Homework2

Coding Howework (submmite the pdf file to Canvas)

Let X be a Possion(λ) random variable. We have seen in class that

$$\mathbb{E}(X) = \text{Var}(X) = \lambda.$$

Suppose that we do not know the true value of λ and want to estimate it from observed data $\{x_1, x_2, \ldots, x_n\}$. There are two possible ways to do estimate λ:

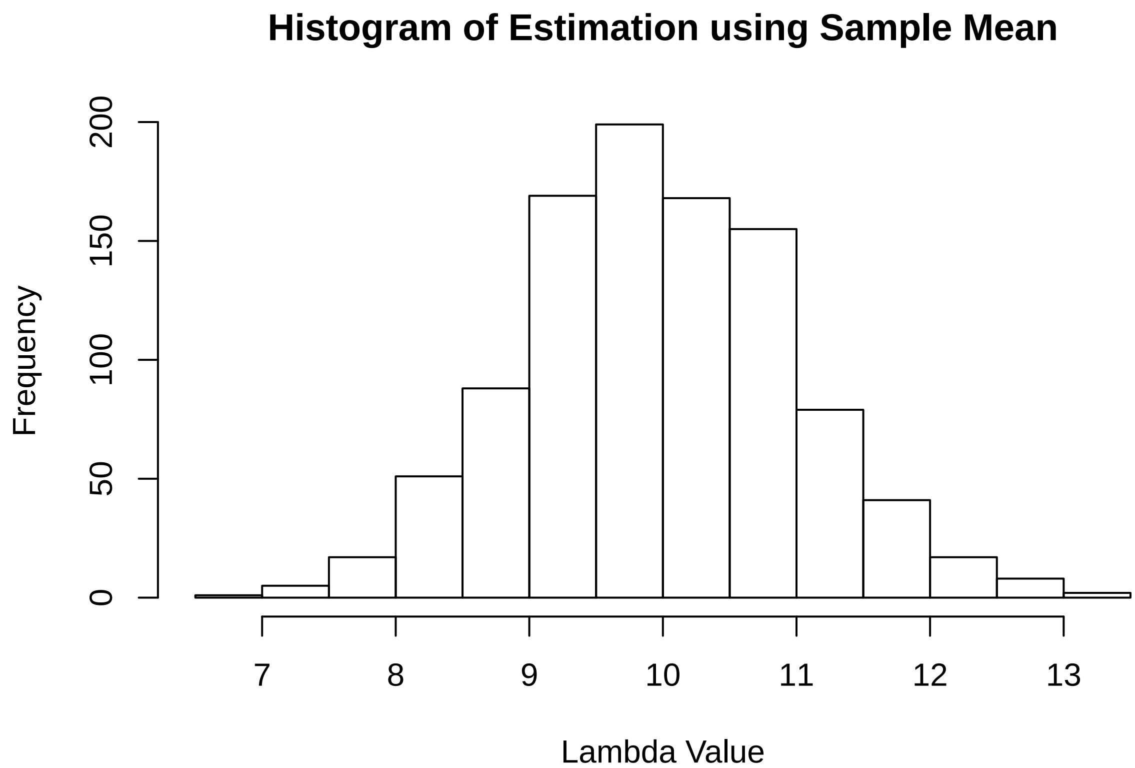
- use the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- use the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Please note that in sample variance, the denominator is $n-1$ instead of n.

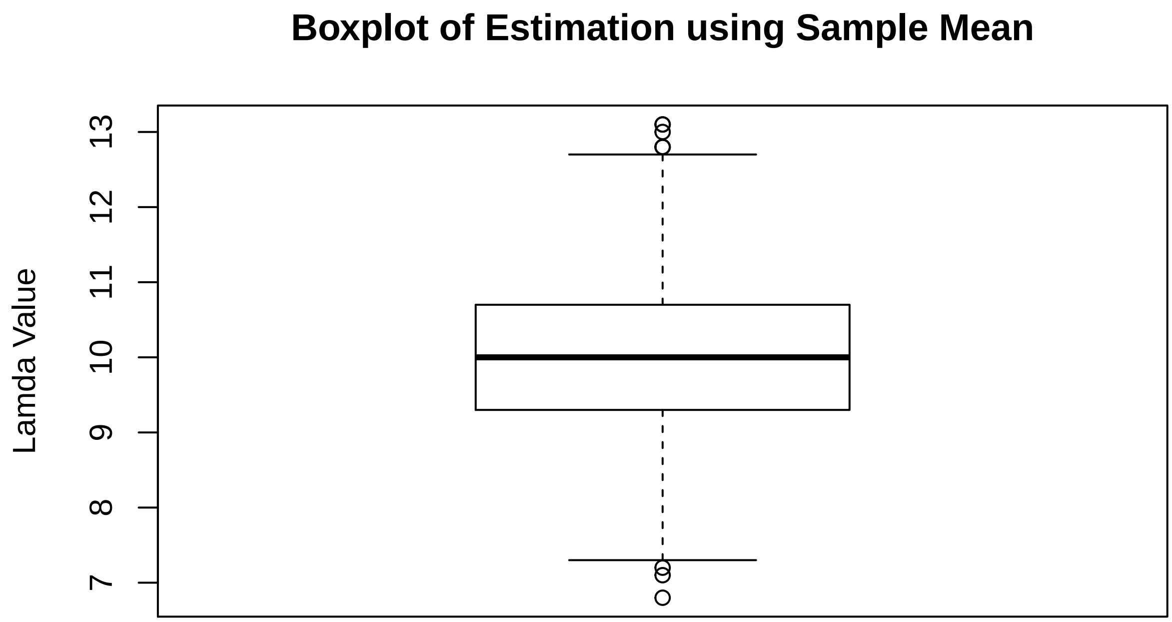
In this assignment, you will compare the two estimators. **In the following questions, we assume that $\lambda = 10$.**

1. Generate $n = 10$ independent Poisson(λ) random variables, calculate the sample mean (you can use `rpois(n = ,lambda = )` function in `R`, where `n` is the total number of random varables generated and `\lambda` is the parameter λ). Do the above 1000 times, then you have 1000 observations of the *sample mean* (each of them is calculated from $n = 10$ independent Poisson (λ) random variables.) **Generate the boxplot and histogram of the 1000 observation of sample means.**

```r
# input your r code here
x1 <- c()
for (i in 1:1000){
  x1[i] = mean(rpois(n = 10, lambda = 10))
}
hist(x1,
     main = "Histogram of Estimation using Sample Mean",
     xlab = "Lambda Value")
```

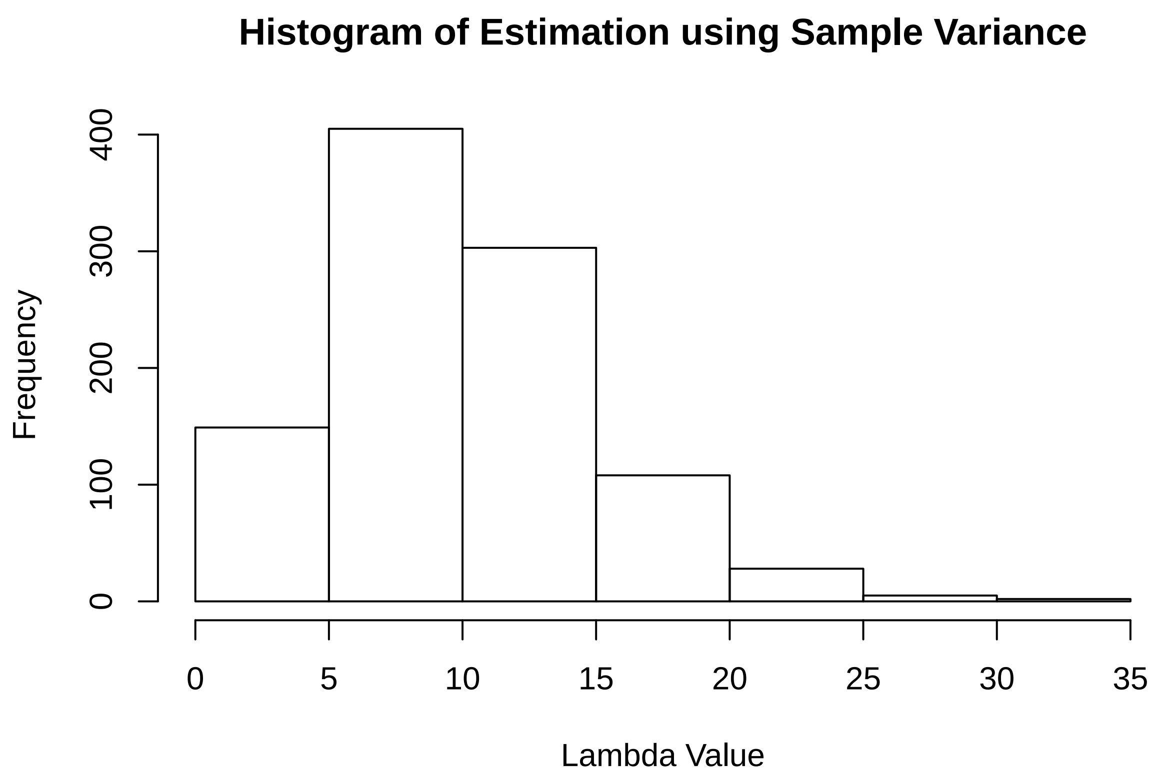



```
boxplot(x1,
        main = "Boxplot of Estimation using Sample Mean",
        ylab = "Lamda Value" )
```

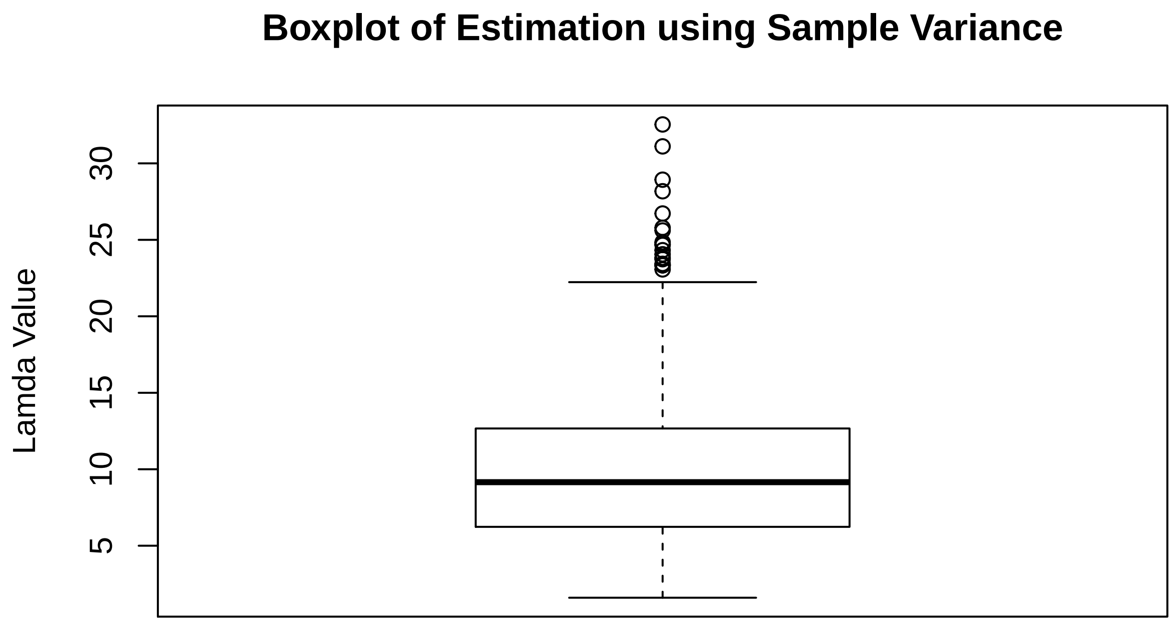


2. For $n = 10$, **repeat Part 1 with the *sample variance*.**

```
# input your r code here
x2 <- c()
for (i in 1:1000){
  x2[i] = var(rpois(n = 10, lambda = 10))
}
hist(x2,
     main = "Histogram of Estimation using Sample Variance",
     xlab = "Lambda Value")
```




```r
boxplot(x2,
        main = "Boxplot of Estimation using Sample Variance",
        ylab = "Lamda Value")
```



3. Compare the boxplot and histogram you obtained from Part 1 and 2. **Comment on the difference between them.** (Hint: measure of dispersion)

```r
# input your r code here
med1 = median(x1)
ran1 = max(x1) - min(x1)
std1 = sd(x1)
qrt1 = quantile(x1)

med2 = median(x2)
ran2 = max(x2) - min(x2)
std2 = sd(x2)
qrt2 = quantile(x2)


cat("Sample Mean as the Estimator", '\n',
    "Median \t \t: ", med1, '\n',
    "Range  \t \t: ", ran1, '\n',
    "Standard Deviation \t: ", std1, '\n',
    "Quartile (Q1)\t \t: ", qrt1[2],'\n',
    "Quartile (Q3)\t \t: ", qrt1[4], '\n')
```

```
## Sample Mean as the Estimator
##  Median       :  10
##  Range        :  6.3
```

```
##  Standard Deviation  :  1.01463
##  Quartile (Q1)       :  9.3
##  Quartile (Q3)       :  10.7
```

```r
cat("Sample Variance as the Estimator", '\n',
    "Median \t \t: ", med2, '\n',
    "Range  \t \t: ", ran2, '\n',
    "Standard Deviation \t: ", std2, '\n',
    "Quartile (Q1)\t \t: ", qrt2[2],'\n',
    "Quartile (Q3)\t \t: ", qrt2[4])
```

```
## Sample Variance as the Estimator
##  Median      :  9.155556
##  Range       :  30.94444
##  Standard Deviation  :  4.844849
##  Quartile (Q1)       :  6.233333
##  Quartile (Q3)       :  12.66667
```

Write down your comments here From the histogram of (1) estimation using 'sample mean'and (2) estimation using 'sample variance', we can compare respectively that:

- The shape of the first histogram almost show a perfectly simmetrical distribution, but the second histogram is positively skewed.
- It also can be seen that the first Standard Deviation of the first histogram is a lot smaller than the second histogram.

Meanwhile, from the boxpot of (1) estimation using 'sample mean' and (2) estimation using 'sample variance' we can analyze the 5-Number Summary:

- The Median of the first boxplot shows the same or almost the same value as the real value which is λ=10, while the median of the second boxplot is more inaccurate compare to the first one.
- The first boxplot have a lot smaller range (max value - min value) compare to the second.
- The first and third quartile on the first boxplot also have a closer value to the real value compare to the second boxplot.

In conclusion, estimating the value of λ using sample mean is more accurate compare to using sample variance.