CS Lab45

Suhani Ariga

12/2/2020

Library Setup

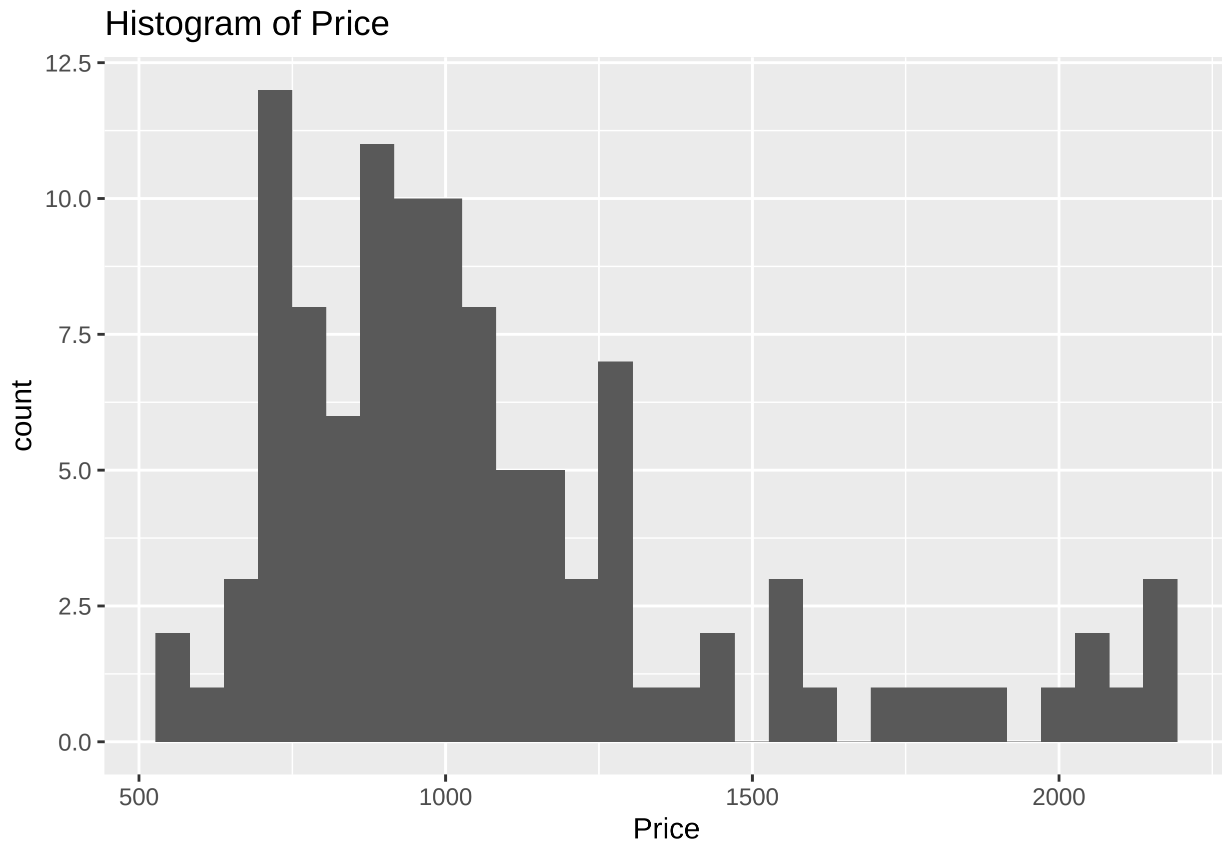
```
library(readxl)
library(ggplot2)
library(boot)
```

Assignment 2

1. Plot the histogram of Price

```
data = read_xls("prices1.xls")

ggplot(data = data, aes(Price)) +
  ggtitle("Histogram of Price") +
  geom_histogram(bins = 30)
```



It reminds us of Gamma distribution as it is skewed towards the right.

```
cat(paste("Mean price is:", mean(data$Price)))
```

```
## Mean price is: 1080.47272727273
```

2. Bootstrap

Bias corrected mean estimate

$$\hat{T} = 2T(D) - \frac{1}{B}\sum_{i=1}^{B} T(D_i^*)$$

Variance

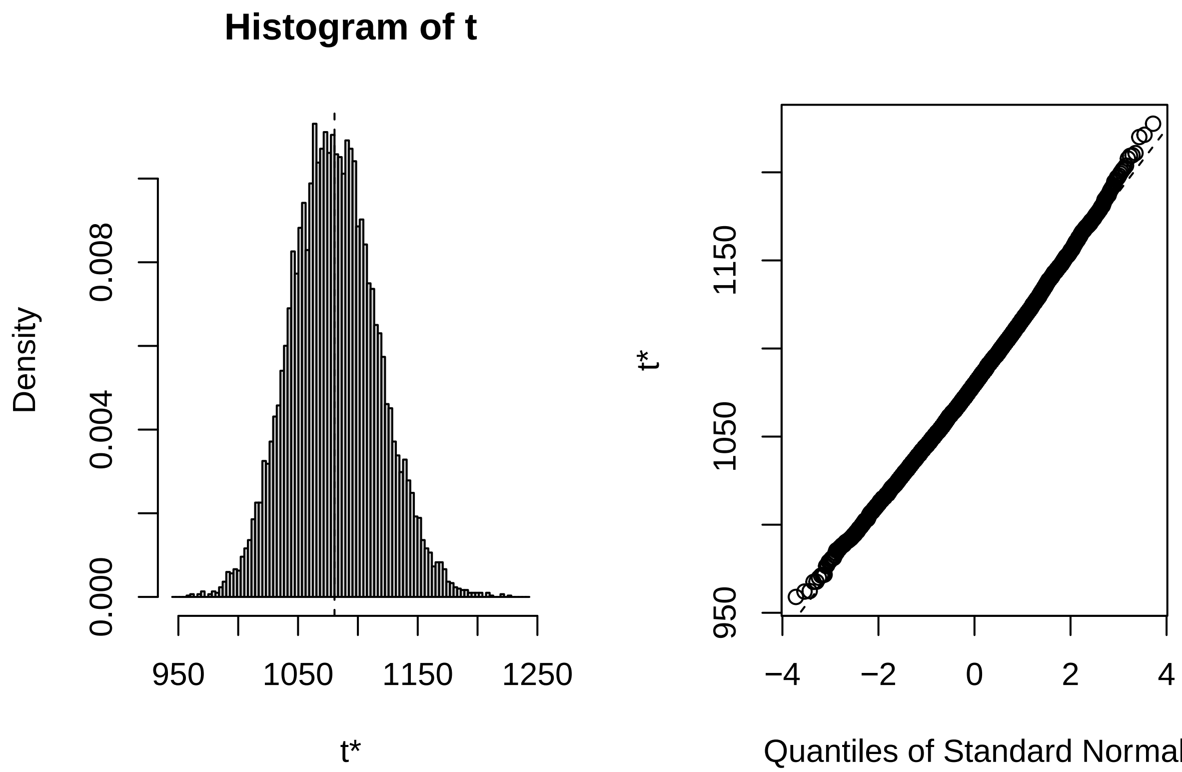
$$Var\hat{[T(.)]} = \frac{1}{B-1}\sum_{i=1}^{B}\Big(T(D_i^*) - \overline{T(D^*)}\Big)^2$$

where B is the number of bootstrap samples, $T(D_i^*)$ the statistic(mean) for each sample and $\overline{T(D^*)}$ is the mean of all the values after using the statistic for each sample.

```
boot_mean = function(data,index){
  return(mean(data[index]))
}
```

```r
set.seed(12345)
boot_obj = boot(data$Price, boot_mean, R = 10000)

plot(boot_obj)
```




```r
# bias correction estimator
bias_cor = 2*mean(data$Price) - mean(boot_obj$t)
cat(paste(" The bootstrap bias-correction is:", bias_cor))
```

```
##  The bootstrap bias-correction is: 1080.21334
```

```r
#variance of estimator
var_est = sum((boot_obj$t-mean(data$Price))^2)/(nrow(boot_obj$t)-1)
cat(paste("The variance of the mean price is :",var_est))
```

```
## The variance of the mean price is : 1295.85514282007
```

```r
# 95% confidence interval for the mean price using bootstrap percentile, bootstrap BCa, and first-order

CI = boot.ci(boot_obj,type = c("norm","perc", "bca"))
CI
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
```

```
## Based on 10000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot_obj, type = c("norm", "perc", "bca"))
##
## Intervals :
## Level      Normal              Percentile            BCa
## 95%   (1010, 1151 )   (1013, 1153 )   (1017, 1159 )
## Calculations and Intervals on Original Scale
```

3. Jackknife

Variance

$$Var[\hat{T(.)}] = \frac{1}{n(n-1)}\sum_{i=1}^{n}(T_i^* - J(T))^2$$

where,

$$T_i^* = nT(D) - (n-1)T(D_i^i),$$

$$J(T) = \frac{1}{n}\sum_{i=1}^{n}T_i^*$$

```
n = nrow(data)
Ti = c()
for (i in 1:n) {
  Ti[i] = n*mean(data$Price) - (n-1)*mean(data$Price[-i])
}

J = (1/n) * sum(Ti)

#  variance of the mean price
Var_jackknife = sum((Ti - J)^2)/(n*(n-1))
cat(paste("Variance of the mean price is", Var_jackknife))
```

```
## Variance of the mean price is 1320.91104405187
```

The variance of the mean price using the jackknife is 1320.911 and the variance of the mean price got from the bootstrap is 1295.855. The difference between the two estimates is 25.056, which is small. Considering the tendency that Jackknife overestimate variance, the difference seems reasonable.

4. Compare the Confidence Intervals

```
Normal = c(CI$normal[2], CI$normal[3], CI$normal[3] - CI$normal[2], ((CI$normal[3] + CI$normal[2])/2))
Percentile = c(CI$percent[4], CI$percent[5], CI$percent[5] - CI$percent[4], ((CI$percent[5] + CI$percen
BCa = c(CI$bca[4], CI$bca[5], CI$bca[5] - CI$bca[4], ((CI$bca[5] + CI$bca[4])/2))
combine = rbind(Normal, Percentile, BCa)

colnames(combine) = c("From", "To", "Length", "Mean")
knitr::kable(combine, caption = "Compare the confidence intervals")
```

Table 1: Compare the confidence intervals

	From	To	Length	Mean
Normal	1009.660	1150.766	141.1059	1080.213
Percentile	1013.455	1153.490	140.0348	1083.473
BCa	1016.601	1159.000	142.3992	1087.800