**Landfilled and WTE in MN**

By: Pa Yang & Kasiab Lor

For this project, we have selected Wastedata2, which has data collected from 1991-2017, as our primary dataset for testing. Our research focuses on two key questions: first, whether there is a significant difference in average landfill between the top three counties, and the second, whether there is a significant difference in the average amount of waste energy for each country between the years 1991, 2000, and 2017. To answer these questions, we needed to create new data sets by manipulating the original data for testing purposes. In the following section, we describe how we created our new datasets, how we tested our questions, and the findings of our analysis.

For first investigation of “whether there is a significant difference in average landfill between the top three counties.” The variables used for independent variables (IV) are County (counties in MN) and dependent variable (DV) is Landfilled (amount of landfilled in tons). The method that was used for this data was the one-way ANOVA, since this data is investigating one IV that affects the average landfill. The null hypothesis is Ho: mu (Hennepin) = mu (Dakota) = mu (Ramsey). The alternative hypothesis is Ha: at least two differ. The test statics for is shown Text

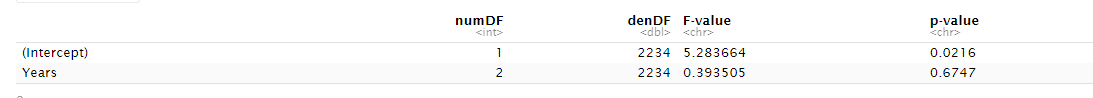
Description automatically generatedText Box

in Figure 1, which shows us that our p-value is lower than the significance level of 0.05, hence, we reject Ho. There is enough evidence that the average landfill is significantly different between the top three counties. Based on this null hypothesis we can tell that at least two of the top three counties average landfill are different and those two counties were between Hennepin-Dakota and Ramsey-Hennepin. The top three counties (Hennepin, Dakota, and Ramsey) have the most populations in MN, which is the reason why they have the most landfills. Since Hennepin has the most average landfilled of about 318517.9 tons, this is the reason that Dakota and Ramsey are more like each other than to Hennepin.

To investigate the second research question regarding the potential difference in the average amount of waste to energy (WTE) for each country between the years 1991, 2000, and 2017, we selected the variables year, county and WTE as essential data points. The independent variables for this research question are year and county, while WTE is the dependent variable. We used the one-ways repeated ANOVA to test this question, as the years were repeated throughout the data set. Our null hypothesis stated that the WTE amount for each county was equal throughout all years, and our alternative hypothesis suggested that at least two years had different amounts of WTE.

Graphical user interface, chart, box and whisker chart

Description automatically generated



Despite having closely clustered point graphs for the years, we still ran the test, obtaining a p-value of 0.67. As this value is greater than our significance level of 0.05, we failed to reject the null hypothesis, implying that there is a high probability that the countries had almost the same amount of WTE throughout the three years. This outcome was surprising, as we expected to observe differences in the amount of WTE over the three years, given the population growth and advancements in technology.

Githubs:

* <https://github.com/Kasiab-Lor/ANOVA-STATS.git>
* <https://github.com/PaDataStories/R-R-markdown-/blob/main/Recycling%20Project>