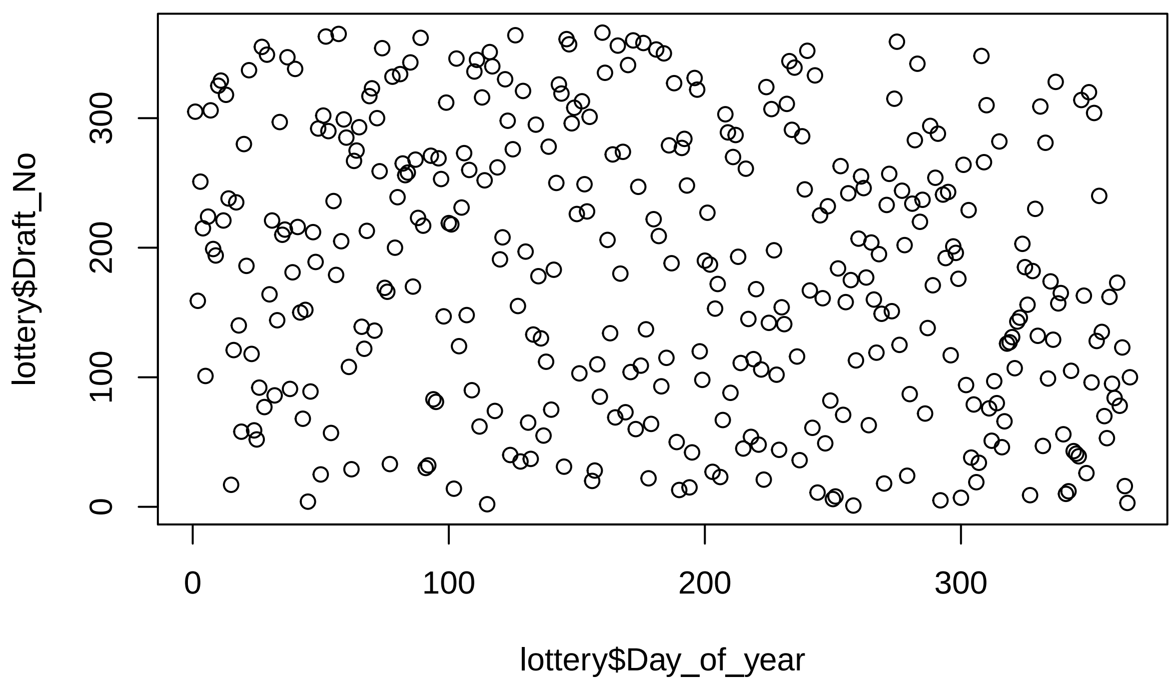
Bootstrap,Jacknife & CI

Sreenand Sasikumar

01/03/2020

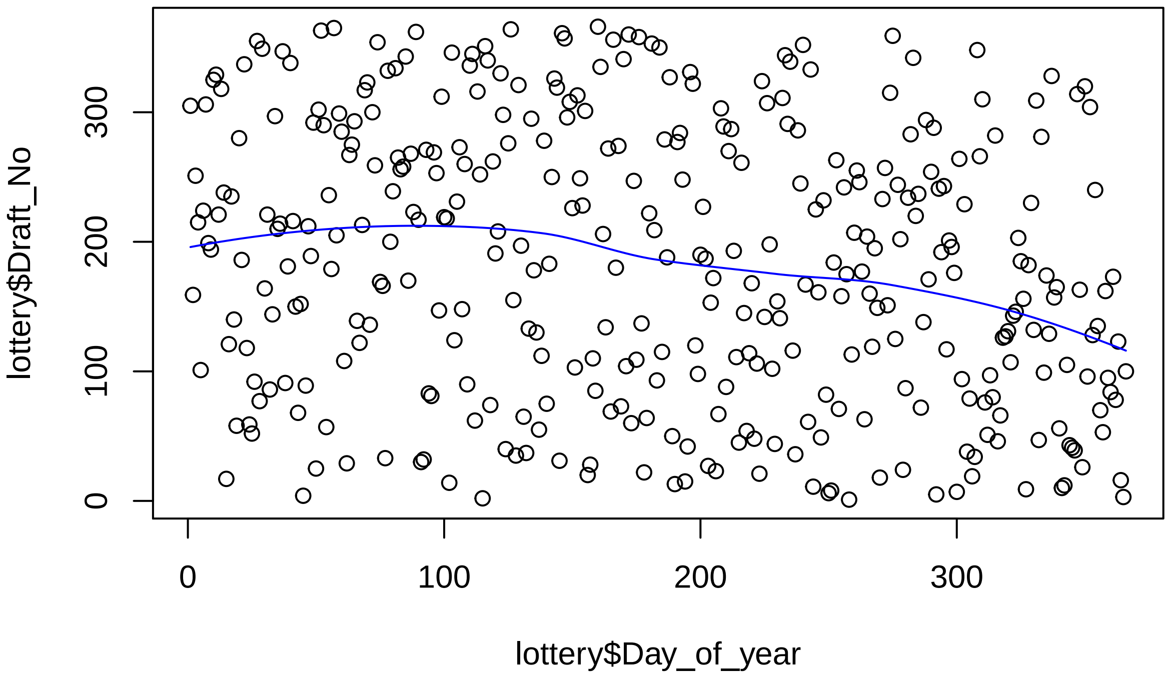
Question 1: Hypothesis testing

1.1



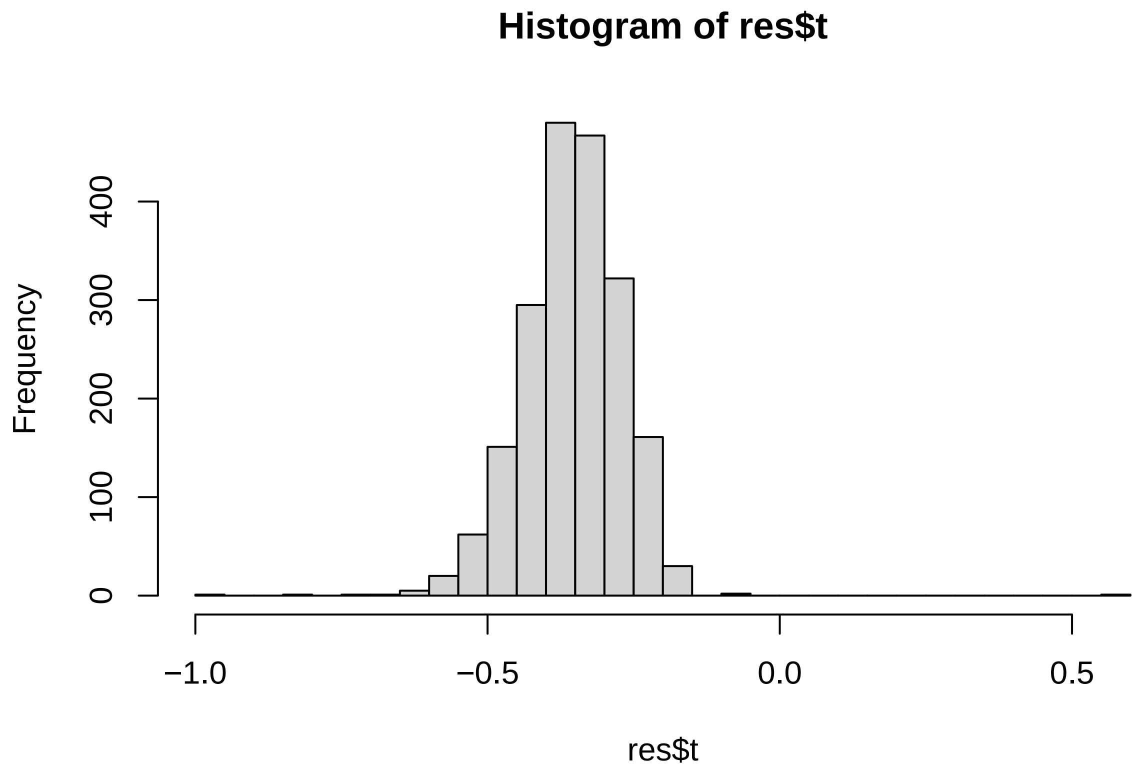
From the plot it is evident that, the lottery looks random as there is no specific pattern or shape.

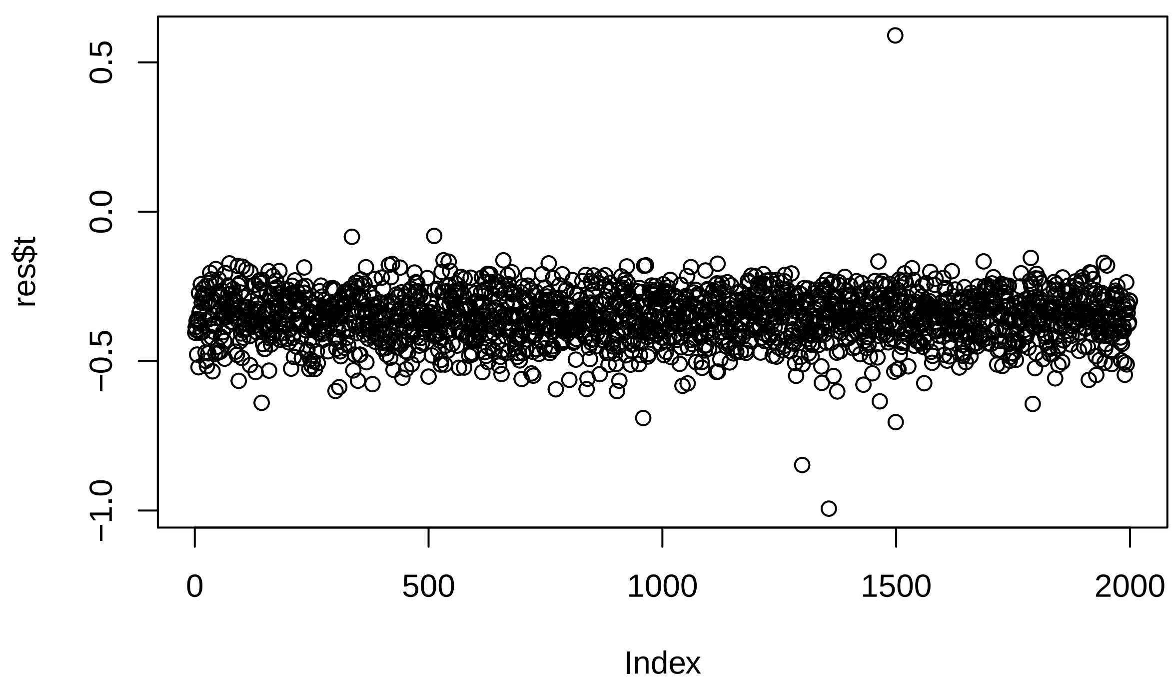
1.2



From the plot we can see that the Y_cap in blue color seems to follow a downward pattern which is negative.At this stage the lottery cannot be proclaimed as random.

1.3


Histogram of res$t
Frequency
0
100
200
300
400
−1.0
−0.5
0.0
0.5
res$t

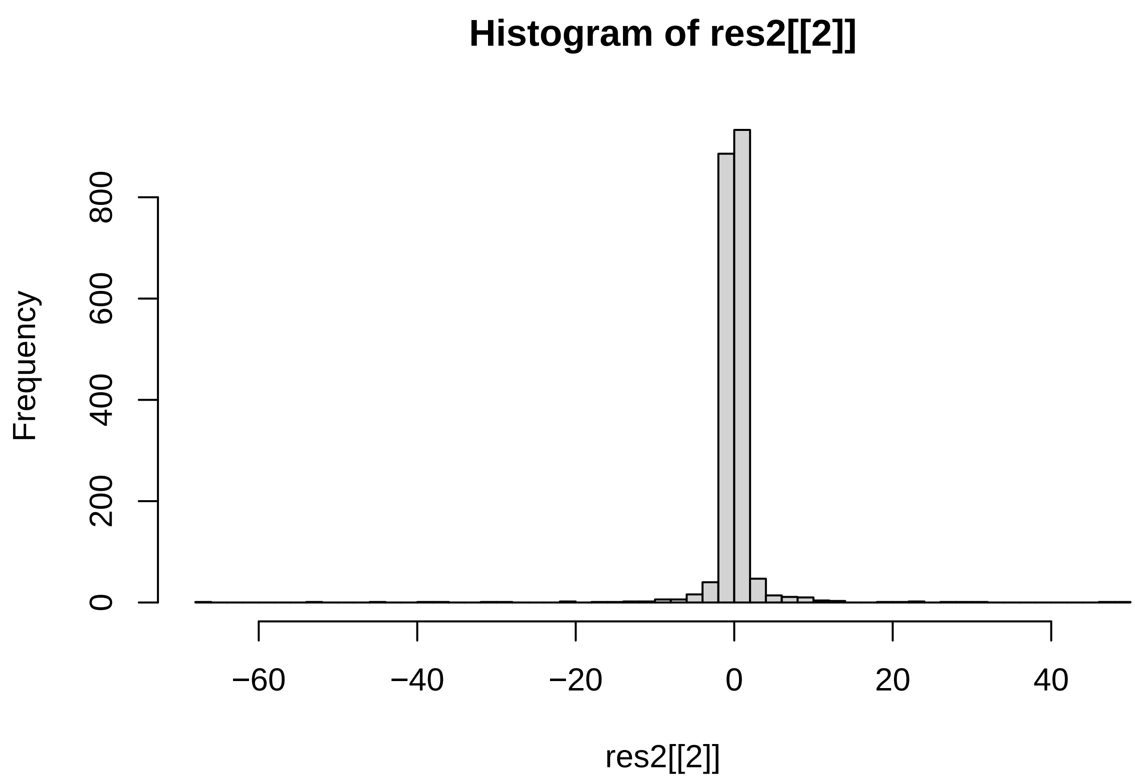

res$t
−1.0
−0.5
0.0
0.5
0
500
1000
1500
2000
Index

```
## The pvalue of the test is found to be :
## [1] 5e-04
```

From the histogram it is quite evident that most of the test values are less than 0 and also obtained pvalue is also less than 0.05, hence we reject the Null hypothesis and conclude that the data is not random.

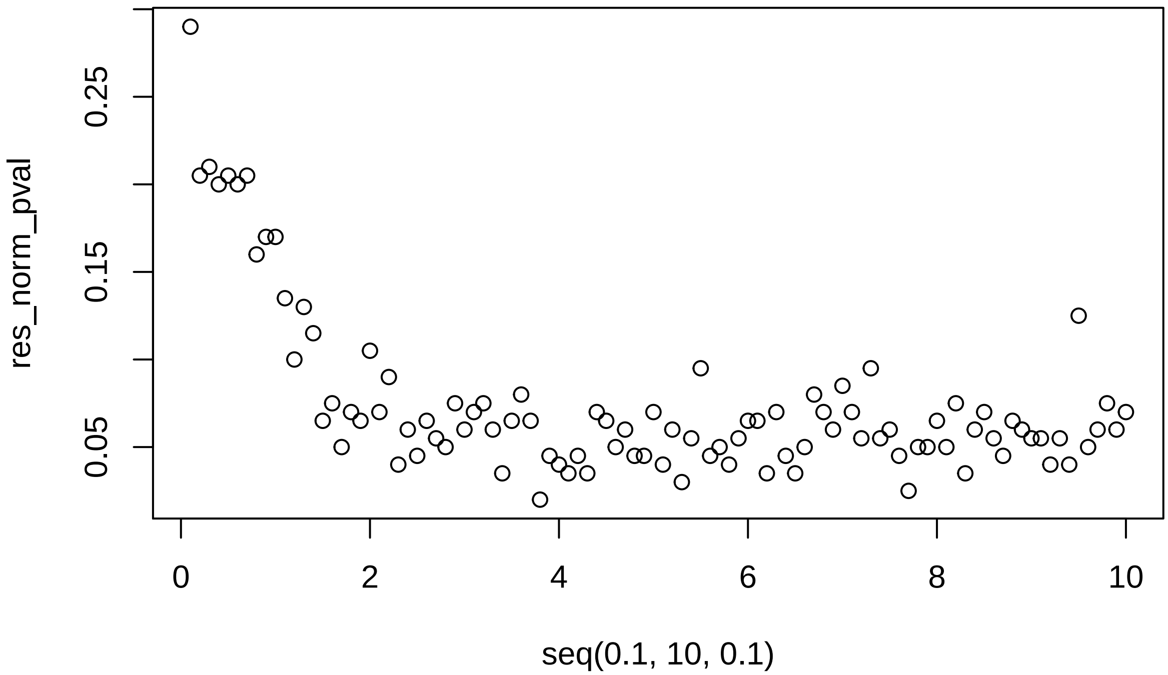
1.4

```
## The pvalue obtained by using permutation test is found to be :
## [1] 0.469
```



Since the obtained Pvalue from Permutation test is greater than 0.05, we cannot reject the Null hypothesis in favour of Alternate hypothesis. Hence we can conclude that we cannot deny that Lottery is random which is contradictory to previous tests.

1.5

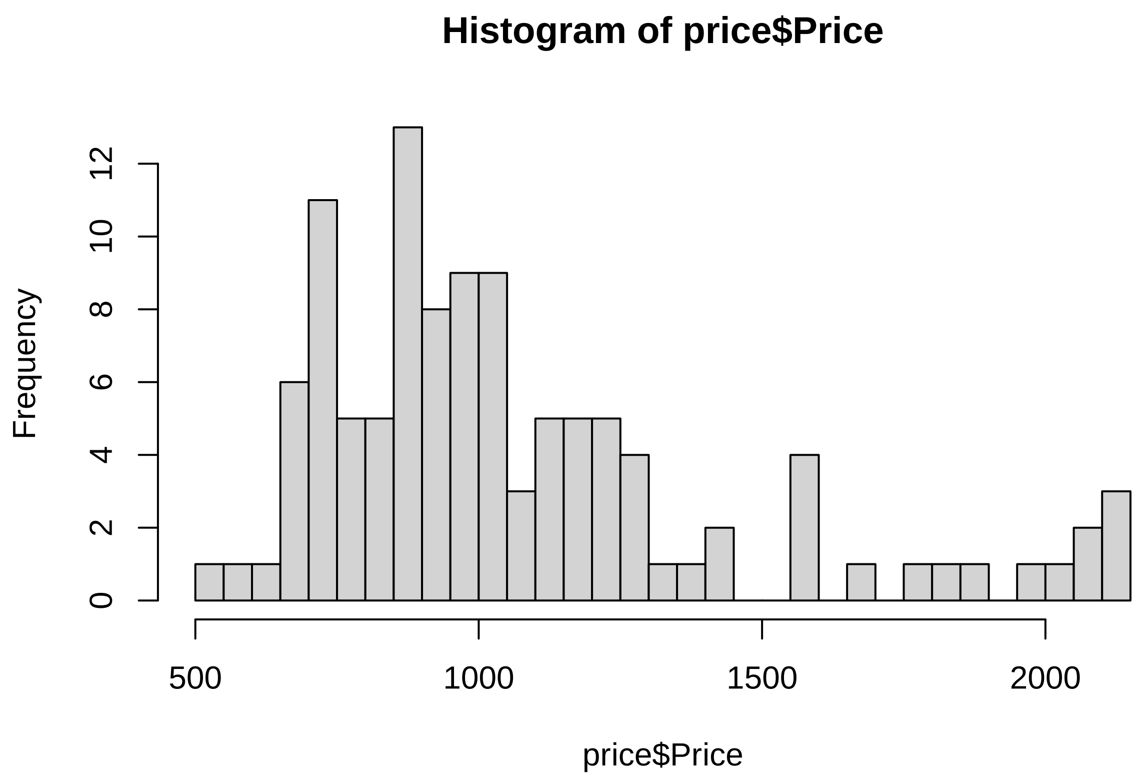



```
## After computing Test statistic for different values of alpha ranging from (0.1 to 10), we obtained
##  24  rejected values, that is pvalue < 0.05. Hence the Crude estimate of the power constructed is cal
```

From the obtained power value we can conclude that the Quality of the fit is not upto the mark, since we randomizing only one variable.

Question 2: Bootstrap, jackknife and confidence intervals

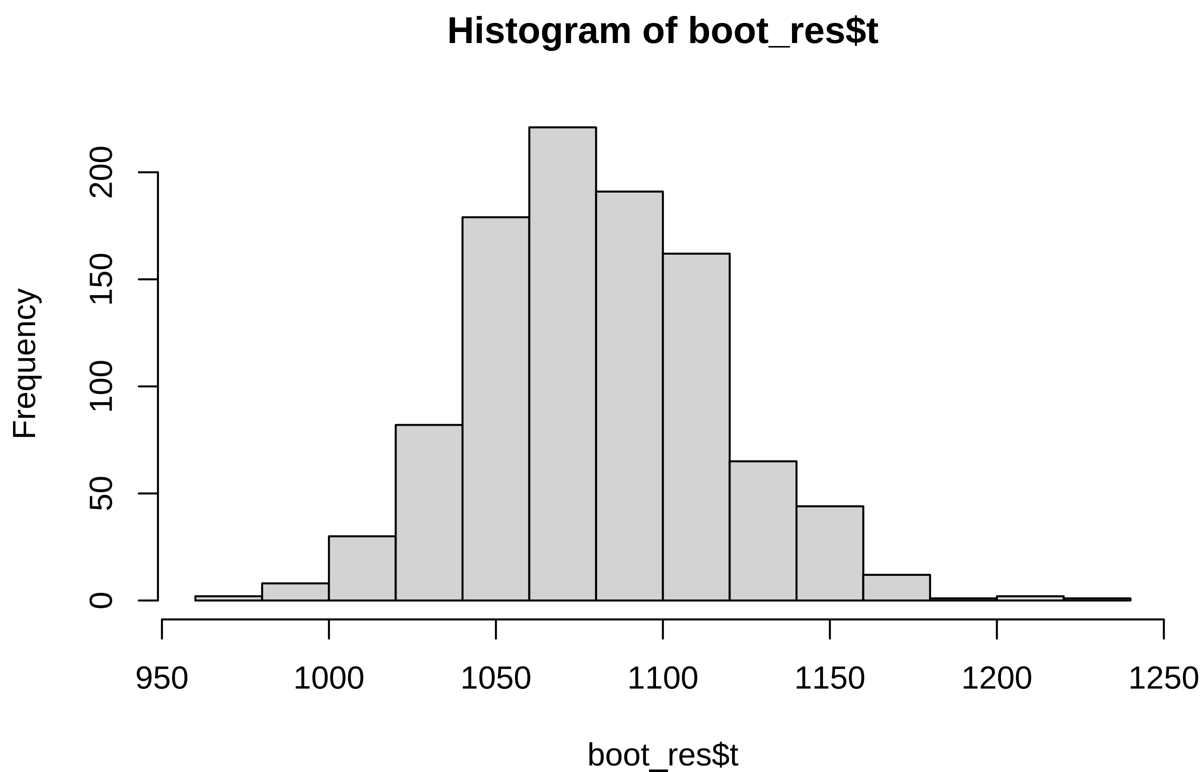
2.1




```
## The mean price is found to be:
## [1] 1080.473
```

From the above Histogram plot, we can see that the plot resembles a Right skewed Normal Distribution.

2.2




```
## The Variance of the mean price is found to be:

## [1] 1304.338

## The Bias-Correction of the mean price is found to be:

## [1] 1080.789

## The 95% Confidence Interval for the mean price using Bootstrap Percentile, Bootstrap BCa
##  and First order Normal approximation are found to be:

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot_res, conf = 0.95, type = c("perc", "bca",
##     "norm"))
##
## Intervals :
## Level      Normal              Percentile            BCa
## 95%   (1010, 1152 )   (1014, 1155 )   (1020, 1161 )
## Calculations and Intervals on Original Scale
```

From the histogram plotted, we can assume that the mean price could be a Normal distribution. Since this is non paramteric bootstrap we cannot be certain about the distribution.

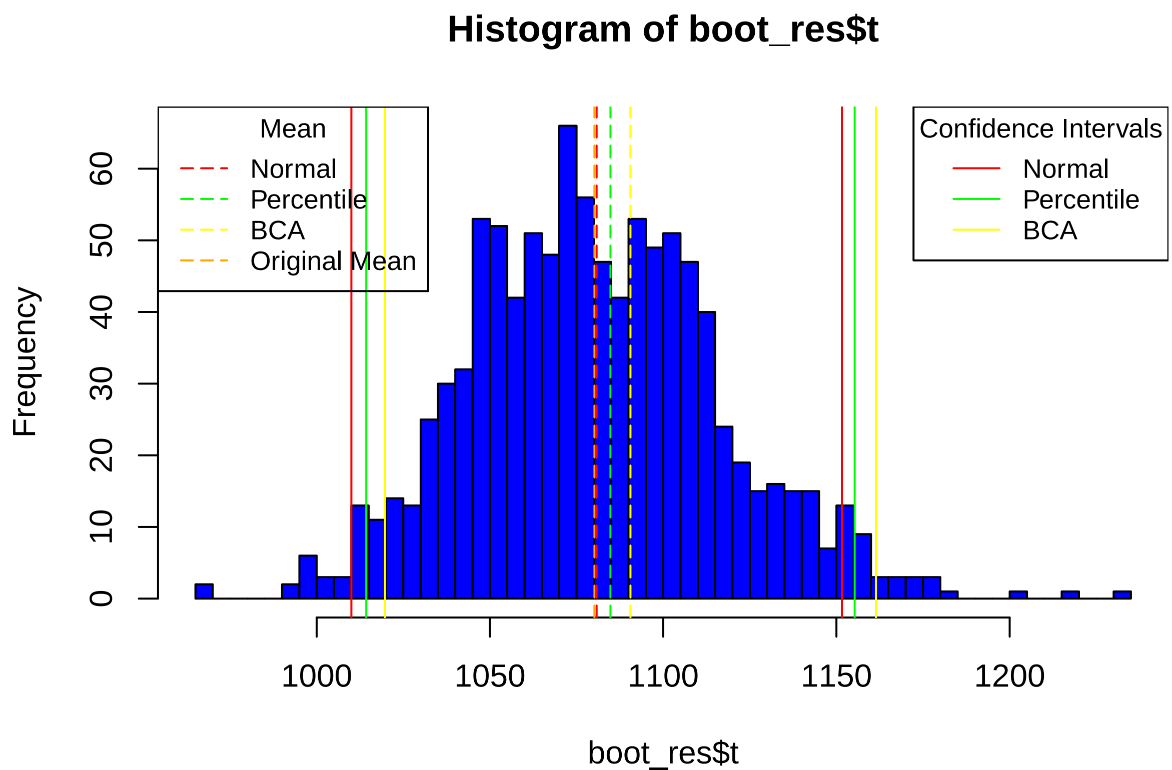
2.3

```
## The estimated Variance of mean price using Jackknife method is found to be:

## [1] 1320.911

## The Variance obtained using Bootstrap method is found to be  1304.338
##  and the Variance obtained using Jackknife method is found to be  1320.911 .
## We can see that Jackknife variance is higher than Bootstrap variance.
```

2.4



From the plot, we can see that the length of Percentile method interval is less compared to other two and the rest two methods have almost same interval lengths. The Original mean and mean obtained from Normal approximation are very close to each other where as BCA mean is the farther than the rest two.

Appendix

```
knitr::opts_chunk$set(echo = TRUE,warning = FALSE)
lottery <- read.csv2(file.choose())
plot(lottery$Day_of_year,lottery$Draft_No)
loe_fit <- loess(Draft_No ~ Day_of_year, data = lottery)
Y_cap <- predict(loe_fit)
plot(lottery$Day_of_year,lottery$Draft_No)
lines(lottery$Day_of_year,Y_cap,col = "blue")
stat1 <- function(data,n)
{
  data1 <- data[n,]
```

```
  fit <- loess(Draft_No ~ Day_of_year, data = data1)
  Y_cap <- predict(fit)
  Xa <- data1$Day_of_year[which.min(Y_cap)]
  Xb <- data1$Day_of_year[which.max(Y_cap)]
  Tes_stas <- (max(Y_cap) - min(Y_cap)) / (Xb - Xa)
  return(Tes_stas)

}
library(boot)
res <- boot(lottery,stat1,2000)

hist(res$t,50)
plot(res$t)
cat("The pvalue of the test is found to be :")
mean(res$t >0)
lott <- function(data,B)
{
  fit <- loess(Draft_No ~ Day_of_year, data = data)
  Y_cap <- predict(fit)
  Xa <- data$Day_of_year[which.min(Y_cap)]
  Xb <- data$Day_of_year[which.max(Y_cap)]
  Tes_stas <- (max(Y_cap) - min(Y_cap)) / (Xb - Xa)

  perm_stas <- vector()
  n <- dim(data)[1]
  for(i in 1:B)
  {
    sam <- sample(data$Day_of_year,n)
    data2 <- data.frame("Day_of_year" = sam,"Draft_No" = data$Draft_No)
    fit_p <- loess(Draft_No ~ Day_of_year, data = data2)
    Y_cap_p <- predict(fit_p)
    Xa_p <- data2$Day_of_year[which.min(Y_cap)]
    Xb_p <- data2$Day_of_year[which.max(Y_cap)]
    perm_stas[i] <- (max(Y_cap_p) - min(Y_cap_p)) / (Xb_p - Xa_p)
  }
  pval <- mean(abs(perm_stas) > abs(Tes_stas))
  return(list("pvalue" = pval,"Permutation Statistic" = perm_stas))
}
res2 <- lott(lottery,2000)
cat("The pvalue obtained by using permutation test is found to be :")
res2[[1]]
hist(res2[[2]],50)
X <- lottery$Day_of_year
Y <- lottery$Draft_No
res_norm_pval <- vector()
k <- 0
for(i in seq(0.1,10,0.1))
{
  Y_norm_ext <- vector()
  for(j in 1:length(X))
  {
    beta <- rnorm(1,183,10)
    Y_norm_ext[j] <- max(0,min(i * X[j] + beta,366))
```

```
  }
  k <- k+1
  data4 <- data.frame("Day_of_year" = X,"Draft_No" = Y_norm_ext)
  res_norm <- lott(data4,200)
  res_norm_pval[k] <- res_norm$pvalue

}
# data4 <- data.frame("Day_of_year" = X,"Draft_No" = Y_norm_ext)
# res_norm <- lott(data4,200)
# res_norm$pvalue

rejected <- length(which(res_norm_pval < 0.05))
plot(seq(0.1,10,0.1),res_norm_pval)
crude_estimate <- rejected/length(res_norm_pval)

cat("After computing Test statistic for different values of alpha ranging from (0.1 to 10), we obtained

price <- read.csv2(file.choose())
hist(price$Price,50)
cat("The mean price is found to be:")
mean(price$Price)
bootstrap_func <- function(data,n)
{
  data1 <- data[n,]
  test_stat <- mean(data1$Price)
  return(test_stat)
}
library(boot)
boot_res <- boot(price,bootstrap_func,1000)
#boot_res$t
hist(boot_res$t)
cat("The Variance of the mean price is found to be: ")
variance <- sum((boot_res$t - mean(boot_res$t))^2) / (1000-1)
variance
cat("The Bias-Correction of the mean price is found to be: ")
bias_corr <- (2 * mean(price$Price)) - ((sum(boot_res$t))/1000)
bias_corr

conf_int <- boot.ci(boot_res,conf = 0.95,type = c("perc","bca","norm"))
cat("The 95% Confidence Interval for the mean price using Bootstrap Percentile, Bootstrap BCa  \n and F
conf_int
jack_kni <- function(data)
{
  test_stas <- vector()
  n <- length(data$Price)
  for(i in 1:n)
  {
    samp_data <- data[-i,]
    test_stas[i] <- mean(samp_data$Price)
  }
  return(test_stas)
}
n <- length(price$Price)
```

```
jac_res <- jack_kni(price)
Ti_star <- (n * mean(price$Price)) - ((n-1) * jac_res)
jack_var <- sum((Ti_star - mean(Ti_star))^2) / (n * (n-1))
cat("The estimated Variance of mean price using Jackknife method is found to be: ")
jack_var

cat("The Variance obtained using Bootstrap method is found to be ",variance,"\n and the Variance obtain
range_Normal <- conf_int$normal[c(2,3)]
range_percentile <- conf_int$percent[c(4,5)]
range_bca <- conf_int$bca[c(4,5)]
hist(boot_res$t,50,col = "blue")
abline(v=range_Normal,col = "red")
abline(v=range_percentile,col = "green")
abline(v=range_bca,col = "yellow")
legend("topright", legend=c("Normal", "Percentile","BCA"),lty =1, col=c("red", "green","yellow"), cex=0
abline(v=mean(boot_res$t),col = "orange",lty = 5)
abline(v=mean(range_Normal),col = "red",lty = 5)
abline(v=mean(range_percentile),col = "green",lty = 5)
abline(v=mean(range_bca),col = "yellow",lty = 5)
legend("topleft", legend=c("Normal", "Percentile","BCA","Original Mean"),lty =5, col=c("red", "green","
```