A4 - Q3 (Ram Yoogesh Gopu - 20867060)

loading the data

```r
pandemic <- read.csv("pandemic.csv")
#pandemic
## the data set has 100 rows with 3 columns a, b and c respectively
```

(A)

(i)

```r
t.test(pandemic$A, mu = 94)
```

```
##
##  One Sample t-test
##
## data:  pandemic$A
## t = 7.7033, df = 99, p-value = 1.031e-11
## alternative hypothesis: true mean is not equal to 94
## 95 percent confidence interval:
##  94.51227 94.86773
## sample estimates:
## mean of x
##     94.69
```

```r
t.test(pandemic$B, mu = 94)
```

```
##
##  One Sample t-test
##
## data:  pandemic$B
## t = 2.306, df = 99, p-value = 0.0232
## alternative hypothesis: true mean is not equal to 94
## 95 percent confidence interval:
##  94.10646 95.41954
## sample estimates:
## mean of x
##    94.763
```

```r
t.test(pandemic$C, mu = 94)
```

```
##
##  One Sample t-test
##
## data:  pandemic$C
## t = 4.8342, df = 99, p-value = 4.897e-06
## alternative hypothesis: true mean is not equal to 94
## 95 percent confidence interval:
##  94.90024 96.15376
## sample estimates:
## mean of x
##    95.527
```

So here, we pass 2 arguments for t.test(). In which first argument defines the columns from the dataset (Typically A, B and C - which are just vector values with respect to the 100 cities) . Second argument defines mean value. we ended up choosing 94 using the value provided in the question (6%, whih is the fatality rate). hence it leaves us 94.

from test A, we have a really low p-value and t = 7.703. from this we can infer that the rate of recovery is high. Its advisable to do test A rather than doing nothing

from test B, p-value is quite high and t = 2.306. from this we can infer that the rate of recovery is low. Its better to do test B rather than doing nothing , though it is not highly recomended as test A.

from test C, we have a low p-value and t = 4.834. From this we can infer that the rate of recovery is high. Henc, its advisable to do test C rather than doing nothing.

From the above 3 tests (A, B and C), its highly advisable to carry out test A since it has really low p-valu e(WHich means that the recovery rate is pretty high). Followed by A, we can do B and C. Though C is not highly advisbale, its better than nothing. But we must consider the following orderA > B > C > Doing nothing.

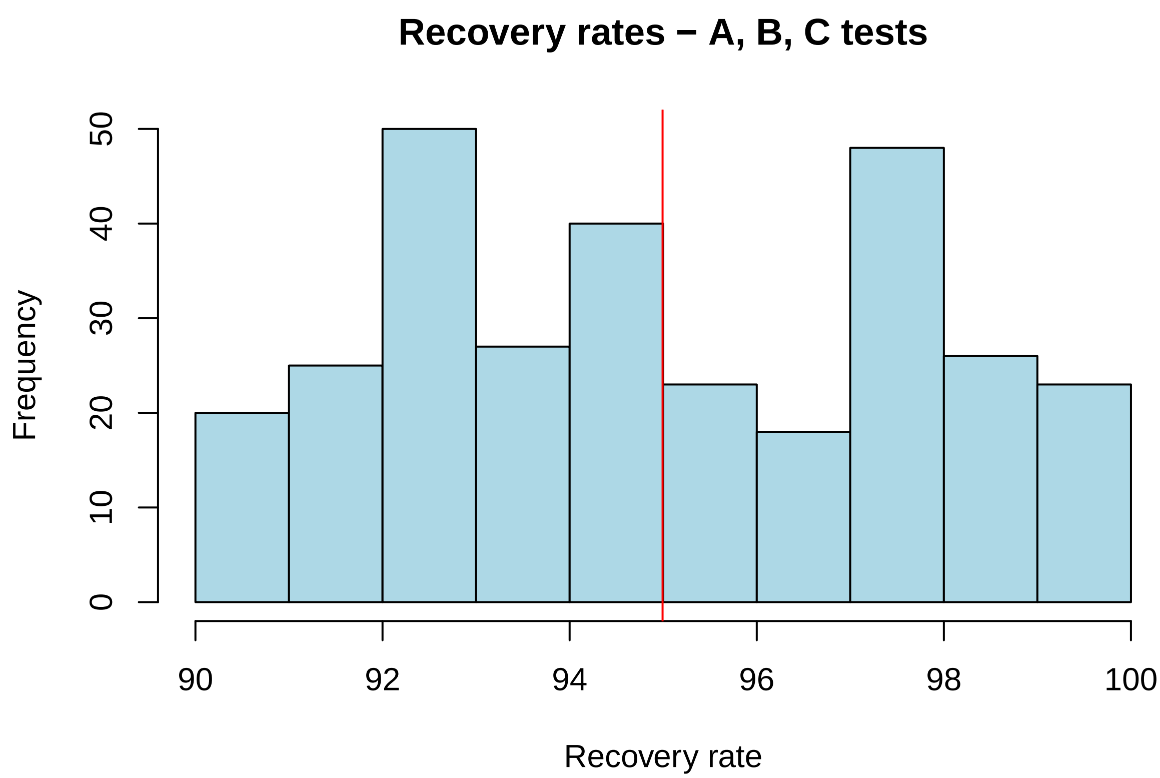
(ii)

From the previous order, i would highly suggest doing test A and it will work more practically. Since A is more significant in saving lives, the number of lives saved will be high when compared to the other 2 tests irrespective of the fatality rate.

(iii)

```r
recovery <- c (pandemic$A, pandemic$B, pandemic$C)


hist(recovery, main = "Recovery rates - A, B, C tests", xlab = "Recovery rate", col = "lightblue" )
abline(v = mean(recovery), col = 'red', lwd = 1, lty = 1)
```



(iv)

I would like to know more about the underlying health states (Whether they have issues such as heart problems,lung issues, kidney stones, blood pressure, obesity, lipids level), based on all these we could try to find a suaitable correlation and also can come to a conclusion taht which of the tests would have a better performance.

(b)

(i)

```
fraction <- sum(pandemic$A > pandemic$B) * (100 / length(pandemic$A))
print("fraction of cities which have higher recovery rates for A than for B" )
```

```
## [1] "fraction of cities which have higher recovery rates for A than for B"
```

```
print(fraction)
```

```
## [1] 66
```

From the above value, this propotion suggest us to choose A since it seems to be more affective than B in many cities.

(ii)

```
fraction <- sum(pandemic$B > pandemic$C) * (100 / length(pandemic$B))
print("fraction of cities which have higher recovery rates for B than for C" )
```

```
## [1] "fraction of cities which have higher recovery rates for B than for C"
```

```
print(fraction)
```

```
## [1] 59
```

This implies that we should choose treament B over C.

(iii)

```
fraction <- sum(pandemic$C > pandemic$A) * (100 / length(pandemic$C))
print("fraction of cities which have higher recovery rates for C than for A" )
```

```
## [1] "fraction of cities which have higher recovery rates for C than for A"
```

```
print(fraction)
```

```
## [1] 70
```

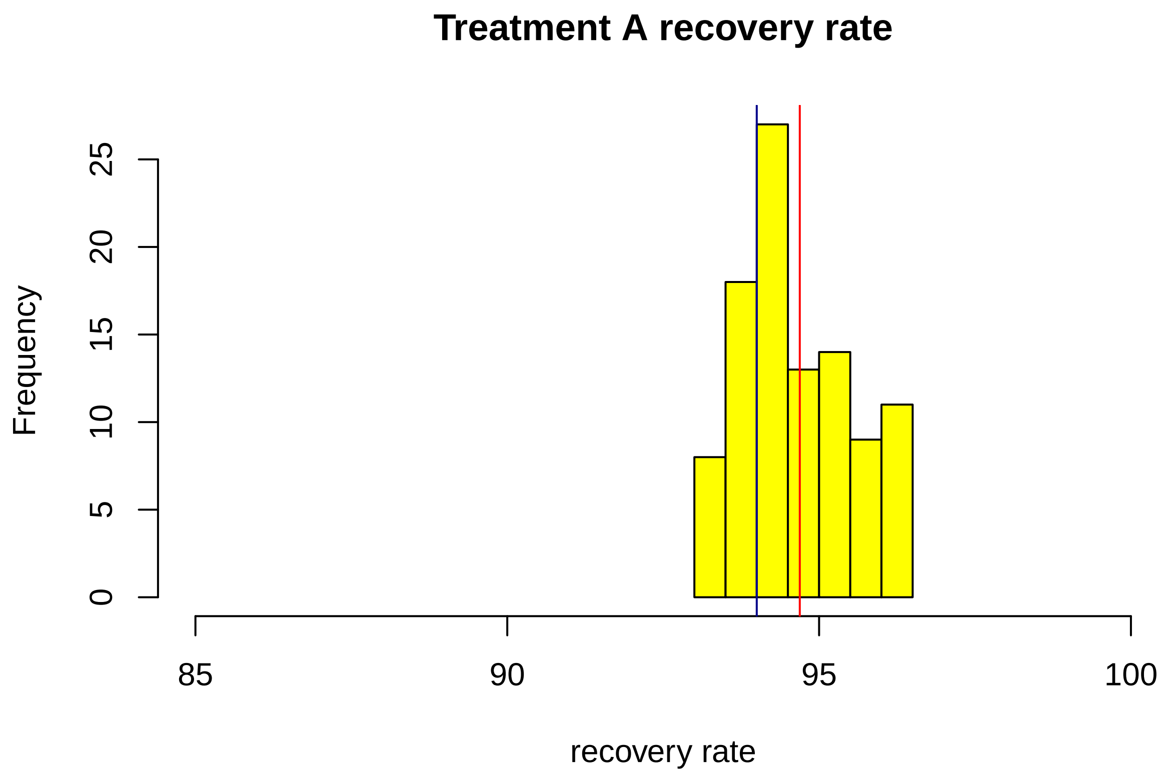
This implies that, its advisable to use treatment C over A

(iv)

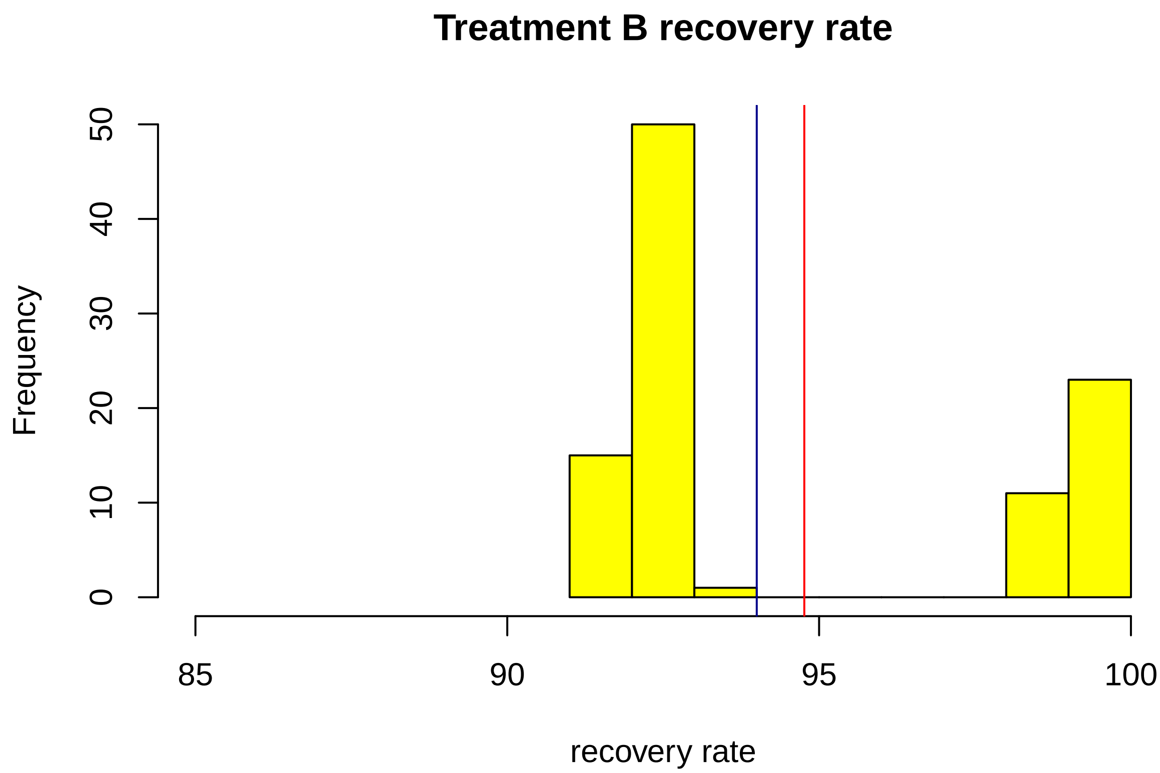
I would like to conclude that treatment C is tend to be more effective when compared to the treatments A and B. I would rank the treatments in the folowing order 1)C 2)B 3)A, where 1 being the highest

(C)

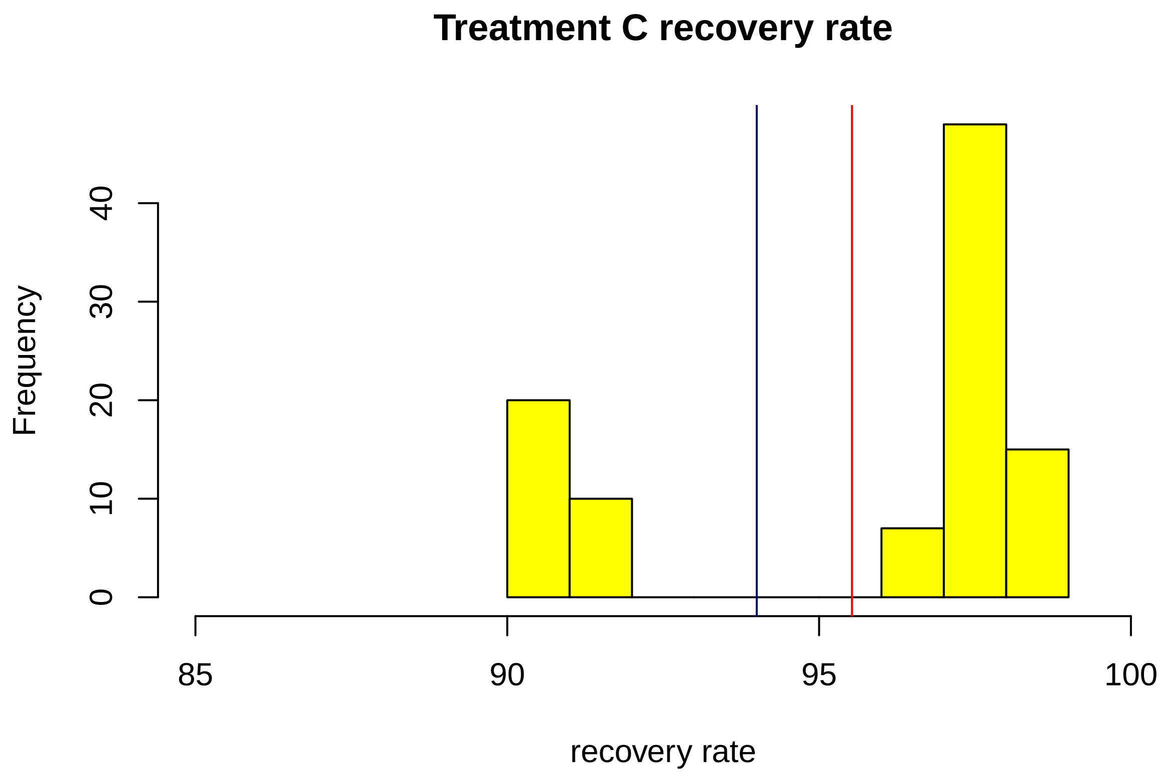
```
hist(pandemic$A, main = "Treatment A recovery rate", xlab = "recovery rate", col ="yellow", xlim = c(85
abline (v = mean(pandemic$A), col='red', lty = 1, lwd = 1)
abline(v = 94, col = 'darkblue')
```




```
hist(pandemic$B, main = "Treatment B recovery rate", xlab = "recovery rate", col ="yellow", xlim = c(85
abline (v = mean(pandemic$B), col='red', lty = 1, lwd = 1)
abline(v = 94, col = 'darkblue')
```




```
hist(pandemic$C, main = "Treatment C recovery rate", xlab = "recovery rate", col ="yellow", xlim = c(85
abline (v = mean(pandemic$C), col='red', lty = 1, lwd = 1)
abline(v = 94, col = 'darkblue')
```



(D)

From the above numbers and graphical representations i would suggest the health scientists worldwide to adopt a treatment based on thier region. Some region can witness C working well , other can witness A working well. So there is no perfect treatment to suggest as a common one for all the cities. But from the evidence, we can say that C tends to be far more better than A and B(but it is not a perfect choice for all the regions in common). So if the scientists are unsure what they have to do, they can choose C on the first place and monitor the recovery rates. If the recovery rates are slow, then they can choose the other methods.