

StatsAss6

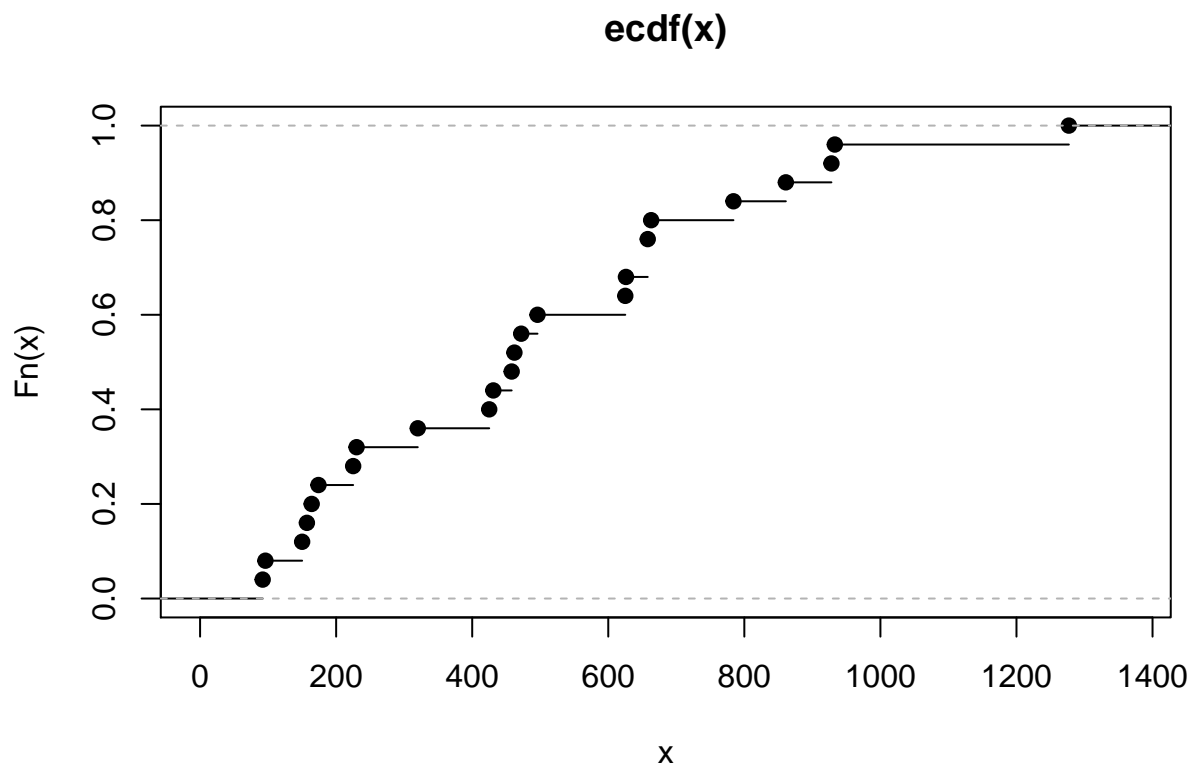
Sumeet Mishra

2/27/2019

```
x = scan("http://pages.iu.edu/~mtrosset/StatInfer/Data/sample771.dat")
```

Q1.(a) Empirical Plot of x

```
plot(ecdf(x))
```



1.(b)

Mean

```
mean(x)
```

```
## [1] 494.6
```

Var

```
var=mean(x^2) - mean(x)^2  
var
```

```
## [1] 91078.72
```

1.(c)

Median

```
median(x)
```

```
## [1] 462
```

Interquartile range

```
iq=quantile(x,0.75)-quantile(x,0.25)
iq
```

```
## 75%
```

```
## 433
```

1.(d)

plug-in estimate of the interquartile range to the square root of the plug-in estimate of the variance

```
iq/sqrt(var)
```

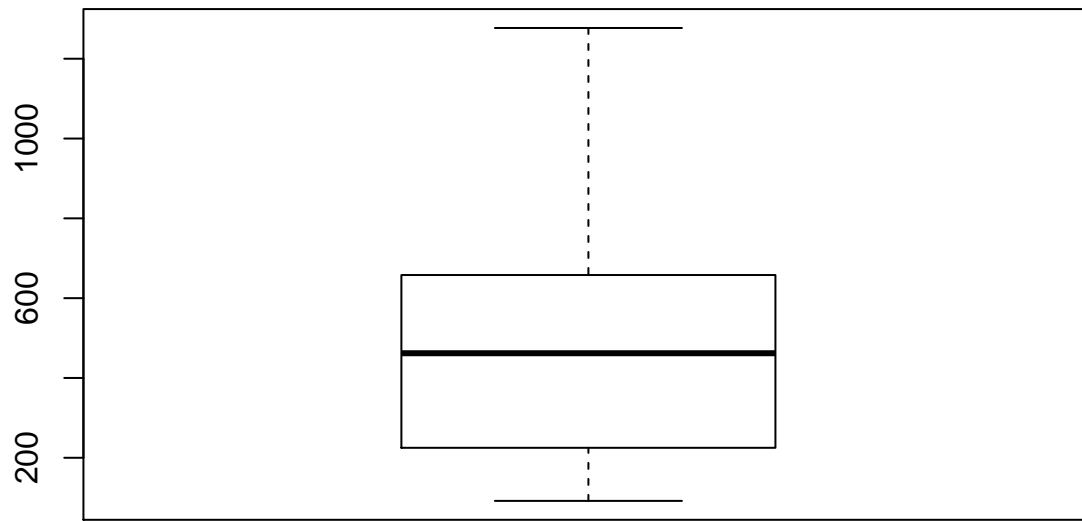
```
## 75%
```

```
## 1.434761
```

1.(e)

boxplot

```
boxplot(x)
```

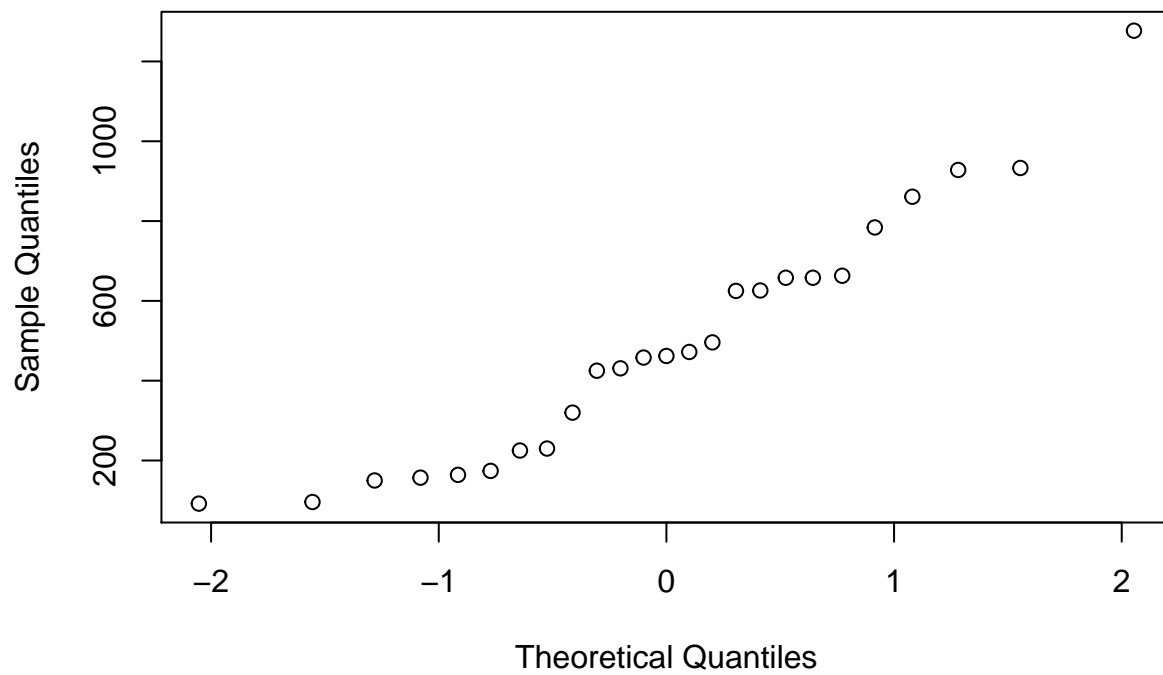


1.(f)

normal probability plot

```
qqnorm(x)
```

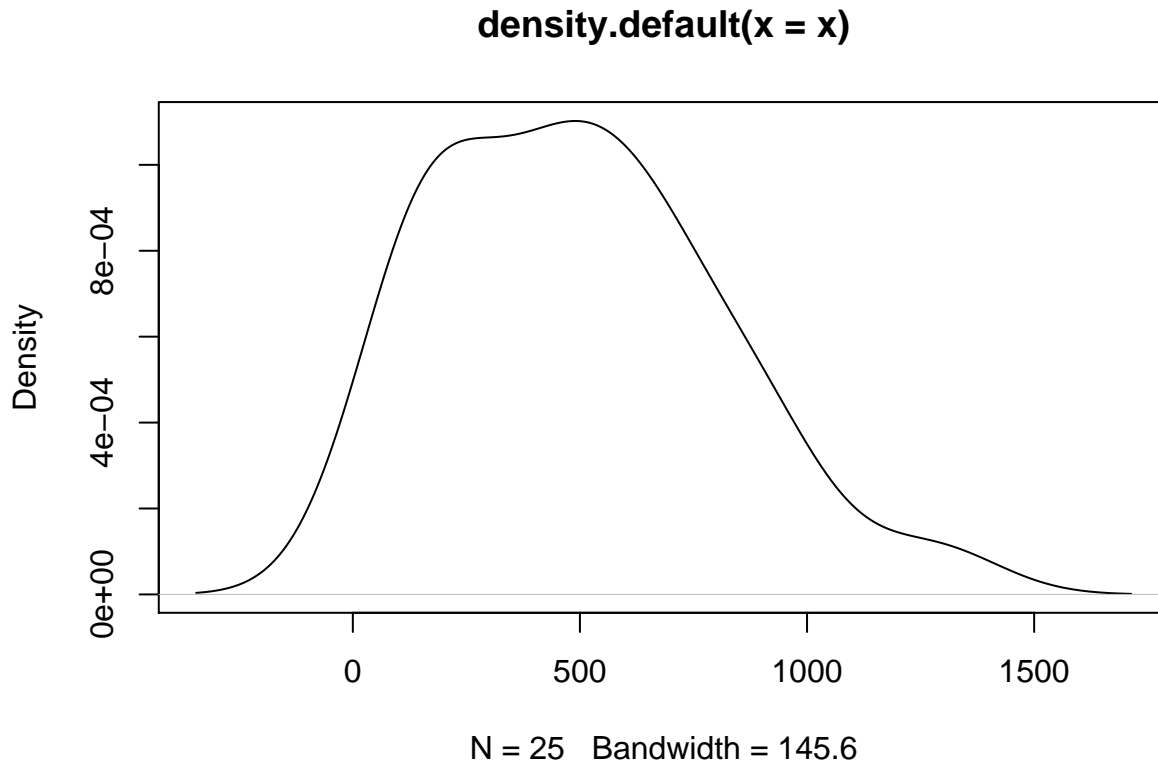
Normal Q-Q Plot



1.(g)

kernel density estimate

```
plot(density(x))
```



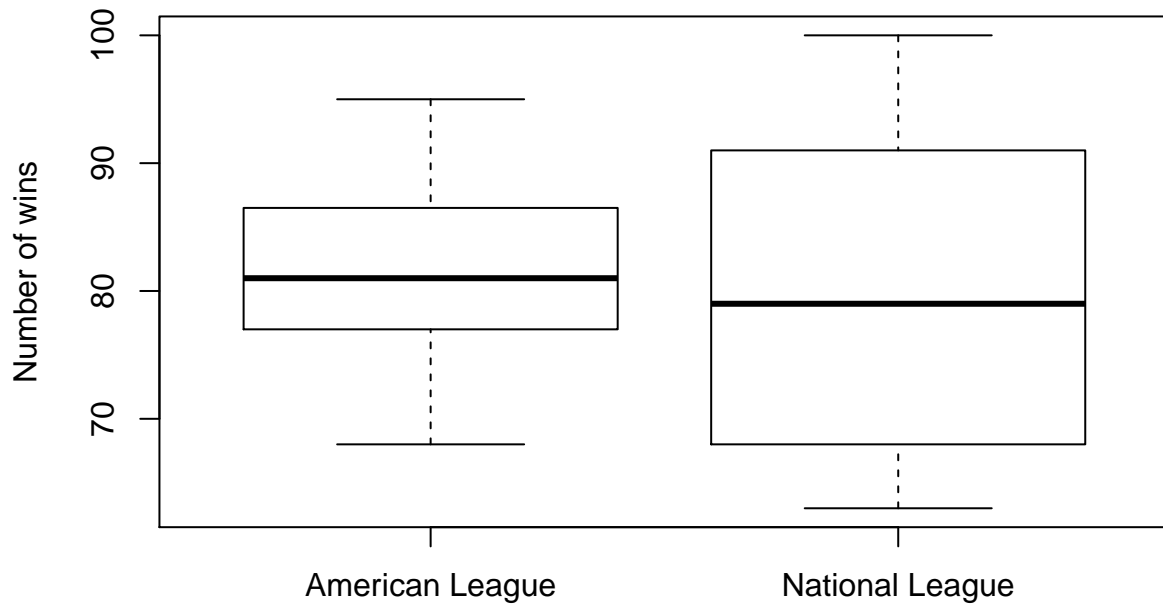
1.(h)

The sample was not drawn from normal distribution as it has two lumps and it's not symmetric(right skewed).

Q.2.(a)

```
am=read.csv('American_League.csv')
nat=read.csv('National_League.csv')
boxplot(am$W,nat$W,main = 'Boxplots of wins in 2015 for 2 leagues',
ylab='Number of wins',
names = c('American League', 'National League'))
```

Boxplots of wins in 2015 for 2 leagues



(b)

Distribution of American league

```
summary(am$W)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  68.00   77.00   81.00   82.07   86.50   95.00
```

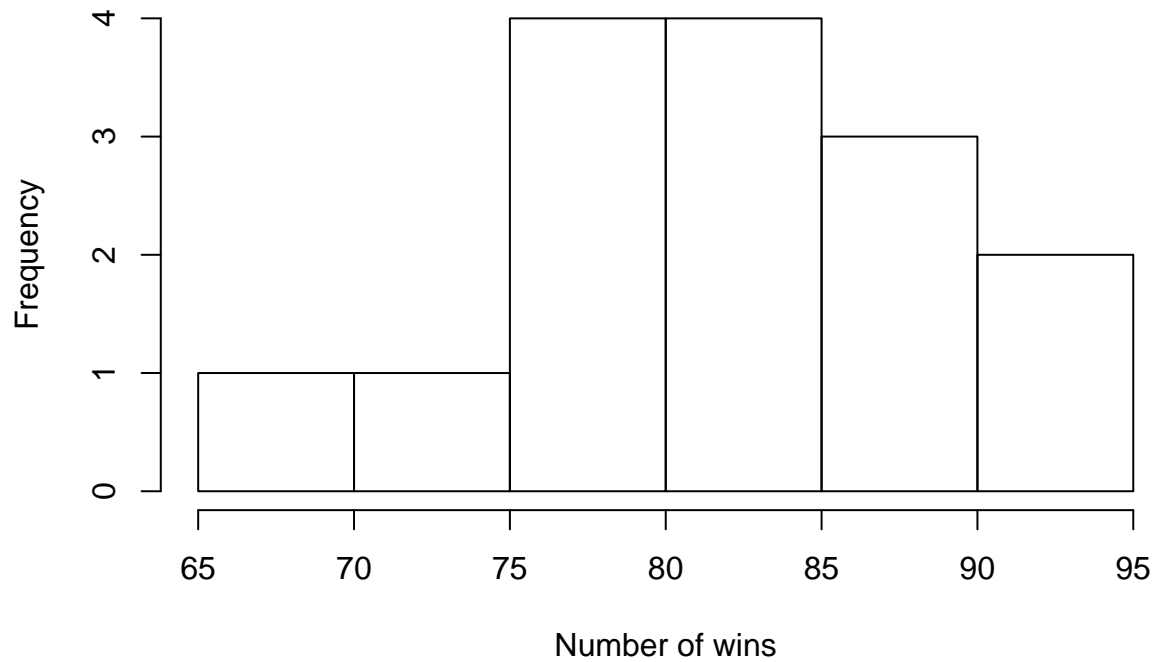
Distribution of National league

```
summary(nat$W)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  63.00   68.00   79.00   79.87   91.00  100.00
```

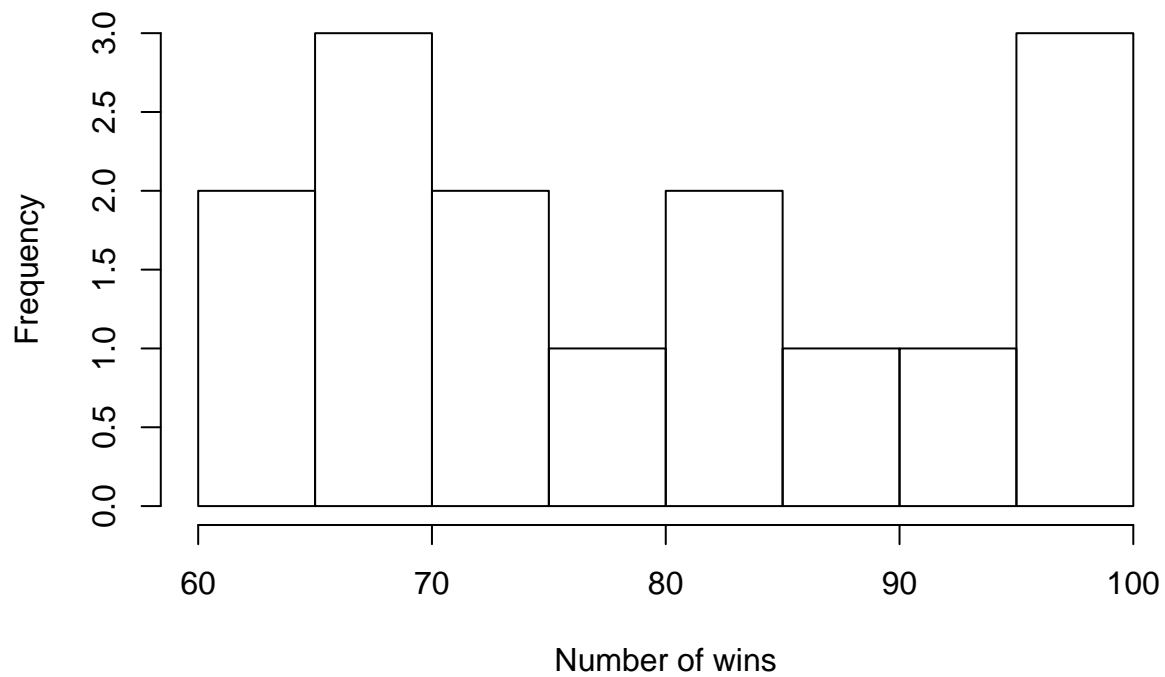
```
hist(am$W,main = 'Histogram of number of wins for American league',xlab='Number of wins')
```

Histogram of number of wins for American league



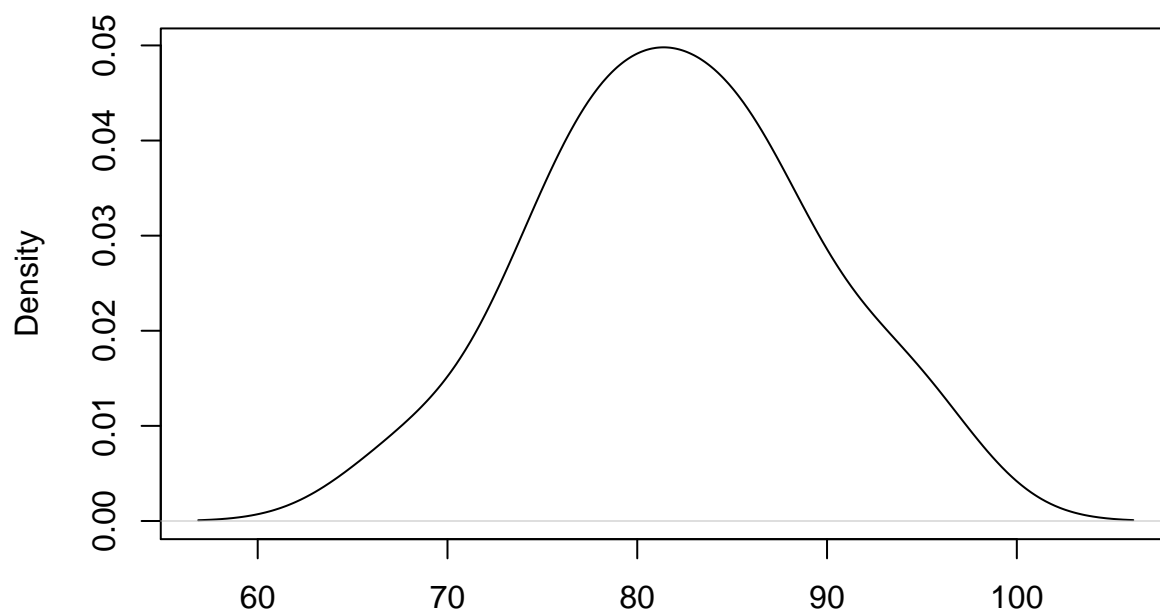
```
hist(nat$W,breaks=10,main = 'Histogram of number of wins for National league',xlab='Number of wins')
```

Histogram of number of wins for National league



```
plot(density(am$W), main='Kernel Density plot of American League')
```

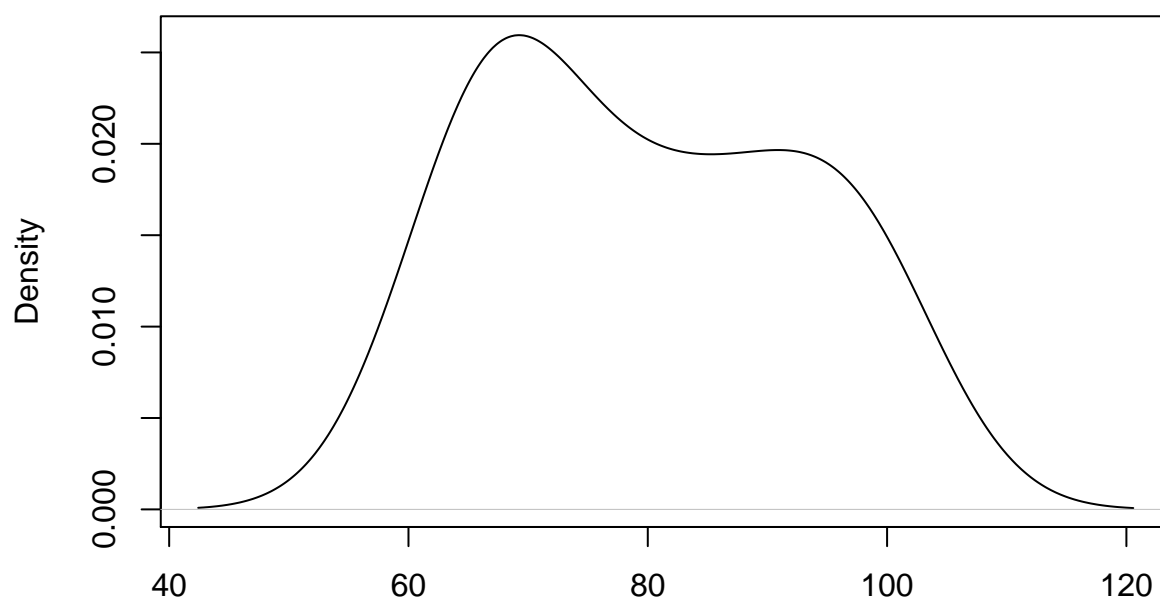
Kernel Density plot of American League



N = 15 Bandwidth = 3.712

```
plot(density(nat$W), main='Kernel Density plot of National League')
```

Kernel Density plot of National League



N = 15 Bandwidth = 6.858

From the above graphs we can see the distributions of the wins of both American and National league. None of them are normal distributions but wins of National league is a bimodal distribution as it has 2 peaks in the density plots. Also from summary of both the distributions it is clear that National league has more wins than American league.