Data Analytics, Project 1

October 17, 2019

T. Nguyen

B. Kaczmarczik

M. Pollari

S. Roedl

E. Mandell

Minnesota Drinking Water Quality Analysis

The goal of our project was to ask who in Minnesota has access to safe drinking water. We approached this question by using the EPA database to select the levels of 3 contaminants in municipal water supplies throughout the state in 2017. We then took the poverty rates by county, ranked the counties in order of poverty level, and binned them into five groups. The contaminant levels by poverty level grouping were then analyzed. We addressed two of the following questions:

* **Question 1**: Is there a significance difference between the mean concentration of a particular contaminant in a group compared to another?
* **Question 2**: Is there a significance difference between the mean concentration of a particular contaminant in the wealthiest counties compared to the poorest?

Statistical Methods:

ANOVA (addresses the first question):

* **Null Hypothesis:** the mean concentration of contaminant is the same across the grouped counties
* **Alternative Hypothesis:** There are significant differences in contaminants between any counties.

t-test (addresses the second question):

* **Null Hypothesis:** the mean concentration of contaminant is the same among the wealthiest and poorest grouped counties
* **Alternative Hypothesis:** There are significant differences in contaminants between the wealthiest and poorest grouped counties.

Observations for Arsenic: There is no statistically significant difference in the levels of arsenic across income groups. The levels recorded in testing across the state are significantly below EPA safety standards, and do not vary significantly either within or between poverty level groupings.

F = 1.642

p-Value = 0.172

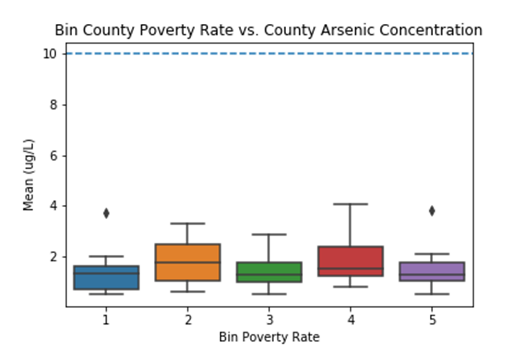
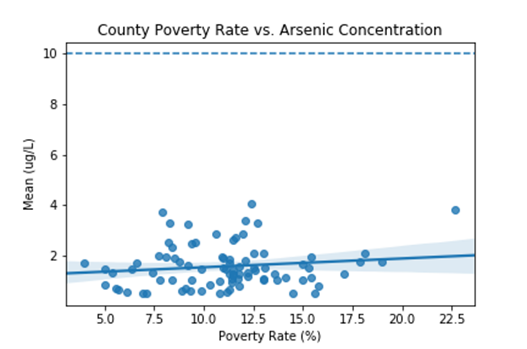
* No significant evidence that there is a difference in concentration of arsenic among the five bins.

*T-test:*

T=-0.488

p-Value = 0.629

* No significant evidence that there is a difference in concentration of arsenic between the wealthiest and poorest counties (bin 1 vs. bin 5).



Observations for Nitrates: There is no statistically significant difference in the levels of nitrates across income groups. The levels recorded in testing across the state are significantly below EPA safety standards, and do not vary significantly either within or between poverty level groupings.

*ANOVA:*

F = 0.343

p-Value = 0.848

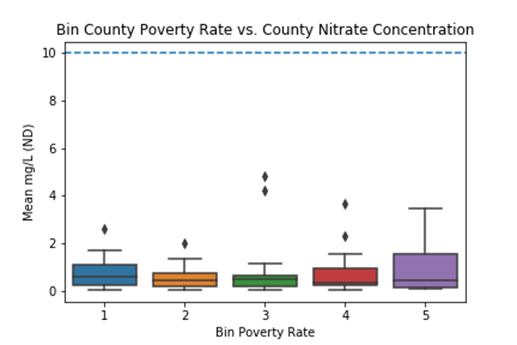
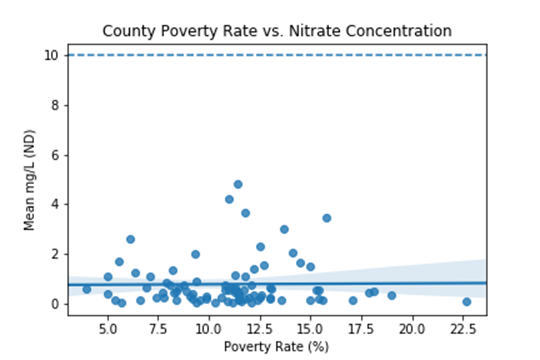
* No significant evidence that there is a difference in concentration of arsenic among the five bins.

*T-test:*

T=-0.5155

p-Value = 0.6105

* No significant evidence that there is a difference in concentration of arsenic between the wealthiest and poorest counties (bin 1 vs. bin 5).



Observations for Uranium: There is no statistically significant difference in the levels of nitrates across income groups, and they do not vary significantly within groupings. The levels recorded in testing across the state are in a few cases above the EPA safety limit, however.

*ANOVA:*

F = 1.0407

p-Value = 0.399

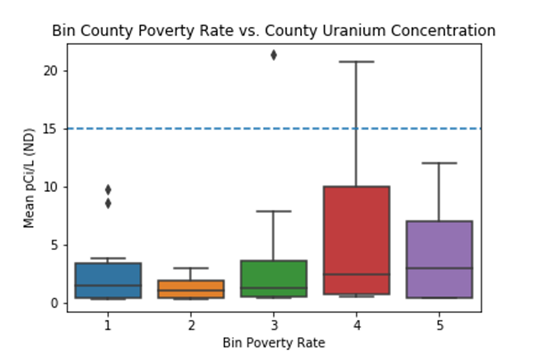
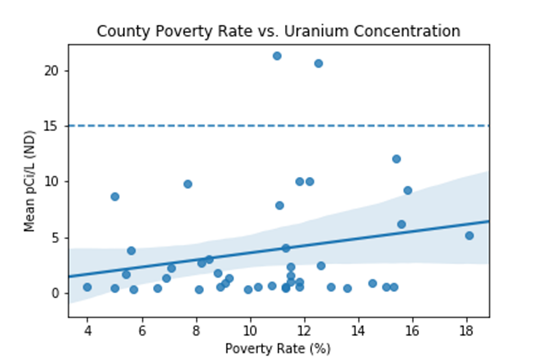
* No significant evidence that there is a difference in concentration of arsenic among the five bins.

*T-test:*

T=-0.7478

p-Value = 0.4678

* No significant evidence that there is a difference in concentration of arsenic between the wealthiest and poorest counties (bin 1 vs. bin 5).



Comments: According to 2017 EPA testing data, the drinking water sourced from municipal water supplies throughout the state of Minnesota are largely safe. Levels of arsenic and nitrate are well below EPA limits, while a small number of counties have levels of uranium which are slightly above the safety limit. Sorting the data frame by mean concentration shows Rock and Yellow Medicine counties have levels of uranium above the EPA limit

**Conclusions:** We fail to reject both the Primary and Secondary null hypotheses. We were not able to show that low income areas have poorer water quality than wealthier areas in Minnesota, or that there is a significant difference in water quality between any of the county groupings. We can also examine the water quality by the city-level instead of by county-level if given more time and available resources. Performing the ANOVA on grouped cities might or might not unveil more findings about their population means. If at least one population mean is different from the others, further testing (ad hoc test) will be needed to identify which groups are different from another.