Maine Lakes – Study on Mercury Levels in Fish

The Government of Maine

Parikshit Patil

26th February, 2024

Executive Summary

The purpose of this study is to assist the Government of Maine in understanding whether the lakes of Maine play a role in allowing toxic chemicals to enter the food chain, specifically mercury, and some factors possibly affecting this. The dataset provided is first preprocessed and cleaned to a more usable format. To perform the analysis, a 3-way framework was used involving visual exploration, quantitative analysis, and statistical analysis. This framework is successful in providing clear and valid conclusions to each of the problem statements posed by the government of Maine. Finally, my reflection on the effect of such methods on public policy is also included.

Introduction

There is a concern at present revolving around toxic chemicals entering the food chain. Mercury levels in fish consumed by people is an area of interest, and the Government of Maine has requested an analysis of the freshwater lakes data they possess. The data provided contains information on mercury levels in fish at a sample of freshwater lakes in their state. Among the 18 features the dataset provides information on, the study and analysis utilize the following three which are relevant to the task at hand: Mercury level in parts per million (ppm), lake type, and presence of human infrastructure.

The task was to investigate 3 areas related to the mercury levels in the fish of the lakes in Maine, using the data provided.

1. To investigate and aim to provide any insight into whether or not the mercury levels at the lakes of Maine should be of concern
2. To investigate if the presence of human water flow infrastructure is related to mercury levels in fish of Maine's lakes
3. To find out if the mercury levels vary by lake type. Specifically, whether it is true or not that oligotrophic lakes have the highest mercury levels and eutrophic lakes have the lowest mercury levels.

Methods

To perform all my analysis, I used software (python) to do all the data processing, statistical analysis, and data visualisation. To begin with, I found that the lake named "PLEASANT", and the ponds named "LONG", "NORTH", "OTTER", and "ROUND" all have different lakes/ponds with the same name. To avoid any issues with the analysis, I chose to differentiate these duplicated names by adding a "1" to the end of the duplicated names. Additionally, I found that 10 values of data were missing from the data set, out of which only 3 were a part of the data I required, which were part of only 2 lakes. Since we have a large number of lakes in the dataset (120), I decided to remove these two lakes from the data thereby avoiding any problems with statistical analysis due to missing data. Finally, for each of the three tasks to analyse, I followed a common approach of first looking at the data visually to identify any trends, viewing the quantitative values of the data relevant to the task, and lastly performing statistical analysis to either confirm or reject the visual trends observed.

Analysis & Results

For task 1 which focused on identifying whether the mercury levels in the lakes of Maine are high enough to be a concern, I created a histogram plot showing the number of lakes at different mercury levels (Figure 1). Through visual analysis, the observable trend is that initially, the number of lakes increases for higher mercury levels, peaks around the same level as Maine's warning level, and then gradually reduces. This suggests that most lakes are within the safe level of mercury set by the Government of Maine. By quantitative analysis (Table 1), I found that 53 lakes have mercury levels above the warning level set by Maine. 53 lakes may seem like a large number, yet small when compared to the total number of lakes (120). Therefore, I employed a statistical analysis method known as a one-sample t-test (Appendix) which statistically compares the lakes' mercury levels to Maine's warning threshold of 0.43. The result of this test conveys that there isn't enough evidence to conclude that the mercury level in Maine lakes exceeds the government's warning threshold. As a result, I determined that the data suggests that mercury levels at the Maine lakes are not high enough to be a concern.

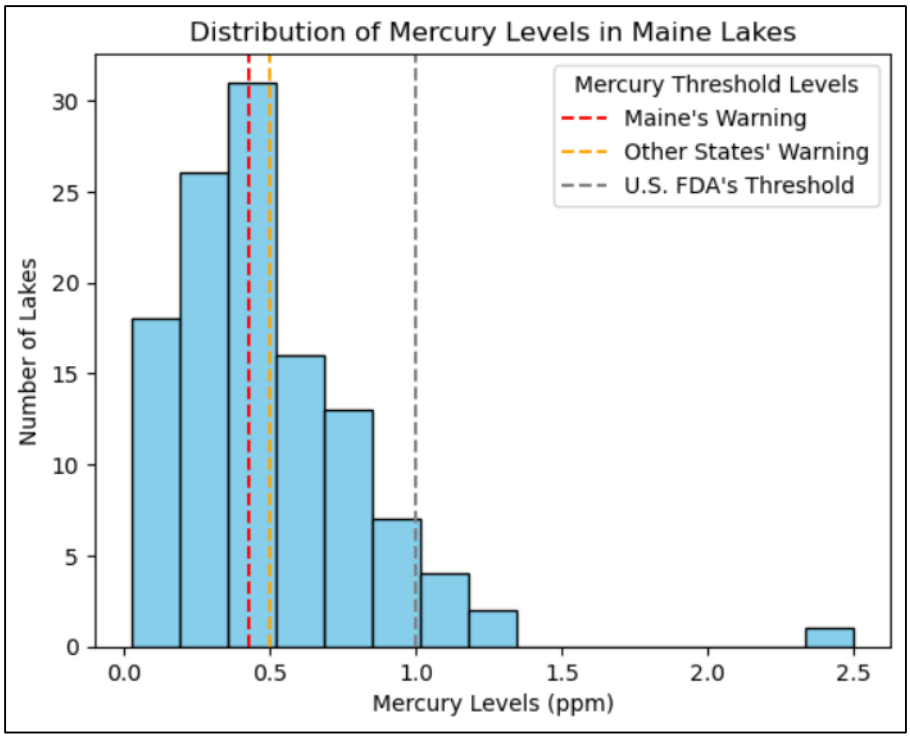

Figure 1 – Distribution of Mercury Levels in Maine Lakes showing the number of lakes at different mercury levels.

Number of lakes with mercury level > 1ppm	7
Number of lakes with mercury level > 0.5ppm	44
Number of lakes with mercury level > 0.43ppm	53

Table 1 – Number of lakes above the different mercury level thresholds

Now consider task 2, which aims to identify if the presence of a dam or other human infrastructure affects the mercury levels of the Maine lakes. Figure 2 depicts the boxplot used to analyze visually the mercury levels for lakes with human infrastructure vs those without human infrastructure. The plots do not reflect any differentiating trend since for both situations the range of mercury levels is similar. Using quantitative data (Table 2) for the number of lakes, we see that there are 25 and 28 lakes without and with human infrastructure respectively whose mercury levels are greater than 0.43. To confirm the analysis thus far, the statistical analysis of a two-sample t-test for mercury levels with and without human infrastructure (Appendix) adds validity by also showing that there is indeed no significance of human infrastructure being present or not in affecting mercury levels.

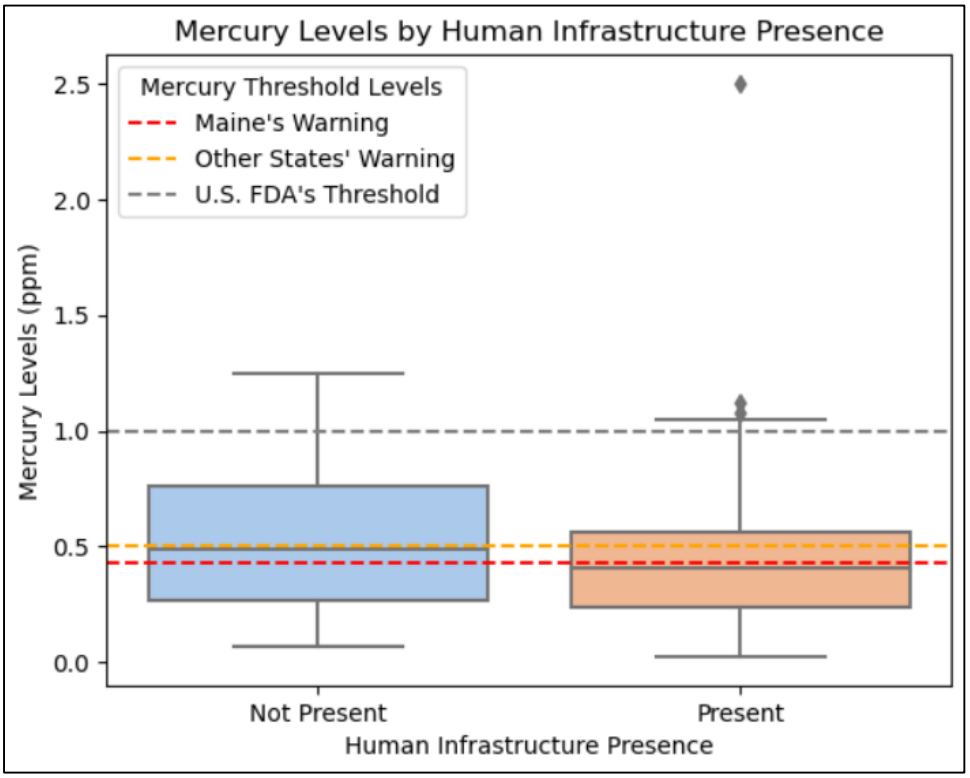

Figure 2 – Box plots showing Mercury Levels in Maine Lakes with nearby human infrastructure not present and present.

Number of Lakes without Human Infrastructure & Mercury Level > 1ppm	3
Number of Lakes without Human Infrastructure & Mercury Level > 0.5ppm	23
Number of Lakes without Human Infrastructure & Mercury Level > 0.43ppm	25
Number of Lakes with Human Infrastructure & Mercury Level > 1ppm	4
Number of Lakes with Human Infrastructure & Mercury Level > 0.5ppm	21
Number of Lakes with Human Infrastructure & Mercury Level > 0.43ppm	28

Table 2 – Number of lakes above the different mercury level thresholds based on the presence of human infrastructure

Finally, for task 3, we aim to find out if the mercury levels vary by the type of lake and to address the validity of previous research which claims oligotrophic lakes have the highest mercury levels and eutrophic lakes have the lowest levels of mercury. We can again use a boxplot to understand the different mercury levels based on lake type (Figure 3). A clear observation is that some eutrophic lakes have higher levels of mercury than oligotrophic lakes. Quantitatively (Table 3), the data tells us that there are 5 oligotrophic lakes and 28 eutrophic lakes with mercury levels greater than 0.43, which is very small compared to the total number of lakes (120). This still is sufficient to disprove the past research since we can see that more eutrophic lakes have higher levels of mercury than oligotrophic lakes. Now to check whether the lake type itself directly has any effect on mercury levels, an ANOVA statistical test (Appendix) can be performed for variation of mercury levels by lake type. The results from this test show that there is not enough evidence from the data to conclude that the lake type affects the mercury levels.

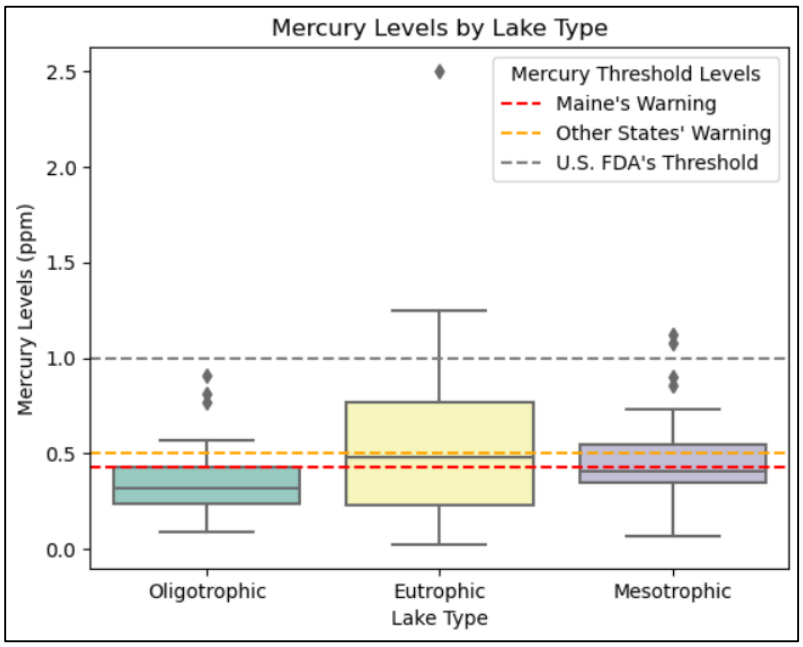

Figure 3 - Box plots showing Mercury Levels in Maine Lakes based on the lake type.

Number of Oligotrophic Lakes with Mercury Level > 1ppm	0
Number of Oligotrophic Lakes with Mercury Level > 0.5ppm	4
Number of Oligotrophic Lakes with Mercury Level > 0.43ppm	5
Number of Eutrophic Lakes with Mercury Level > 1ppm	5
Number of Eutrophic Lakes with Mercury Level > 0.5ppm	26
Number of Eutrophic Lakes with Mercury Level > 0.43ppm	28
Number of Mesotrophic Lakes with Mercury Level > 1ppm	2
Number of Mesotrophic Lakes with Mercury Level > 0.5ppm	14
Number of Mesotrophic Lakes with Mercury Level > 0.43ppm	20

Table 3 – Number of lakes above the different mercury level thresholds based on the lake type

Conclusion

To summarize our results, we can say that the data suggests the mercury levels are almost high enough to be a concern, however, not enough to be a concern yet. Additionally, there is clear evidence that the presence of dams or any other man-made infrastructure has no significant effect towards the mercury levels in fish of the lakes in Maine. Furthermore, opposite to the results suggested by past research, eutrophic lakes seem to have higher levels of mercury than oligotrophic lakes. However, we conclude that based on the data there is still not enough evidence to claim that the lake type affects the mercury level directly.

I believe that such data modelling can have a significant impact on public policy since it reveals insights that cannot be understood merely by observation of data. Moreover, data modelling raises new decisions such as whether to create future-proof policies based on the current analysis and trends. As seen in the Maine lakes examples, data modelling also helps strengthen and add confidence to the observed trends, helping the policymakers quickly transition from a research and planning phase to an implementation phase.

References:

SciPy. *Scipy.stats.ttest_1samp*. scipy.stats.ttest_1samp - SciPy v1.12.0 Manual. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_1samp.html

SciPy. *Scipy.stats.ttest_ind*. scipy.stats.ttest_ind - SciPy v1.12.0 Manual. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html

SciPy. *Scipy.stats.f_oneway*. scipy.stats.f_oneway - SciPy v1.12.0 Manual. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html

SciPy. *Scipy.stats.chi2_contingency*. scipy.stats.chi2_contingency - SciPy v1.12.0 Manual. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chi2_contingency.html

GPT, Chat (2024, February 25). *Python code error correction*.

Appendix

All the data visualisations, quantitative data analysis, and statistical analysis of data are implemented through software coding using Python. This code can be found at the following link for a more detailed understanding: https://github.com/parikshit13/Maine-Lakes-mercury-level-analysis