

IE266 Engineering Statistics

Case Study I

Due date: 09.05.2022 Monday 23:59

In answering all questions, please state your assumptions clearly.

The city council of the municipality wants to take steps on solid waste separation and recycling and a decision is made to hold a tender. As a result, 36 regions of the city were tendered to 2 different companies. Company A was allocated to operate in 16 of these regions, and company B in the remaining 20 regions. The municipality did not want the companies' operations to be limited to only collecting wastes and recycling these wastes, but also aimed to benefit from this initiative after the agreements. Thus, the companies are expected to work to raise awareness on this issue in the regions they operate.

Suppose firms have been engaged in solid waste management activities in these regions for some time. The municipality thinks that the companies have achieved a standard in solid waste separation by now and wants to examine their efficiency in this field. In accordance with the agreement to which they are subject, companies provided two types of data for the last 35 weeks to the municipality. The first type of data is the weekly amount of waste collected from bins. Companies use three types of bins for three different solid waste types (plastic, glass, and aluminum), and these weekly amounts are collected for each type of bin (measured in m^3). The second type is the weekly amount of waste thrown into the plastic/glass/aluminum bins even though they are not plastic/glass/aluminum. Similar to the first type, these weekly amounts are collected for each type of bin (in m^3). The attached file, sw_data.xlsx, contains the information on the companies operating in the regions and the data mentioned above for each region. One can see the first rows of the provided data in the below table. Columns 3-5 contain the first data type, and columns 6-8 contain the 2nd type of data.

Table 1

Week	Region	total waste from plastic recycling bins	total waste from glass recycling bins	total waste from aluminum recycling bins	non-recyclable waste from plastic recycling bins	non-recyclable waste from glass recycling bins	non-recyclable waste from aluminum recycling bins
1	1	694.13	165.33	361.68	228.90	20.56	80.79
1	2	613.08	201.79	349.93	243.47	25.97	55.39

Assume that the amount of waste (separated or not separated) generated in a region is independent of that another region, and the amount of waste generated in a region in a week is independent of that in another week. Also, the long-run average percentage of the non-recyclable waste accidentally thrown into recycle bins is stationary over time.

1.Descriptive statistics

- a. Use appropriate descriptive statistics (graphical and numerical tools) to compare the weekly waste of regions for a given company. Compare the weekly workload of the companies (in terms of the aggregate amount of recyclable waste) with each other. Is there a reason to doubt that a company spends more effort than the other in collecting the recyclable waste? in separating the non-recyclable waste?
- b. Is it possible to say that weekly plastic, glass, and aluminum wastes are correlated? Use appropriate descriptive statistics (graphical and numerical tools) to support your answer.
- c. Draw appropriate plots to visually check the normality of weekly plastic waste in a region. Compare the plot with the weekly total waste of the same region, and the weekly total waste collected in all regions for which a company is responsible.

2. Statistical inference

- a. Compare the average weekly amount of plastic waste collected by Company A with that of Company B at 0.05 significance level. Do the same analysis for glass waste. (Before the analysis, first statistically verify whether population variances are identical). Finally compare weekly plastic waste collected by company A with that of glass waste collected by the same Company, at 0.05 significance level to check whether there are any differences. Use confidence intervals for comparison purposes. To make statistically sound comparisons which assumptions are needed?
- b. The city council thinks that the weekly ratio of non-recyclable waste (**in the total waste**) is a good indicator of the awareness of the people in the city about solid waste recycling. In each region, in the recycle bins the long-run average ratio of non-recyclable waste to total waste should not exceed %18. Looking at weekly ratios for each region, the council decides whether there is sufficient evidence to conclude that long-run average ratio is above %18, to possibly initiate a campaign for that region to increase awareness. Construct a CI at 0.05 significance level to check

whether initiating a campaign is necessary for each region. If city council checks the ratio for the whole city by looking at weekly data aggregated over the regions (i.e., design a campaign either for the whole city or none), would your answer to initiate a campaign change?

- c. Companies state that when the ratio of wastes that are not suitable for the containers exceeds a certain threshold for each type of waste (as indicated in Table 2), there is a significant loss of efficiency in the processes after the waste is collected, in the recycling processes at the facilities, and they describe these weeks as “lost” weeks for that type of waste. When allocating the regions to the companies, the municipality has ensured that the proportions of lost weeks are identical for both companies on expectation. For each solid waste type, compare the number of “lost” weeks for the companies to check whether there is sufficient statistical evidence to falsify the municipality’s claim. Use confidence intervals at 0.05 significance level. Comment on your findings. State any necessary assumptions to make statistically decisive comparisons.

Table 2 Threshold ratios for non-recyclable waste for different solid waste types. If for a waste type ratio in a week exceeds the threshold, it is a “Lost” week for that waste type

Plastic	Glass	Aluminum
0.20	0.13	0.25

Format and Organization:

- Please write in proper font size 12 and 1.5 paragraph spacing with reasonable margins.
- You do not need to include an introduction, conclusion, and appendix section in the report.
- Number and title report sections properly.
- The format and organization of the report will be considered in grading.
- This homework is for a group of four. Working in collaboration with other groups is not allowed. You can discuss the problem with your partners only.
- Submit your work under Case Study 1 on ODTÜClass. Upload a single zip file, including your report and all the files you use to answer the questions (Minitab, R, and Excel files).
- You should use either Minitab or R to carry out your statistical analysis. R is recommended and it will bring a 5% bonus.
- Please use comments in your R scripts to make your codes readable.