Question 1

1. First the sample mean and sample variance is calculated,

```
covidcases <- read.csv("daily.covid.aug1to7.csv")

mean_hat <- mean(covidcases$daily.covid.cases)

var_hat <- var(covidcases$daily.covid.cases)
```

which yields a result of 7359.571 average cases and a variance of 4108399.952 cases.

After that, we retrieve the student-t value for α = 0.05:

```
size <- length(covidcases$daily.covid.cases)

t_value <- qt(p = 1-0.05/2, df=size-1)
```

Where we get $t_{\alpha/2,n-1}$ = 2.446912.

By the formula, we can calculate the confidence interval of normal mean and unknown variance.

```
mean_hat - t_value*sqrt(var_hat/size)

mean_hat + t_value*sqrt(var_hat/size)
```

Therefore, the approximate confidence interval of the sample data is (5484.984, 9234.159).

From the data, we estimate the average daily reported case numbers of people in Victoria, Australia infected with the novel coronavirus (Covid-19) to be 7359.571 cases, and we are 95% confident that the population mean reported cases in Australia is between 5484.984 cases and 9234.159 cases.

2. First the sample mean and the sample variance for the second dataset is calculated,

```
covidcases2 <- read.csv("daily.covid.aug8to14.csv")

mean_hat2 <- mean(covidcases2$daily.covid.cases)

var_hat2 <- var(covidcases2$daily.covid.cases)
```

which yields a result of 4879.000 average cases and a variance of 1286109.333 cases.

The size of the dataset is also obtained:

```
size2 <- length(covidcases2$daily.covid.cases)
```

The difference in average daily reported cases between the two samples is,

```
mean_diff <- mean_hat - mean_hat2
```

Where we get $\mu_A - \mu_B = 2480.571$ cases.

The 95% confidence z-value is 1.96.

By the formula, we can calculate the confidence interval for difference of normal means.

```
mean_diff - 1.96*sqrt(var_hat/size+var_hat2/size2)

mean_diff + 1.96*sqrt(var_hat/size+var_hat2/size2)
```

Therefore, the approximate confidence interval for $\mu_A - \mu_B$ is (759.959, 4201.184).

We can summarize that the estimated difference in daily reported case numbers of people in Victoria, Australia infected with the novel coronavirus (Covid-19) between the first 7 day block in August (sample size = 7) and the second 7 day block in August (sample size = 7) is 2480.571 cases. We are 95% confident that the population mean difference in daily reported cases is between 759.959 cases and 4201.184 cases. It is suggestive of a positive difference at population level.

3. We are testing for the hypothesis $H_0 : \mu_x = \mu_y$, where μ_x is the average daily reported cases of people in Victoria, Australia infected with the novel coronavirus (Covid-19) in the first 7 day block in August and μ_y is the average daily reported cases in the second 7 day block in August.

First we calculate the test statistic, $z_{(\mu x - \mu y)}$ by the formula,

```
z <- (mean_hat - mean_hat2)/sqrt(var_hat/size + var_hat2/size2)
```

and we get a result of z = 2.825.

Then we can find the approximate p-values using $p \approx 2P(Z < -|z_{(\mu x - \mu y)}|)$,

```
p <- 2*pnorm(-abs(z))
```

where we get the result p = 0.00471.

Since p < 0.01, we have strong evidence against the null, which is the population average daily reported case numbers between the two seven-day blocks is the same. The p-value suggests that the difference in average reported daily case numbers between the two seven-day blocks is not 0.

Question 2

1. First a set of y values from 0 to 25 is created.

```
y = (0:25)
```

Then, the values for v and r are declared, and the negative binomial probability for each value of y is calculated from the formula.

```
v1 = 0

r1 = 1

p1 = choose(y+r1-1,y)*(r1^r1)*((exp(v1)+r1)^(-r1-y))*exp(y*v1)
```

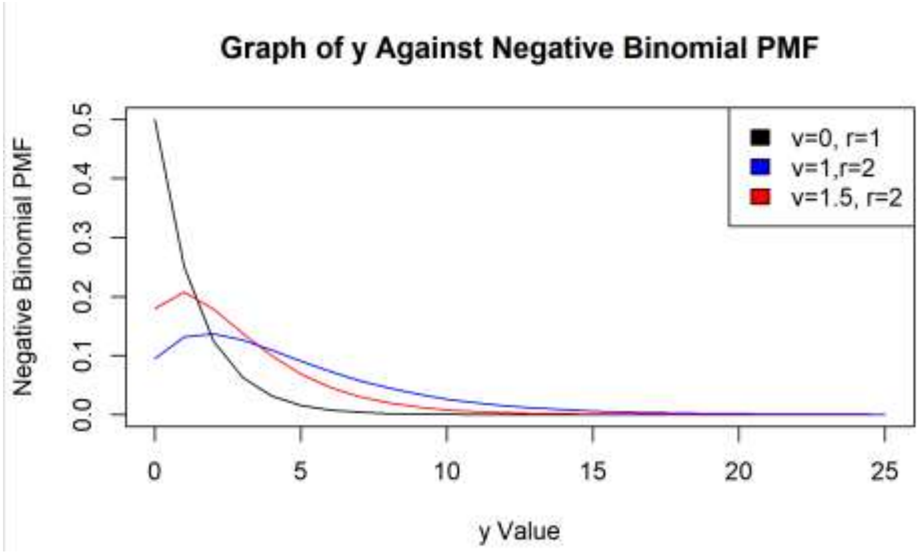
The graph is plotted using the plot() function.

```
plot(y,p, xlab="y Value", ylab="Negative Binomial PMF", main="Graph of y Against Negative Binomial PMF", type="l")
```

The steps are repeated for (v=1, r=2) and (v=1.5, r=2), and the lines and a legend are added onto the graph,

```
v2 = 1

r2 = 2

p2 = choose(y+r2-1,y)*(r2^r2)*((exp(v2)+r2)^(-r2-y))*exp(y*v2)

lines(y,p2, col="red")

v3 = 1.5

r3 = 2

p3 = choose(y+r3-1,y)*(r3^r3)*((exp(v3)+r3)^(-r3-y))*exp(y*v3)

lines(y,p3, col="blue")

legend(x = "topright", c("v=0, r=1", "v=1,r=2", "v=1.5, r=2"), fill=c("black","blue","red"))
```

Which yields the following graph.



2. L = $\prod_{i=1}^{n}\binom{yi+r-1}{yi}$ r^{r} $(e^{v} + r)^{-r-yi}$ e^{yi*v}

= $\binom{y1+r-1}{y1}$ r^{r} $(e^{v} + r)^{-r-y1}$ e^{y1*v} $*$ $\binom{y2+r-1}{y2}$ r^{r} $(e^{v} + r)^{-r-y2}$ e^{y2*v} $*$ $\binom{y3+r-1}{y3}$ r^{r} $(e^{v} + r)^{-r-y3}$ e^{y3*v} $*$ … * $\binom{yn+r-1}{yn}$ r^{r} $(e^{v} + r)^{-r-yn}$ e^{yn*v}

= $\prod_{i=1}^{n}\binom{yi+r-1}{yi}$ * [$\prod_{i=1}^{n}(\mathrm{e}^\wedge\mathrm{v}\ +\ \mathrm{r})^\wedge(-\mathrm{r}-\mathrm{yi})$] * [$\prod_{i=1}^{n}(\mathrm{e})^\wedge(\mathrm{yi}*\mathrm{v})$] * $(r^{r})^{n}$

= $\prod_{i=1}^{n}\binom{yi+r-1}{yi}$ * $(e^{v}+ r)$ ^ ($\sum_{i=1}^{n} -r - yi$) * e^ ($\sum_{i=1}^{n} yi * v$) * r^{rn}

= $\prod_{i=1}^{n}\binom{yi+r-1}{yi}$ * $(e^{v}+ r)$ ^ (-nr - $\sum_{i=1}^{n} yi$) * e^ (v * $\sum_{i=1}^{n} yi$) * r^{rn}

3. -log(L) = - log ($\prod_{i=1}^{n}\binom{yi+r-1}{yi}$ * $(e^{v}+ r)$ ^ (-nr - $\sum_{i=1}^{n} yi$) * e^ (v * $\sum_{i=1}^{n} yi$) * r^{rn})

= -log ($\prod_{i=1}^{n}\binom{yi+r-1}{yi}$) – log ((e^{v}+ r) ^ (-nr - $\sum_{i=1}^{n} yi$)) – log (e^ (v * $\sum_{i=1}^{n} yi$)) – log (r^{rn})

= -log ($\prod_{i=1}^{n}\binom{yi+r-1}{yi}$) - (-nr - $\sum_{i=1}^{n} yi$) log (e^{v}+ r) - (v * $\sum_{i=1}^{n} yi$) log e – rn log r

= -log ($\prod_{i=1}^{n}\binom{yi+r-1}{yi}$) + (nr + $\sum_{i=1}^{n} yi$) log (e^{v}+ r) - (v * $\sum_{i=1}^{n} yi$)(1) – rn log r

= -log ($\prod_{i=1}^{n}\binom{yi+r-1}{yi}$) + (nr + m) log (e^{v}+ r) – vm – rn log r [m = $\sum_{i=1}^{n} yi$]

4. d(-log(L))/dv = [(nr+m)/(e^{v}+r)] * e^{v} – m

Let d(-log(L))/dv = 0

[(nr+m)/(e^{v}+r)] * e^{v} = m

$nre^{v} + me^{v} = me^{v} + mr$

$nre^{v} = mr$

$e^{v} = m/n$

v log e = log (m/n)

v(1) = log m – log n

$\hat{v}$ = log m – log n [m = $\sum_{i=1}^{n} yi$]

$d/dx(\log(x)) = 1/x$

$d^2/dx^2(\log(x)) = -(1/x^2)$

5. $b(\hat{v}) = E[\ \hat{v}(Y)\ ] – v$

$= E[\log (nY/n)] - v$

$= E[\log(Y)] - v$

$= \log(\mu_X) + (\ d^2(\log(\mu_X))/d\mu_X^2\ ) * V[Y]/2 - v$

$= \log (E[Y]) – 1/ E[Y]^2 * V[Y]/2 - v$

$= \log (e^v) – 1/e^{2v} * e^v (e^v + r)/2r - v$

$= v \log (e) – (e^{2v} + e^v r)/2r*e^{2v}) - v$

$= v – (e^v + r)/2re^v – v$

$= (-e^v-r)/2re^v$

$V[\hat{v}(Y)] = [d(\log(\mu_X))/d\mu_X]^2\ V[Y]$

$= (1/ E[Y])^2 * e^v (e^v + r)/r$

$= e^v (e^v + r)/re^{2v}$

$= (e^v + r)/re^v$

Question 3

1. First the mean and variance of the data are calculated:

mean = $E[X] = 176/240 = 0.733$

variance = $E[X^2] – E[X]^2 = 0.733 – 0.537 = 0.196$

We get a mean of 0.733, and a variance of 0.196.

The t-value is calculated using the data, which is used to obtain the minimum and maximum 95% confidence intervals.

```
student_t = qt(p= 1-0.05/2, df=n)

CI_min = mu-student_t*sqrt(var/n)

CI_max = mu+student_t*sqrt(var/n)
```

From there we get a confidence interval of (0.6769852, 0.7896815).

The estimate of the preference for humans turning their heads to the right when kissing is 0.733. We are 95% confident that the population mean of people tilting their head to the right is between 0.6769852 and 0.7896815.

2. We are testing for the hypothesis $H_0 : \theta = \theta_0$, where θ_0 is 0.5 (i.e. there is no preference in humans for tilting their head to one particular side when kissing).

The estimate of the population success probability is

$\hat{\theta} = m/n = 176/240$

And the test statistic is the approximate z-score:

$z = (\theta - \theta_0)/\text{sqrt}(\theta_0*(1- \theta_0)/n)$

using R to calculate the p-value:

```
p = 2*pnorm(-abs(z))
```

Yields a p-value of 4.845e-13.

Since $p < 0.01$, we have strong evidence against the null, which suggests that there is a preference in humans for tilting their head to one particular side when kissing.

3. Using R, we can calculate the p-value using the t.test() function.

```
t.test(data, mu=0.5)
```

Which gives an exact p-value of 1.94977e-14. This p-value suggests that we have strong evidence against the null, meaning that there is a preference in humans for tilting their head to one particular side when kissing.

4. We are testing for the hypothesis $H_0 : \theta_x = \theta_y$, where θ_x is the population success probability of people turning their heads to the right when kissing; and θ_y is the population success probability of people being right-handed.

$\hat{\theta}_x = 176/240$, $\hat{\theta}_y = 210/240$

A pooled estimate of θ, $\hat{\theta}_p$ is calculated:

$\hat{\theta}_p = (m_x + m_y)/ (n_x + n_y) = (210+176)/(240+240) = 0.804$

The test statistic is:

$z = (\hat{\theta}_x - \hat{\theta}_y)/\text{sqrt}(\hat{\theta}_p (1 - \hat{\theta}_p)(1/n_x + 1/n_y))$

which gives us a value of z = -3.911

Calculating the p-value with R:

p = 2*pnorm(-abs(z))

Yielding a p-value of 9.191477e-05.

Since $p < 0.01$, we have strong evidence against the null, which suggests that the preference for head turning to the right/left is not a product of right/left-handedness.