Case Study I

IE266 Engineering Statistics

**Group 16**

Burak Baki Bakır 2530103

İlkay Koç 2444792

Zühal Ceren Aydın 2444370

**DESCRIPTIVE STATISTICS**

**Question 1:**

**a)**

Firstly, the data is divided into Company A and Company B, then the total weekly waste of regions for each company is determined weekly by adding up the total wastes of all the recycling bins that are plastic, glass, and aluminum recycling bins for each week separately and using this data 16 different bar graphs for Company A and 20 different bar graphs for Company B were obtained to compare the weekly waste, as shown in Figure 1 and Figure 2 respectively. When it’s analyzed the Figure 1 and Figure 2, it’s seen that there is no common trend between the graphs. Hence, there is not a specific relationship between the regions or weeks.

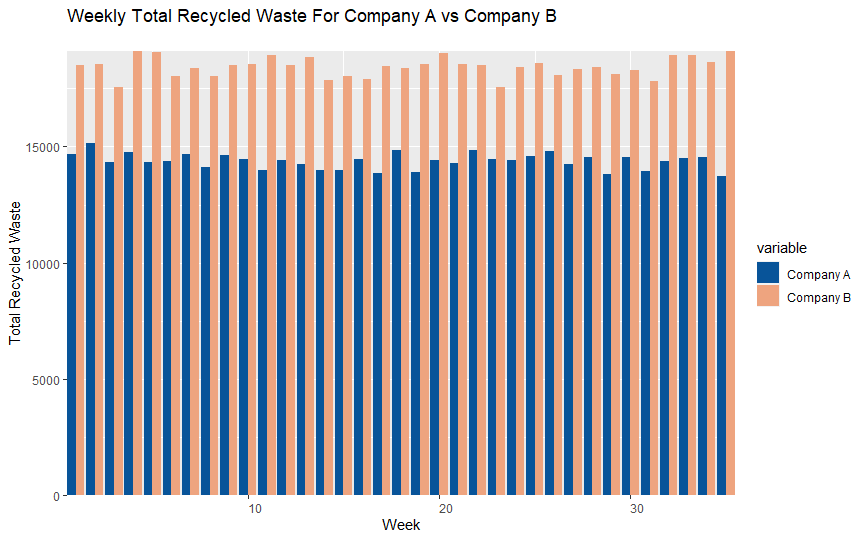


**Figure 1**



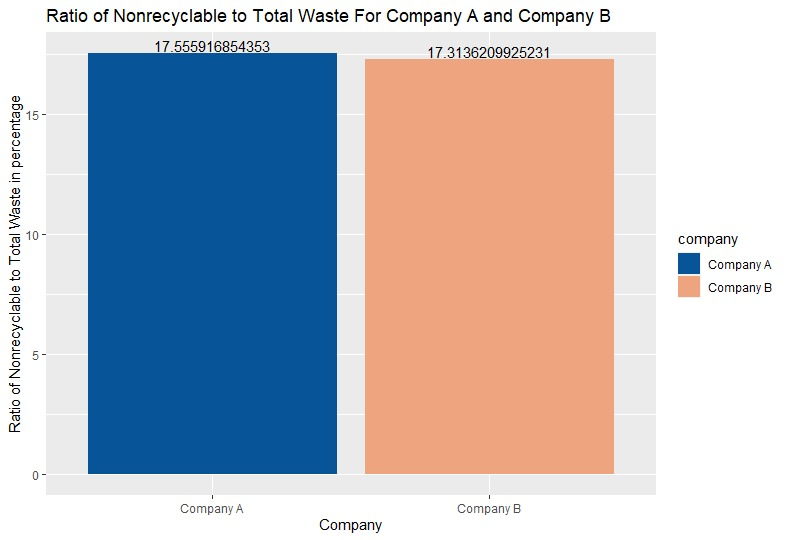
**Figure 2**

It was assumed that a comparison of the weekly workload could be found by comparing the aggregate amount of recyclable waste weekly. Moreover, the total amount of recyclable waste is found by subtracting the amount of non-recyclable waste values from the total waste values separately for plastic, glass, and aluminum. After this operation, the values of plastic, glass, and aluminum are summed up for Company A and Company B. As a result, the total amount of recyclable values represent the new columns in the graph created. As seen in the Figure 3, the total recyclable waste of Company A is higher than Company B every week. That means that the workload of Company A is more than Company B.



**Figure 3**

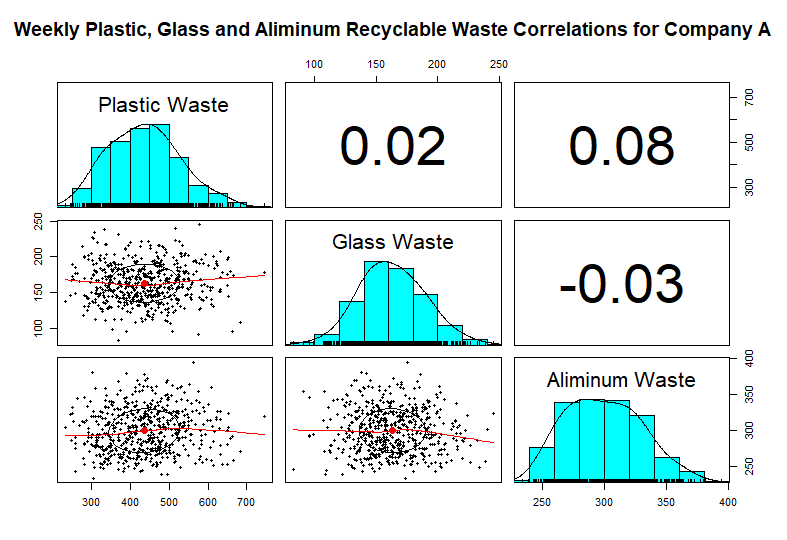
However, only a comparison of the total recycled waste of each company can create a biased result since Company B is responsible for more regions than Company A. Moreover, it directly shows that Company B has more workload than Company A. Thus, it is used the idea of the separation of the non-recyclable wastes from the recycling bins to compare the efforts of each company to avoid this situation. While calculating the effort for each company, the ratio of the non-recyclable waste summation from plastic, glass, and aluminum bins to the total waste of summation of plastic, glass, and aluminum for each company is used and made into bar graphs. As a result of this calculation, it is found that Company A has 17.56% effort, and Company B has 17.31% effort. As shown in Figure 4, when the ratios of these two companies are compared, it is observed that their efforts are so close, and there is no significant difference between them.



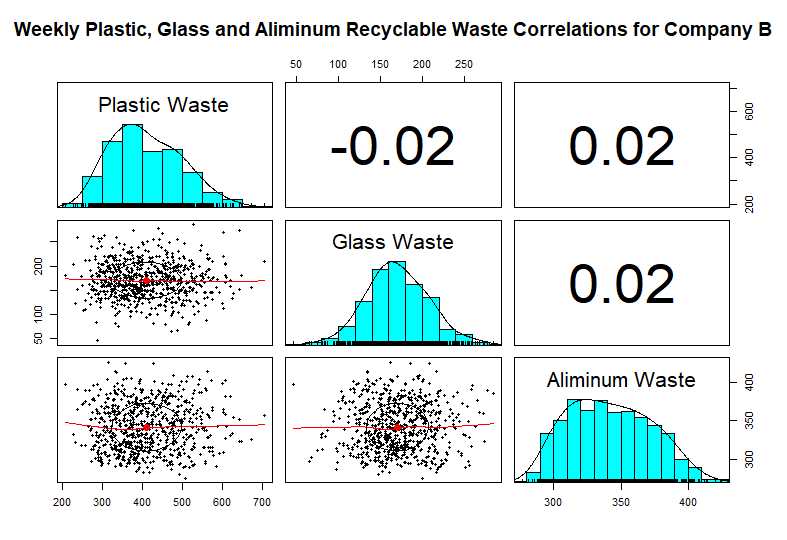
**Figure 4**

**b)**

As shown in the Figure 5,it is observed that there is a 0.02 positive correlation between weekly plastic and glass, 0.04 positive correlation between weekly plastic and aluminum, and a 0.03 negative correlation between weekly glass and aluminum for the Company A. However, it can be said that there is no correlation between these values because the correlation coefficients are insignificant. Moreover, as shown in the Figure 6, similar comments can be made for Company B. It is observed that there is a 0.02 negative correlation between weekly plastic and glass and a 0.02 positive correlation between weekly plastic and aluminum; glass and aluminum for Company B. Since, the correlation coefficients are not significant, it can be said that there is no correlation for Company B, too.

****

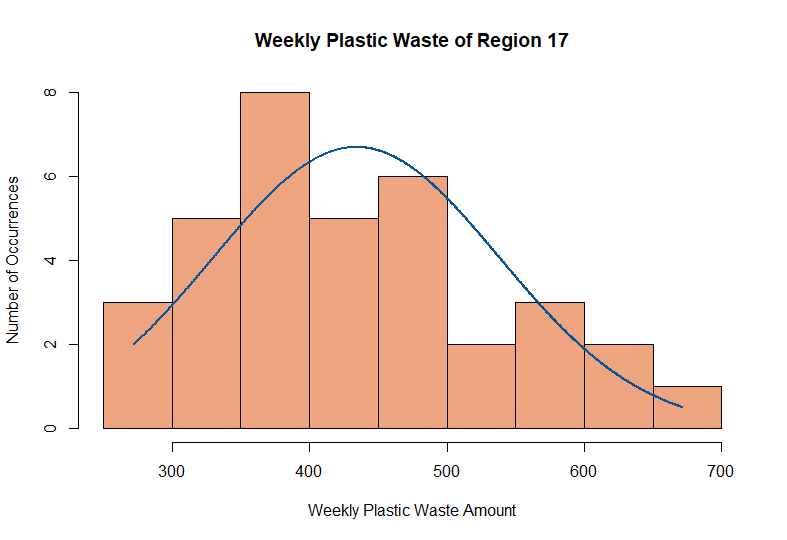
**Figure 5**

****

**Figure 6**

**c)**

Figure 7 shows the graph of the weekly plastic waste collected in Region 17 that is chosen randomly. The figure shows a drawn plot of the weekly plastic waste in Region 17 with the normal line to check the normality of weekly plastic waste in Region 17.If it is visually checked the normality of weekly plastic waste, it can be seen that it is right-skewed, and **the distribution does not quite approximate the normal distribution as shown in the figure**. Moreover, most of the occurrences happened between 350-400, which is the mode of the data. Additionally, the median and mean are between 400-450.



**Figure 7**

Figure 8 shows the graph of the weekly total waste collected in Region 17. As shown in the figure, it is drawn with the normal line to check the normality of weekly total waste in Region 17. It is found that the weekly total waste collected in Region 17 is not skewed, and **the distribution approximates the normal distribution more than the distribution of the weekly plastic waste.** However, even though it is not skewed, the weekly total waste graph also does not quite approximate the normal distribution. Moreover, most of the occurrences happened between 1000-1050 and 1150-1200, which means that the distribution of the weekly total waste collected in Region 17 has two modes. In addition to that, the median and mean are between 1100-1150.

Chart, histogram

Description automatically generated

**Figure 8**

Figure 9 shows the graph of the weekly total waste collected in all regions for which Company B is responsible. The graph of Company B is drawn as Region 17 is in the responsibility of Company B. As shown in the Figure 9, it is drawn with the normal line to check the normality of weekly total waste collected in all regions operated by company B**.** When the total waste collected from all regions operated by Company B and the weekly total waste for Region 17 is compared, it can be seen that the total waste collected in all regions approximates normal distribution significantly more. The reason is this CLT states that when sample size increases, the distribution of sample means approximates the normal distribution regardless of the population distribution. Additionally, mode, mean and median is between 22200-22400.

Chart, histogram

Description automatically generated

**Figure 9**

**STATISTICAL INFERENCE**

**Question 2:**

**a)**

Firstly we accepted that when the population is modeled by a normal distribution, the tests and intervals described in this section are applicable to compare the average weekly amount of plastic waste collected by Company A with that of Company B at 0.05 significance level. Then, it should be verified that the population variances are identical. To check their identicalness, the formula

S12/ S22\* F1-𝛂/2, n2- 1, n1- 1 ≤ 𝛔12/ 𝛔22 ≤ S12/ S22\* F𝛂/2, n2- 1, n1- 1

S1= Standard deviation of the average weekly plastic amount of Company A

S2= Standard deviation of the average weekly plastic amount of Company B

n1= Sample size of weekly averages of plastic waste for Company A

n2= Sample size of weekly averages of plastic waste for Company B

is used to determine the ratio of variances. As shown in the Figure 10, the lower bound is found as 0.73, and the upper bound is found as 2.86 by using the formula. Since 1 is in this interval, and there isn’t any evidence showing that they are not equal, there is a possibility that their ratio is equal to 1. Thus, it can be assumed that population variances are equal. After assuming population variances equal, it can be compared the average weekly amount of plastic waste collected by Company A with that of Company B at 0.05 significance level by using confidence intervals. To make this comparison, the formula

x̄1- x̄2- t𝛂/2, n1+n2-2Sp ≤ μA- μB ≤ x̄1- x̄2+ t𝛂/2, n1+n2-2Sp

where

Sp=

t𝛂/2, n1+n2-2= t-score for sample size n1+n2 with significance level 0.05

S1= Standard deviation of the average weekly plastic amount of Company A

S2= Standard deviation of the average weekly plastic amount of Company B

x̄1= Average weekly plastic amount of Company A

x̄2= Average weekly plastic amount of Company B

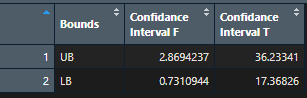
n1= Sample size of weekly averages of plastic waste for Company A

n2= Sample size of weekly averages of plastic waste for Company B

μA= Population mean of plastic waste for Company A’s regions

μB= Population mean of plastic waste for Company B’s regions

can be used. As shown in the Figure 10, by using the formula lower bound is found as 17.36, and upper bound is found as 36.23. Thus, according to these results, the average weekly amount of plastic waste collected by Company A is higher than Company B since it’s checked that μA- μB and the bounds are positive.



**Figure 10**

Firstly, we accepted that when the population is modeled by a normal distribution, the tests and intervals described in this section are applicable to compare the average weekly amount of glass waste collected by Company A with that of Company B at 0.05 significance level. After that, it should be verified that the population variances are identical. To check their identicalness, the formula

S12/ S22\* F1-𝛂/2, n2- 1, n1- 1 ≤ 𝛔12/ 𝛔22 ≤ S12/ S22\* F𝛂/2, n2- 1, n1- 1

S1= Standard deviation of weekly glass amount of Company A

S2= Standard deviation of weekly glass amount of Company B

n1= Sample size of weekly averages of glass waste for Company A

n2= Sample size of weekly averages of glass waste for Company B

is used to determine the ratio of variances. By using the formula lower bound is found as 0.49, and upper bound is found as 1.93 as shown in Figure 11. Since 1 is in this interval, and there isn’t any evidence showing that they are not equal, there is a possibility of their ratio to be equal to 1. Thus, it can be assumed that population variances are equal. After assuming variances equal, it can be compared the average weekly amount of glass waste collected by Company A with that of Company B at 0.05 significance level. To make this comparison, the formula

x̄1- x̄2- t𝛂/2, n1+n2-2Sp ≤ μA- μB ≤ x̄1- x̄2+ t𝛂/2, n1+n2-2Sp

where

t𝛂/2, n1+n2-2= t-score for sample size n1+n2-2 with significance level 0.05

Sp=

S1= Standard deviation of weekly glass amount of Company A

S2= Standard deviation of weekly glass amount of Company B

x̄1= Average weekly glass amount of Company A

x̄2= Average weekly glass amount of Company B

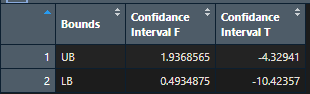
n1= Sample size of weekly averages of glass waste for Company A

n2= Sample size of weekly averages of glass waste for Company B

μA= Population mean of glass waste for Company A’s regions

μB= Population mean of glass waste for Company B’s regions

can be used. By using the formula lower bound found as -10.42 and upper bound found as -4.32 as shown in Figure 11. Thus, according to these results, the average weekly amount of plastic waste collected by Company B is higher than Company A since it’s checked that μA- μB and the bounds are negative.



**Figure 11**

Finally, we accepted that when the population is modeled by a normal distribution, the tests and intervals described in this section are applicable to compare weekly plastic waste collected by Company A with that of glass waste collected by the same company, at 0.05 significance level, first it should be verified that the population variances are identical. To check their identicalness, the formula

S12/ S22\* F1-𝛂/2, n2- 1, n1- 1 ≤ 𝛔12/ 𝛔22 ≤ S12/ S22\* F𝛂/2, n2- 1, n1- 1

S1= Standard deviation of weekly plastic amount of Company A

S2= Standard deviation of weekly glass amount of Company A

n1= Sample size of weekly averages of plastic waste for company A

n2= Sample size of weekly averages of glass waste for company A

is used to determine the ratio of variances. By using the formula lower bound is found as 0.04 and upper bound is found as 0.172 as shown in Figure 12. Since 1 is not in this interval, and there isn’t any evidence showing that they are equal, there is no possibility of their ratio to be equal to 1. Thus, it can be assumed that population variances are not equal. After taking variances as not equal, it can be compared the average weekly amount of glass waste collected by Company A with that of Company B at 0.05 significance level

again the formula

x̄1- x̄2- t𝛂/2, n1+n2-2 ≤ μ1- μ2 ≤ x̄1- x̄2+ t𝛂/2, n1+n2-2

where

t𝛂/2, n1+n2-2= t-score for sample size n1+n2-2 with significance level 0.05

S1= Standard deviation of weekly plastic amount of Company A

S2= Standard deviation of weekly glass amount of Company A

x̄1= Average weekly plastic amount of Company A

x̄2= Average weekly glass amount of Company A

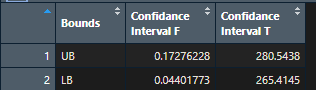
n1= Sample size of weekly averages of plastic waste for Company A

n2= Sample size of weekly averages of glass waste for Company a

μA= Population mean of plastic waste for Company A’s regions

μB= Population mean of glass waste for Company A’s regions

is used. By using the formula lower bound is found as 265.41 and upper bound is found as 280.54 as shown in Figure 12. Thus, according to these results, the average weekly amount of plastic waste collected by Company A is more than the average weekly amount of glass waste collected by the same company.



**Figure 12**

**b)**

Firstly, we accepted that when the population is modeled by a normal distribution, the tests and intervals described in this section are applicable. Then, the average ratio of non-recyclable waste to total waste is found by summation of the ratios weekly for every region and divided by the total number of weeks, 35. After this operation, the confidence interval at a 0.05 significance level for every region is constructed by using the t-test formula to check whether initiating a campaign is necessary for each region. According to the city council, the long-run average ratio of non-recyclable waste to total waste should not exceed %18. Thus, the lower confidence interval determines, the lower bound of the weekly average ratio of non-recyclable waste to total waste for every region and the lower confidence interval is found by the formula

For Company A

x̄- t𝛂, n1-1 ≤ μ1

t𝛂, n1-1= t-score for sample size n1 with significance level 0.05

S1= Standard deviation of Weekly ratio of non-recyclable waste for Company A

x̄1= Average Weekly ratio of non-recyclable waste for Company A

n1= Sample size of weekly = 16

μ1= Population means

For Company B

x̄- t𝛂, n2-1 ≤ μ2

t𝛂, n1-1= t-score for sample size n1 with significance level 0.05

S2= Standard deviation of Weekly ratio of non-recyclable waste for Company B

x̄2= Average Weekly ratio of non-recyclable waste for Company B

n2= Sample size of weekly = 20

μ2= Population means

But because sample size is large enough, we used z-score instead of T-score

And used formula

For company A

x̄- z𝛂,  ≤ μ1

z𝛂= z-score with significance level 0.05

S1= Standard deviation of Weekly ratio of non-recyclable waste for Company A

x̄1= Average Weekly ratio of non-recyclable waste for Company A

n1= Sample size of weekly = 35

μ1= Population means

For company B

x̄2- z𝛂,  ≤ μ2

z𝛂= z-score with significance level 0.05

S2= Standard deviation of Weekly ratio of non-recyclable waste for Company B

x̄2= Average Weekly ratio of non-recyclable waste for Company b

n2= Sample size of weekly = 35

μ2= Population means

As shown in Figure 13, Figure 13x, Figure 14 and Figure 14y every lower confidence interval at a 0.05 significance level for every region is represented for both company A and company B in the tables and graphs, respectively.

When we look at the data, the lower bounds of 5th,9th,12th,13th, 16th, 20th,27th,29th regions are exceed the 18%. Therefore, initiating a campaign is necessary for only these regions.

Chart, box and whisker chart

Description automatically generated **Figure 13**

Graphical user interface

Description automatically generated with medium confidence

**Figure 13x**

Chart, box and whisker chart

Description automatically generated **Figure 14**

A picture containing graphical user interface

Description automatically generated

**Figure 14y**

**For Whole City**

x̄3- z𝛂,  ≤ μ3

z𝛂= z-score with significance level 0.05

S3= Standard deviation of Ratio of non-recyclable waste for for all regions

x̄3= Average ratio of non-recyclable waste for all regions

n3= Sample size of regions = 36

μ3= Population means

When the city council checks the ratio for the whole city, the ratio is found by summation of the weekly non-recyclable waste for every region and divided by the total waste of the whole city for every week. It is used to find the lower bound of the whole city to determine whether that city council changes a campaign or not. The lower bound is found as 0.1711467, as shown in Figure 15. As a result, we can say that the campaign does not need to initiate a campaign since the lower bound is found to be smaller than 0.18.

Graphical user interface, application

Description automatically generated

**Figure 15**

**c)**

In the first place, we accepted that when the population is modeled by a normal distribution, the tests and intervals described in this section are applicable. Ho, the null hypothesis, is determined as that p1 is equal to p2, and H1, the alternative hypothesis, is determined as that p1 is not equal to p2. The confidence interval on the difference of two proportions (𝑝1,𝑝2) from two independent populations, Company A and Company B, was used to reject the equality of p1 and p2, and the formula in Figure 16 was utilized to make this calculation. In the formula, p1 represents the proportions of lost weeks for Company A, and p2 represents the proportions of lost weeks for Company B**.** As shown in the Figure 17, the lower bound was found -0.2332123, and the upper bound is found 0.1189266 for plastic, by using the formula. According to these results, the proportions of lost weeks of plastic waste for Company A and Company B are found as equal to each other. This was reached by observing that the result of p1- p2 can be equal to zero since 0 is within the found range. As a result, it means that the H0 fails to reject.

Logo

Description automatically generated with low confidence

Z𝛂/2= z-score with significance level 0.05

p̂1= Proportion of lost weeks to all weeks for waste Plastic, Glass, Aluminum of Company A

p̂2= Proportion of lost weeks to all weeks for waste Plastic, Glass, Aluminum of Company B

n1= Sample size of weekly proportion of lost weeks to all weeks for waste Plastic, Glass, Aluminum = 35 of Company A

n2= Sample size of weekly proportion of lost weeks to all weeks for waste Plastic, Glass, Aluminum = 35 of Company B

p1= Proportion of lost weeks in the population for waste Plastic, Glass, Aluminum of Company A

p2= Proportion of lost weeks in population for waste Plastic, Glass, Aluminum of Company B

**Figure 16**

Graphical user interface

Description automatically generated

**Figure 17**

Besides, the upper and lower bound values for non-recyclable glass and aluminum waste are found to be equal to each other and 0. So, after calculating the confidence interval on the proportions of each company's non-recyclable glass and aluminum waste, the upper and lower bounds are found as 0. Hence, glass and aluminum waste values do not directly affect the result.