Waste Data ANOVA Analysis

Anthony Davis

STAT301

**Introduction**

This paper outlines an analysis on waste data in Minnesota from the years 2013 until 2017. This comes from a dataset titled Wastedata1, which encapsulates the amount of residential and commercial recycling in tons from various counties in Minnesota. All the data takes place between 1991 and 2017, but for the sake of obtaining more meaningful results, this analysis will only encapsulate a smaller range of the included years. The analysis was conducted to determine if year and recycling type had an effect or difference on the total recycled tons in Minnesota.

The first variable used was the numerical independent variable of year in which the amount of recycling was calculated. The second variable was the categorical independent variable of recycling type, which was created by combining the variables “Res Tons” and “CII Tons” (residential or commercial recycling) and creating a single variable to distinguish the two. The last variable used was the numerical dependent variable of recycling tons, the total amount of recycled materials of the respective type and year in tons.

The null and alternative hypothesis for recycling type were:

The null and alternative hypothesis for year were:

**Method**

The two-way ANOVA test was used to conduct this analysis. The reason for this test is that the data used was balanced, had two independent variables and one dependent variable, and had no repeating measures. All calculations done in R, and all decisions were based on a significance level of .05 (5%).

**Conclusion**

The two-way ANOVA test in R gave the following results:

A black text on a white background

Description automatically generated

With this, we conclude that there is a significant main effect from recycling type on recycled tons with an F-value of 49.359 and a p-value of nearly 0. There is no main effect from year on recycled tons from 2013 to 2017 with an F-value of 1.025 and a p-value of 0.311. Lastly, there is no significant interaction between the recycling type and year, with an F-value of 0.031 and a p-value of 0.860.

An interaction plot was also produced to visualize these results, as seen below:

A graph with lines and numbers

Description automatically generated

The dotted line, for commercial recycling, which for every point (year) is noticeably larger than that of the solid line, which represents residential recycling. This makes sense given that our testing showed that the recycling type does have a significant main effect on recycling tons. So from the years of 2013 to 2017, Minnesota saw more recycling commercially than from residential zones.

One interesting point as seen on the graph is at the year 2014, where a large spike in commercial recycling is visible, and then falling almost to the 2013 level or recycling in the year 2015. For residential recycling, no such pattern is noticeable. Despite this, the test saw no main effect from year nor any significant interaction between recycling type and year. Perhaps if further testing was done, this could be discerned as a potential outlier.

One limitation of this analysis is the lack of knowledge of any confounding variables. Given that this dataset was taken from an outside source, information on anything that could have interfered with data collection, or skewed results is left unknown. This information could potentially describe why we saw such a jump in commercial recycling in 2014, but since we lack that information, it cannot be determined.

GitHub: <https://github.com/AnthonyDavisJ/STAT301Project>