**Dissecting Recycling Dynamics: A Two-Way ANOVA Analysis of Trends in Minnesota’s County-Level Waste Management from 1991 to 2017**

**Introduction**

The project is based on the dataset named “waste1data”. This dataset contains recycling data from 86 counties in Minnesota (MN) spanning from 1991 to 2017. It consists of 49,019 rows and 5 columns, which indicate the year of data collection, the county name, the category of recycling, and the weights of residential (Res) and commercial, institutional, and industrial (CII) recycling in tons.

***Research Questions***

There are two research questions that mainly aim to answer two primary questions related to waste recycling. The first question wants to understand whether there is a significant difference in the mean weight of waste recycled between two sources: residential areas and commercial, industrial, and institutional sources.

The second question intends to explore if there are significant differences in the mean weight of waste recycled across various recycling categories, including Paper, Metal, Glass, Plastic, Hazardous, Other, and Organic materials. These questions will help to identify specific patterns and possibly improve recycling processes based on the source and type of materials recycled.

***Given Variables***

The given dataset, “waste1data,” comprises 5 variables and 49,019 rows. These variables are Year, County, Category, Res Tons, and CII Tons. The Year variable is a numerical and discrete data type representing the mean year in which the recycling data was collected. The County variable is a categorical and nominal data type pertaining to the name of the county in Minnesota where the recycling data was collected.

The Category variable is also categorical and nominal, referring to the type of recycling material, which is categorized into 7 levels: Paper, Metal, Glass, Plastic, Hazardous, Other, and Organic. This variable is crucial for understanding the distribution of recycling activities across different material types. Category is considered as independent variable.

The Res Tons variable is a numerical and continuous data type that indicates the weight of residential recycling collected, measured in tons. This variable quantifies the volume of recycling from residential sources. It is considered that dependent variable.

The CII Tons variable is a numerical and continuous data type concerning the weight of recycling collected from commercial, institutional, and industrial sources, also measured in tons. This variable reflects the volume of recycling from commercial, institutional, and industrial sources. It is considered that dependent variable.

***Variables Retained and Transformed for Analysis***

The Year and County variables are excluded from the analysis. The Category variable, which is categorical and nominal, is retained. This variable categorizes the types of materials recycled into 7 levels: Paper, Metal, Glass, Plastic, Hazardous, Other, and Organic. It is crucial for understanding the distribution of recycling activities across different material types and is used as an independent variable.

The Res Tons and CII Tons variables have been retained and transformed from wide to long format, resulting in two new variables: Type and Recycle Tons. The Type variable is categorical and nominal, denoting the source of the recycling which is either residential or commercial, institutional, and industrial sources. It is used as an independent variable.

Recycle Tons, a numerical and continuous variable, represents the combined weight of recycling collected from both residential and commercial, institutional, and industrial sources, measured in tons. This variable, which serves as a dependent variable, quantifies the total volume of recycling collected.

***Null and Alternative Hypotheses***

* **Main Effect of Category**
  + **Null Hypothesis ():** The means across all categories are equal.

* + - : = = = = = =
  + **Alternative Hypothesis ():** Not all category means are equal.
    - **:** At least two means are different.
* **Main Effect of Type**
  + - **Null Hypothesis ():** The mean weight of recycled tons is the same for residential and commercial/industrial/institutional (CII) collections.: =
  + **Alternative Hypothesis ():** The mean weight of recycled tons for residential collection is not equal to that for commercial/industrial/institutional (CII).
    - **:**
* **Interaction Between Type and Category** 
  + : There is no interaction between Type and Category
  + **:** There is interaction between Type and Category

**Methods**

***Preliminary Steps***

Before conducting the statistical analysis, preliminary steps of data exploration, visualization, and transformation are undertaken. These initial steps are crucial for ensuring the quality and integrity of the dataset. They include checking for and addressing any missing values, outliers, or anomalies that could impact the results. Additionally, data is transformed from wide to long format.

***Appropriate Analysis Chosen***

This project employs a two-way ANOVA to analyze the dataset.

***Reason for Choosing Two-Way ANOVA***

The two-way ANOVA is selected for some compelling reasons that align with the goals and structure of the dataset. For specifics, the primary aim of our research questions is to compare mean differences across two independent factors: Type and Category. Therefore, the two-way ANOVA is the most suitable statistical test for this purpose.

Moreover, beyond examining the individual effects of source type (Residential vs. CII) and recycling category (Paper, Metal, Glass, Plastic, Hazardous, Other, Organic) on the amount of waste recycled, two-way ANOVA allows for the examination of interaction effects between these factors. Thus, this is crucial for understanding if the impact of one factor depends on the level of the other, and it provides a more nuanced insight into recycling behaviors across different materials and source types.

**Conclusion**

***Test Statistics, P-Values, and Decisions Based on P-Values***

The test statistics and p-values provided by the two-way ANOVA give informative insights for the Category, Type, and their interaction. The p-values for each variable and their interaction are highly significant, well below the conventional alpha threshold of 0.05. Specifically, as shown in Figure 1, the F value for the Category variable is 342.1, and the p-value is less than 2e-16, which is significantly below 0.05. This strongly suggests a significant main effect of the category on the amount of recycling. Consequently, we reject the null hypothesis ​ for Category, indicating there is evidence that the mean weights of recycled tons significantly vary among the different categories (Paper, Metal, Glass, Plastic, Hazardous, Other, Organic).

In terms of the Type variable, which distinguishes between Residential and Commercial/Industrial/Institutional (CII) sources, the reported F value is 143.4 with a p-value less than 2e-16, again below 0.05. This result leads us to reject the null hypothesis ​ for Type, providing evidence that the mean weights of recycled tons significantly differ between Residential and CII sources.

Additionally, the interaction between Category and Type also shows significant statistical values, with an F value of 109.3 and a p-value less than 2e-16, which is less than 0.05. Rejecting the null hypothesis ​ for the interaction effect indicates there is significant evidence that the impact of one factor varies depending on the level of the other. This confirms that the relationship between category and type significantly influences recycling behaviors and that these factors do not operate independently.

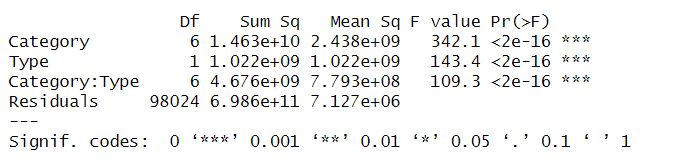
****

Figure 1: Output from Two-Way ANOVA

Since the p-values for Category, Type, and the interaction between Category and Type are all significantly less than 0.05, we reject the null hypotheses for each of these factors. This decision indicates significant differences in the means of the weight of Recycle Tons within Types, as confirmed by Figures 3, and that at least two means of the weight of Recycle Tons are statistically significantly different when comparing different levels within Categories and their interactions, as confirmed by Figure 2.

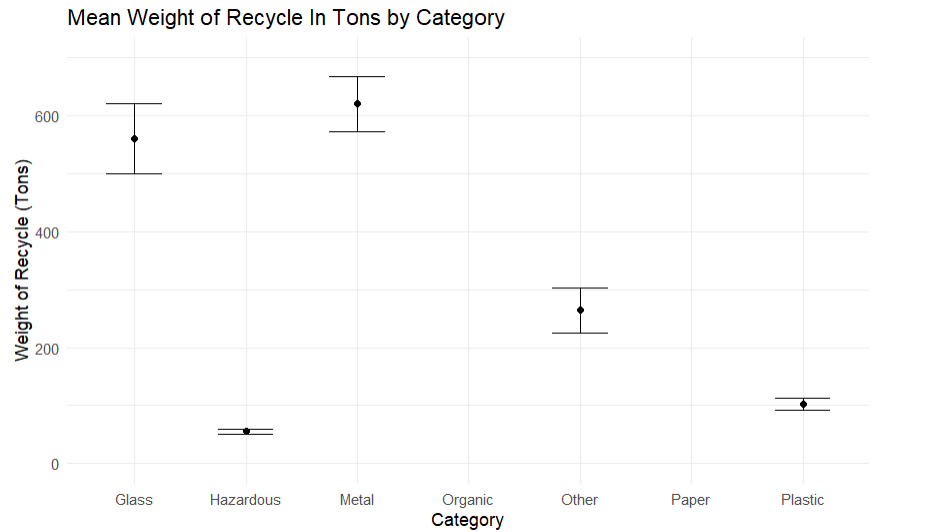
****

Figure 2: Mean Weight of Recycle in Tons by Category

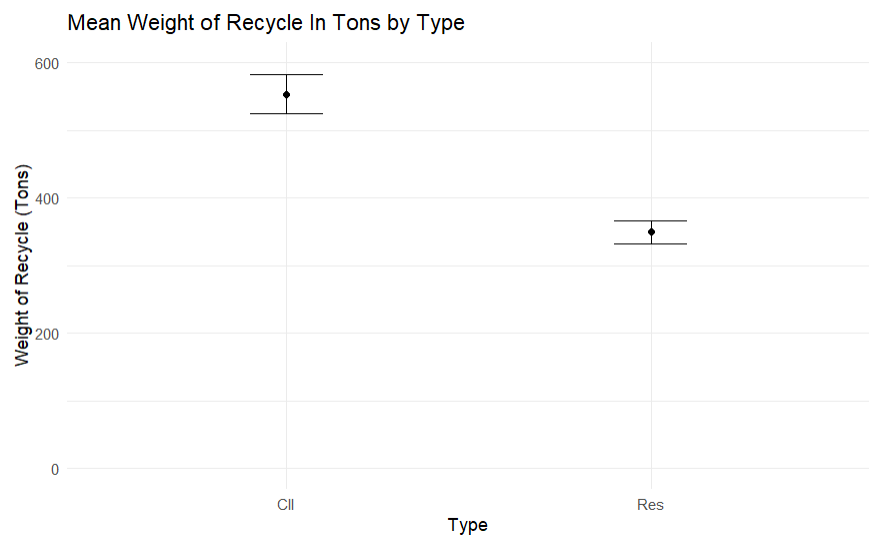
****

Figure 3: Mean Recycle in Tons by Type

***Explanation of Decision***

The decision to reject the null hypothesis suggests that both the source type of the recycled material (Residential and CII) and the category of the material (Paper, Metal, Glass, Plastic, Hazardous, Other, Organic) have statistically significant effects on the weight of the recycled material. Additionally, the significance of the interaction term implies that the effect of one independent variable on the dependent variable is not consistent across the levels of the other variable. In practical terms, this might mean that the effectiveness of recycling efforts varies not only by material type but also by whether the material originates from residential or from commercial, institutional, and industrial (CII) sources.

***Most Interesting Finding***

Perhaps the most interesting finding is the significance of the interaction effect, which could indicate complex dynamics between the source of the waste (Res and CII) and the Category of material (Paper, Metal, Glass, Plastic, Hazardous, Other, Organic) being recycled. This suggests that recycling programs might need to be focused specifically on different categories of recycling (Paper, Metal, Glass, Plastic, Hazardous, Other, Organic) and their Type of source (Res or CII) to maximize efficiency.

To gain a better understanding of the interaction effects between the Category and Type variables within the dataset, a more focused analytical approach was adopted. The data were filtered to retain only one Type at a time, and individual one-way ANOVAs were conducted for each, specifically for Residential (Res) and Commercial, Institutional, and Industrial (CII). This method provides a clearer analysis of the differences within material categories under each type setting.

***Analysis within Residential Type***

Initially, the data was filtered to include only the Residential type. A one-way ANOVA was then performed to test for significant differences among the material categories (Paper, Metal, Glass, Plastic, Hazardous, Other, Organic) within residential settings. Significant differences were detected among several pairs of categories. However, no significant differences were observed between Paper and Glass, between Plastic and Hazardous, and between Organic and Metal. These findings suggest variable recycling rates for different materials within residential settings but indicate similar treatment for certain materials.

***Analysis within Commercial/Industrial/Institutional Type***

Similarly, the dataset was filtered to include only entries classified under CII. A one-way ANOVA was conducted test for significant differences among the material categories (Paper, Metal, Glass, Plastic, Hazardous, Other, Organic) within Commercial, Institutional, and Industrial (CII) settings. While significant differences were prevalent among most pairs, no significant differences were observed between Other and Glass, and between Plastic and Hazardous. This pattern highlights diverse recycling behavior in CII settings, with some materials exhibiting comparable recycling weight.

***Limitations***

There are some limitations of this analysis. The ANOVA used assumes homogeneity of variances. If this condition is not met, the validity of the test results could be compromised. Additionally, the analysis fails to consider potential confounding variables that were not included in the model but might still influence the amount of waste recycled. Lastly, the dataset is limited to data from Minnesota, which may not be generalizable to other regions or states.

Top of Form