MIDDLE EAST TECHNICAL UNIVERSITY INDUSTRIAL
ENGINEERING DEPARTMENT

Mesude GÜRÜN 2519262

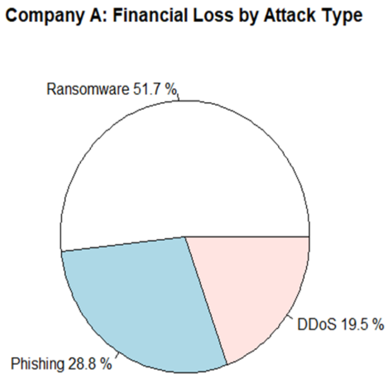
Part A

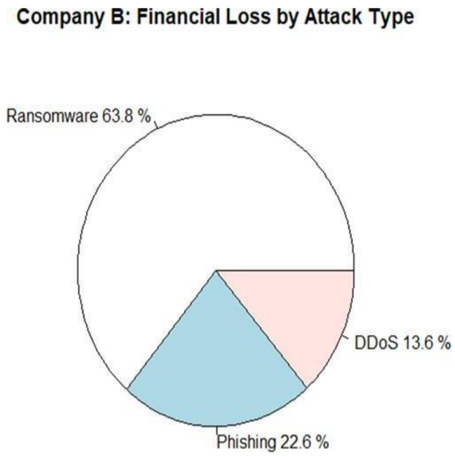
Q1)

According to the analysis have been made the financial loss that is caused by the responsibility of Company A over 52 weeks is $901,997.33 İn total. The losses caused by Ransomware, DDoS, and phishing attacks are 51.7%, 19.5%, and 28.8% of total loss accordingly as it can be observed from the pie chart.

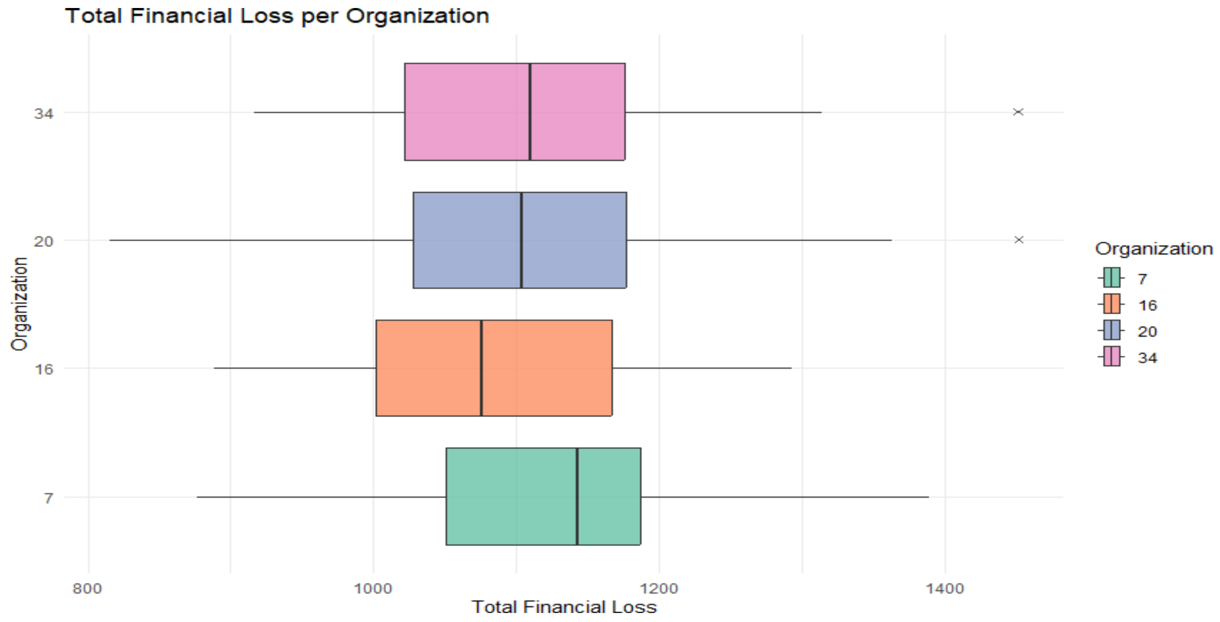
According to analysis have been made the financial loss that is caused by the responsibility of Company B over 52 weeks is $1,139,685.61 in total. The losses caused by Ransomware, DDoS, and phishing attacks are 63.8%, 13.6%, and 22.6% of total loss accordingly as it can be observed from the pie chart.

For both companies, ransomware is the dominant cyberattack in terms of causing financial loss with almost ⅔ of total attacks that are on the responsibility of Company B and more than half of the total on the responsibility of Company A. Furthermore, it can be observed that while Company B is more vulnerable than Company A to ransomware type of attacks, it is vice-versa for DDoS and phishing types of attacks. The pie chart is used in order to demonstrate the percentage contributions of total attacks for companies.





Q2

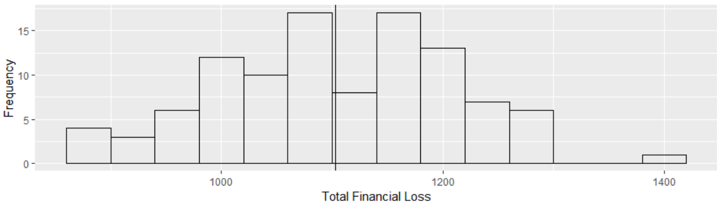


In this question, first we discuss Total Financial Loss for each selected organization, and then compare Company A and Company B.

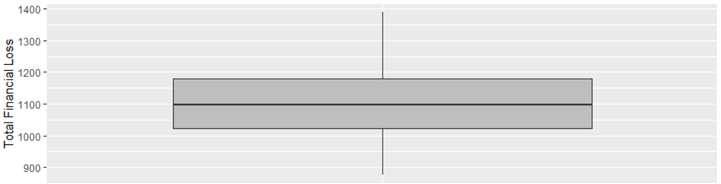
Organization 7 has the highest median, highest typical weekly loss, and Organization 16 has the smallest median, lowest typical weekly loss. Median values of Organizations which belongs to Company B (20 and 34) fall between Company A's organizations' median values (7 and 16).

Organization 16 has the smallest standart deviation which implies that its weekly financial loss vary less, and Organization 20 has the greatest standart deviation, greatest variability in weekly financial loss.

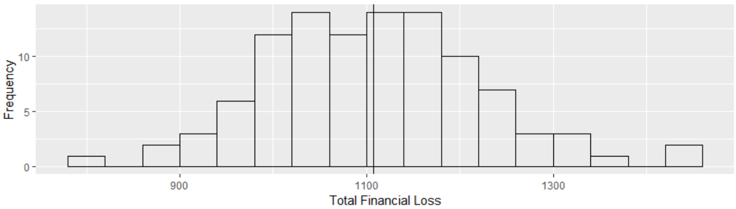
Company B also has the smallest minimum value and highest maximum value, which also contributes to its variability.



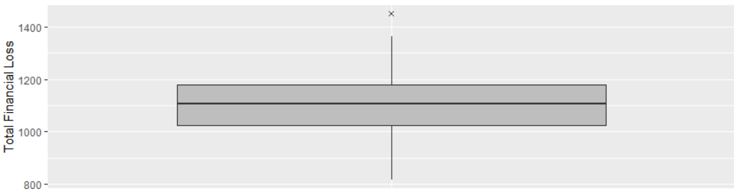
Company A - Histogram



Company A - Boxplot



Company B - Histogram



Company B - Boxplot

Total losses for company A for selected organizations (organizations 7 and 16) is $114,714.60and total losses for company B (organizations 20 and 34) is $115,201.50. The total loss of company B is slightly higher than that of company A, so it can be said that company A looks better at first sight.

In terms of median and mean, Company B also has a slightly higher median and mean compared to Company A, which implies that Company B loses more on average and in a typical week. Company B has a slightly higher mean compared to Company A, which implies that its average weekly loss is greater than Company A. Even though Company B has an outlier, it still has a slightly higher median compared to Company A.

Company B has both lower minimum and higher maximum losses, and also one outlier above the upper whisker, which is also supported by a higher standard deviation. So, after comparing median and mean values, it can be concluded that company A looks better. Also, lower standard deviation of Company A contributes to that idea since it implies less volatility.

Q3)

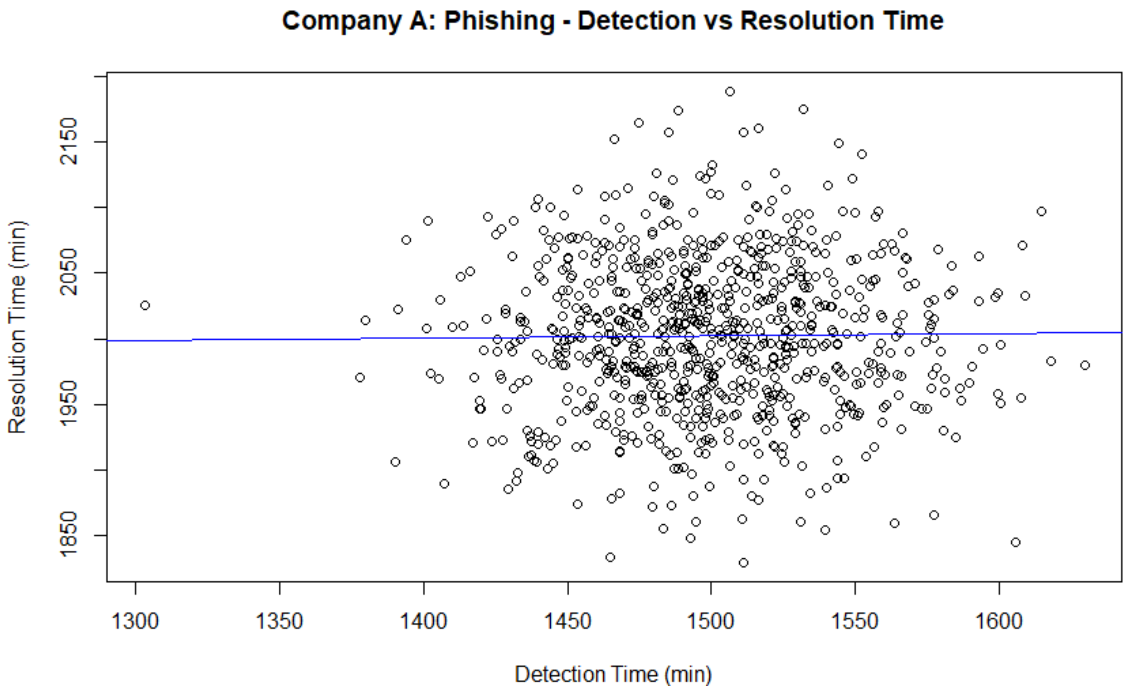
In this section, a correlation analysis has been conducted for both Company A and Company B to investigate the correlation between the detection time and the resolution time for phishing attacks. The weekly data for 36 organizations over 52 weeks has been used for this investigation.

Company A:

A scatterplot has been drawn to visualize the correlation between the detection time and the resolution time for phishing in Company A. However, there is no strong linear pattern observed among data points on the scatterplot at first sight.

Then, the initial impression is supported by the Pearson Correlation Coefficient, which is found to be as low as 0.016. Also, the regression line overlaid on the scatterplot is found to be nearly horizontal with a very small positive angle degree with the x-axis.

Consequently, our investigations indicate that there is a very small positive correlation between detection and resolution time for phishing attacks which can be negligible in Company A.

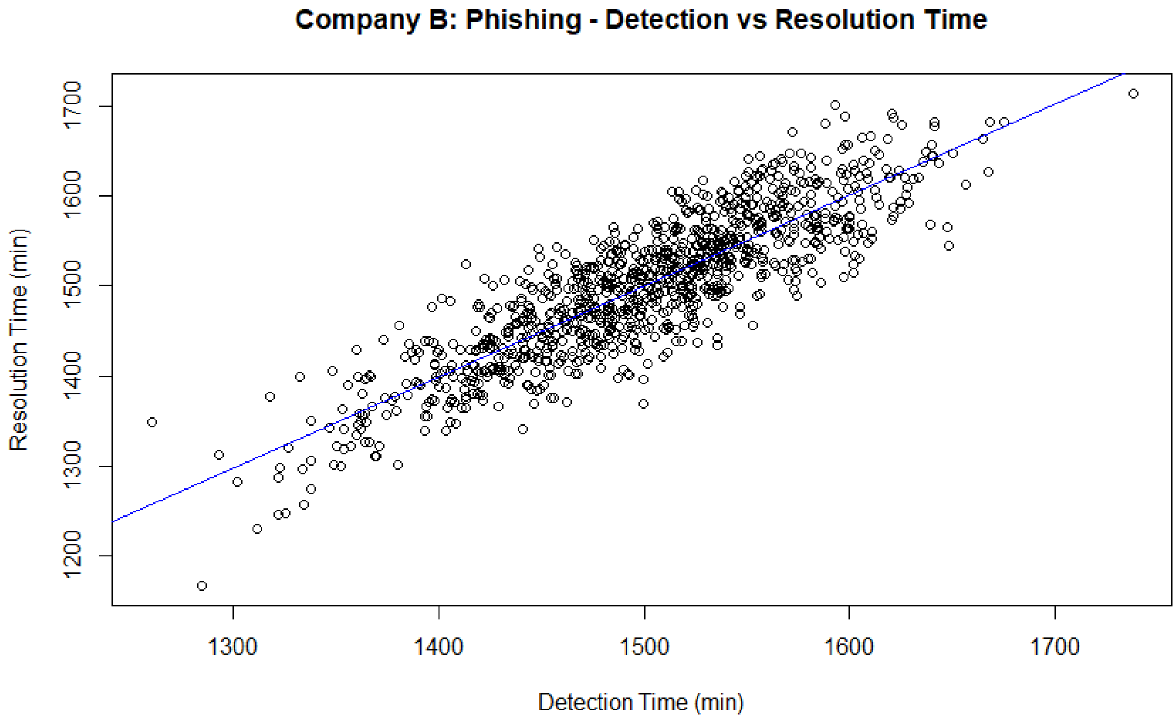


Company B:

A scatterplot has been drawn to visualize the correlation between the detection time and the resolution time for phishing in Company B. As a result, there is a strong linear pattern observed among data points on the scatterplot at first sight.

Then, the initial impression is supported by the Pearson Correlation Coefficient, which is found to be as high as 0.869. Also, the regression line overlaid on the scatterplot is found to be nearly diagonal with a degree that is a little lower than 45 degrees with the x-axis.

Consequently, our investigations indicate that there is a positive correlation between detection and resolution time for phishing attacks in Company B.

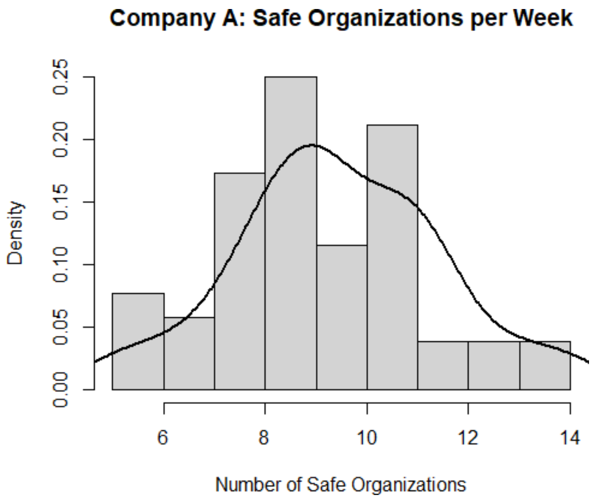


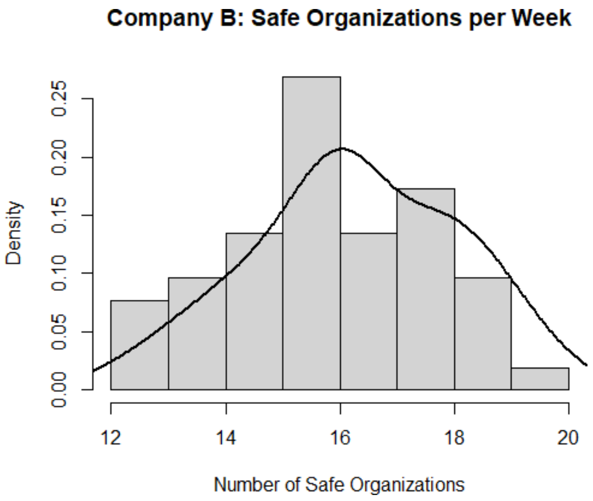
Q4)

In this question, the statistic, the average number of safe organizations per week for Company A and Company B will be discussed. We would expect normal distribution since the sample size is large enough. Distributions will be compared based on kurtosis and skewness.

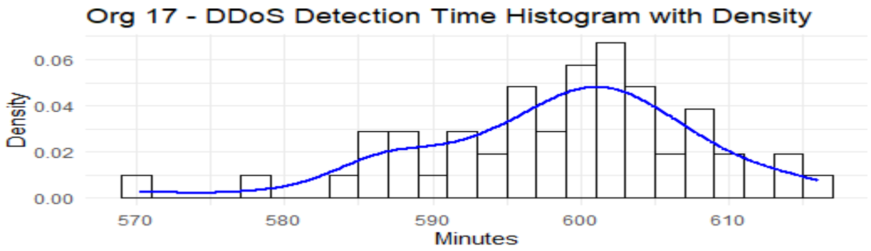
As it can be seen from the histogram and with the help of the density line, the data is spread and slightly right-skewed. Kurtosis value for Company A equals 3.09, very close to 3, which is the kurtosis value of normal distribution. Therefore, we can say that distribution obtained from the data supports the conjecture.

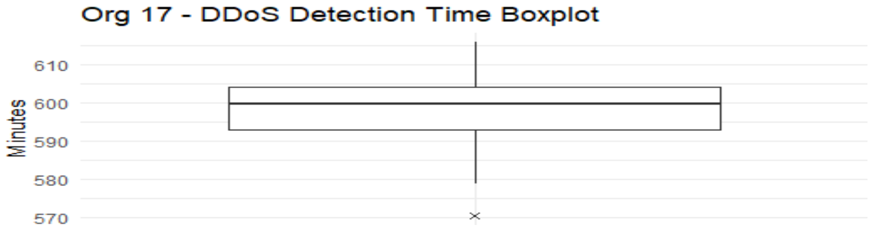
On the other hand, for Company B, the data is slightly left skewed, with a kurtosis value of 2.54. This value indicates that Company B has more flat distribution than normal distribution and also Company A's distribution.





Q5)

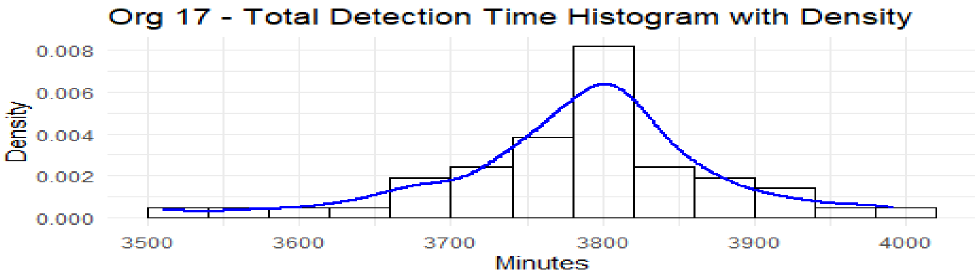


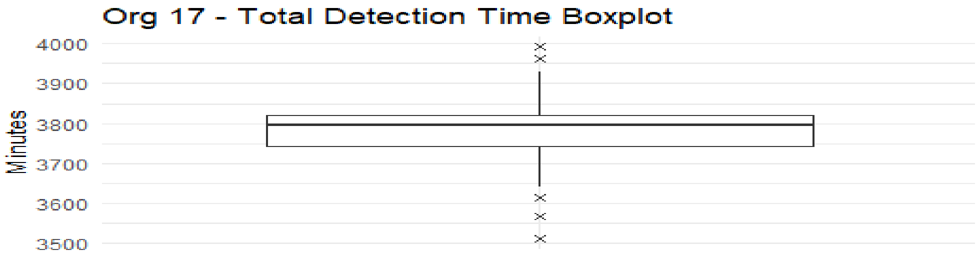


For total detection time DDoS attacks, the median is slightly higher than the median value. Yet, the mean and the median values are very close, with mean = 598.5 and median = 599.8, which indicates that the data set is approximately symmetric distribution with a mild left skew. The kurtosis value of the distribution is calculated as 3.61, which indicates that the peak is sharper than the normal distribution.

The Interquartile Range is 11.1 minutes, which implies half of the weeks lie within an 11.1-minute span. The difference between minimum and maximum values equals 45.8 minutes, which indicates that most extreme weeks differ by 45.8 minutes.

As it can be seen from the histogram with the help of the density line, the distribution of the weekly total detection time DDoS Attacks is left skewed, and not visually similar to normal distribution. Hence, the normality assumption fails for total detection time DDoS attacks.

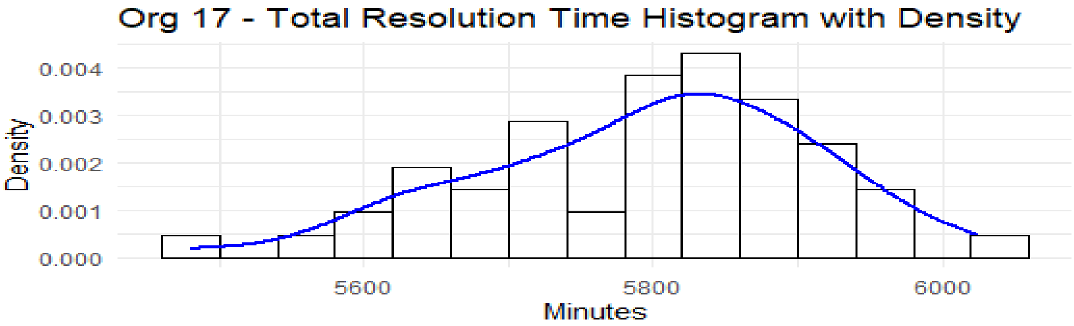


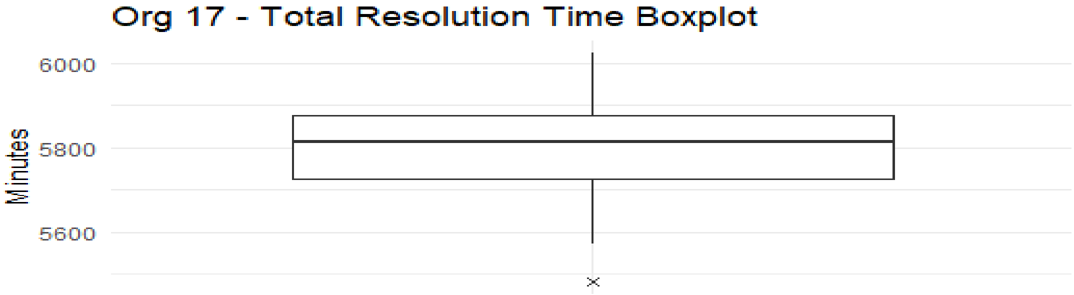


For Total Detection Time, the median is slightly higher than the mean value, yet again, the mean and the median values are very close, which indicates that the data set is approximately symmetric distribution with a mild left skew. The kurtosis value of the distribution is calculated as 4.16, which indicates that the peak is sharper than the normal distribution.

The Interquartile Range is 80 minutes, which implies half of the weeks lie within an 80- minute span. The difference between minimum and maximum values equals 481 minutes, which indicates that most extreme weeks differ by more than 8 hours.

As it can be seen from the histogram with the help of the density line, the distribution of the weekly total detection time is slightly left skewed, yet it is visually close to normal distribution. Hence, the normality assumption holds for total detection time aggregated across all three attack types (ransomware, phishing, and DDoS).





For Total Resolution Time, the median is slightly higher than the mean value, yet again, the mean and the median values are very close, which indicates that the data set is approximately symmetric distribution with a mild left skew. The kurtosis value of the distribution is calculated as 2.87, which indicates that the peak is slightly flatter than the normal distribution.

The Interquartile Range is 153 minutes, which implies half of the weeks lie within a 153- minute span. The difference between minimum and maximum values equals 543 minutes, which indicates that most extreme weeks differ by more than 9 hours.

As it can be seen from the histogram with the help of the density line, the distribution of the total resolution time is left skewed, and not visually similar to normal distribution. Hence, the normality assumption fails for total resolution time.

Part B

Q1)

In order to compare means, null hypothesis and alternative hypothesis are formed. Null hypothesis suggests that mean of Company A and Company B are equal. Alternative hypothesis suggests opposite, means are not equal. Confidence intervals of the difference of means are formed with the aim of comparison. If the interval includes 0, it is failed to reject null hypothesis. Otherwise, null hypothesis is rejected.

According to Confidence Intervals that are formed at 0.01 significance level, meaning that 99% Confidence Interval, difference of means of Company A and Company B's yearly total detection time per organization for ransomware attacks is in the range of [5003.9, 5651.8]. Therefore, it can be concluded that their means of yearly total detection time per organization for ransomware attacks are not same since it differs in between 5003.9 and 5651.8. Similarly, difference of means of Company A and Company B's yearly total detection time per organization for DDoS attacks is in the range of [7554.5, 7837.5]. Again, it can be concluded that their means of yearly total detection time per organization for DDoS attacks are not same and it differs in between 7554.5 and 7837.5.

To statistically sound, it is required to make assumptions like independence and normality. It is assumed that detection times for Company A and Company B are independent of each other and normally distributed.

Q2)

In this case we analyse if the response time of Company C is lower than that of Company B, using a significance level of 0.05. Difference of Company B and Company C's response time's confidence interval calculated as µB - µC : [0.2, 2.5], with a significance level of 0.05. It is observed that confidence interval is completely positive which means response time of Company B is always larger than Company C with a significance level of 0.05. Hence, Company C is faster than Company B with a probability of %95.

Q3)

In this section, an analysis has been conducted to determine whether there is sufficient evidence to conclude that the probability of Company A’s “successfully detecting” (in any week for any organization) exceeds 55%. We apply hypothesis testing for a population proportion by using a confidence interval and sample proportion ($\hat{p}$). The required assumptions are as follows:

1) The observations are assumed to be independent and identically distributed (i.i.d.) Bernoulli random variables.

2) Random sampling is assumed to be applied.

3) Sample size is large enough for normal approximation to hold.

Sample proportion is $\hat{p} = 0.590$, and the 95% confidence interval is [0.556, 0.624]. Since $\hat{p}>0.55$, and 0.55 is not inside the interval [0.556, 0.624], we can conclude that there is sufficient evidence to conclude that the probability of Company A’s “successfully detecting” (in any week for any organization) exceeds 55%.