**ANOVA Project for Minnesota Waste Data**



**Bethelhem Kassa & Hodo Ibrahim**

**STAT 301| Analysis of Variance and**

**Multivariate Analysis**

**Project Outline**

1. Introduction

* Research Questions
* Variable Chosen
* Null and alternative hypothesis

1. Methods
2. Conclusion
3. Bibliography

**Tables**

1. Table 1. Statistical summary of landfilled in tons for the four counties

**Figures**

1. One-way ANOVA test result
2. TukeyHSD test result
3. Error\_bar to show the significant mean landfilled difference among the four counties
4. One-way repeated measure test result
5. Posthoc test result

**Minnesota Waste Data Analysis**

**Introduction**

For this Project, we have used the watedata2, which lists Minnesota waste management: recycling, organics, onsite, WTE, and Landfilled in tons from 1991 to 2017 for around 85 counties. However, we mainly focused on “Recycling” for certain years and “Landfilled” in tons for four counties. This analysis will answer the following research questions raised.

**Research Question:**

1. Is there a significant mean landfilled tons difference between the four counties (Anoka, Dakota, Hennepin, and Ramsey)?
2. Is there a significant mean recycling difference for the years 1991, 2000, 2009 & 2017?

**Variables Chosen**

In order to answer these questions and run the appropriate analysis, we have used the variables county, recycling, Landfilled, and year.

For the first question, we need to use “Landfilled” as the dependent variable whereas “County” is the independent variable. And for the second question, “Recycling” is the dependent variable, and “Year” is the independent variable.

Landfilled: the amount of landfilled in tons

Recycling: the amount of recycling in tons

Year: the year when data are collected

County: county in MN where data are collected

Since there are around 85 counties, we only chose Anoka, Dakota, Hennepin, and Ramsey counties because of their population size. According to Minnesota Demographics (2021):

* Hennepin - 1,270,283
* Ramsey – 549,379
* Dakota – 435,863
* Anoka – 360,773

According to the Minnesota Pollution Control Agency, recycling benefits the environment and economy. It slows the filling of landfills and saves energy. Lower energy use means reduced air pollution and greenhouse gas emissions.

In order to show the recycling significant difference over the years, we chose every 9-year interval starting from 2001.

**Null and Alternative Hypothesis**

For question 1:

*H0: μA= μD= μH= μR;*

*where μA = mean landfilled of Anoka County μH = mean landfilled of Hennepin County,*

*μD = mean landfilled of Dakota County, μR = mean landfilled of Ramsey County,*

*Ha: at least two of the means of landfilled are different between the four counties*

For question 2:

*H0:μ (1991) = μ (2000) = μ (2009) = μ (2017);*

*whereμ (1991) = Recycling in tons for 1991 μ (2000) = Recycling in tons for 2000*

*μ (2009) = Recycling in tons for 2009 μ (2017) = Recycling in tons for 2017*

*Ha: at least two years have a difference in means of recycling*

**Methods**

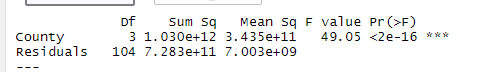
In order to answer the questions raised for this project, we have used *one-way ANOVA* and *one-way repeated measure* respectively. We chose *One-way ANOVA* because our independent variable (County) has four levels and we were comparing means for the four groups. The reason for choosing *one-way repeated measures* was that we had one independent variable, so that told us it would be a one-way test and since the same counties were in each group it would be a repeated measure (repeated more than two times). The assumptions were met because the counties were independent of each other, and the sample size was large enough to assume a normal distribution. Before testing the hypothesis, we used the filter function to choose the four counties and four years. Statistical analysis for the “Landfilled” variable has been performed. Also, we tried to see if there were missing values but didn’t find one. And the statistical significance level used for the analysis is 0.05.

|  |  |
| --- | --- |
| **Landfilled in tons** | |
| **Minimum** | **0** |
| **1st Quartile** | **69161** |
| **Median** | **169531** |
| **Mean** | **172792** |
| **3rd Quartile** | **217423** |
| **Maximum** | **513092** |

***Table 1. Statistical summary of Landfilled in tons for the four counties***

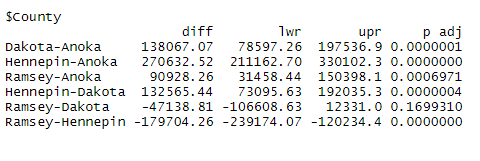
**Conclusion**

During the analysis of the first question “Is there a significant difference between the mean landfilled tons of the four counties?” We got an F statistic of 49.05 and a p-value of almost 0.

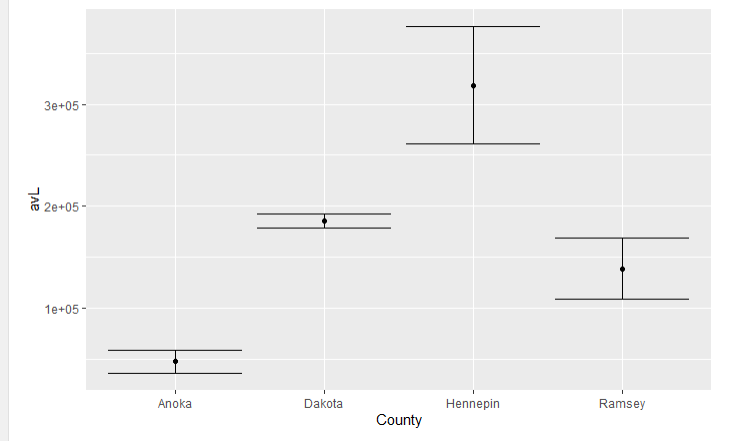


***Figure 1, one-way ANOVA test result***

So, we reject the null hypothesis. Therefore, there is sufficient evidence to support the claim that there is a significant difference between the mean landfilled tons of the four counties. In order to see which counties, have a significant difference we used the TukeyHSD function. And this showed us significant differences among all the counties except between Ramsey and Dakota. We can also see the mean difference among these counties in the error bar graph, figure 3, below. The interesting finding from this analysis is that Hennepin’s mean landfilled is significantly different from the other counties, this might be due to the high population difference between Hennepin and the other counties.



***Figure 2, TukeyHSD() result***



***Figure 3. Error bar to show the significant mean landfilled difference among the four counties.***

For the second question,” Is there a significant mean recycling difference for the years 1991, 2000, 2009 & 2017?”. Upon running the test, using statical software, we got an F statistic of 15.36 and a p-value of less than .0001.

Table

Description automatically generated

F***igure 4, One-way repeated measure test result***

Using a significance level of 0.05, we can make the decision to reject the null hypothesis. Our p-value is significant, so there is evidence that there is a significant mean recycling difference between the years 1991, 2000, 2009, and 2017. With the test being significant, we wanted to know what groups had significant mean differences between them. After running a posthoc test, we learned that there was a significant mean difference between groups 1991 & 2000, 1991 & 2009, 1991 & 2017, 2000 & 2017, and 2009 & 2017.

A screenshot of a computer

Description automatically generated with medium confidence

***Figure 5, posthoc test result***

This helped us have a deeper understanding of our test. The most interesting finding was learning that there was no significant difference between 2000 & 2009, but every other group had a significant mean difference between them. We wonder what factors could have contributed to that result. Our analysis was limited because our data only went as far back as 1991. We also only included 4 groups in our analysis but at the same time, we didn’t want a large number of groups that could make our analysis more complex.

**Bibliography**

1. <https://www.minnesota-demographics.com/counties-by-population>
2. Minnesota Pollution Control Agency

<https://www.pca.state.mn.us/air-water-land-climate/recycling-in-minnesota>