**Mercury Contamination of Fish in Ontario Bodies of Water**

STAT3240 Term Project

Sasha Ross-Jewell, Angel Rios, Ahmed Salem, Shahana Lalmahomed, Simona Singh

Problem Statement

In humans, mercury was first connected to medical concerns in the mid 20th century, especially in populations that would eat substantial amounts of fish. One of the most notable cases in history is the city Minamata, Japan, which was severely impacted by mercury poisoning in the 1950s (Environment Canada, 2003). Thousands were affected by the mercury they consumed from fish, resulting in devastating neurological impacts especially for pregnant and nursing women and young children, due to its neurotoxicity (Environment Canada, 2003). As more information came out regarding the effects of contaminants on fish, and their effects on the human body, tracking those levels became a priority.

The problem we have studied was if there is a relationship between mercury contamination in fish and their weight, species, and sex. As mercury bioaccumulates in tissue, we hypothesized that fish with greater weight will also have a higher mercury level, as the level increases up the aquatic food chain. We also speculate that if weight is related to mercury level, then larger species of fish will have higher weight and therefore more mercury contamination. We are aware that in many species, female fish also tend to be larger than male fish so we also hypothesize female fish will have higher mercury levels.

Data Collection

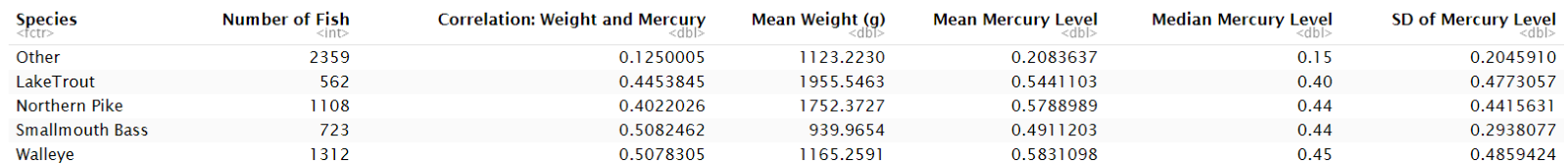
Since the 1960's, data has been collected about the amount of contamination detected in fish of Ontario bodies of water, and this information is used for advisories regarding safety of fish for consumption and trends of the contaminants over time.

The data came from the Ontario government, and this dataset specifically included information gathered from 2005 to 2018. As this data was collected on 125,000 fish, we decided to narrow it down by only using data from 2018, the fish being collected using large gill nets that would attempt to capture a wide variety of the population. Small muscle samples would be taken, frozen and then analyzed for contaminants.

Instead of including other contaminants also studied, such as polychlorinated biphenyls, we just looked at mercury contamination. There were 23 variables in the original dataset, so we removed all except the variables we would study; leaving mercury level (in micrograms per gram), weight (in grams), species and sex. Some of the fish were missing data such as species and weight, and we removed those from the dataset to simplify the analyses we would be doing. There were many species included in the study, so we decided to look at the four most prevalent species, and group the rest into "Other," which will be used as our reference category.

Analyses

Data Exploration

We did descriptive statistics on the data, specifically looking at the correlation between weight and mercury, and the mercury level mean, median and standard deviation grouped by species.

As displayed by this output, there appears to be a moderately strong positive linear correlation (ranging from 0.40 to 0.508) between weight and mercury level, at least for the individual species studied; Lake Trout, Northern Pike, Smallmouth Bass and Walleye. “Other” did not show this correlation, likely as it is many species combined into one category.

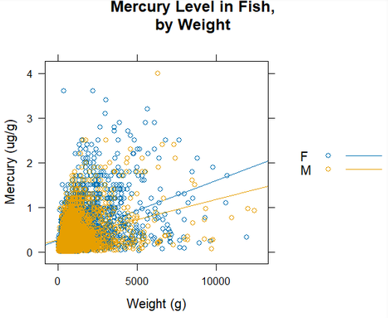
The mean and median of the mercury levels are similar across the four specific species, ranging from 0.49 to 0.58 µg/g and 0.4 to 0.45 µg/g respectively. There is a large decrease from those ranges to the values for the “Other” species – with a mean of 0.208 µg/g and a median of 0.15 µg/g, indicating lower mercury levels in general in the species not specifically analyzed. There is a smaller standard deviation, and therefore lower amount of variation, in mercury level for “Other” and Smallmouth Bass at 0.20 and 0.29 µg/g respectively. The small SD of the Others is surprising as we expected a wide variety of species to vary more than the specific species themselves. The remaining three species, Walleye, Lake Trout and Northern Pike all had similar SD ranging from 0.44 to 0.48 µg/g.

Histogram of Mercury Level

A graph of a graph

Description automatically generatedWe made a histogram of the Mercury Level (µg/g), with a solid vertical line displaying the level of mercury deemed safe to consume by Health Canada at 0.5 µg/g (Environment Canada, 2003). Though the greatest number of fish lie below the toxic level, it was surprising how many greatly exceeded that level, with the highest at 2.5 µg/g.

Scatterplot of Mercury Level Against Weight

The scatterplot displays, as we gathered from our previous data exploration, that there is a moderate positive linear correlation between mercury level and weight. As shown in the plot, when sex is taken into consideration, there appears to be a more acute positive linear correlation between mercury and weight in females than males (steeper slope).

Boxplot of Mercury Level and Species

A chart with different colored squares

Description automatically generatedThe boxplot displays the mercury level across the species we looked at. As with the above scatterplot, this result is expected from our data exploration, displaying our previous observations on the mean, median and standard deviations visually.

Initial Regression Model

A screenshot of a computer

Description automatically generatedWe did an initial regression of mercury level against weight, species and sex that was additive, first-order and had no interactions. The low p-values displayed in this output indicate all the predictor variables we chose should indeed be in the model. We did analyses on this initial model to see if it should be altered.

Initial Model Analyses

A diagram of a box plot initial model

Description automatically generated

We made residual plots and based on the initial model's QQ plot and box plot being strongly right skewed, and the box cox indicating a transformation in mercury level could improve our model, we performed the transformation of Mercury Level0.2 which gave much better results, as will be displayed. However, before transforming mercury level, we wanted to look at possible interactions between our predictor variables.

Finding Interactions

We figured there would be an interaction between weight and species, and possibly weight and sex. Looking at a model first with an interaction between weight and sex, and then a model with an interaction between weight and species, both were better than the original model that had no interaction.

Following this, we wanted to know if having both interactions would be beneficial. When we looked at the Anova comparing a model with interaction of weight and sex, then adding weight and species, there was a very small p-value indicating both was better than weight and sex alone.

However, when we switched it around and looked at whether the model would be improved by adding the interaction between weight and species, then adding weight and sex, we got a relatively high p-value of 0.1344, indicating the second interaction would not improve our model. This must mean that enough of the variation explained by the sex and weight interaction is already explained by the species and weight interaction. As this model is simpler (less predictors as there is only one interaction), we decided to go with it and include only the interaction of weight and species.

A screenshot of a computer

Description automatically generated

Transformation on Interaction Model

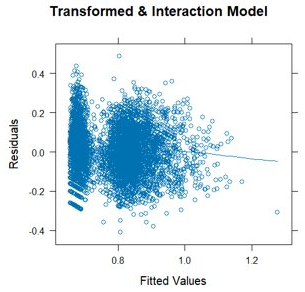
Now that we knew the interaction we wanted to include, we transformed the mercury level and re-did our residual plots to see if they were improved. The QQ plot indicates the residuals are much more normal than they had been, as does the box plot with a more even spread (though there are still outliers). The box cox indicates the transformation we performed to mercury was effective as the 1 is within the 95% confidence level.

A graph with a line and a rectangular object

Description automatically generated with medium confidence

When looking at the plots of the residuals, their variation has also improved with our new model, resulting in them being more evenly centered around 0 and a narrower spread (though there is still fanning in that, ironically, looks like a fish).

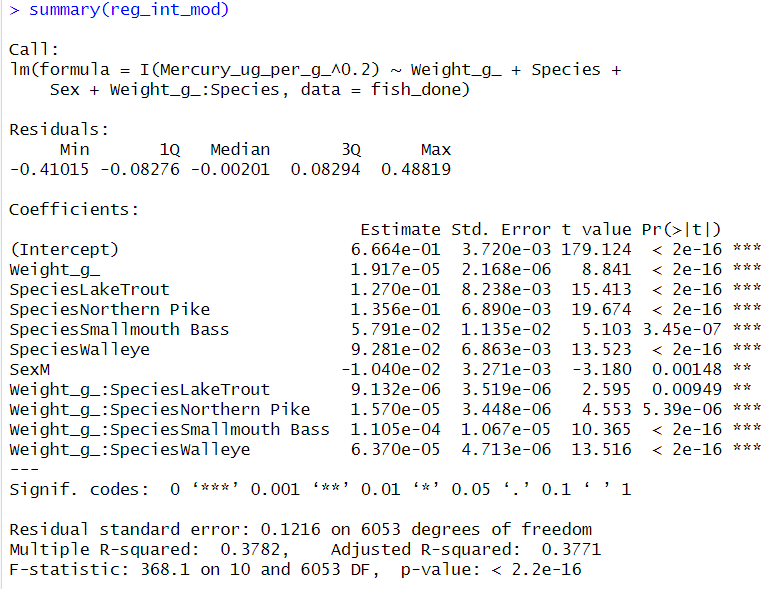
A diagram of a model

Description automatically generated  

To try to find why this shape was happening, we tested for multicollinearity using the VIF values, which indicated we did not have substantial multicollinearity as the values produced were just above 1. We also found there were outliers with large leverage, and tried removing fish with a Cook’s distance greater than 3x the mean to see if our model improved- it did not make much of a difference as the adjusted R2 went from 0.377 in the original model to 0.40 when the outliers were removed, and the residuals plot was still fish shaped.

Results

We did the regression of our final model, with Mercury Level0.2 against weight, species, sex and the interaction between weight and species.



The intercept coefficient will not tell us much in this case due to extrapolation as we would not have a fish with a weight of 0g. As displayed by the above output:

* Males tend to have 0.00104 µg/g Mercury Level0.2 less, on average, than females, keeping species and weight constant
* For Lake Trout, when weight is increased by 1g, we tend to see a 2.83x10-5 µg/g, on average, increase in Mercury Level0.2 (found by 1.917x10-5 + 9.132x10-6), keeping sex constant
* For Northern Pike, when weight is increased by 1g, we tend to see a 3.49x10-5 µg/g, on average, increase in Mercury Level0.2, keeping sex constant
* For Smallmouth Bass, when weight is increased by 1g, we tend to see a 1.30x10-4 µg/g, on average, increase in Mercury Level0.2, keeping sex constant
* For Walleye, when weight is increased by 1g, we tend to see an 8.29x10-5 µg/g, on average, increase in Mercury Level0.2, keeping sex constant

As the adjusted R2 of 0.377 on the regression displays, 37.7% of the variation in mercury level can be explained by this linear model.

We did an anova test on the final model (transformed & interaction term included) and got the following output:

A screenshot of a computer

Description automatically generated

Each of the p-values are exceedingly small, indicating that each term added reduces a substantial amount of variability in mercury level, given the terms already present.

Conclusions

Our hypothesis that as weight increases, mercury level tends to increase was correct! Our hypothesis that males tend to have lower mercury levels compared to females of the same weight and species, was also correct.

It was, however, surprising to see that smaller species tend to accumulate mercury at a higher rate than larger species. For example, when we looked at mean weight of fish for the specific species studied, from smallest to largest they are; Smallmouth Bass, Walleye, Northern Pike, Lake Trout. However, based on the above results for estimated mercury level change when 1g of weight is increased for the specific species, the greatest to least estimated rates are; Smallmouth Bass (1.30x10-4 µg/g), Walleye (8.29x10-5 µg/g), Northern Pike (3.49x10-5 µg/g), Lake Trout (2.83x10-5 µg/g).

A chart of different colored dots

Description automatically generated

Discussion

This data is specific to Ontario bodies of water, and thus the analyses are relevant mainly for populations of freshwater sport fish. However, different areas may be more likely to have marine mercury contamination than others depending on nearby anthropogenic sources. Though the dataset itself included location, we did not study that variable in this report, and it would be an interesting parameter to include if similar analyses were conducted. Regarding the dataset itself, we believe it would be beneficial for researchers to include approximate age or life stage of the fish, as size of the fish and mercury level is likely related to their age- a small species may be lighter but have a higher mercury level due to their age.

It must also be noted that, though researchers attempted to get a broad range of species, there is no guarantee that they were able to, which may indicate our samples do not fully represent the population of fish in the bodies of water.

References

Environment Canada. (2003). *Mercury: Fishing for Answers*. Water Policy and Coordination Directorate. https://publications.gc.ca/collections/Collection/En14-1-2002E.pdf

Government of Ontario. (2021). *Fish Contaminants* [Data set]. [https://files.ontario.ca/ moe\_mapping/downloads](https://files.ontario.ca/moe_mapping%20/downloads)/2Water/fish\_contaminant\_data/Fish%20Contaminant%20Data%202005-2018.xlsx