Does the Water Quality Vary Between Modoc County, CA and Santa Clara County, CA?

Aditya Kiran Aswin Kumar: addyk@umd.edu

Pranav Tejasvi Adiraju: padiraju@umd.edu

Alanna Anderson: aander23@umd.edu

Vijay Arni: varnis@umd.edu

University of Maryland

Professor Antoun

INST 627 Fall 2022

December 10, 2022

1.Introduction

The research question is: Does the water quality vary between Modoc County, CA and Santa Clara County, CA? We chose this research question because we wanted to know if water quality would vary between a rich and poor region in the United States. Modoc County, CA and Santa Clara County, CA represent our chosen economic focus because of their respective differences in median home price. Modoc County, CA has a median home price of $192,800 while Santa Clara County has a median home price of $1,296,980.

The research question is important because it allows us to understand matters of inequities in the United States. People living in areas with poor water quality might suffer from poor health outcomes and increased rates of illnesses. We believe that people deserve to know if the water quality in their area is insufficient and that people deserve to know if low-income people will bear the brunt of these inequities due to their places of residence.

2. Methods and Data

2.1 Data Sources

National Association of Realtors
We collected the median home price by county from this data source. The data source contains the median home price of all the counties in the US where we picked one of the highest median home price counties and the lowest median home price county in the state of California.

California Open Data Portal
We collected the quality of water data from this data source. The quality of water dataset contains DWR-collected, current and historical, chemical, and physical parameters found in drinking water, groundwater, and surface waters throughout California.

2.2 Analysis
The two-sample t-test (also known as the independent samples t-test) is a method used to test whether the unknown population means of two groups are equal or not. Usually, a two-sample t-test is used when we want to compare two independent groups or populations to see if their means are different.

We are using a two-sample t-test here because we have two samples (one sample being the sample collected from Modoc and the other from Santa Clara) where we are trying to determine if they are equal or not. The preliminary answer to the research question is that water quality does vary between Modoc County, CA and Santa Clara County, CA. The null hypothesis is that water quality does not vary between Modoc County, CA and Santa Clara County, CA.

To determine whether we can reject the null hypothesis, we examined four dependent variables to determine water quality. The dependent variables are: specific conductance, dissolved oxygen, turbidity and pH. Our independent variable is median home price.

2.3 Specific Conductance
Specific conductance (also called specific conductivity or just conductivity) is a measure of the ability of water to conduct an electrical current. Specific conductance is a crucial water-quality measurement because it estimates the dissolved material in the water. Our analyses and descriptive statistics for specific conductance are shared below:
- Mean= 444.68
- Standard Deviation=284.51
- Count(n)=822
- The sample 1 mean (340.25) is significantly different to sample 2 mean (507.26)(t=-8.15), (p<0.01).

2.4 Dissolved oxygen
Dissolved Oxygen is the amount of oxygen present in the water. Water containing high levels of organic matter can make water oxygen deficient, which in turn degrades the quality of water. Our analyses and descriptive statistics for dissolved oxygen are shared below:
- Mean=9.097941
- Standard Deviation=2.61294
- Count(n)=3537
- The sample 1 mean (8.763291) is significantly different to sample 2 mean (9.287785)(t=-6.0741), (p<0.01).

2.5 Turbidity
Turbidity is a measure of the relative clarity of the water. Large particles such as minerals, dirt, plankton, bacteria and clay can cause cloudiness. Turbidity is an essential water-quality measurement because it estimates the number of large particles in the water. Our analyses and descriptive statistics for turbidity are shared below:
- Mean: 20.741
- Standard Deviation: 28.11348
- Count(n)=299
- The sample 1 mean (24.43195) is significantly different to sample 2 mean (8.016418), p=0.00002).

2.6 pH
PH levels impact most chemical reactions in the water. It is one of the critical elements of drinking water. pH levels under five and beyond ten are considered harmful to humans as they are acidic and alkalic, respectively.

- Mean=7.755301
- Standard Deviation=0.4806247
- Count(n)=5156
- The sample 1 (7.78754) mean is significantly different to sample 2 mean (7.75360) (t=2.4728) (p-value = 0.1345).

Since the average pH for water quality in Santa Clara and Modoc Counties lie within the range of 6.5-8 pH levels. We cannot determine if water quality has any impact on either county.

3. Results

We used specific conductance, dissolved oxygen, turbidity and pH to measure the quality of water. Our findings are reported in the following sections:

3.1 Specific conductance

The sample 1 mean (340.25) is significantly different to sample 2 mean (507.26)(t=-8.15), (p<0.01). We rejected the null hypothesis and concluded that specific conductance varies from county to county depending on median property value.
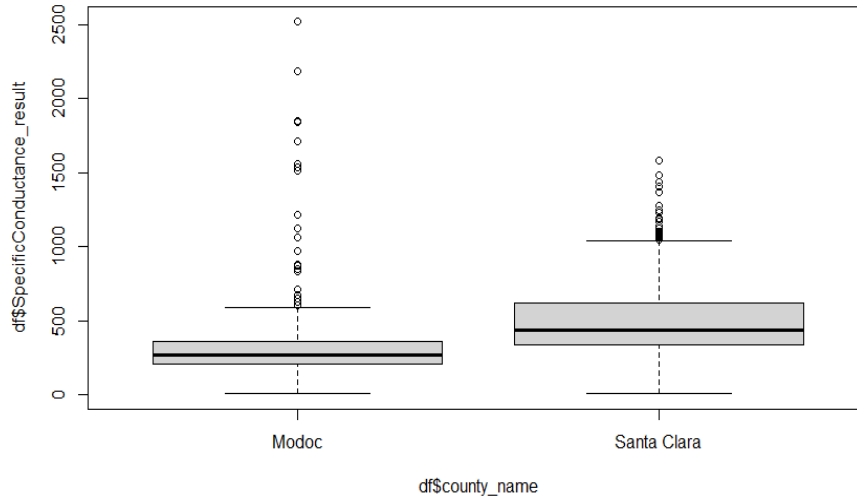

Figure 1: Box Plot for Specific Conductance

3.2 Turbidity

The sample 1 mean (8.763291) is significantly different to sample 2 mean (9.287785)(t=-6.0741), (p<0.01). We rejected the null hypothesis and concluded that turbidity varies from county to county depending on median property value.

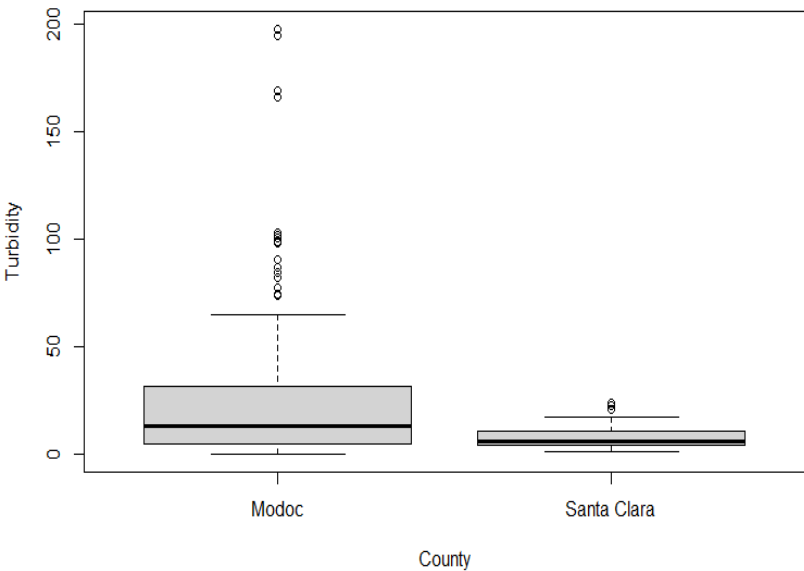Figure 2: Box Plot for Turbidity

3.3 Dissolved Oxygen

The sample 1 mean (24.43195) is significantly different to sample 2 mean (8.016418), p=0.00002). We rejected the null hypothesis and concluded that turbidity varies from county to county depending on median property value.
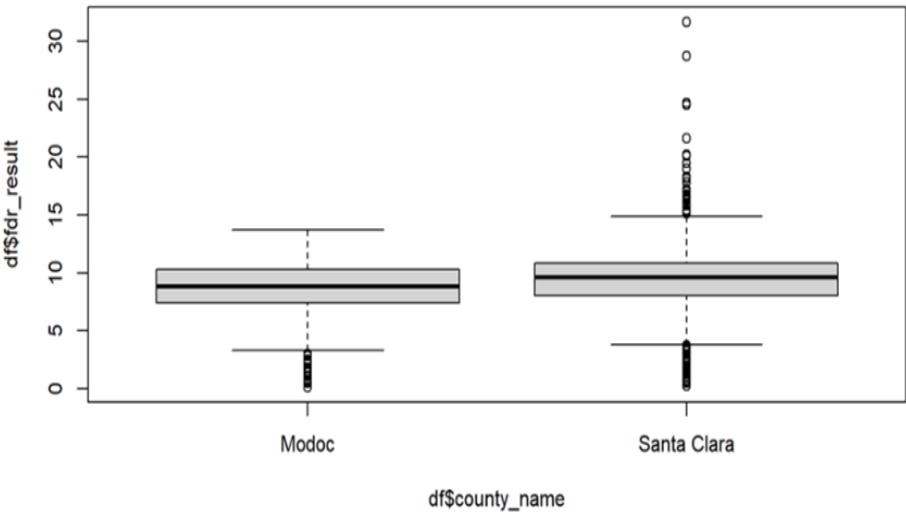

Figure 3: Box Plot for Dissolved Oxygen

3.4 pH

The sample 1 (7.78754) mean is significantly different to sample 2 mean (7.75360) (t=2.4728) (p-value = 0.1345). Since the average pH in Santa Clara and Modoc County lies within the range of 6.5-8 pH levels, we cannot determine if water quality has any impact on either county. Therefore, we cannot reject the null hypothesis or conclude that pH varies from county to county depending on median property value.
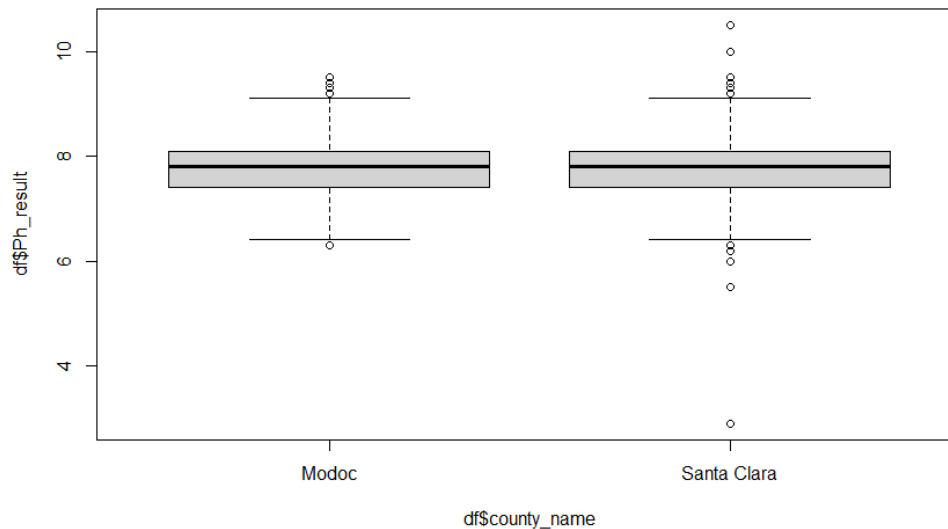

Figure 4: Box Plot for PH

From the outputs of the two-sample t-test, it can be seen that the quality of water significantly varies from a rich county like Santa Clara to a poor county like Modoc in the state of California.

4. Conclusion

4.1 Limitations

The first limitation was discovered in the pH parameter analysis. Although there seems to be a significant difference between the samples of the two counties, the means are not that different. Modoc's mean pH is 7.78 while Santa Clara's mean pH is 7.75. These values are not that impactful in trying to determine the water quality and thus, pH cannot be used as a valid parameter.

The second limitation occurred in the turbidity analysis. The dataset contained an outlier at the extremity of the mean, specifically 3-sigma greater. In the final analysis, the outlier was

removed. Even in the absence of the outlier, the p-value showed a statistically significant relationship. It is possible that the outlier was present because of the geographic differences between Modoc County, CA and Santa Clara County, CA. Modoc County is a more rural area, which means turbidity water samples might have been gathered from more rivers and surface water than in Santa Clara County, which could have resulted in an outlier.

4.2 Research Conclusion

To conclude, the practical implication of the research is that the quality of water is different between Modoc County, a poor county, and Santa Clara County, a rich county. The findings can be used to study why certain water quality parameters are worse in these counties. The California State Water Resource Control Board must look into this matter and ensure all residents are provided with the same quality of water regardless of the county they live in. For further research, we recommend first looking at comparisons between other counties in California, and then pursuing other states in the United States.
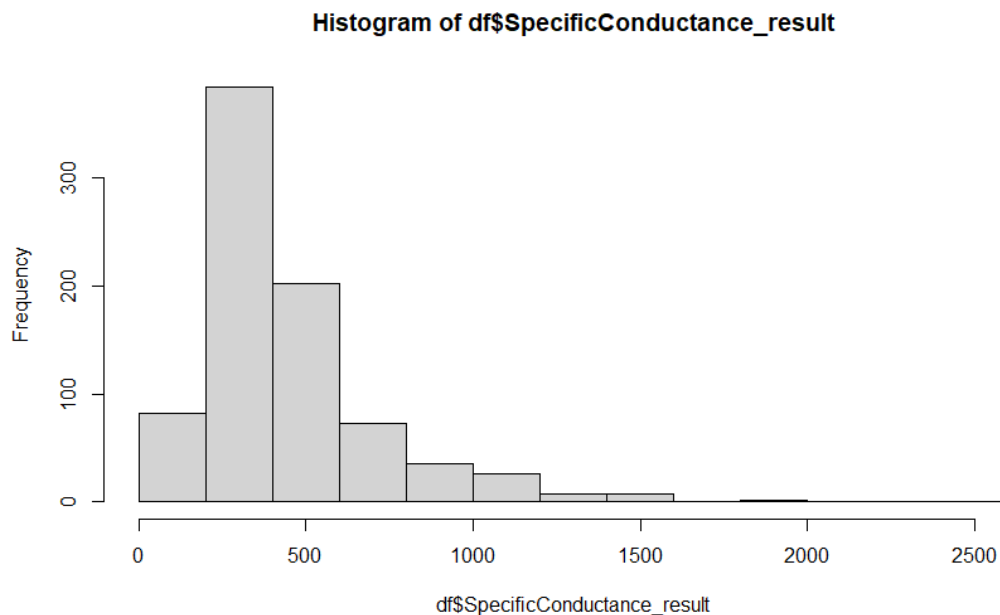
<u>5. Appendix</u>
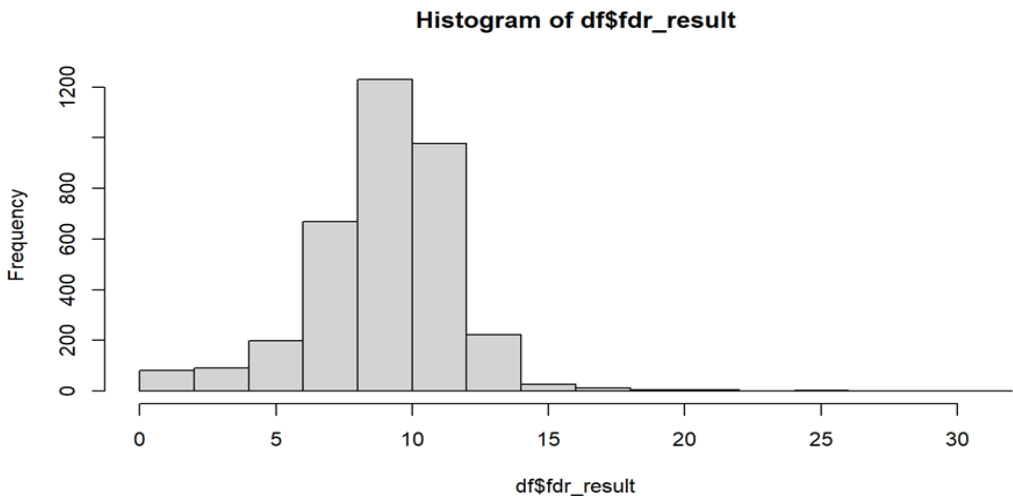

Figure 5: Histogram for Specific Conductance

Histogram of df$fdr_result

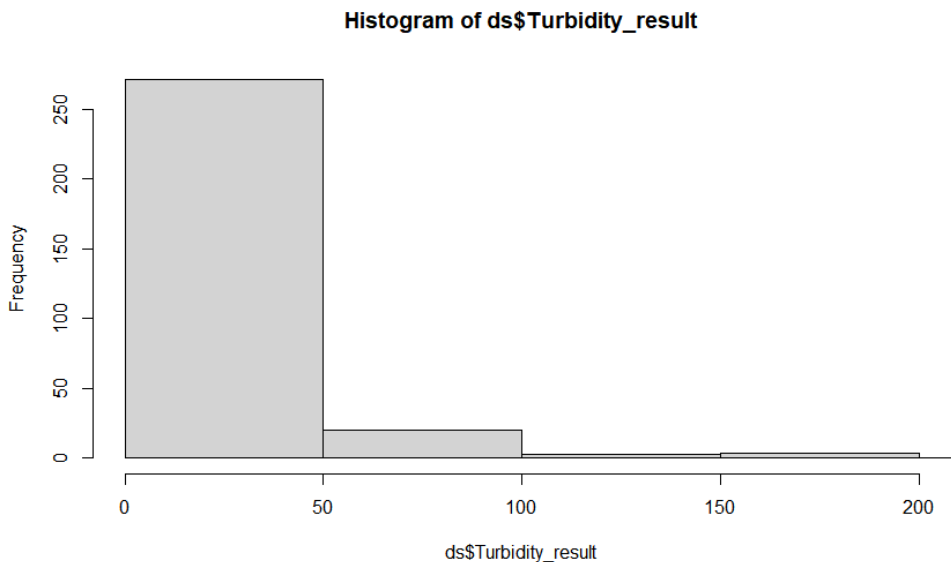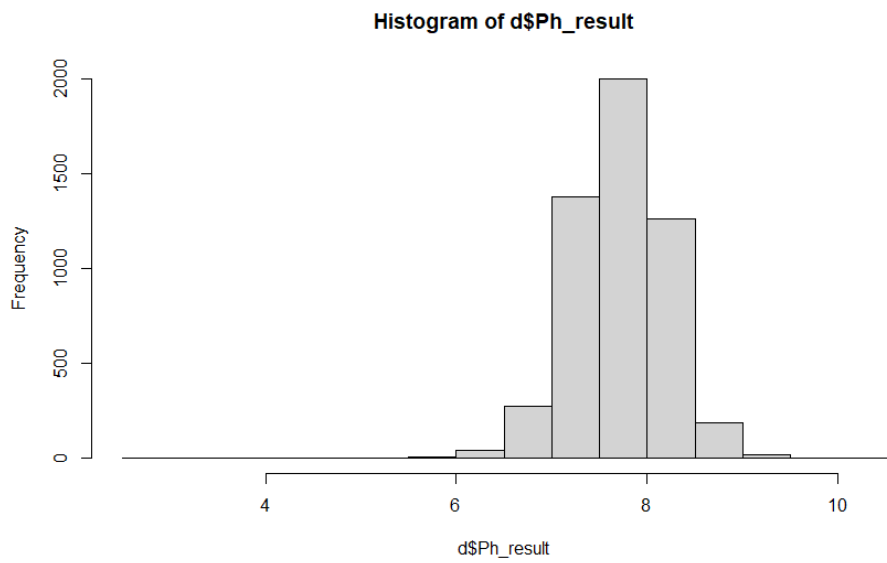Figure 6: Histogram for Dissolved Oxygen

Histogram of ds$Turbidity_result

Figure 7: Histogram for Turbidity

Figure 8: Histogram for PH