Pstat 105 Lab2

Kendall Brown 8564403

Fall 2017

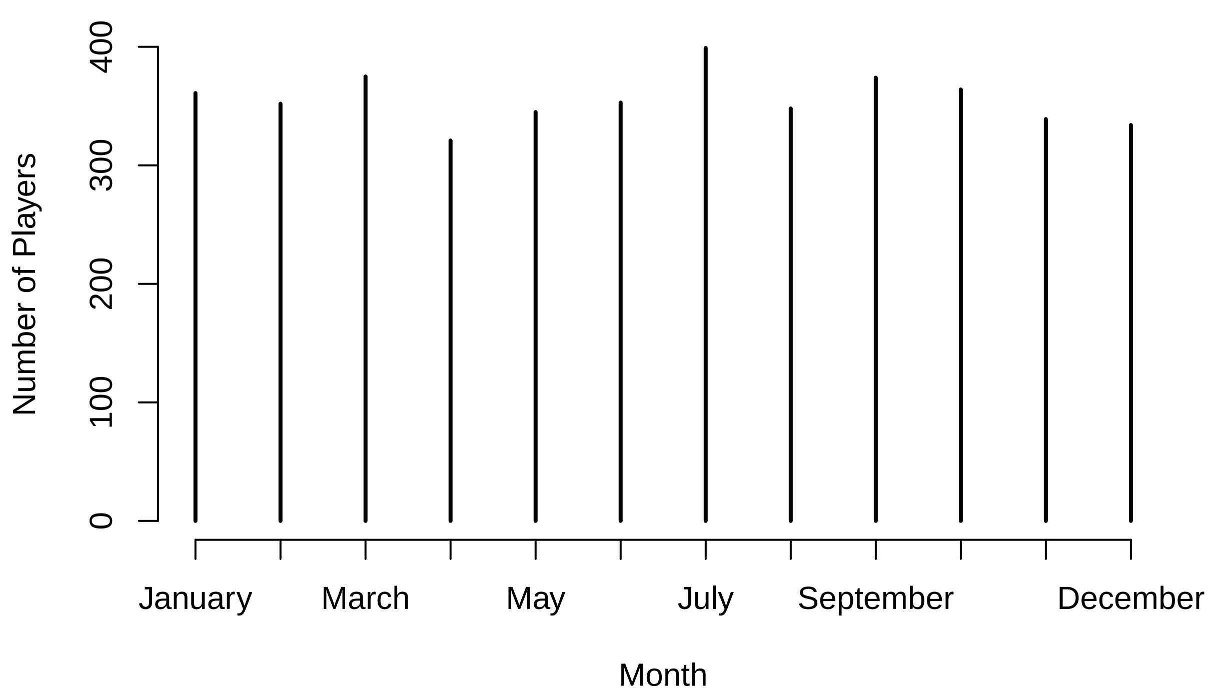
Q1a. Table for all athletes in the data set. Ho:Professional basketball players are born similarly to the general population. Ha:Professional basketball players are born during a certain time of year. Level=5%

```
Bdays <- read.table("BBallBDays.txt")
Bday.month <- Bdays$Month
Bday.year <- Bdays$Year
Month.names=c("January","February","March","April","May","June","July",
"August","September","October","November","December")
Bday.month=factor(Bday.month,levels =Month.names,ordered = TRUE)
```

```
mt=table(Bday.month)
mt
```

```
## Bday.month
##   January  February     March     April       May      June      July
##       361       352       375       321       345       353       399
##    August September   October  November  December
##       348       374       364       339       334
```

```
plot(mt,xlab="Month",ylab = "Number of Players")
```



Q1b.Chi-Squared test for all athletes. Ho: P(Jan)=P(Feb)=P(Mar)=…=P(Dec). Ha: Each month has a different probability of birthing a professional basketball player.

```
MEL=rep(1,12)/12
chisq.test(mt,p=MEL)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  mt
## X-squared = 13.581, df = 11, p-value = 0.2571
```

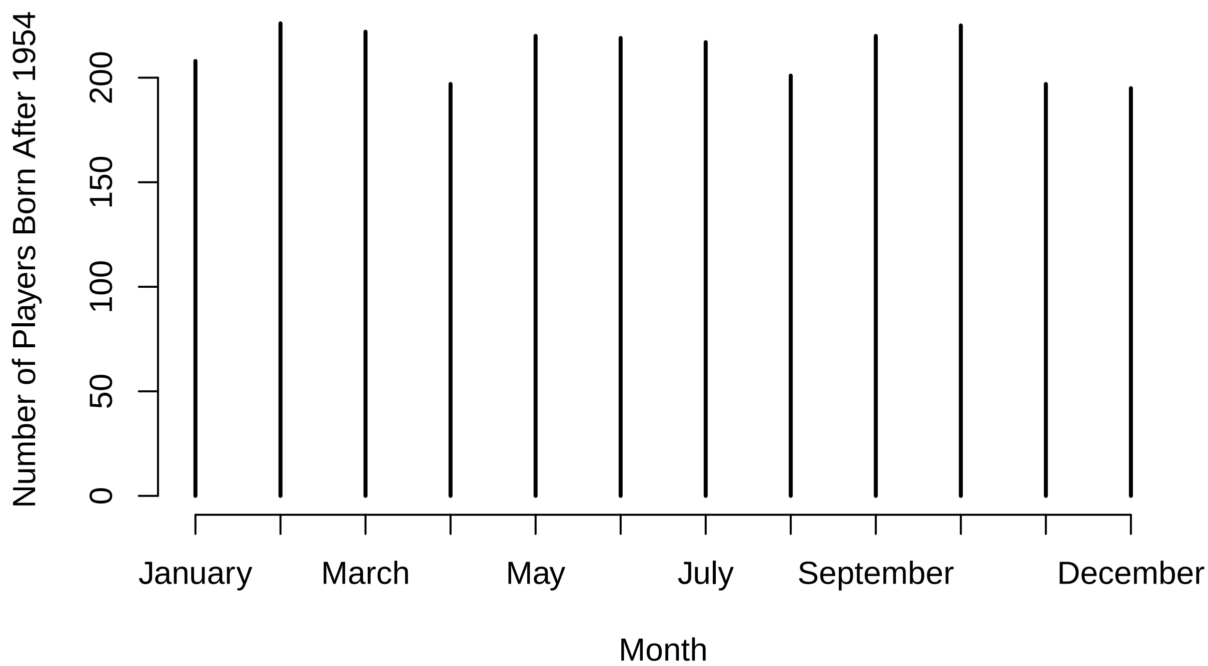
We observe a p-value of ~.26, indicating that there is not enough evidence to reject the null hypothesis at a 5% level. There is not enough evidence to suggest that basketball players are not equaly likely to be born in each month.

Q2c.Table, Plot, and Chi-Squared test and table for athletes born after 1954. Ho: For year>1954, P(Jan)=P(Feb)=P(Mar)=…=P(Dec). Ha: Each month has a different probability of birthing a professional basketball player.

```
Bday.1955=subset(Bdays,Year>1954)
Bday.month.1955=Bday.1955$Month
Bday.month.1955=factor(Bday.month.1955,levels = Month.names,ordered = TRUE)
mt1955=table(Bday.month.1955)
mt1955
```

```
## Bday.month.1955
##   January  February     March     April       May      June      July 
##       208       226       222       197       220       219       217 
##    August September   October  November  December 
##       201       220       225       197       195
```

```
plot(mt1955,xlab="Month",ylab="Number of Players Born After 1954")
```




```
chisq.test(mt1955,p=MEL)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  mt1955
## X-squared = 7.2662, df = 11, p-value = 0.7771
```

We observe a p-value of ~.78, indicating that there is not enough evidence to reject the null hypothesis at a 5% level. There is strong evidence to suggest that professional basketball players born after 1954 are equally likely to be born in each month

Q1d. Chi-Squared Test considering days per month. Ho:Probabilty a professional basketball player is born in a certain monthis proportional to the number of days in that month. Ha:Probability of a professional basketball player being born in a given month is independent of the number of days in that month.

```
PPM=summary(Bday.month.1955)
DPM=c(31,28.25,31,30,31,30,31,31,30,31,30,31)
PDPM=DPM/365.25
chisq.test(PPM,p=PDPM)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  PPM
## X-squared = 10.746, df = 11, p-value = 0.4648
```

We observe a p-value of ~.46, indicating that there is not enough evidence to reject the null hypothesis at a 5% level. Professional basketball players appear to have a birthday distribution proportional to the average number of days in a given month over the average number of days in a given year.

Q1e. Chi-Squared test with CDC date. Ho:Professional basketball players have birth months according to the CDC data. Ha:Professional basketball players have birth months dissimilar to the CDC data.

```
EPPMCDC=c(.0815,.0752,.0837,.0816,.0859,.0813,.0883,.0892,.0866,.0849,.0787,.0830)
chisq.test(x=PPM,p=EPPMCDC,rescale.p = TRUE)
```

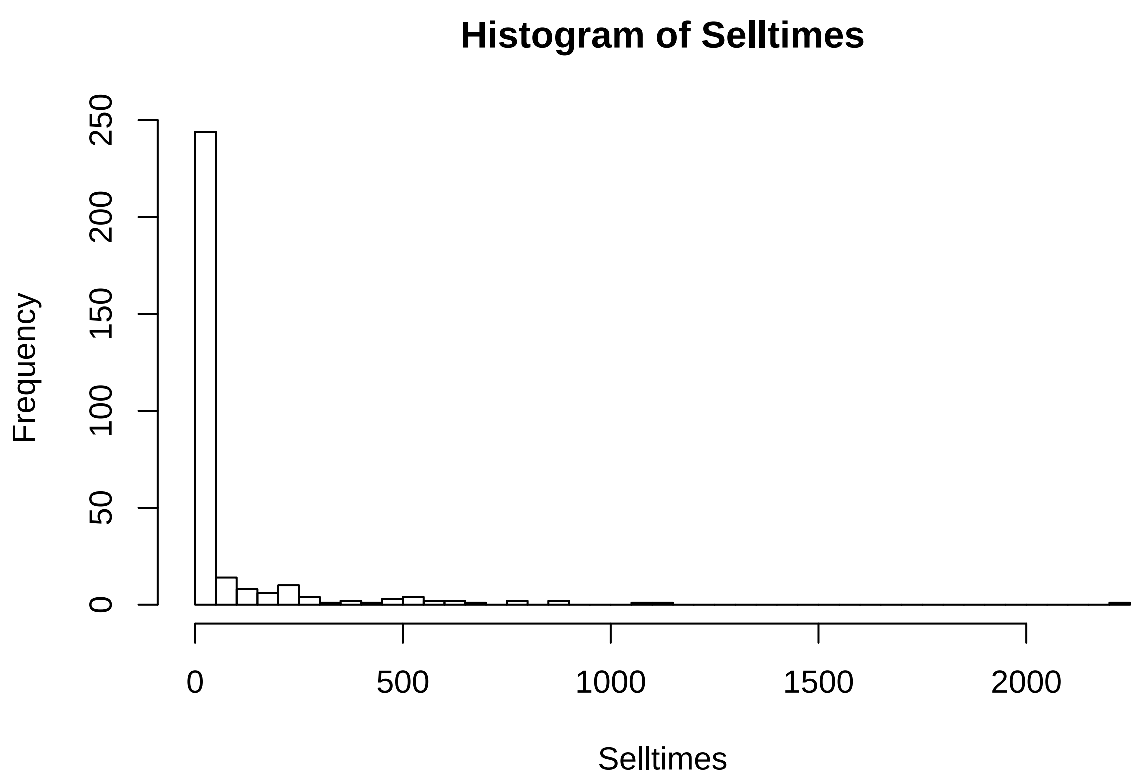
```
##
##  Chi-squared test for given probabilities
##
## data:  PPM
## X-squared = 12.81, df = 11, p-value = 0.3059
```

We observe a p-value of ~.3, indicating that there is not enough evidence to reject the null hypothesis at a 5% level. Professional basketball players appear to have a birthday distribution similar to the CDC estimates.

Q1f. From the tests conducted, we fail to reject the null hypothesis at a 5% level. Professional basketball players appear to share a similar birthday distribution to the general population.

Q2a. Histogram of Selltimes dataset.

```
selltimes <- scan("Selltimes.txt")
hist.sell=hist((selltimes),breaks=35, main="Histogram of Selltimes",xlab = "Selltimes")
```



Q2b. MLE of Lambda

```
lh=1/(mean(selltimes))
lh
```

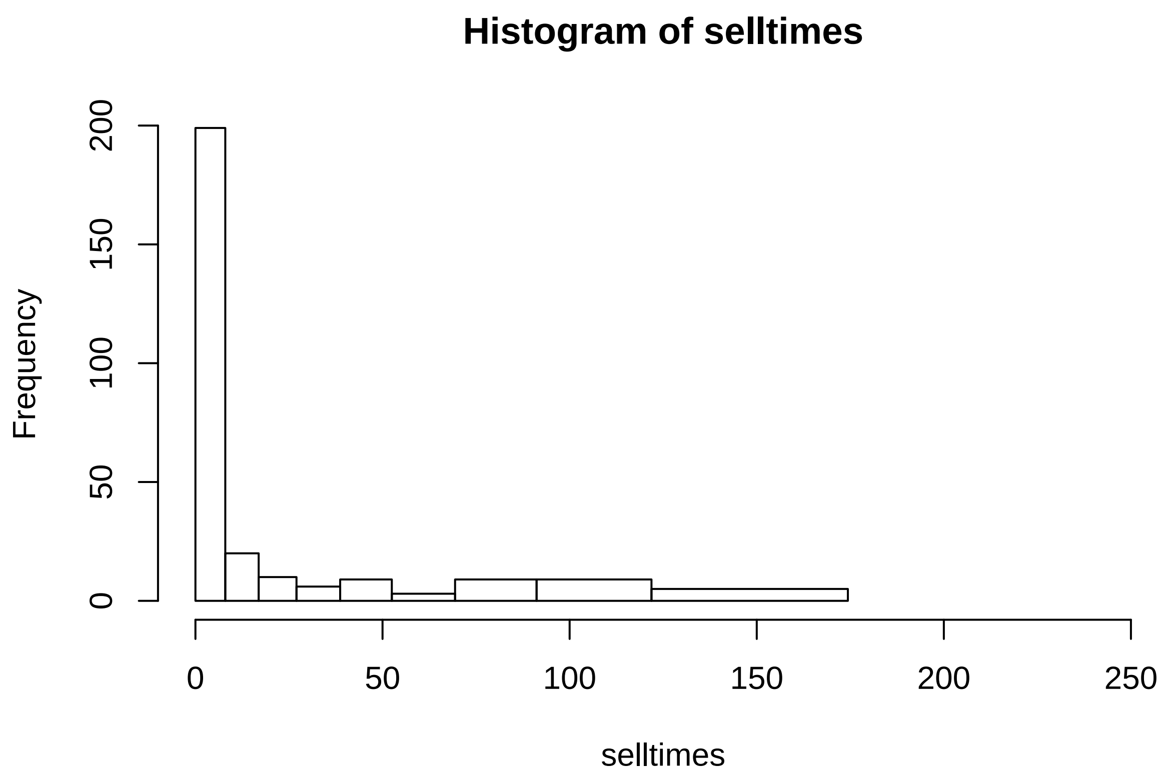
```
## [1] 0.01320614
```

Q2c. Histogram of Selltimes using MLE of Lambda.

```
#breakpoints chosen such that each box is expected to hold 10% of the data
bp=(log(seq(1,0,-.1)))/(-lh)
bp
```

```
##  [1]   0.000000   7.978144  16.896951  27.008260  38.680910  52.486724
##  [7]  69.383676  91.167634 121.870400 174.357124        Inf
```

```
hist.sell.lh=hist(selltimes,breaks=bp,freq=T,xlim = c(0,250))
```



Q2d.Number of observations in each interval.

```
hist.sell.lh$counts
```

```
##  [1] 199  20  10   6   9   3   9   9   5  39
```

Q2e. Chi-squared test.

```
chisq.test(hist.sell.lh$counts)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  hist.sell.lh$counts
## X-squared = 1048.1, df = 9, p-value < 2.2e-16
```

Calculated p-value is near zero implying we reject the null hypothesis that the data follows an exponential distribution.

Q2f. Observed vs Expected counts

```r
hist.sell.lh$counts
```

```
##  [1] 199  20  10   6   9   3   9   9   5  39
```

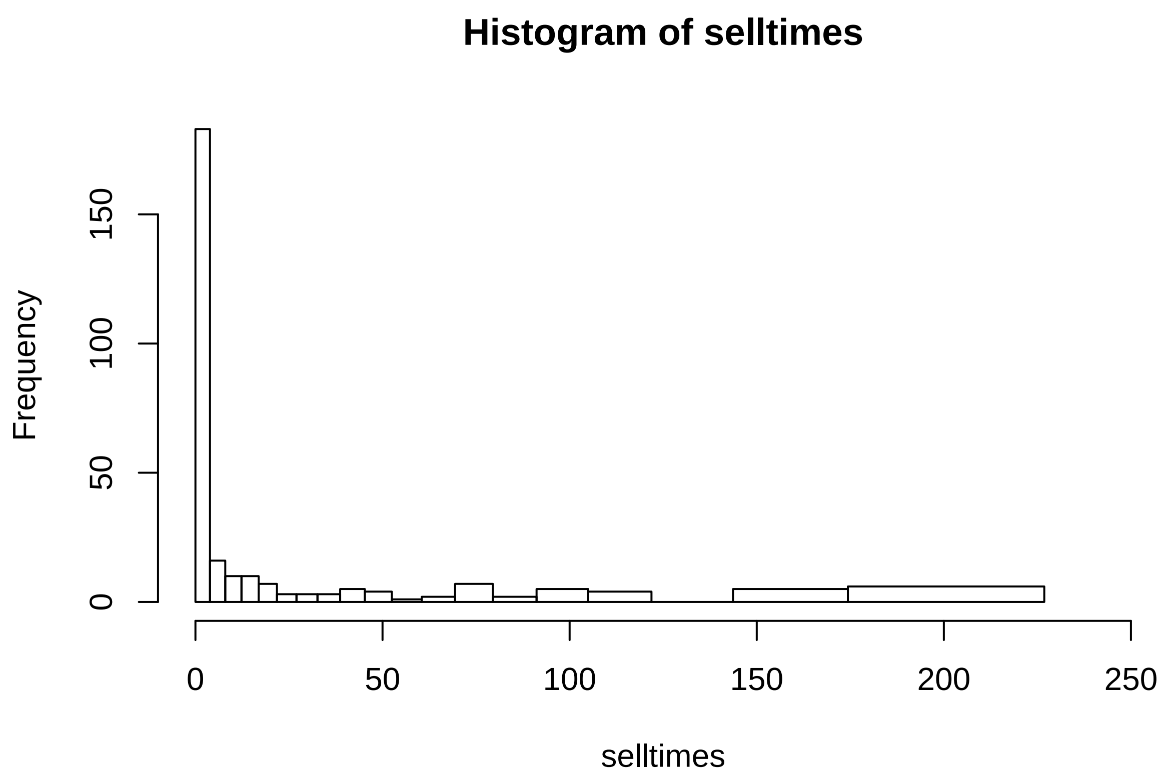
```r
chisq.test(hist.sell.lh$counts)$expected
```

```
##  [1] 30.9 30.9 30.9 30.9 30.9 30.9 30.9 30.9 30.9 30.9
```

It appears that, in comparison to our exponential, much of the observed data is skewed towards smaller values. Moreover, it is clearly shown that just under 2/3 of our observed values lay below our exponential distribution's lower 10% quantile.

Q2e. Testing various number of breaks

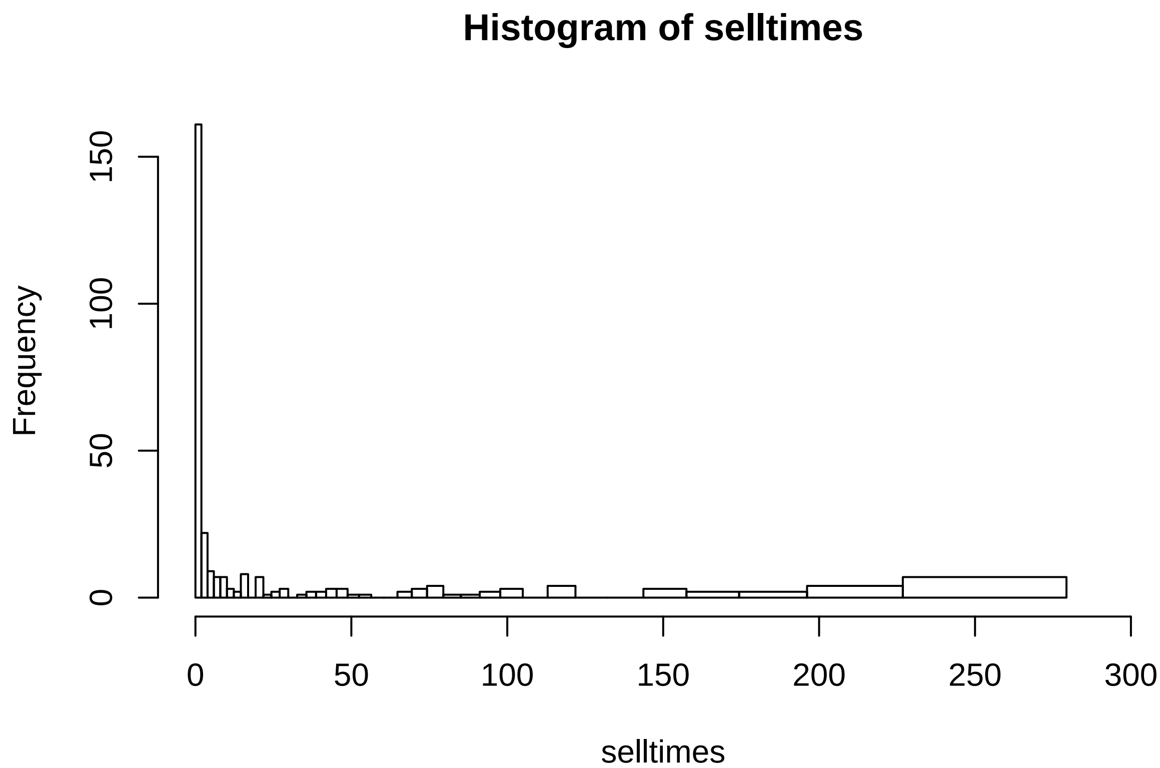
```r
bp20=(log(seq(1,0,-.05)))/(-lh) #20 Breaks
hist.sell.lh.20=hist(selltimes,breaks=bp20,freq=T,xlim = c(0,250))
```




```r
chisq.test(hist.sell.lh.20$counts)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  hist.sell.lh.20$counts
## X-squared = 1976.5, df = 19, p-value < 2.2e-16
```

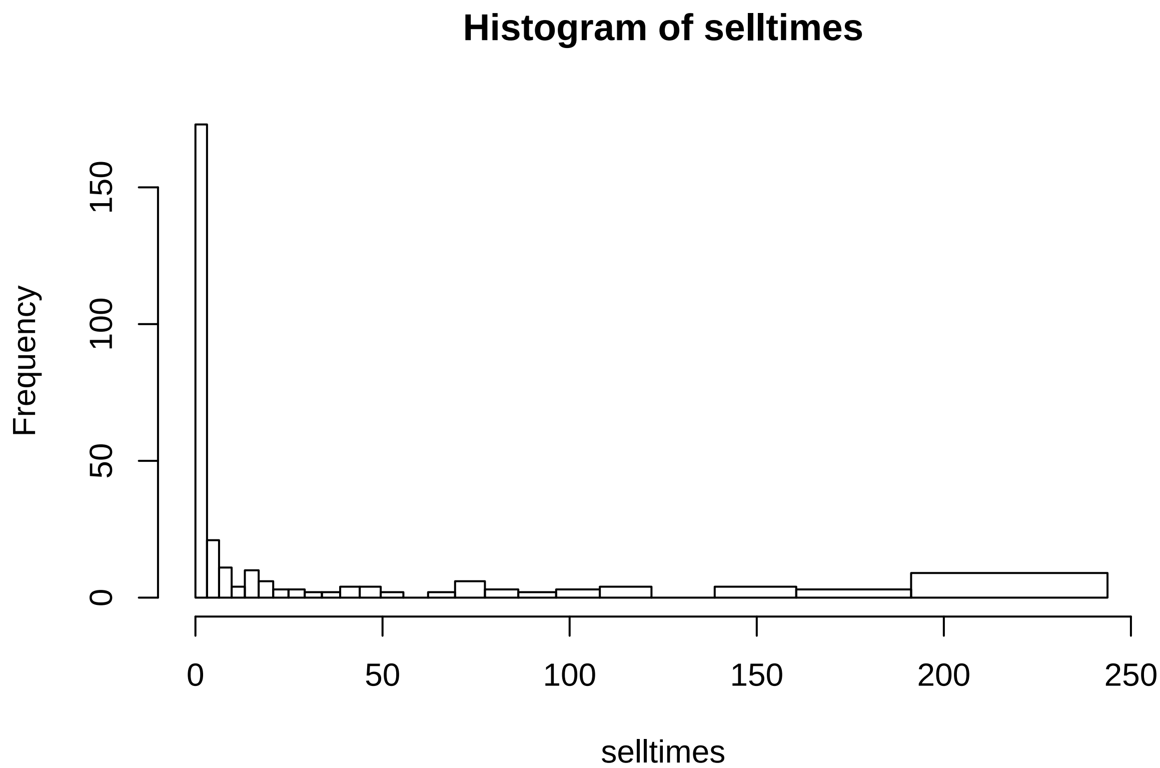
```r
bp40=(log(seq(1,0,-.025)))/(-lh) #40 Breaks
hist.sell.lh.40=hist(selltimes,breaks=bp40,freq=T,xlim = c(0,300))
```




```
chisq.test(hist.sell.lh.40$counts)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  hist.sell.lh.40$counts
## X-squared = 3260.1, df = 39, p-value < 2.2e-16
```

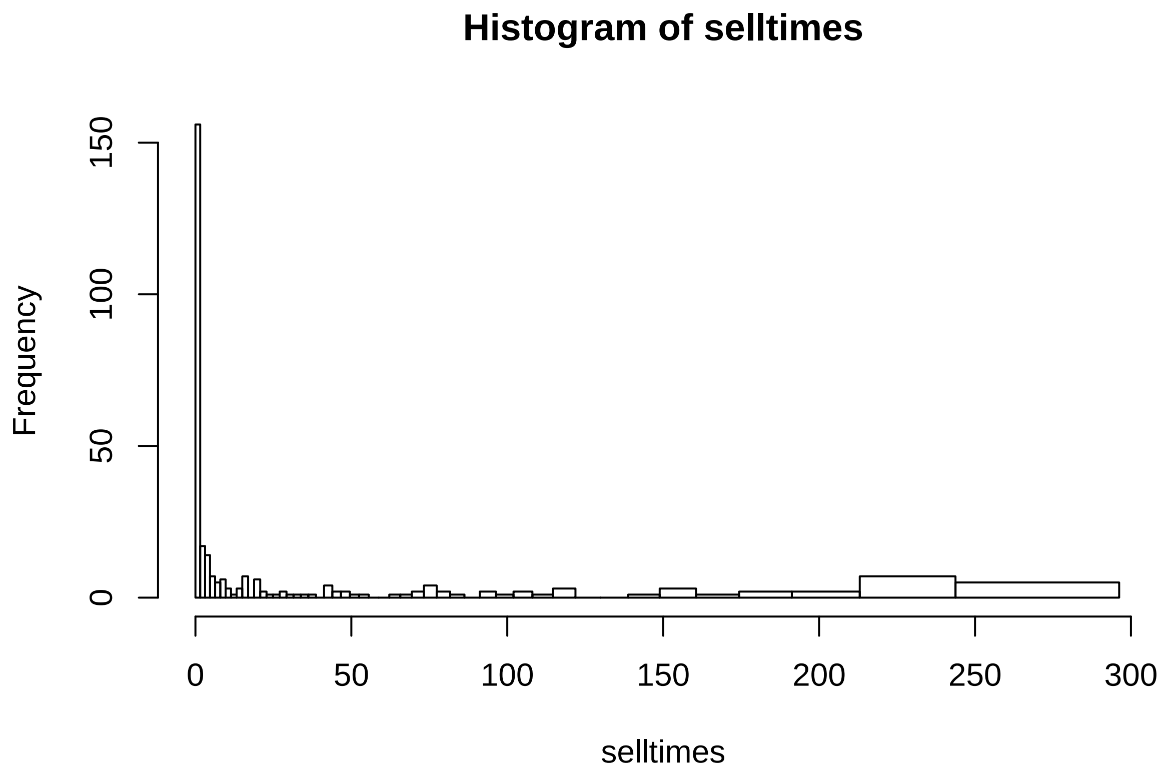
```
bp25=(log(seq(1,0,-.04)))/(-lh) #25 Breaks
hist.sell.lh.25=hist(selltimes,breaks=bp25,freq=T,xlim = c(0,250))
```




```
chisq.test(hist.sell.lh.25$counts)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  hist.sell.lh.25$counts
## X-squared = 2253.5, df = 24, p-value < 2.2e-16
```

```
bp50=(log(seq(1,0,-.02)))/(-lh) #50 Breaks
hist.sell.lh.50=hist(selltimes,breaks=bp50,freq=T,xlim = c(0,300))
```




```r
chisq.test(hist.sell.lh.50$counts)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  hist.sell.lh.50$counts
## X-squared = 3856.5, df = 49, p-value < 2.2e-16
```

It appears as we increase the number of breaks in our histogram, the larger X-Squared test statistic we acquire. However, there is overwhelming evidence given here. We continue to reject the null hypothesis. The selltimes data set does not appear to follow an exponential distribution with parameter lambda = 0.01320614.