
##
```

```
## 3) Mean and Median:
```

```
Mean <- mean(myTreeGirth) # Output mean of Girth vairable to console
Median <- median(myTreeGirth) # Output median of Girth vairable to console
cat(" Mean/average of the variable Girth is :",Mean,"\n Median or 50th Quantile of the
    variable Girth is :",Median)
```

```
## Mean/average of the variable Girth is : 13.24839
```

```
## Median or 50th Quantile of the
```

```
## variable Girth is : 12.9
```

```
cat("\n-----\n")
```

```
##
```

```
## -----
```



```
cat("\n4) Shape: Uniform distribution of 31 observations:\n")
```

```
##
```

```
## 4) Shape: Uniform distribution of 31 observations:
```

```
cat("\n-----\n")
```

```
##
```

```
## -----
```

```
cat("\n5) 2.5% and 97.5 quantiles of the distribution represents 95% of  
the population distribution. i.e. between red and blue lines from above histogram represent  
the central region of the distribution\n")
```

```
##
```

```
## 5) 2.5% and 97.5 quantiles of the distribution represents 95% of  
## the population distribution. i.e. between red and blue lines from above histogram represent  
## the central region of the distribution
```

```
cat("\n***** Chapter 3, Exercise 4 - Ends *****\n")
```

```
##
```

```
## ***** Chapter 3, Exercise 4 - Ends *****
```

## Chapter 3, Exercise 5

Write R code that will construct a sampling distribution of means from the girth data (as noted above, if you did exercise 3 you can use `myTreeGirth` instead of `trees$Girth`). Make sure that the sampling distribution contains at least 1,000 means. Store the sampling distribution in a new variable that you can keep using. Use a sample size of  $n = 7$  (sampling with replacement) (2 pts). Show a histogram of this distribution of sample means. Then, write and run R commands that will display the 2.5% and 97.5% quantiles of the sampling distribution on the histogram with a vertical line (1 pt).

### 1) Sampling Distribution with replacement

- sampling distribution of means from the girth data with replacement is created (`sampleGirth`) using the R code below. In addition, it takes 10,000 means as mentioned is created on another variable called `samplingDistribution`. With 10,000 trials of mean value of Sample size 7 with replacement ensures that the rule of large numbers and central limit theorem is kept intact that the next draw has the equal chance of pulling out any member of the population.
- `sampleGirth <- function(n) {sample(myTreeGirth, size=n, replace=TRUE)}`
- `samplingDistribution <- replicate(10000, mean(sampleGirth(7)))`

## 2) Histogram & Quantiles

- Histogram: `hist(samplingDistribution)` plots a normal distribution of sample means with bell-shaped curve having symmetric curves on each side of the median.
- quantiles: `quantile(samplingDistribution,prob=0.025)` and `quantile(samplingDistribution,prob=0.975)` function returns the 2.5% and 97.5% quantiles of the distribution
  - 2.5% Quantile of the `samplingDistribution` is : 11.07143 | represents low end of the left tail of the density curve.
  - 97.5% Quantile of the `samplingDistribution` is : 15.58571 | represents the low end of right tail of the density curve.
  - Region between these two ranges suggests the possibility of 95% of the distribution of observations.
  - 50% quantile (median) is 13.253 is almost closer to mean value 13.253 of the distribution
  - Min. 1st Qu. Median Mean 3rd Qu. Max.
  - 9.629 12.429 13.200 13.253 14.057 17.257

Please find more details from the R code below,

```
cat("\n***** Chapter 3, Exercise 5 - Begins *****\n")

##
## ***** Chapter 3, Exercise 5 - Begins *****

set.seed(772) # Control randomization

# function to create random sample of myTreeGirth distribution with n observations with replacement
sampleGirth <- function(n) {sample(myTreeGirth, size=n, replace=TRUE)}

#Store the sampling distribution to a variable with 10,000 means
#Store the sampling distribution to a variable
samplingDistribution <- replicate(10000, mean(sampleGirth(7)))

cat("\nmean of the myTreeGirth is:",mean(myTreeGirth))

##
## mean of the myTreeGirth is: 13.24839

cat("\nmean of the samplingDistribution is:", mean(samplingDistribution))

##
## mean of the samplingDistribution is: 13.25298

cat("\nsummary samplingDistribution is:\n")

##
## summary samplingDistribution is:
```

```
summary(samplingDistribution)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.629 12.429 13.200 13.253 14.057 17.257
```

```
cat("\n-----\n")
```

```
##
## -----
```

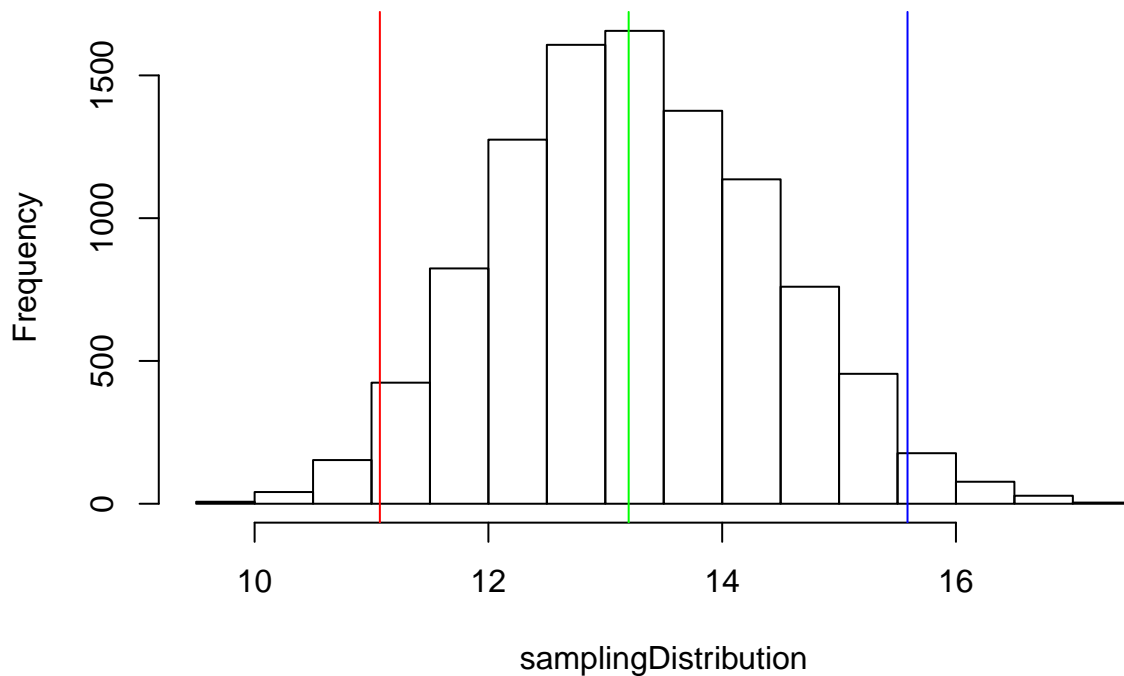
```
cat("\n***** Chapter 3, Exercise 5 - Continues *****\n")
```

```
##
## ***** Chapter 3, Exercise 5 - Continues *****
```

```
# cat("\n1) Histogram of samplingDistribution (normal distribution) is:\n")
```

```
hist(samplingDistribution)
abline(v=quantile(samplingDistribution,prob=0.025),col="red")
abline(v=quantile(samplingDistribution,prob=0.975),col="blue")
abline(v=quantile(samplingDistribution,prob=0.50),col="green")
```

## Histogram of samplingDistribution



```
cat("\n***** Chapter 3, Exercise 5 - Ends *****\n")
```

```
##
## ***** Chapter 3, Exercise 5 - Ends *****
```

## Chapter 3, Exercise 6

*If you did Exercise 4, you calculated some quantiles for a distribution of raw data. If you did Exercise 5, you calculated some quantiles for a sampling distribution of means. Briefly describe, from a conceptual perspective and in your own words, what the difference is between a distribution of raw data and a distribution of sampling means (2 pts). Finally, comment on why the 2.5% and 97.5% quantiles are so different between the raw data distribution and the sampling distribution of means (2 pts).*

### 1) Why distributions differ - mainly because of the distribution of raw data and distribution of sample population of its mean values.

- `hist(myTreeGirth)` represents an uniform frequency distribution of values across 31 observations. Shape has consistent high and low bars across. Values 10 to 12 ranges in high frequencies while the ~18 having the least frequency.
- `hist(samplingDistribution)` - plots a normal distribution of sample means with bell-shaped curve having symmetric curves on each side of the median.

Difference in shape of these datasets is due to the fact that the `samplingDistribution` dataset is With 10,000 trials of sample means of Sample size 7 with replacement. Even though there is high and low values in the sample means; however, with 10,000 values of sample means the distribution is centered across value almost matching the mean of `myTreeGirth` with 31 observations. Thus proving the law of large numbers rule.

### 2) Why quantiles differ - mainly because of the distribution of raw data and distribution of sample population of its mean values.

- quantiles: `quantile(samplingDistribution,prob=0.025)` and `quantile(samplingDistribution,prob=0.975)` function returns the 2.5% and 97.5% quantiles of the distribution
  - 2.5% Quantile of the `samplingDistribution` is : 11.07143 | represents low end of the left tail of the normal distribution curve.
  - 97.5% Quantile of the `samplingDistribution` is : 15.58571 | represents the low end of right tail of the normal distribution curve.
  - summary `samplingDistribution` is:
    - Min. 1st Qu. Median Mean 3rd Qu. Max.
    - 9.629 12.429 13.200 13.253 14.057 17.257
  - quantiles: `quantile(myTreeGirth,prob=0.025)` and `quantile(myTreeGirth,prob=0.975)` function returns the 2.5% and 97.5% quantiles of the distribution
    - 2.5% Quantile of the raw data is : 8.525 | represents low end of the left tail of the raw data/frequency distribution curve.
    - 97.5% Quantile of the raw data is : 18.65 | represents the low end of right tail of the raw data/frequency distribution curve.
  - Summary of 1st variable `Girth` is:
    - Min. 1st Qu. Median Mean 3rd Qu. Max.
    - 8.30 11.05 12.90 13.25 15.25 20.60
  - The Min and max values from both distribution is different because when we take mean of 7 values with replacement for 10,000 trials converges the values to its center value eliminating the lowest and highest values.

Please find more details from the R code below,

```
cat("\n***** Chapter 3, Exercise 6 - Begins *****\n")

##
## ***** Chapter 3, Exercise 6 - Begins *****

cat("\n1) Comparison by Mean values:")

##
## 1) Comparison by Mean values:

cat("\nmean of the raw data is:",mean(myTreeGirth))

##
## mean of the raw data is: 13.24839

cat("\nmean of the samplingDistribution is:", mean(samplingDistribution))

##
## mean of the samplingDistribution is: 13.25298

cat("\n-----\n")

##
## -----

cat("\n2) Comparison by Quantiles:\n")

##
## 2) Comparison by Quantiles:

rawQ2.5 <- quantile(myTreeGirth, prob=0.025) #
rawQ97.5 <- quantile(myTreeGirth, prob=0.975) #
cat(" 2.5% Quantile of the raw data is :",rawQ2.5,"\n 97.5% Quantile of the raw data is :",rawQ97.5)

## 2.5% Quantile of the raw data is : 8.525
## 97.5% Quantile of the raw data is : 18.65

cat("\n-----\n")

##
## -----
```

```
samplingQ2.5 <- quantile(samplingDistribution, prob=0.025) #
samplingQ97.5 <- quantile(samplingDistribution, prob=0.975) #
cat(" 2.5% Quantile of the samplingDistribution is :",samplingQ2.5,"\n 97.5% Quantile of the
    samplingDistribution is :",samplingQ97.5)

## 2.5% Quantile of the samplingDistribution is : 11.07143
## 97.5% Quantile of the
## samplingDistribution is : 15.58571

cat("\n-----\n")

##
## -----

cat("\n***** Chapter 3, Exercise 6 - Ends *****\n")

##
## ***** Chapter 3, Exercise 6 - Ends *****
```

## Chapter 3, Exercise 7

*Redo Exercise 5, but this time use a sample size of  $n = 70$  instead of the original sample size of  $n = 7$  used in Exercise 5. (1 pt) Explain why the 2.5% and 97.5% quantiles are different from the results you got for Exercise 5 (1 pt). As a hint, be sure to specify what about a sample makes it “better.” (1 pt)*

### 1) New Sampling distribution

- `NewsamplingDistribution <- replicate(10000, mean(sampleGirth(70)))` - this code creates a new sampling distribution with  $n=70$  and assigns into vector “NewsamplingDistribution”
- `hist(NewsamplingDistribution)` - plots a normal distribution of sample means with bell-shaped curve having symmetric curves on each side of the median.

### 2) Why distributions differ?

\* mainly because of the increase in sample size of raw data and distribution of sample population of it.

- `summary samplingDistribution` is:
- Min. 1st Qu. Median Mean 3rd Qu. Max. \*9.629 12.429 13.200 13.253 14.057 17.257
- `summary NewsamplingDistribution` is:
- Min. 1st Qu. Median Mean 3rd Qu. Max.
- 11.75 12.99 13.24 13.25 13.49 14.66
- The Min and max values from both distribution is evident that the change (increase) in sample size converges the values to its center value eliminating the lowest and highest values.

### 3) Why quantiles differ -

- As above case, summary of the both distribution values it evident that the lower and upper bound are adjusted and centered towards the mean; thus, the significant increase in 2.5%, 25% 50% values and significant decrease in 3rd and 4th quartiles. Hence, proving the central limit theorem. Central region that constitutes 95% of the observations is densely populated around mean and 50th percentile.
- quantiles: `quantile(samplingDistribution,prob=0.025)` and `quantile(samplingDistribution,prob=0.975)` function returns the 2.5% and 97.5% quantiles of the distribution
  - 2.5% Quantile of the `samplingDistribution` is : 11.07143 | represents low end of the left tail of the normal distribution curve.
  - 97.5% Quantile of the `samplingDistribution` is : 15.58571 | represents the low end of right tail of the normal distribution curve.
- quantiles: `quantile(NewsamplingDistribution,prob=0.025)` and `quantile(NewsamplingDistribution,prob=0.975)` function returns the 2.5% and 97.5% quantiles of the distribution
  - 2.5% Quantile of the `NewsamplingDistribution` is : 12.53143
  - 97.5% Quantile of the `NewsamplingDistribution` is : 13.99857

### 4) What makes sample better - I think couple of things makes a sample better.

1. Repetitious sampling with replacement ( sampling with replacement) - guarantees the next draw has equal probability
2. Law of large Numbers - sampling distribution with more trials matches the mean of the raw distribution
3. varying sample size with replacement ( when the sample size is increased) also converges the central limit theorem

Please find more details from the R code below,

```
cat("\n***** Chapter 3, Exercise 7 - Begins *****\n")

##
## ***** Chapter 3, Exercise 7 - Begins *****

set.seed(772) # Control randomization

cat("\n1) New Sampling distribution with n=70:\n") #

##
## 1) New Sampling distribution with n=70:

#Store the sampling distribution to a variable with 10,000 means with sample size of 70
#Store the sampling distribution to a variable
NewsamplingDistribution <- replicate(10000, mean(sampleGirth(70)))

cat("\nmean of the myTreeGirth (raw data) is:",mean(myTreeGirth))
```

```
##
## mean of the myTreeGirth (raw data) is: 13.24839

cat("\nmean of the samplingDistribution is:", mean(samplingDistribution))

##
## mean of the samplingDistribution is: 13.25298

cat("\nmean of the NewsamplingDistribution is:", mean(NewsamplingDistribution))

##
## mean of the NewsamplingDistribution is: 13.2451

summary(NewsamplingDistribution)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      11.75  12.99   13.24   13.25   13.49   14.66

cat("\n-----\n")

##
## -----

cat("\n***** Chapter 3, Exercise 7 - Continues *****\n")

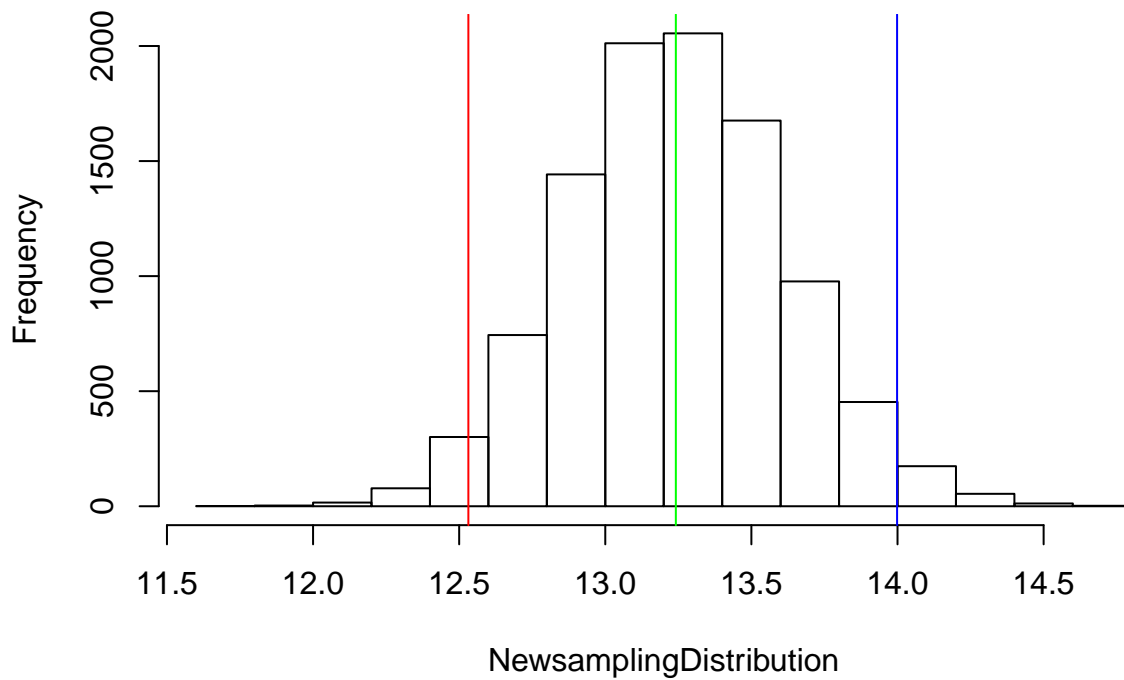
##
## ***** Chapter 3, Exercise 7 - Continues *****

# cat("\n1) Histogram of NewsamplingDistribution is:\n")

hist(NewsamplingDistribution)
abline(v=quantile(NewsamplingDistribution,prob=0.025),col="red")
abline(v=quantile(NewsamplingDistribution,prob=0.975),col="blue")
abline(v=quantile(NewsamplingDistribution,prob=0.50),col="green")
```



## Histogram of NewsamplingDistribution



```
cat("\n2) Comparison by Quantiles:\n")
```

```
##
## 2) Comparison by Quantiles:
```

```
samplingQ2.5 <- quantile(samplingDistribution, prob=0.025)
samplingQ97.5 <- quantile(samplingDistribution, prob=0.975)
cat(" 2.5% Quantile of the samplingDistribution is :",samplingQ2.5,"\n 97.5% Quantile of the
samplingDistribution is :",samplingQ97.5)
```

```
## 2.5% Quantile of the samplingDistribution is : 11.07143
## 97.5% Quantile of the
## samplingDistribution is : 15.58571
```

```
cat("\n-----\n")
```

```
##
## -----
```

```
NewsamplingQ2.5 <- quantile(NewsamplingDistribution, prob=0.025)
NewsamplingQ97.5 <- quantile(NewsamplingDistribution, prob=0.975)
cat(" 2.5% Quantile of the NewsamplingDistribution is :",NewsamplingQ2.5,"\n 97.5% Quantile of
the NewsamplingDistribution is :",NewsamplingQ97.5)
```

```
## 2.5% Quantile of the NewsamplingDistribution is : 12.53143
## 97.5% Quantile of
## the NewsamplingDistribution is : 13.99857
```

```
cat("\n-----\n")
```

```
##
```

```
## -----
```

```
cat("\n***** Chapter 3, Exercise 7 - Ends *****\n")
```

```
##
```

```
## ***** Chapter 3, Exercise 7 - Ends *****
```