QUESTION 1

1.1

```
covid1 = read.csv("daily.covid.aug1to7.csv")
covid1.mean = mean(covid1$daily.covid.cases)
#The sample mean is 7359.571 cases.

qt(p = 1-0.05/2, df = 7-1)
#This result is 2.446912
```

$\hat{\mu}_A = 7359.571$] From R

$t_{\alpha/2, n-1} \approx 2.447$

$\hat{\sigma}^2_A = \frac{1}{6}\sum_{i=1}^{7}(y_i - 7359.571)^2$ $a = 7359.571$

$= \frac{1}{6}\left[(10038-a)^2 + (9060-a)^2 + (8756-a)^2 + (7456-a)^2 + (6274-a)^2 + (5084-a)^2 + (4878-a)^2\right]$

$= \frac{24650399.71}{6} \Rightarrow 4108399.952$

$CI: \left(7359.571 - 2.44\sqrt{\frac{4108399.952}{7}},\ 7359.571 + 2.44\sqrt{\frac{4108399.952}{7}}\right)$

$= (5490.279, 9228.863)$

The estimated mean value for covid cases in August (sample size = 7) is 7359.571 cases. We are 95% confident that the population mean for the cases is between 5490.279 and 9228.863 cases.

1.2

```
covid2 = read.csv("daily.covid.aug8to14.csv")
covid2.mean = mean(covid2$daily.covid.cases)
#The sample mean is 4879 cases.

qt(p = 1-0.05/2, df = 7-1)
#This result is 2.446912
```

$\hat{\mu}_B = 4879$

$t_{\alpha/2, n-1} \approx 2.447$] From R

$\hat{\sigma}_B^2 = \frac{1}{6}\sum_{i=1}^{7}(y_i - 4879)$ $\qquad b = 4879$

$= \frac{1}{6}\left[(6348-b)^2 + (5857-b)^2 + (5518-b)^2 + (5135-b)^2 + (4274-b)^2 + (3419-b)^2 + (3602-b)^2\right]$

$= \frac{7716656}{6} \Rightarrow 1286109.333$

$\hat{\mu}_A - \hat{\mu}_B = 7359.571 - 4879 = 2480.571$

estimated mean difference

CI difference in means:

$$\left(\hat{\mu}_A - \hat{\mu}_B - z_{\alpha/2}\sqrt{\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}},\ \hat{\mu}_A - \hat{\mu}_B + z_{\alpha/2}\sqrt{\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}}\right)$$

$$= \left(2480.571 - 1.96\sqrt{\frac{4108399.952}{7} + \frac{1286109.333}{7}},\ 2480.571 + 1.96\sqrt{\frac{4108399.952}{7} + \frac{1286109.333}{7}}\right)$$

$= (759.959, 4201.183)$

The estimated difference in mean of covid cases between August 1 to 7 and 8 to 14 (both size $n = 7$) is 2480.571. We are 95% confident that the population mean difference in covid cases between these 2 groups is between 759.959 and 4201.183. As both ends of the interval are positive and far from 0, there are definitely significance in covid cases after 7th of August.

1.3

$H_0 : \mu_x = \mu_y$

vs

$H_1 : \mu_x \neq \mu_y$,

H_0: Population average daily reported cases between the two groups is the same.
H_1: Population average daily reported cases between the two groups is not the same.

$$Z = \frac{\hat{\mu}_1 - \hat{\mu}_y}{\sqrt{\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}}} = \frac{7480.271}{\sqrt{\frac{4108399.952}{7} + \frac{1286109.333}{7}}} = 7.825$$

$2P(Z < -2.825) = 2 * \text{pnorm}(-2.825)$ ← in R

$\simeq 0.00472$

```
#p-value
2*pnorm(-2.825)
#p-value = 0.00472
```

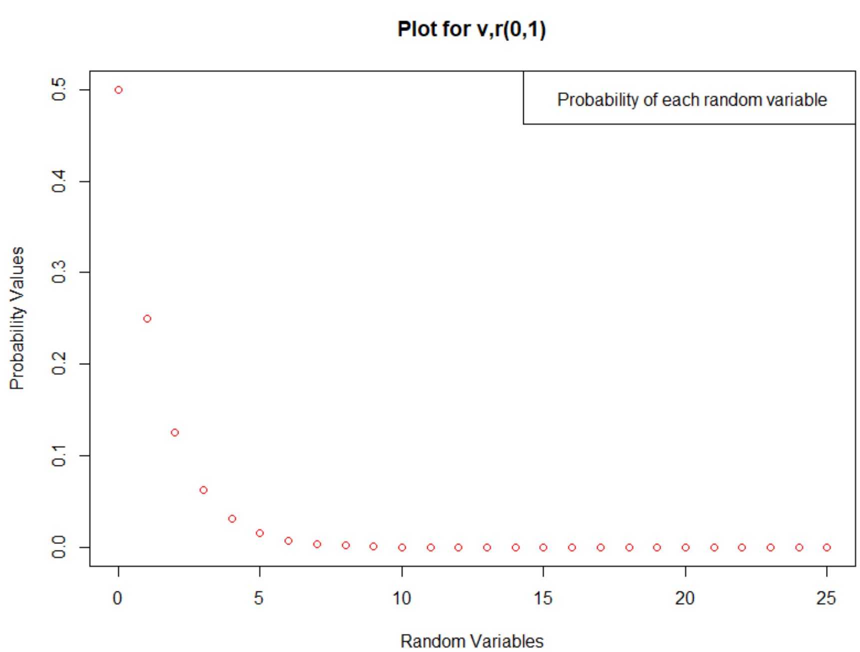
$0.00472 < 0.05$ → Strong evidence against null hypothesis that there is no difference between population

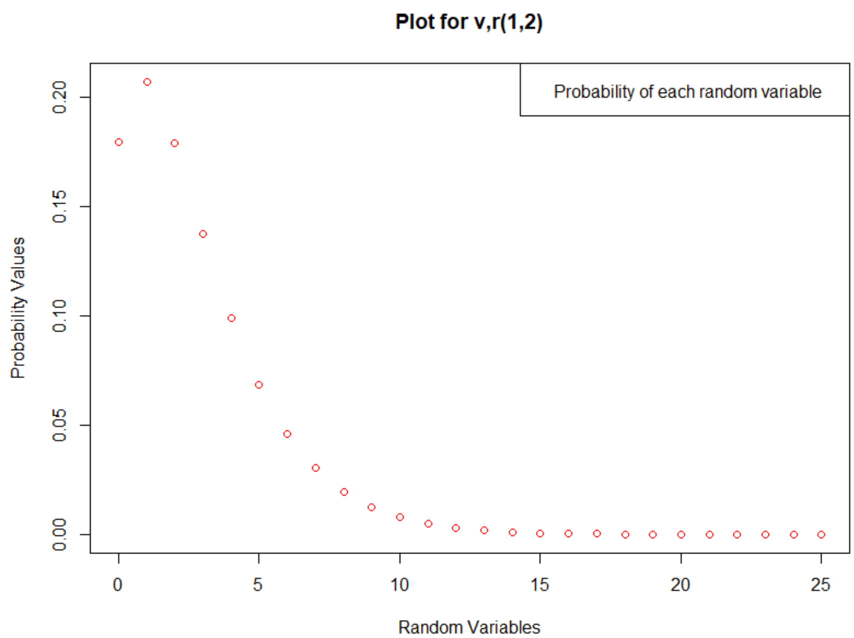
This p-value indicates that the changes in the daily reported cases is incredibly unlikely to happen just by chance. We accept the alternative hypothesis, that the average daily reported cases between the two groups are not the same.

QUESTION 2

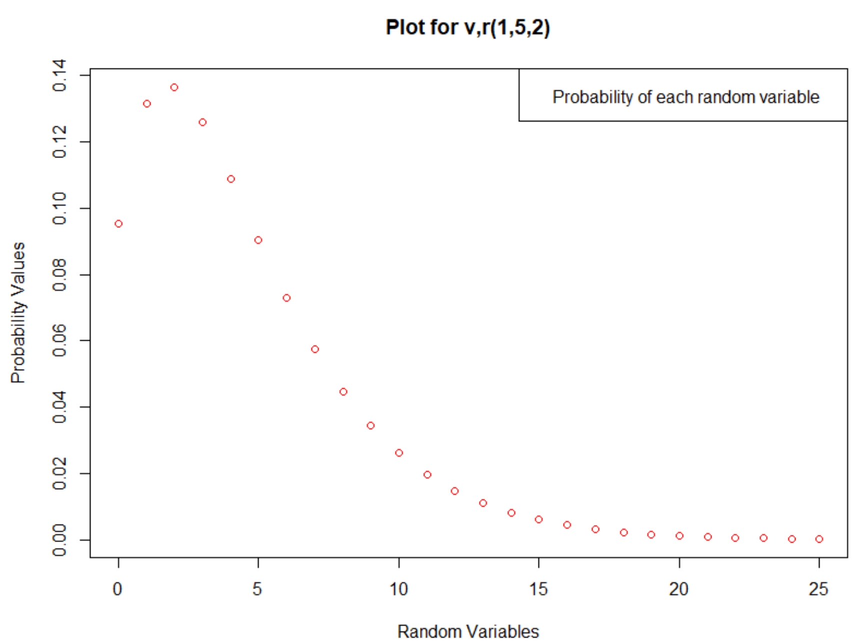
2.1

```r
#QUESTION 2.1

x=seq(0,25)#The x values
y=rep(0,26)  #The y values

#For v,r(0,1)
v = 0
r = 1
for(i in 1:26)
{
  y[i] = choose((x[i]+r-1),x[i])*(r^r)*((exp(v)+r)^(-r-x[i]))*exp(x[i]*v)
}

plot(x,y,main="Plot for v,r(0,1)", ylab = "Probability Values", xlab = "Random Variables", col= "red")
legend(x = "topright",legend = "Probability of each random variable")

#For v,r(1,2)
v = 1
r = 2

for(i in 1:26)
{
  y[i] = choose((x[i]+r-1),x[i])*(r^r)*((exp(v)+r)^(-r-x[i]))*exp(x[i]*v)
}

plot(x,y,main="Plot for v,r(1,2)", ylab = "Probability Values", xlab = "Random Variables", col= "red")
legend(x = "topright",legend = "Probability of each random variable")

#For v,r(1.5,2)
v = 1.5
r = 2

for(i in 1:26)
{
  y[i] = choose((x[i]+r-1),x[i])*(r^r)*((exp(v)+r)^(-r-x[i]))*exp(x[i]*v)
}

plot(x,y,main="Plot for v,r(1,5,2)", ylab = "Probability Values", xlab = "Random Variables", col= "red")
legend(x = "topright",legend = "Probability of each random variable")
```





Plot for v,r(1,2)
Probability of each random variable
Probability Values
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
25
Random Variables



Plot for v,r(1,5,2)
Probability of each random variable
Probability Values
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0
5
10
15
20
25
Random Variables

2.2

All samples are independent and identically distributed

$$Pr[Y_1=y_1]\cdot Pr[Y_2=y_2]\cdot Pr[Y_3=y_3]\cdot \ldots \cdot Pr[Y_n=y_n]$$

$$=\left[\binom{y_1+r-1}{y_1} r^r (e^v+r)^{-r-y_1} e^{y_1 v}\right]\cdot \ldots \cdot\left[\binom{y_n+r-1}{y_n} r^r (e^v+r)^{-r-y_n} e^{y_n v}\right]$$

$$=\prod_{i=1}^{n}\left[\binom{y_i+r-1}{y_i} r^r\cdot (e^v+r)^{-r-y_i} e^{y_i v}\right]$$

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

2.3

$$L(y\mid v,r) = -\log\left[r^{nr}\cdot (e^v+r)^{-nr-\sum_{i=1}^{n} y_i} e^{v\sum_{i=1}^{n} y_i} \prod_{i=1}^{n}\binom{y_i+r-1}{y_i}\right]$$

$$= -\left[\log(r)^{nr} \log(e^v+r)^{-nr-\sum_{i=1}^{n} y_i} \log(e)^{v\sum_{i=1}^{n} y_i} \log\left(\prod_{i=1}^{n}\binom{y_i+r-1}{y_i}\right)\right]$$

$$= -\left[nr\log(r) - nr\log(e^v+r) - \sum_{i=1}^{n} y_i \log(e^v+r) + v\sum_{i=1}^{n} y_i + \log\left(\prod_{i=1}^{n}\binom{y_i+r-1}{y_i}\right)\right.$$

$$L(y\mid v,r) = -nr\log(r) + nr\log(e^v+r) + \sum_{i=1}^{n} y_i \log(e^v+r) - v\sum_{i=1}^{n} y_i - \log\left(\prod_{i=1}^{n}\binom{y_i+r-1}{y_i}\right)$$

2.4

let $\sum_{i=1}^{n} y_i = n\bar{y}$

$$L(y|v,r) = -nr\log(r) + nr\log(e^v+r) + n\bar{y}\log(e^v+r) - n\bar{y}v - \log\left(\prod_{i=1}^{n}\binom{y_i+r-1}{y_i}\right)$$

$$\frac{\partial L(y|v,r)}{\partial v} = 0$$

$$0 + \frac{\partial}{\partial v}\left(nr\log(e^v+r)\right) + \frac{\partial}{\partial v}\left(n\bar{y}\log(e^v+r)\right) - \frac{\partial}{\partial v}(n\bar{y}v) - 0 = 0$$

$$\frac{nre^{\hat{v}}}{e^{\hat{v}}+r} + \frac{n\bar{y}e^{\hat{v}}}{e^{\hat{v}}+r} - n\bar{y} = 0$$

$$nre^{\hat{v}} + n\bar{y}e^{\hat{v}} - n\bar{y}(e^{\hat{v}}+r) = 0$$

$$nre^{\hat{v}} + \cancel{n\bar{y}e^{\hat{v}}} - \cancel{n\bar{y}e^{\hat{v}}} - rn\bar{y} = 0$$

$$nr(e^{\hat{v}} - \bar{y}) = 0$$

$$e^{\hat{v}} - \bar{y} = 0$$

$$e^{\hat{v}} = \bar{y}$$

$$\hat{v} = \ln(\bar{y})$$

2.5

$$\text{Bias} = E[\hat{\theta}(Y)] - \theta$$
$$= E[\hat{v}] - v$$

Variance: $V[\hat{v}]$

QUESTION 3

3.1

$\hat{\theta}_{ML} = \frac{176}{240} = 11/15 \qquad (1-\hat{\theta}_{ML}) = 1 - 11/15 = 4/15$

$n = 240$

$$CI = \left(\frac{11}{15} - 1.96\sqrt{\frac{\frac{11}{15}\left(\frac{4}{15}\right)}{240}} \, , \, \frac{11}{15} + 1.96\sqrt{\frac{\frac{11}{15}\left(\frac{4}{15}\right)}{240}} \right)$$

$$= (0.677, 0.789)$$

In our sample of $n=240$, the observed probability of humans turning their head to the right when kissing is 11/15. We are 95% confident that the true population probability of someone turning their head to the right is between 0.677 and 0.789 (which are biased towards turning right). This shows that there is a large probability of people turning their head to the right when kissing.

3.2

$H_0: \theta = 0.5$

$H_1: \theta \neq 0.5$

H_0: There is no preference in humans for tilting their head to a side when kissing.

H_1: There is a preference in humans for tilting their head to a particular side when kissing.

$$Z_\theta = \frac{\frac{176}{240} - \frac{1}{2}}{\sqrt{\frac{\frac{1}{2}(1-\frac{1}{2})}{240}}} \approx 7.23$$

approximation p-value in R:

$2P(Z < -7.23) = 2 * \text{pnorm}(-\text{abs}(7.23))$

```
> 2*pnorm(-abs(7.23))
[1] 4.82994e-13
```

$\approx 4.82994 \times 10^{-13}$

p-value of 4.82994×10^{-13} which means that if the null hypothesis is true, and $n = 240$, then 4.82994×10^{-11} % of the time 176 or less people turn right or 64 or more turning right when kissing. This is very strong evidence against the null.

3.3

```
> binom.test(176, 240, p=0.5)

        Exact binomial test

data:  176 and 240
number of successes = 176, number of trials = 240, p-value = 2.854e-13
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.6726355 0.7881673
sample estimates:
probability of success
             0.7333333
```

Use binom.test in R

The same conclusion as in Q3.2 but the p-value is $2.854e^{-13}$ instead of 4.82994×10^{-13}. This is also a strong evidence against the null hypothesis. The larger the sample size will result in the two p-values to be closer because normal approximation would be better.

3.4

$m_x = 176 \quad m_y = 210 \quad \hat{\theta}_x = 176/240$

$n_x = 240 \quad n_y = 240 \quad \hat{\theta}_y = 210/240$

$$\hat{\theta}_p = \frac{176+210}{240+240} \approx 0.783$$

$$Z(\hat{\theta}_x - \hat{\theta}_y) = \frac{176/240 - 210/240}{\sqrt{0.783(1-0.783)\left(\frac{1}{176/240} + \frac{1}{214/240}\right)}} \approx 0.731$$

use pnorm in R

p-value = 2 * pnorm(-0.731) $\approx$ 0.465

```
> 2*pnorm(-0.731)
[1] 0.4647792
```

$H_0 : \theta_x = \theta_y$, handedness does not affect the head turn while kissing

$H_1 : \theta_x \neq \theta_y$, handedness does affect the head turn while kissing

The p-value suggests that if we repeated the experiment and handedness does not affect the head turn. then approximately 46.5% of the 240 people's outcome would result in a difference as great as we observed, or greater, just by chance.
This means we have strong evidence against the null hypothesis.