STAT 403 HW 4

Chongyi Xu

May 2, 2018

Question 1

Asssume we observe 100 random variables $X_1, \cdots, X_{100} \sim$Pois(4).

(a) What are the bias and variance of the MLE of the rate parameter λ using these 100 observations.

$$\begin{aligned}
\text{Bias}(\hat{\lambda}) &= \text{E}[\frac{1}{n}\sum X_i] - \lambda \\
&= \text{E}[X_i] - \lambda \\
&= \lambda - \lambda = 0 \\
\text{Var}(\hat{\lambda}) &= \text{Var}[\frac{1}{n}\sum X_i] \\
&= \frac{1}{n^2}\text{Var}[\sum X_i] \\
&= \frac{1}{n^2} \cdot n\lambda = \frac{\lambda}{n}
\end{aligned}$$

(b) Write down a 90% confidence interval of λ using these 100 observations.

$$\frac{1}{100}\sum_{i}^{100} X_i \pm 1.645 \cdot \sqrt{\frac{\sum(X_i - \bar{X}_i)^2}{100-1}}$$

where $\bar{X}_i$ is the sample mean of the observations.

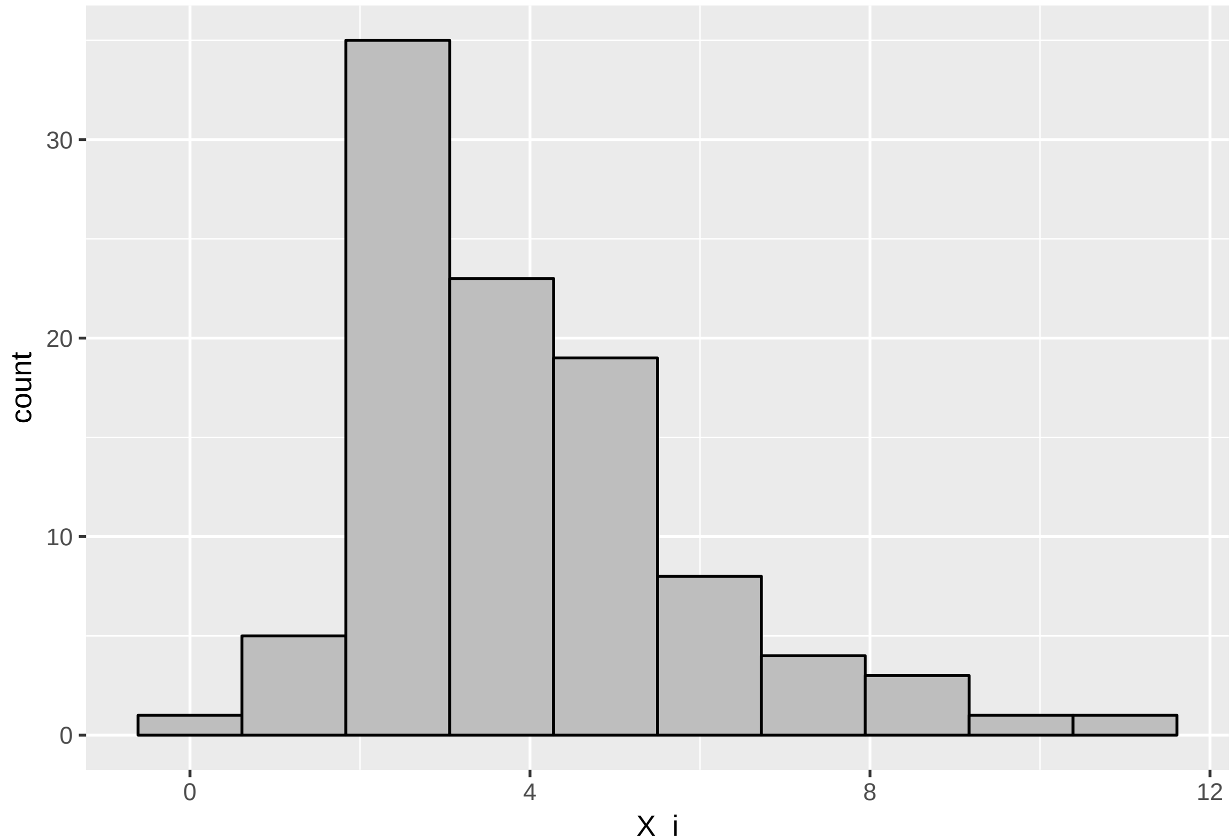
(c) Use R to generate 100 IID random points from Pois(4), show the histogram. What is the value of MLE using these 100 points?

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

```
set.seed(403)
n <- 100
dat <- rpois(n, lambda=4)

ggplot() + geom_histogram(aes(dat),
                          bins=10,
                          color="black",
                          fill="grey")+ xlab('X_i')
```



Compute for the value of MLE

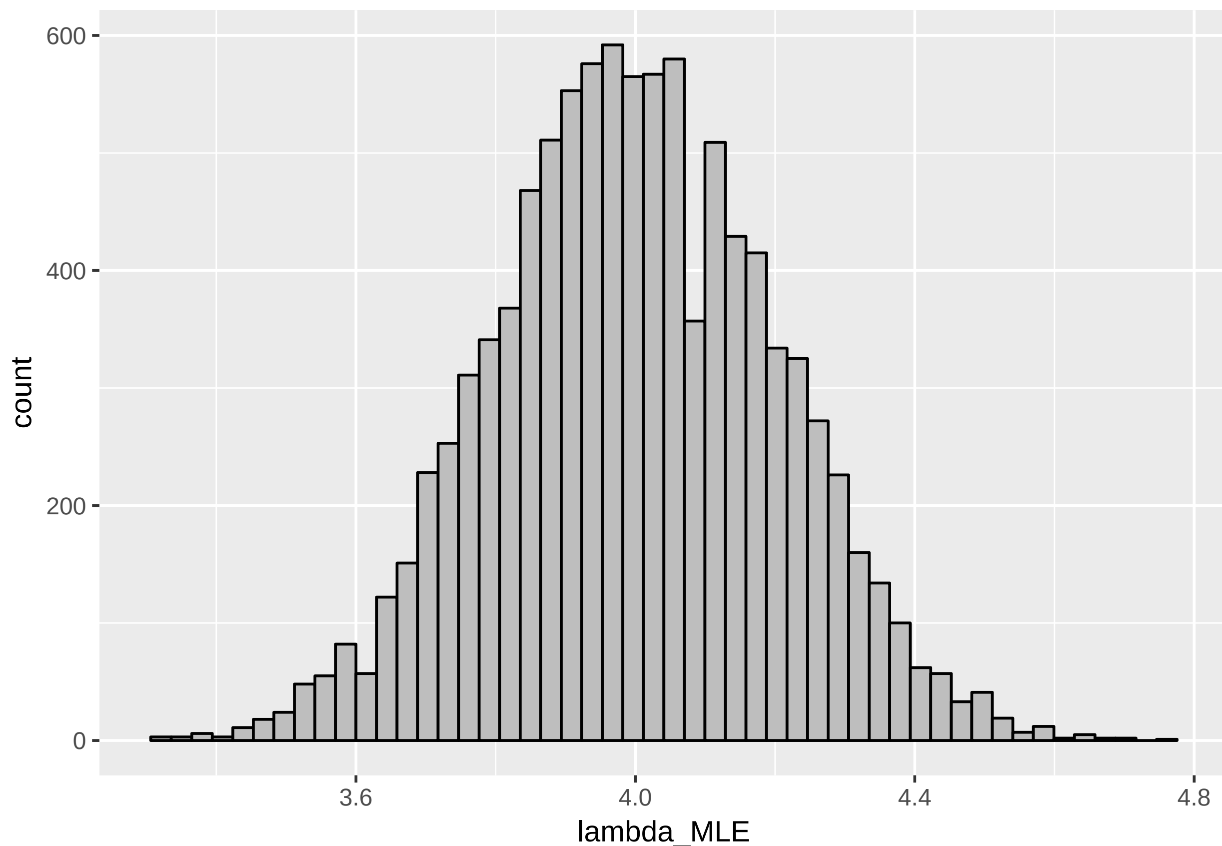
```
lambda_MLE <- mean(dat)
print(paste('The value of MLE using these 100 points is ', lambda_MLE))
```

```
## [1] "The value of MLE using these 100 points is  4.04"
```

(d) Use R to run $N = 10000$ Monte Carlo Simulations of obtain the MLE of a size 100 random sample. Plot the histrogram of these 10000 realizations. Does the fitted value looks like a normal distribution?

```
N=10000
size <- 100
lambda_MLE <- rep(0, N)
set.seed(403)
for (i in 1:N) {
  dat <- rpois(size, lambda=4)
  lambda_MLE[i] <- mean(dat)
}

ggplot() + geom_histogram(aes(lambda_MLE),
                  bins=50,
                  color="black",
                  fill="grey")
```

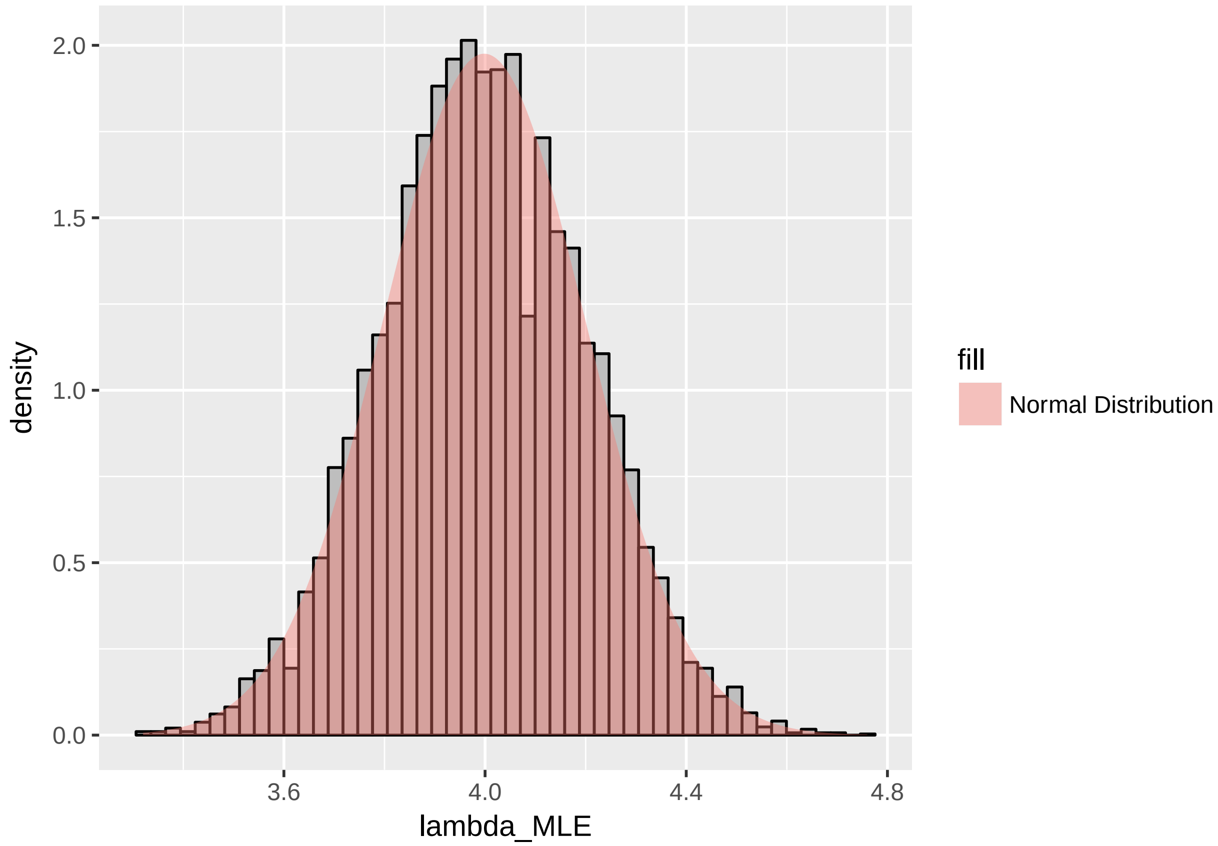


We now want to check if plot does look like a normal distribution.

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.4.4
```

```
fit <- fitdistr(lambda_MLE, densfun='normal')
x <- seq(from=min(lambda_MLE),to=max(lambda_MLE),length.out=N)
dn <- dnorm(x, fit$estimate[1], fit$estimate[2])
ggplot() + geom_histogram(aes(lambda_MLE, y=..density..),
                  bins=50,
                  color="black",
                  fill="grey") +
  geom_area(aes(x,dn,
                fill='Normal Distribution'),
            alpha=0.4)
```



From the plot above, we can see that the fitted value looks like a normal distribution.

(e) What fraction of the realizations of the MLE are within the interval [3.5,4.5]? Can you come up with an explanation of this?

```
fraction <- length(lambda_MLE[lambda_MLE<=4.5 & lambda_MLE>=3.5]) / N
print(paste('The fraction of the realizations of the MLE are within the interval [3.5,4.5] is ', fracti
```

```
## [1] "The fraction of the realizations of the MLE are within the interval [3.5,4.5] is  0.9889"
```

The reason might be that [3.5, 4.5] is a 99% confidence interval for the $\hat{\lambda}_{MLE}$. In the other word, we are about 99% sure that teh real λ will lie in this region.

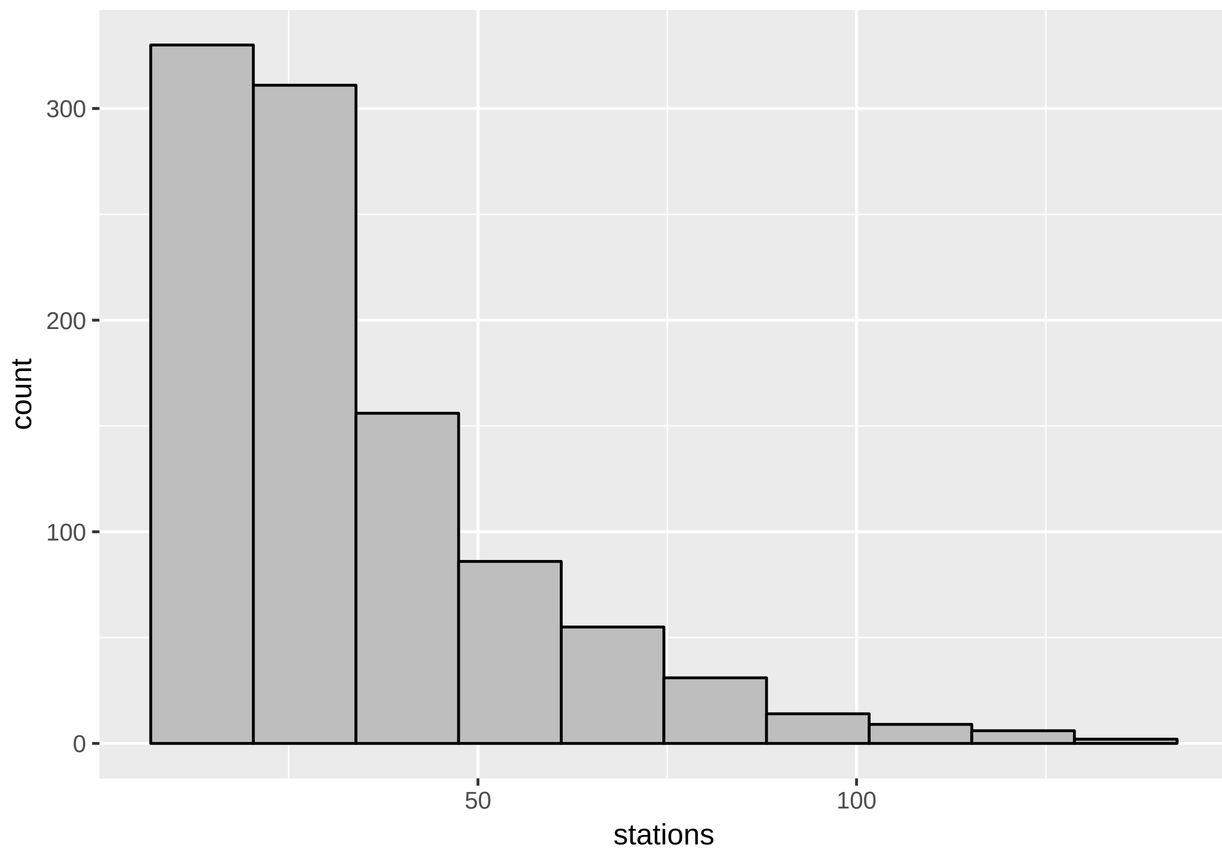
Question 2

Load the dataset `Earthquake in Fiji`.

```
dat <- read.table("http://www.stat.cmu.edu/~larry/all-of-nonpar/=data/fijiquakes.dat", header=TRUE)
```

(a) Show the histogram of the variable `stations`.

```
ggplot() + geom_histogram(aes(dat$stations),
                          bins=10,
                          color="black",
                          fill="grey") + xlab('stations')
```

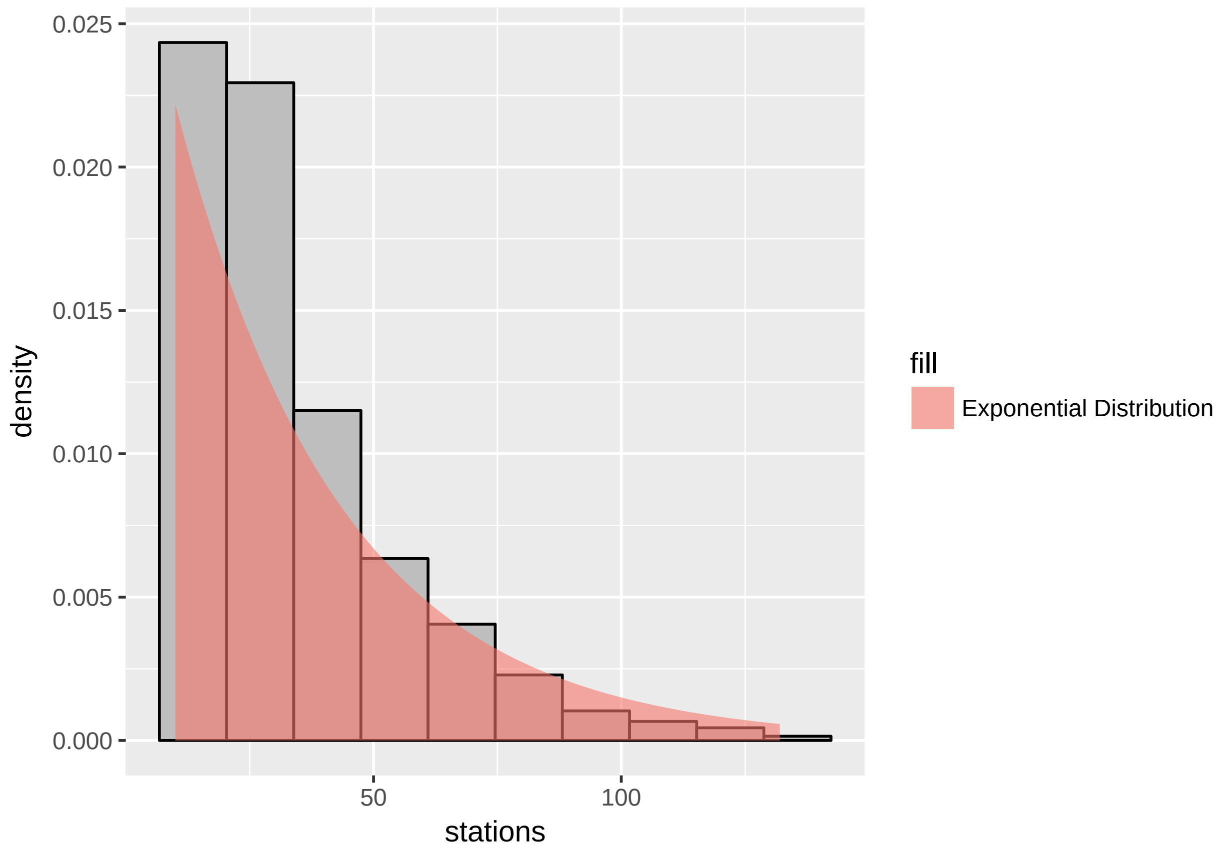


(b) We now fit an exponential distribution to the variable `stations`. What is the fitted value of the rate parameter?

```
fit <- fitdistr(dat$stations, densfun='exponential')
print(paste('The fitted value of the rate parameter is ', fit$estimate))
```

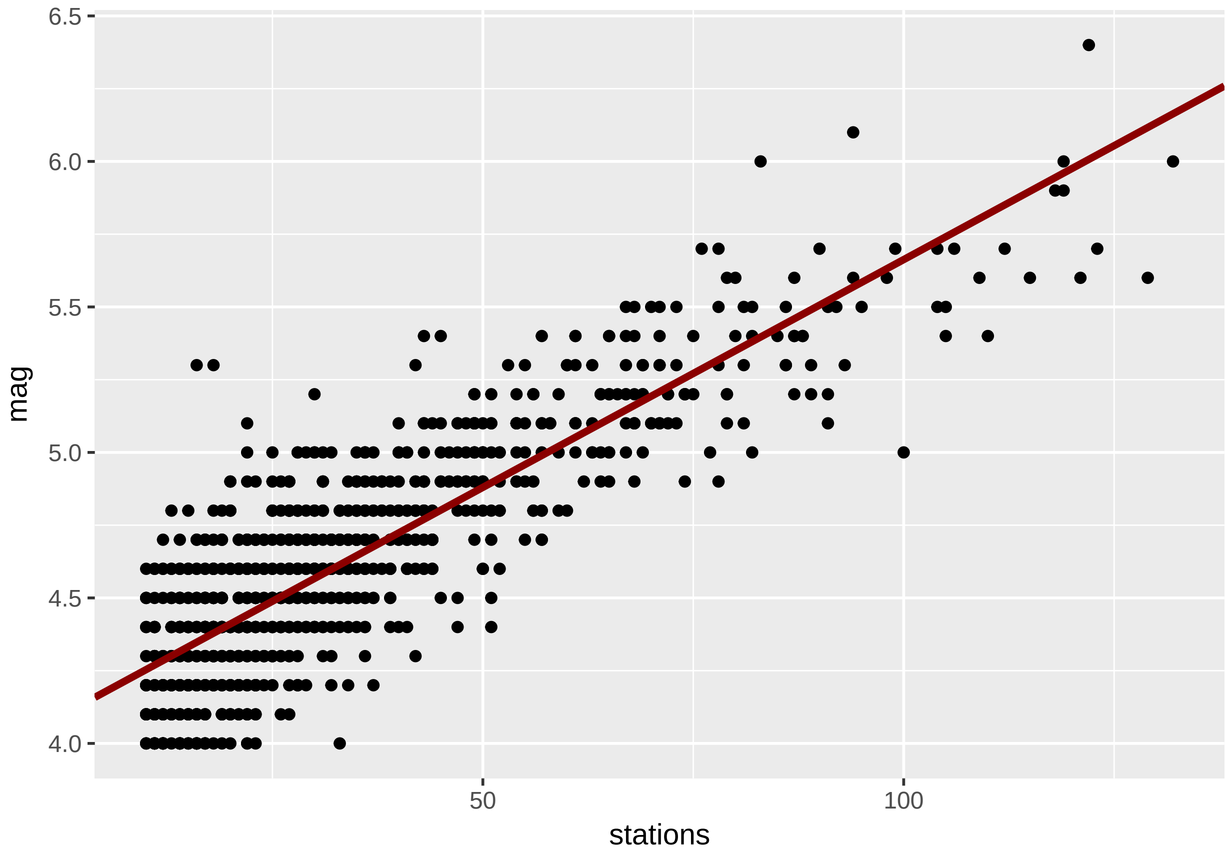
```
## [1] "The fitted value of the rate parameter is  0.0299239930576336"
```

```
x <- seq(from=min(dat$stations), to=max(dat$stations), length.out=length(dat$stations))
de <- dexp(x, rate=fit$estimate)
ggplot() + geom_histogram(aes(dat$stations, y=..density..),
                          bins=10,
                          color="black",
                          fill="grey") +
  xlab('stations') +
  geom_area(aes(x, de,
                fill='Exponential Distribution'),
            alpha=0.6)
```



(c) Fit a linear model iwth the response variable `Y` being `mag` and the covariate `X` being `stations`. WHat are the fitted slope? Show the scatter plot and attach the fitted regression line.

```
model <- lm(data=dat, mag~stations)
ggplot() + geom_point(data=dat,
                      aes(stations, mag)) +
  geom_abline(intercept=model$coefficients[1],
              slope=model$coefficients[2],
              col='darkred', size=1.3)
```




```r
print(paste('The fitted slope is ', model$coefficients[2]))
```

```
## [1] "The fitted slope is  0.0156542115038549"
```

```r
print(paste('The fitted intercept is ', model$coefficients[1]))
```

```
## [1] "The fitted intercept is  4.09726755996418"
```

(d) What is a 95% confidence interval of the fitted slope?

```r
interval <- confint(model, 'stations', level=0.95)
print(paste('The 95% confidence interval of the fitted slope is [', interval[1], ',', interval[2], '].'
```

```
## [1] "The 95% confidence interval of the fitted slope is [ 0.0150545997847491 , 0.0162538232229607 ].
```