**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 1.1 Statistical summary of landfilled in tons for the four counties

**Figures**

1. Figure 1.1 One-way ANOVA test result
2. Figure 1.2 TukeyHSD test result
3. Figure 1.3 Error\_bar to show the significant mean landfilled difference among the four counties
4. Figure 2.1 One-way repeated measure ANOVA test result
5. Figure 2.2 Posthoc test result

**Minnesota Waste Data Analysis**

**Introduction**

We will use the Minnesota waste management data from 1991 – 2017 labeled Wastedata2, to perform two different analyses. We have chosen to focus on two types of waste management. We will be using “Recycling” and “Landfilled” amounts in tons to formulate our research questions.

**Research Question**

1. Is there a significant difference between the mean landfilled amount in tons for 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**

To answer these questions and run the appropriate analysis, we will be using the variables County, Recycling, Landfilled, and Year.

For question 1, “Landfilled” will be the dependent variable and “County” will be the independent variable. For question 2, “Recycling” will be the dependent variable, and “Year” will be the independent variable.

Landfilled: landfill amount in tons

Recycling: recycling amount in tons

Year: year when the data was collected

County: county in MN where data was collected

Since there are about 87 counties in the state of Minnesota, we have chosen to focus on Anoka, Dakota, Hennepin, and Ramsey counties for question 1, 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.

We also chose 1991 as a baseline for question 2 and retested after every 8-9 years to see if there is a significant mean difference in recycling over the years.

**Null and Alternative Hypothesis**

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 means are different*

Question 2:

*H0: μ (91) = μ (00) = μ (09) = μ (17);*

*Where μ (91) = mean recycling in tons for the year 1991*

*μ (00) = mean recycling in tons for the year 2000*

*μ (09) = mean recycling in tons for the year 2009*

*μ (17) = mean recycling in tons for the year 2017*

*Ha: at least two means are different*

**Methods**

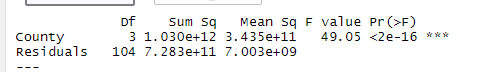
To answer the research questions above, we will be using One-way ANOVA and One-wayrepeated measure ANOVA. We chose One-way ANOVA for question 1, because our independent variable (County) has four levels, and we are comparing means for the four different groups. We chose One-way repeated measures ANOVA for question 2 because we have one independent variable that has four levels, and since the same counties are in each group, the test would be a repeated measure. The assumptions are met because the counties are independent of each other, and the sample size was large enough to assume a normal distribution. Before testing the hypothesis, we used the appropriate functions to prepare our dataset for the tests that were about to be performed. Statistical analysis for the “Landfilled” variable has been performed and displayed below.

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

***Table 1.1 Statistical summary of Landfilled in tons for the four counties***

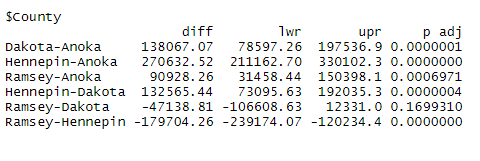
**Conclusion**

For the analysis of question 1, we used a significance level of 0.05 to test the claim “Is there a significant difference between the mean landfilled amount in tons for the four counties?”. We got an F statistic of 49.05 and a p-value of almost 0.

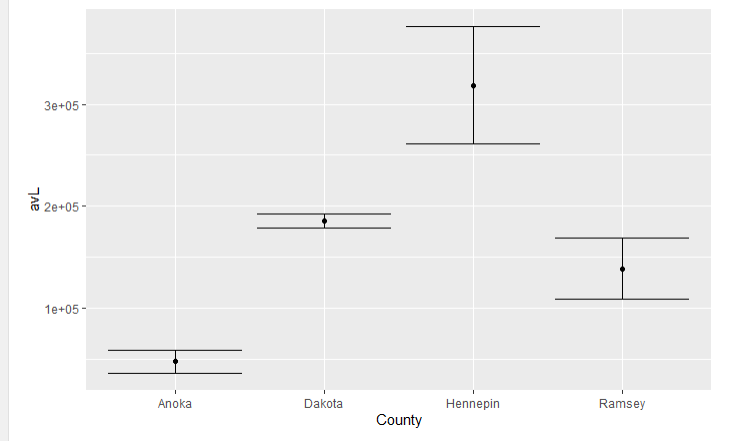


***Figure 1.1 One-way ANOVA test result***

With these results, we reject the null hypothesis. Therefore, there is sufficient evidence to support the claim that there is a significant difference between the mean landfilled in tons of the four counties. To see which counties have a significant difference between them, we used the TukeyHSD function. This showed us significant differences between all the counties except Ramsey and Dakota. We can also see the mean difference among these counties in the error bar graph, figure 1.3, below. The interesting finding from this analysis is that Hennepin’s mean landfilled in tons is significantly different from the other counties, this might be due to the high population difference between Hennepin and the other counties.



***Figure 1.2 TukeyHSD() result***



***Figure 1.3 Error bar to show the significant mean landfilled difference among the four counties.***

For question 2, is there a significant mean recycling difference for the years 1991, 2000, 2009 & 2017? Upon performing the tests, using statistical software, we got an F statistic of 15.36 and a p-value of less than .0001.

Table

Description automatically generated

F***igure 2.1 One-way repeated measure ANOVA 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 a significant test, we want to know what groups had significant mean differences between them. After performing a posthoc test, we learned that there is 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 2.2 Posthoc test result***

This allowed us to have a deeper understanding of our tests. 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>