Date: Friday 6th September 2019

To: Dr. Nokuthaba Sibanda (STAT501 Course Coordinator), School of Mathematics and Statistics, Victoria University of Wellington.

Report on whether or not the amount of fish consumed has an effect on mercury levels in the hair of fishermen of Kuwait.

Minh Chau Van Nguyen

300490894

Victoria University of Wellington

Based on the following article:

N.B. Al-Majed and M.R. Preston. Factors influencing the total mercury and methyl mercury in the hair of the fishermen of Kuwait. Environmental Pollution. 109: 239-250.

**Abstract**

The purpose of this report was to analyze the mecury data and determine whether the amount of fish consumed has an effect on mercury levels in the hair of fishermen of Kuwait. The total sample size is one-hundred and thirty-five, with fishermen being the critical group (sample of size n1 = 100) and a control group of construction workers (sample of size n2 = 35). The individuals of these two groups were assumed to be independent and randomly picked. The dataset originally had a highly skewed and non-random distribution. As a result, log-transformation has been done to get a roughly log-normal distribution of Total Hg levels data which is shown to be less skewed after the procedure. The statistical method used was regression analysis done entirely in R Studio. After log-transforming the data, the results appeared to be more consistent from both the regression model and the ANOVA table in comparison to the results under the same analysis from when log-transformation has not been done. Total Hg concentrations (TotHg) in human hair are found to be uncorrelated to the amount of fish consumed (fishmlwk) for fishermen group. However, amount of fish is found to be statistically significant and positively correlated to Total Hg levels in control group. The residual plots showed why it might be that fishmlwk is not statistically significant with the regression line being a poor fit on the dataset. It was shown that the regression line has violated several regression assumptions. One of the reasons being the outliers (extreme values) that concerns the slope of the regression line.

*Keywords: Mercury; Total mercury; Fishermen; Hair samples; Fish consumption; Control group; Log-transformation; Log-Normal distribution; Regression; ANOVA; Outliers;*

**Introduction**

Mercury (Hg) for a long time has been infamous for its highly toxic compounds and a dangerous health hazard to humans, with main source of exposure being fish consumption. Fish is therefore the main subject of this study for which it has been proven in several past studies that mecury concentration levels showed a positive correlation with fish consumed.

Mercury comes in different forms, but the most common form that is most likely associated with mercury poisoning is methylmercury (organic mercury in water-soluble form). In fact, nearly all of the mercury in fish muscle occurs as methylmercury (M.M. Storelli et al., 2002). A study by Gearhart et al. (1995) showed that the values for TotHg in fish vary between 0.3 g/g fresh weight set by the US Environmental Production Agency (EPA) (with not more than 66% in the form of organic Hg), which means methylmercury level in fish was still above average and concerning.

ROMPE (1998) investigated whether different species of fish has different mercury levels in the Arabian (Persian) Gulf region, where Kuwait fishing village is located. As an example, total Hg values in fish of the region were estimated to range between 0.007 g g (wet wt.) for Mediterranean amberjack (*Seriola dumerlii*) and 0.540 g/g (wet wt.) in Karanteen sea beam (*Crenidens crenidens*). Al-Majed and Rajab (1998) also reported total Hg and MeHg levels in 105 samples of 23 different species. The levels of total Hg varied between 0.123 g g (dry wt.) in river shad (*Hilso ilisha*) of 450 g weight and 35 cm length to 4.500 g/g in Indian flathead (*Platycephalus indicus*) of 1200g weight and 47 cm length. The corresponding MeHg values were 0.015 and 3.862 g/g, respectively.

The relationship between fish consumption and mecury levels is believed to be dependent on other factors such as edible parts of fish, sizes of fish, etc., Augier et al. (1993) have outlined different concentrations of total Hg in different organs (liver, lung, kidney, muscle, heart and brain) of same species of fish and reported that the highest levels of Hg were found in the brain and other fatty tissues. Furthermore, P. Houserova et al. (2006) agreed that the total mercury concentration levels in fish is not related only to the mercury content in the sediments, but also to the diet composition of the fish, and to the other chemical and biological characteristics of the aquatic ecosystem.

This study was initialized to investigate the relationship between the amount of fish intake and mercury levels for the critical group, fishermen living in Doha Fishing Village, Kuwait. I The dataset of Hg levels taken from the hair of the fishermen of Kuwait derived from the article of interest will be used for statistical analysis. Among Kuwait residents, Bou-Olayan and Al-Yakoob (1994) reported mean Hg concentrations in hair of 4.054.40 g/g in 68 females and 5.55.33 g/g in 38 males. The present study is going to discuss about male fishermen only since Hg concentration in hair has appeared to be more significant than that of female (Bou-Olayan and Al-Yakoob,1994).

**Objectives and Data**

The mercury data were collected with the aim of investigating the factors that influence total mercury and methylmercury in the hair of the fishermen of Kuwait. It is expected that methylmercury level is the highest out of all the other mercury species with respect to total mercury. In this report I am mainly discussing whether there are some sort of relationship between the two variables: Total Hg concentration levels (TotHg) vs. Fish meals per week (fishmlwk), where fishmlwk represents the amount of fish consumed in order to answer the research question. I am going to split the dataset into two, one for the fishermen group and one for the control group to ensure independent results for both groups. In addition, I am talking briefly about whether there is a positive linear relationship between MeHg and TotHg, i.e. MeHg increases as TotHg increases; if there is a difference between different categories of fishmlwk (group 3, group 4, group 7, group 14 and group 21 ranked from lowest fishmlw to the highest in that order); and most importantly answering the research question by comparing the mercury levels between fishermen group and control group with respect to the amount of fish consumed and discuss other factors that might have affected the relationship using an appropriate statistical method.

**Materials and Methods**

*Section 1. Sampling*

*1.1. Sampling idea*

One hundred human hair samples were collected from fishermen (age range 25 – 60), living in Doha, Fishing village, Kuwait (target group). Thirty-five additional samples were taken from a control group working in a local construction company (age range 26 – 35). All participants were men.

*1.2. Data Collection Method*

The samples were taken from several sites of the scalp of each individual using clean stainless-steel scissors. The samples weigh 2-3 g each were then put in a polyethylene plastic sampling bag. Each bag was sealed, labelled separately and stored in a deep freezer until the time of analysis.

*1.3. Questionnaire*

A questionnaire was completed for each volunteer (N = 135 volunteers in total) in order to assess his dietary habits. Questions related to dietary habits included number of fish meals week (variable of interest), quantity of fish meal (not included in the dataset), source of fish (not included in the dataset), edible parts of fish, etc...

*1.4. Sample preparation and analysis*

Hair samples were cut into short segments and washed successively with acetone and water. Samples were separated by centrifugation and dried in a laminar flow hood (UNEP, 1987). *Quality Assurance.* The accuracy of total Hg analysis was checked by running four samples of Sample Reference Material (SRM) with each batch of samples (set of ten samples), and the accuracy of MeHg analysis was checked by running two SRMs with each batch of samples (set of eight samples).

*Section 2. Regression Analysis*

*2.1. The idea of Regression Analysis*

This report’s analysis is conducted through the use of Linear Regression. It is used to predict the value of an independent (response) variable *y* based on one or more independent (predictor) variables *x*. The general model is :

**yi = 0 + 1\*xi + i**

In this case, the dependent variable y is TotHg (total Hg) or MeHg(MeHg) and the independent variable x is fishmlwk (fish meals per week). There are two observations for group 3, twelve observations for group 4, seventy observations for group 7, five observations for group 14 and eleven observations for group 21. I have decided to make some changes to the fishmlwk variable. By treating fishmlwk as a categorical variable, group 3 and 4 were combined together, and group 14 and 21 were also combined together so there would be 3 groups in total, i.e. group 1 (lowest fish consumption) for fishmlwk = 3 & fishmlwk = 4, group 2 (medium fish consumption) for fishmlwk = 7, group 3 (highest fish consumption) for fishmlwk = 14 and fishmlwk = 21.

As a result, there are two independent variables, **x2** for group 2, **x3** for group 3, and group 1 is specified as the baseline group. The reason smaller groups were combined into one large group is because there are only 2 observations for fishmlwk = 3, and same reason for combining group 14 and group 21 together (small sample size could lead to invalid results).

The model is then:

**yi = 0 + 1\*x2i + 2\*x3i + i**

There are now fourteen observations for group 1, seventy observations for group 2 and sixteen observations for group. The mean of MeHg is 4.025 g/g and the mean of TotHg is 4.181 g/g for the fishermen group. We can see that the mean of MeHg represents approximately 96% the mean of TotHg.

Figure 1. Bar plots of mean of MeHg and TotHg

*A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated*

As mentioned previously, MeHg represents most of species in TotHg. It is seemed to be the case by looking at the bar plots showing the concentration level of MeHg is almost the same as TotHg. Therefore, I’d done my analysis only on one dependent variable and my pick was TotHg. It can be seen from the bar plot for control group that it also has three categories for fishmlwk, group 1 : fishmlwk = 0, group 2 : fishmlwk = 1, group 3 : fishmlwk = 2.

*2.2. Preliminary Analysis of the data*

Before analyzing the data, I tried to understand the variables better by using a scatter plot to visualize the relationship between the independent and the dependent variable.

Figure 2. Fishermen preliminary analysis scatter plot and boxplot for TotHg

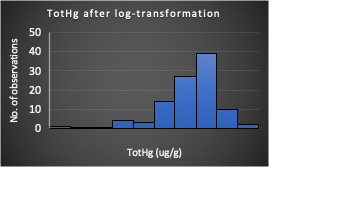
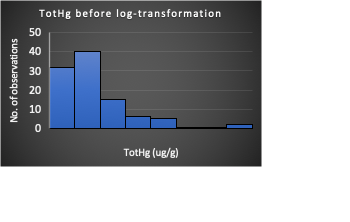
A close up of a white background

Description automatically generatedA screenshot of a cell phone

Description automatically generated

The scatter plot for TotHg isn’t suggesting linearity and showing a non-random pattern instead. This explains the low correlation between TotHg and fshmlpw (correlation = 0.192). Note : if the correlation is closer to 1 then a linear relationship is implied. The box plot is suggesting that there is no differences between groups of fishmlwk, at least not significant.

Figure 3. Histogram of before and after log-transforming TotHg



The histogram on the left shows that the distribution of TotHg is not normally distributed but is instead right-skewed. So, I’d decided to log-transform the data in order to make the current highly skewed distribution less skewed and more normally distributed. The TotHg data now has a log-normal distribution. This can be valuable for making patterns in the data more interpretable help meet the assumption of constant variance in the context of linear modeling. Overall this could potentially help make a non-linear relationship more linear.

**Results**

By comparing the results of before and after log-transforming the data analyses, I am going to decide whether or not TotHg has a relationship with fishmlwk.

*TotHg against fishmlwk before log-transforming the data*

The model : = 5.0911 – 1.4328\*x2 + 0.5795\*x3

Figure 4. Scatter plot of the fitted regression line

A screenshot of a cell phone

Description automatically generated

The intercept (0 **=** 5.0911) in this model is also the lowest fish consumption group, i.e. group 1. Though the p-value for the intercept is significant it just means that the baseline group is significant which it always is. Hence significant intercept does not imply that group 1 is significant. The p-values for the rest of the groups of fishmlwk are also not significant, with p-value = 0.123 for group 2 and p-value = 0.616 for group 3. However, the p-value from ANOVA table is significant for fishmlwk at 5% level (p-value = 0.0394), which is not consistent to the regression model’s result. This could be due to the within group differences, i.e. every fisherman in group 1 has significantly different TotHg levels, same for group 2 and group 3. In conclusion, the model overall is not clearly determining whether or not fishmlwk has a significant effect on TotHg. It’s also inappropriate to run test hypotheses on 1and 2 parameters since the residual plots showed that the assumptions for regression has not been met.

*TotHg against fishmlwk after log-transforming the data*

New model : **log() = 1.22160 – 0.13877\*x2 – 0.03757\*x3**

Figure 5. Scatter plot the fitted regression line after log transformation

A screenshot of a cell phone

Description automatically generated

The p-values for both group 2 (p-value = 0.616) and group 3 (p-value = 0.193) are not significant, which agrees with the p-value = 0.8415 of fishmlwk from the ANOVA table. The R-squared value is 0.003551. With such large p-value the regression line is not a good fit to the data since there is only ~0.35 % of the effect of fishmlwk has on TotHg. Overall, the regression model and the ANOVA table from the log-transformed data confirmed that there is no relationship between fshmlwk and TotHg for the fishermen group.

*Residual plots of the log-transformed data*

I had a careful look at the residual plots for the log-transformed model to see why it might be that fishmlwk and TotHg are not in a linear relationship. Looking at the Residuals vs. Fitted plot it’s obvious that there exists a non-linear relationship between TotHg and fishmlwk since the red line is not close to the dotted line. The Residuals vs. fishmlwk identifies the data as not random (a pattern that is not random suggest lack of independence) meaning the observations are not independent of each other. Though the Normal Q-Q shows the data to be roughly normally distributed (the slope curves downward but still tends to a normal distribution) and the Scale-Location plot assumes a constant uniform variance, the linear regression line is still not a good fit to the data for still violating several linear regression assumptions. Thus TotHg does not share a relationship with fishmlwk. Cook’s distance indicates that there are three observations (3, 85 and 96) that could potentially have large influence in the model. From residuals vs.Leverage plot these observations are identified as outliers with most of the dots are not inside the Cook’s distance lines. The outliers potentially have large effect on the slope of the regression line fitting the data (influential points) and hence affect the linear relationship between TotHg and fishmlwk. Since only one assumption out of four has met it is inappropriate to perform test hypotheses for 1and 2 parameters’ significance.

**Conclusion**

In contrast to the fishermen group, it seems that as the fishmlwk increases (i.e. from fishmlwk = 0 to fishmlwk = 2), the mean of Total Hg also increases.

Figure 6. Bar plot of mean of log(TotHg) vs. fishmlwk for control group

The p-value for fishmlwk is also significant, so there’s evidence that fishmlwk is significant (with respect to the residual plots assumptions). In conclusion, there is a positive linear relationship between TotHg and fishmlwk for the control group. The two groups were significantly different in their fish diet habits (i.e. frequency, quality and type of fish consumed), with fish form a significant part of the fishermen’s diet so it was expected that the amount of fish consumed for fishermen group would be more significantly related to mecury levels than the control group. However, my results has proved that there is no linear relationship between the amount of fish consumed and mecury levels at all. It may be added the health risk of eating fresh fish maybe less than that associated with eating canned tuna fish.

Since fish meals per week (fishmlwk) is not clearly defined (is 3 categorized for 3 fish meals per week or is it 3 fish per meal per week, and so on), it is not a good representative of the independent variable of interest which is the amount of fish consumed. I am assuming fishmlwk is categorized as number of meals that included fish per week. Even so, the fishermen could consume more than just one fish per meal. For example, for fishmlwk = 3 some fishermen could have different amounts of fish per meal for 3 meals, and other fishermen could have different and less fish per meal for 14 meals with fishmlwk = 14.

The amount of fish consumed is also dependent on edible parts of fish and type of fish as well. Hence even if some fishermen consume lots of fish per week, but the fish they are eating is the type with the least mecury concentration levels and the fish parts they eat also contains the least mecury concentration levels then their observed TotHg levels will be low.

Another reason why fishmlwk is not related to TotHg could be because there is no drastically significant differences between the fishmlwk groups as shown in the preliminary analysis. In another words, the TotHg concentration levels of the lowest consumption group should be smaller than the TotHg concentration levels of the highest consumption group but it is not the case when looking at each observation individually. For example, one fisherman in group 1, his TotHg level is 11.863 which is suspiciously high for a low consumption group. This is also one of the outliers (extreme values) identified in the residual plots (observation number 3).

The outliers are there for many possible reasons, data was entered incorrectly, participants giving wrong answers, variability in the population, etc. I didn’t remove these outliers to get a better fit regression line because I didn’t collect the data and wasn’t sure what caused the outliers. I tried to perform the regression analysis on the log-transformed data (TotHg) with more predictor variables (age, residence time, height, weight, part of fish) all included in my model but I still got insignificant p-value for fishmlwk.

It is with strong evidence that the amount of fish consumed does not have an effect on TotHg levels. Following that, it is fair to say that fishmlwk also has no effect on the levels of MeHg. In conclusion, the amount of fish does not have an effect on mecury levels in the hair of fishermen of Kuwait, based on the analysis of the provided dataset.

**Appendices**

**Residual plots before log-transforming the data**

Figure 7. Residuals vs. Fitted and Residuals vs. fishmlwk plots

**A close up of a map

Description automatically generated**

Figure 8. Normal Q-Q and Scale-Location plots

**A close up of a map

Description automatically generated**

Figure 9. Cook’s distance and Residuals vs. Leverage plots

**A screenshot of a cell phone

Description automatically generated**

**Residual plots after log-transforming the data**

Figure 10. Residuals vs. Fitted and Residuals vs. fishmlwk plots

A screenshot of a cell phone

Description automatically generated

Figure 11. Normal Q-Q and Scale-Location plots

A close up of a map

Description automatically generated

Figure 12. Cook’s distance and Residuals vs. Leverage plots

A close up of a map

Description automatically generated

**References**

Al-Majed, N., Rajab, W., 1998. Levels of mercury in the marine environment of ROPME Sea Area. offshore Environment of the ROPME Sea Area after the war related oil spill (Results of the 1993±94 Umitaka-Maru Cruises, 125±147). Terra Scientific Publishing Company, Tokyo.

Augier, H., Park, W.K., Ronneau, C., 1993. Mercury contamination of the striped dolphin Stenella-coeruleoalba meyen from the French Mediterranean coasts. Marine Pollution Bulletin 26, 306±311.

Bou-Olayan, A., Al-Yakoob, S., 1994. Mercury in human hair: a study of residents in Kuwait. Journal of Environmental Science and Health A29, 1541±1551.

Lipfert, F., Moskowitz, P., Fthenakis, V., Dephillips, M., Viren, J., Saro􏰐, L., 1995. Assessment of adult risk of paresthesia due to mercury from coal combustion. Water, Air, and Soil Pollution 80, 1139±1148.

M. M. Storelli, R. Giacominelli Stuffler, A. Storelli, and G. O. Marcotrigiano, 2002. Total Mercury and Methylmercury Content in Edible Fish from the Mediterranean Sea. Journal of Food Protection, Vol. 66, No. 2, 2003, Pages 300–303.

N.B. Al-Majed and M.R. Preston, 2000. Factors influencing the total mercury and methyl mercury in the hair of the fishermen of Kuwait. Environmental Pollution. 109: 239-250.

P. Houserova, V. Kuban, P. Spurny, P. Habarta, 2006. Determination of total mercury and mercury species in fish and aquatic ecosystems of Moravian rivers. Department of Chemistry and Biochemistry, 2Department of Fishery and Hydrobiology, Mendel University of Agriculture and Forestry, Brno, Czech Republic. Veterinarni Medicina, 51, 2006 (3): 101–110.

ROPME, 1988. Regional Organization for the Protection of the Marine Environment, Results of project number: KA/5102 82 10(2363) rev.4. Survey of mercury in fish and sediments from the ROPME Sea Area (unpublished report). Prepared by IAEA- MESL, Monaco.

UNEP, 1987. The determination of methyl mercury, total mercury and total selenium in human hair (Reference Methods for Mar- ine Pollution Studies No. 46). United Nations Environment Programme, Nairobi.