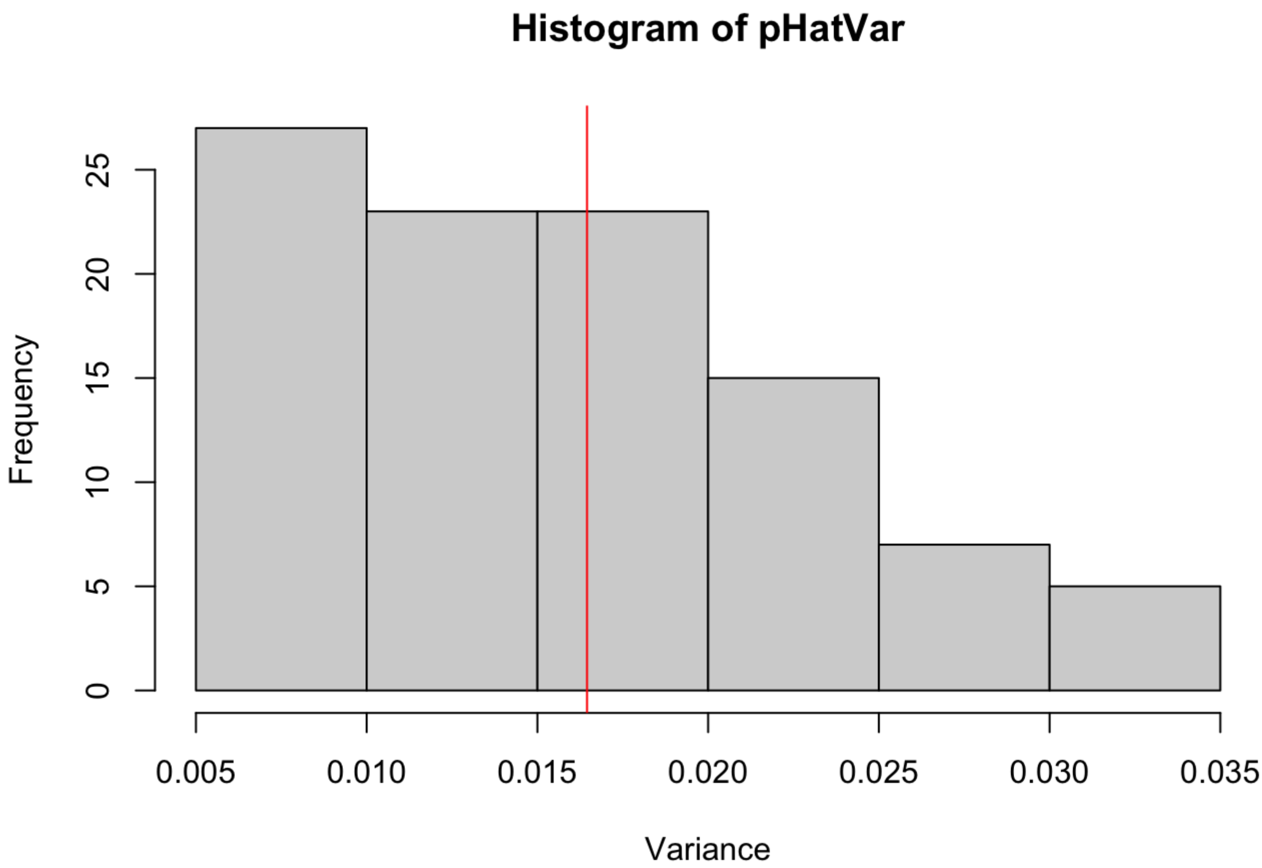
Computer Practical 1

Q1 and Q2

```
n <- 13
p <-0.31
# we will carry out 100 trials in which each trial we observe 10 Binomial(n,p)
sampleSize <- 10
trials <- 100
# varHat stores the true variance of the MLE of p
varHat <- (1/n) * p * (1-p)
print(paste("Variance of pHat(Y) is " ,varHat))
```

```
## [1] "Variance of pHat(Y) is  0.0164538461538462"
```

```
# phat is the MLE of p function
phat <-function(Y) {
  Y/n
}
# we generate 1000 binomials and organise them into 10 rows of 100
# this means each column is a trial
xValues <- rbinom(sampleSize*trials, n, p)
xSamples <- matrix(xValues, nrow=sampleSize) # 10x100
# we pass each row into phat to get the MLE function transformation
pHatSamples<-apply(xSamples, 1, phat) # it becomes 100x10
# we then get the sample variance of each row
pHatVar <- apply(pHatSamples, 1, var) # stays 100x1
hist(pHatVar, xlab = "Variance")
abline(v = varHat, col='red')
```



Q3 (Q4 and Q5 done in working out section)

```
year.data <- read.csv("year_data.csv")
knitr::kable(year.data)
```

year	births	deaths	clinic
1841	3036	237	1
1842	3287	518	1
1843	3060	274	1
1844	3157	260	1
1845	3492	241	1
1846	4010	459	1
1841	2442	86	2
1842	2659	202	2
1843	2739	164	2
1844	2956	68	2
1845	3241	66	2
1846	3754	105	2

```
n1 <- sum(year.data[year.data$clinic==1,]$births)
y1 <- sum(year.data[year.data$clinic==1,]$deaths)
n2 <- sum(year.data[year.data$clinic==2,]$births)
y2 <- sum(year.data[year.data$clinic==2,]$deaths)
print(paste("The number of births in clinic 1 is ", n1))
```

```
## [1] "The number of births in clinic 1 is  20042"
```

```
print(paste("The number of deaths in clinic 1 is ", y1))
```

```
## [1] "The number of deaths in clinic 1 is  1989"
```

```
print(paste("The number of births in clinic 2 is", n2))
```

```
## [1] "The number of births in clinic 2 is 17791"
```

```
print(paste("The number of deaths in clinic 2 is", y2))
```

```
## [1] "The number of deaths in clinic 2 is 691"
```

```
print(paste("The ML estimates of p1hat(y1) is y1/n1 which is 1989/20042 which is",y1/
n1))
```

```
## [1] "The ML estimates of p1hat(y1) is y1/n1 which is 1989/20042 which is 0.0992415
926554236"
```

```
print(paste("The ML estimates of p2hat(y2) is y2/n2 which is 691/17791 which is " , y
2/n2))
```

```
## [1] "The ML estimates of p2hat(y2) is y2/n2 which is 691/17791 which is  0.0388398
628520038"
```

Q6 First Part (second part done in working out section)

```
month.data <- read.csv("month_data.csv")
# removes NA data items
month.data <- month.data[!is.na(month.data$births),]
month.data$rate <- month.data$deaths/month.data$births
month.data$date <- as.Date(month.data$date)
intervention.date <- as.Date("1847-05-15")
plot(month.data$date, month.data$rate, pch=20, main="Mortality rate by month",
     xlab="Date; red line indicated start of intervention period", ylab="Rate ")
abline(v=intervention.date, col="red")
```


```
before.interventions <- month.data[month.data$date < intervention.date,]
after.intervention <- month.data[month.data$date > intervention.date,]

n1 <- sum(before.interventions$births)
y1 <- sum(before.interventions$deaths)
n2 <- sum(after.intervention$births)
y2 <- sum(after.intervention$deaths)
print(paste("The number of births before intervention is ", n1))
```

```
## [1] "The number of births before intervention is  19571"
```

```
print(paste("The number of deaths before intervention is ", y1))
```

```
## [1] "The number of deaths before intervention is  2060"
```

```
print(paste("The number of births after intervention is", n2))
```

```
## [1] "The number of births after intervention is 6595"
```

```
print(paste("The number of deaths after intervention is", y2))
```

```
## [1] "The number of deaths after intervention is 142"
```

```
print(paste("The ML estimates for the mortality rates before intervention is y1/n1 wh
ich is " , y1/n1))
```

```
## [1] "The ML estimates for the mortality rates before intervention is y1/n1 which i
s  0.105257779367431"
```

```
print(paste("The mortality rates after intervention is y2/n2 which is ", y2/n2))
```

```
## [1] "The mortality rates after intervention is y2/n2 which is  0.0215314632297195"
```

Computer Practical 1

Q4) $\mathbb{E}[W] = \mathbb{E}[\hat{p}_1(Y_1) - \hat{p}_2(Y_2)]$

$$= \mathbb{E}[\hat{p}_1(Y_1)] - \mathbb{E}[\hat{p}_2(Y_2)]$$

$$= p - p = 0$$

$$\mathrm{Var}(W) = \mathrm{Var}(\hat{p}_1(Y_1) - \hat{p}_2(Y_2))$$

$$= \mathrm{Var}(\hat{p}_1(Y_1)) + \mathrm{Var}(\hat{p}_2(Y_2))$$

$$= \frac{1}{n_1} p_1(1-p_1) + \frac{1}{n_2} p_2(1-p_2)$$

$$=\left(\frac{1}{n_1}+\frac{1}{n_2}\right)p(1-p)$$

Q5)

We are assuming the true underlying mortallity rates are the same so let

$$p_1 = p_2 = p = \frac{1989+691}{20042+17791} = \frac{2680}{37833}$$

$$= 0.07083763$$

Then $\mathbb{P}\left(|W-\mu_w| \geq \hat{p}_1(y_1)-\hat{p}_2(y_2)\right)$

$$= \mathbb{P}\left(|W| \geq 0.09924159 - 0.0388\right)$$

$$= \mathbb{P}\left(|W| \geq 0.06040173\right)$$

$$\leq \frac{\sigma^2}{(0.06040)^2} = \frac{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)p(1-p)}{(0.06040)^2}$$

$$= \frac{\left(\frac{1}{20042}+\frac{1}{17791}\right)(0.07083)(1-0.07083)}{(0.06040)^2}$$

$$= 0.0019142$$

Thus we have shown that there is a small probability we will see such a large difference between the observed mortality rate under the assumption that the true underlying mortality rates are actually the same.

Q6) let p be true underlying mortallity rate: then as before we assume

$$p = \frac{2060 + 142}{19571 + 6595} = \frac{2202}{26166}$$

$$= 0.08415501$$

Let $W := \hat{p}_1(Y_1) - \hat{p}_2(Y_2)$

Similarly to before,

$$\mathbb{E}(W) = 0$$

$$Var(W) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right) p(1-p)$$

assuming underlying mortallity rate same for both random variables

Then $\mathbb{P}(|w - \mu_w| \geq \hat{p}_1(y_1) - \hat{p}_2(y_2))$

$$= \mathbb{P}(|w| \geq 0.1052777 - 0.021531)$$

$$= \mathbb{P}(|w| \geq 0.08374639)$$

$$\leq \frac{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) p(1-p)}{(0.08374639)^2}$$ by Chebysev's Inequality

$$= \frac{\left(\frac{1}{19571} + \frac{1}{6595}\right)(0.08415501)(1-0.084155)}{(0.08374639)^2}$$

$$= 0.0022278158$$

Thus we have shown there is a small probabillity we will see such a large difference between the observed mortallity rate under the assumption the underlying mortallity rates are the same.

Q7) $L(\theta) \propto \prod_{i=1}^{2} f_{Y_i}(y_i, n_i, x_i; \theta)$