Adrian Parenteau­­

Data 301

04/21/24

ANOVA Analysis of Composting rates:

Corporate vs. Residential Recyclers

**Introduction**

Global Warming and environmental decay are two major problems facing the world today. For this reason and others, services such as recycling have become more important than ever. This report seeks to investigate if recycling rates of commercial recyclers are greater than that of residential recyclers.

The data set used in this report was created by the Minnesota Pollution Control Agency. This data set includes five different variables. These variables include Year (of data collected), County, Category (of Recycling material), Res Tons (residential recycling weight in tons), and CII Tons (commercial, institutional, and industrial recycling weight in tons). County and Category are categorical variables while Year, Res Tons, and CII Tons are numerical.

The first question this report will seek to answer is whether businesses recycle a significantly higher volume of material than residents do. To answer this question, we will use two variables from the data set. The first variable (x1) is Res Tons. As stated previously Res Tons represents the residential recycling weight measured in tons. The second variable (x2) is CII Tons. This variable represents commercial, institutional, and industrial (i.e. businesses) recycling weight measured in tons. I hypothesize that businesses recycle a greater weight than residents. I believe this will be the case due the significantly greater volume of waste created by businesses compared to that of individuals.

To perform our test, we will have to transform these variables to create two new ones. The first new variable will be recycler type (residential or business). This variable will be our independent variable and is categorical. The second will be weight (recycled materials measured in tons). This variable will be the dependent variable and is numerical.

The second question investigated in this report is what the impact of county and category are on recycle weight. To answer this question, we will need three variables: county, category, and weight. County and category will be our two independent variables. Both are already in the data set and are categorical. County represents each county in MN where recycling data was collected. For this test we will select only the three most populous counties: Hennepin, Ramsey, and Dakota. Category is the type of material recycled (e.g. Organics, cans, paper, etc.). The third variable, weight, is our dependent variable and is the same one that we constructed for our first test. This variable measures the weight of material that has been recycled in tons.

**Methods**

To answer the question of whether residents or companies recycle more we will perform a simple t-test. A t-test test compares the variance between two different groups. In this case our two groups are residents (Res) and companies (CII). We want to analyze the variance in recycling rates between these two groups. We chose a t-test because we only have one independent variable: Recycler Type.

For the first our Null and alternative hypothesis are as follows:

H0: µRes = µCII

Ha: µRes < µCII

Our second test is aimed at tackling two independent variables at once. For this reason, we will select a two-way ANOVA. To perform this test, we will have to look at three different test statistics, one for each main effect of the two independent variables and the third for the interaction between the two independent variables.

For the second test our Null and alternative hypotheses are as follows:

*Main effect for County:*

H0: µ(Hennepin) = µ(Ramsey) = µ(Dakota)

Ha: at least two means are different

*Main effect for Category:*

H0: µ(metal) = µ(plastic) = µ(glass) = µ(paper) = µ(organic) = µ(hazardous) = µ(other)

Ha: at least two are different

*Interaction between County and Category:*

H0: There is no interaction between County and Category

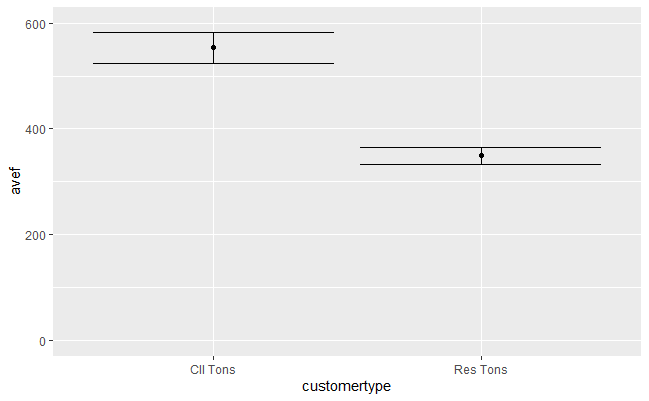
Ha: There is an interaction between County and Category

**Conclusion**

The results of our first test show that there is a significant difference in the recycle weight mean between residents and businesses (t = 11.813, p < 0.001). Therefore, we can conclude that there is enough evidence to reject the null hypothesis. There is enough evidence to suggest that businesses recycle significantly more by weight than residents. A plot of recycle means of both Res Tons and Res CII can be seen in figure 1.1.

Figure 1.1

*Recycling Means by Customer Type*



For the second test we first will look at the results of the main effects for County and Category and then we will look at the interaction between the two variables. The results for the main effect of County are F = 0.356, p = 0.551. Therefore, we do not reject the null hypothesis. These results suggest that there is not a significant difference between in mean recycling weight between counties. Test results for the main effect for Category are, F = 19.084, p < 0.001. Therefore, we reject the null hypothesis, there is a significant difference in mean recycle weight between at least two of the recycle type categories. Finally for the interaction the results are F = 0.952, p = 0.457, therefore we do not reject the null hypothesis. There is not enough evidence to suggest there is a significant interaction between County and Category. An interaction plot can be seen in figure 1.2. Test statistics and p-values for all three tests can be seen in table 1.3.

Figure 1.2

*Interaction plot between County, Category and Weight*

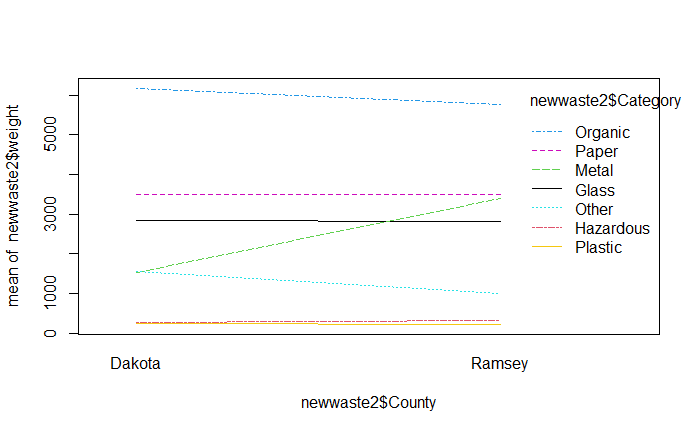


Table1.3

*Test Two Statistics and p-values*

|  |  |  |
| --- | --- | --- |
| **Variable** | **F Value** | **P Value** |
| County | 0.356 | 0.551 |
| Category | 19.084 | <2e-16 |
| County: Category | 0.952 | 0.457 |

From our findings we have learned that businesses recycle significantly more material than residents. In the greater context of the test this result does make sense as figure 1.1 clearly shows that businesses recycle around 100 tons more weight than residents do. Our second test revealed that there is not a significant difference in the amount recycled between counties. However, we did find that there is a significant difference in the amount being recycled between the types of recyclable material. Lastly, we found that there is not an interaction between County and Category. These results make sense as not every material is as common. For instance, most people and businesses use and recycle paper products but not every person or business produces hazardous waste that needs to be recycled.

Overall, I found this project to be very interesting. It is good to know that businesses take recycling as seriously as they produce a large quantity of waste. I found is surprising that county and category do not interact with each other. I would have figured that some counties contain more industries that produced more of a certain kind of waste than others (like a paper factory in on county producing more paper than another county without a paper factory).

One limitation of this study is that we did bot explore every variable in the data set. For instance, Year was not explored in any of our tests. When trying to understand a topic as great recycling rates and patterns of businesses and residents in Minnesota a bigger test investigating all available data may have been a better choice. Another limitation is that we had to narrow the second test down to just three of the moist populous Counties. It could be useful not combine Counties into regions to get a better picture of recycle amount for the entirety of Minnesota.

**GetHub Repository Link:**

https://github.com/AdrianParenteau/Data-301