Date: Friday 7th September, 2019

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

Report on whether or not the amount of fish consumed is related to the 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. 2000. Factors influencing the total mercury and methyl mercury in the hair of the fishermen of Kuwait. Environmental Pollution. 109: 239-250.

This report is intended solely for the purpose of completing STAT501 Assignment 1. Do not use and extract the information from this report as it uses data collected in 1999 hence some of these information may be incorrect and irrelevant with today.

**Abstract**

“Methyl mercury concentrations in human hair showed a positive correlation with fish consumed”.

The purpose of this report is to analyze the data provided on three variables and determine whether or not the amount of fish consumed has an effect on mercury levels in the hair of fishermen in Kuwait.

*Keywords* :

**Background**

* Mercury (Hg) and its compounds are recognized as potentially hazardous material and are rated in the top category of environment pollutants.
* Mercury comes in different forms, but the most common form that is the most 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 (14,15, 23), and fish consumption is a major route for mercury uptake by humans (7, 9).
* A number of workers have investigated the importance of fish as a source of exposure to Hg in the Arabian (Persian) Gulf region.

**Introduction**

**Background**

* Mercury (Hg) and its compounds are recognized as potentially hazardous material and are rated in the top category of environment pollutants.
* Mercury comes in different forms, but the most common form that is the most 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 (14,15, 23), and fish consumption is a major route for mercury uptake by humans (7, 9).
* A number of workers have investigated the importance of fish as a source of exposure to Hg in the Arabian (Persian) Gulf region.

**Objectives and Data**

A “client” has collected data on mercury levels from two different groups: fishermen group (sample of size n1 = 100) and a control group of construction workers (samples of size n2 = 35). All of the participants were men. The 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 the potentially positive linear relationships between the three variables:

* MeHg concentration levels (mg g) vs. Fish meals per week (meals)
* Total Hg concentration levels (mg g) vs. Fish meals per week (meals)

in order to answer the research question.

In addition, I am talking briefly whether there is a positive linear relationship between MeHg and TotHg, i.e. MeHg increases as TotHg increases; and most importantly comparing the mercury levels between the two groups, fishermen and control with respect to their weekly intake of fish.

It is also worth mentioned that though the unit used in all the articles mentioned in this report is g g (microgram) but I am going to be using mg g (milligram) for the unit since the data provided is in mg g.

|  |  |
| --- | --- |
| **1 g g = 0.001 mg g** |  |

**Review of Literature**

Lipfert et al. (1995) determined that the toxicity of Hg raised the greatest concern from a human perspective as a result of several widely published mercury-related disasters.

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 is still above average and concerning.

Different species of fish has different mercury levels. In an extensive International Atomic Energy Agency exercise during 1984-85 (ROPME, 1988), the overall mean value for different species was found to be 0.158 0.145 g g (wet wt.).

Al-Majed and Rajab (1998) performed an exercise on 105 samples of 23 different species and reported Hg and MeHg levels. The overall mean concentrations for all the different species were 0.8930.799 g g and 0.711 0.655 g g for total and MeHg, respectively.

According to IPCS (1990) the reference mean for TotHg in hair is 2.0 g g.

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 5.525.33 g g in 38 males. The present study is going to discuss about male fishermen only since Hg concentration in hair appeared to be more significant than that of female.

The study derived from the article of interest assesses Hg levels in the hair of members of the fishing community in Kuwait who were predicted to represent a critical population for Hg exposure by the marine route. I am going to be using this study to elaborate more on the potential link between mercury levels and the weekly intake of fish from the target group, fishermen of Kuwait.

**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). Randomization of samples is assumed.

*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 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, source 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).

*1.5. Quality Control*

*Section 2. Regression Analysis*

*2.1. The idea of Regression Analysis (more specifically Simple Linear Regression Analysis)*

This report’s analysis is conducted through the use of Simple Linear Regression.

Simple Linear Regression is used to predict the value of an independent (response) variable *y* based on an independent (predictor) variable *x*. The aim is to establish a mathematical equation between the two variables. Note: for each observation, y value can only be predicted when x value is known.

Simple Linear Regression model/equation:

yi = 0 + 1\*xi + i

where

* i = 1st, 2nd, ..., nth observations
* 0, the intercept parameter, is the mean value for yi when xi takes the value 0
* 1, the slope parameter, is the expected change in yi when xi is increased by one unit.

In this case, the dependent variable y is TotHg (total Hg) or MeHg(MeHg) and the independent variable x is fshmlpw (fish meals per week). Note MeHg and TotHg are two separated dependent variables so there are going to be two separated regression models.

*2.2. Software*

My analysis was performed entirely in Microsoft Excel though some bits included are also done in R studio. R Studio is also used for double-checking to ensure consistent hence accurate results.

*2.3. Preliminary Analysis of the data*

But before jumping straight into analyzing the data, let’s try to understand the variables graphically by using a scatter plot to visualize the linear relationship between the independent and dependent variables.

**2.3.1. Fishermen group**

Figure 1. Fishermen preliminary analysis scatter plots for TotHg and MeHgA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

Interpretation: The scatter plots aren’t suggesting linearity. This explains the low correlation between TotHg and fshmlpw (correlation = 0.192), likewise for MeHg and fshmlpw (correlation = 0.186). Note : the closer the correlation is to 1 the more a linear relationship is implied.

The low correlation between the variables could be due to the outliers (data points that differ significantly from other observations) as seen in the scatterplots. The outliers are there for many possible reasons, data was entered incorrectly, participants giving wrong answers, variability in the population, etc.. Another reason could be the counts of the different categories of fshmlpw are drastically different. For example, there are only 5 and 11 observations for the highly fish-consumed groups but there are 70 observations for one of the low fish-consumed groups.

It might be a good idea to remove these outliers before we run the regression analysis, but expect the scatter plots to look almost exactly the same as we are only removing a few values while the rest of data still remain similar.

In order to compare the outputs before and after removing outliers with the aim is to find out whether they are the main reason for low correlation between TotHg/MeHg and fshmlpw, I performed regression analysis on both of the data sets.

**2.3.2 Control group**

I only did preliminary analysis on the fishermen group but not on the control group since fishermen group is the target group, whereas control group is included in the report solely for the comparison between fishermen (who are likely to consume more fish) and construction workers to back up the research question. Moreover, the control group doesn’t appear to have any significant outliers that should concern the analysis overall so preliminary analysis is not necessary.

**Results**

*Section 1. Regression Analysis for fishermen group in Excel*

I am going to be comparing the before-removing-the-outliers result and after-removing-the-outliers for the fishermen group only since the control group appears to have no outliers as mentioned previously.

For both cases (before and after removing outliers) I am going to add up all of the values of MeHg and TotHg for different categories of fshmlpw variable, and then divide the total by their observation counts to find the mean of MeHg and TotHg.

Table 4. Identify the outliers



*1.1. Before removing outliers, n1 = 100*

Table 5. Mean of MeHg and TotHg before removing outliers



Figure 2. Histogram of Mean MeHg and Mean Total Hg

Interpretation: Both the mean values of MeHg and TotHg for fshmlpw = 3 is higher than the mean values for fshmlpw = 21. This could be because one of the outliers lying in this category, plus the number of observations for this group is only 2 which contributes to a greater mean. In addition, both the mean values of MeHg and TotHg for fshmlpw = 4 is higher than mean values for fshmlpw = 7. It is probably of the same reason, with 2 out of the outliers are present in this group (fshmlpw = 4).

Next perform regression analysis in Excel on all 100 observations for TotHg.

Table 6. ANOVA table for TotHg



Interpretation: The alpha level specified is 0.05 for this analysis. Since the p-value for fshmlw variable is 0.055 > 0.05, it is not significant. This implies that fshmlpw is not significant and there is no connection between TotHg and fshmlpw. The significance F shown here gives the probability that the model is wrong so we want significance F to be as small as possible. Here significance F is quite large so this is not a good model.

*1.2. After removing outliers*

Remove 4 outliers specified earlier and perform the same steps to get the mean values for MeHg and TotHg.

Table 6. Mean of MeHg and TotHg



Figure 3 . MeHg and TotHg means vs. Fshmlpw bar graphs

Interpretation: Now that the outliers have been removed, the bar graphs looks better in a sense that both MeHg and TotHg increases as the number of fish meals per week increases.

Regression Analysis on 96 observations for fishermen group after removing the outliers

Table 7. Anova Table for TotHg after removing outliers





Interpretation: The p-value for fshmlpw is significant hence there is an association between TotHg and fsmlpw. Significance F is extremely small (0.0003) which means this is a good model. In conclusion the Anova table agrees with the bar graphs that as fshmlpw increases, the TotHg concentration levels also increases.

Table 8. Anova Table for MeHg after removing outliers





Interpretation: Same as TotHg.

*1.3. Comparison of TotHg and MeHg after removing outliers for fishermen group*

Figure 4.

A screenshot of a cell phone

Description automatically generated

Interpretation: It is obvious that MeHg takes up the most in TotHg, with MeHg increases as TotHg increases.

*Section 2. Regression Analysis on Control group*

*2.1. Analysis of TotHg*

Figure 5. Mean TotHg vs. fshmlpw





Interpretation: It appears that as the fshmlpw increases, i.e. from 0 fish meal to 2 fish meals per week, the mean of Total Hg also increases. The p-value for fshmlpw is also significant, meaning fshmlpw is significant. In conclusion, there is a positive linear relationship between TotHg and MeHg. Significance F is also very small (1.071e-16), implies that this is a very good model.

*2.2. Analysis of MeHg*





Interpretation: It appears that as the fshmlpw increases, i.e. from 0 fish meal to 2 fish meals per week, the mean of Total Hg also increases. The p-value for fshmlpw is also significant, meaning fshmlpw is significant. In conclusion, there is a positive linear relationship between TotHg and MeHg. Significance F is also very small (1.071e-17), implies that this is a very good model.

**Discussion**

*Comparison between fishermen group and control group*

TotHg and MeHg concentration levels are 1.069 3.132 and 1.032 3.042 g g (after fitting the regression line) in that order for fishermen group. TotHg and MeHg concentration levels are 0.609 1.176 and 0.568 1.124 g g respectively.

**Conclusion and Limitations**

**Appendices**

Table 1. Fishermen data

|  |  |  |
| --- | --- | --- |
| fshmlpw | MeHg | TotHg |
| 3 |  |  |
| 4 |  |  |
| 7 |  |  |
| 14 |  |  |
| 21 |  |  |

Table 2. Fishermen data after removing outliers

|  |  |  |
| --- | --- | --- |
| fshmlpw | MeHg | TotHg |
| 3 |  |  |
| 4 |  |  |
| 7 |  |  |
| 14 |  |  |
| 21 |  |  |

Table 3. Control group data

|  |  |  |
| --- | --- | --- |
|  | MeHg | TotHg |
| 0 |  |  |
| 1 |  |  |
| 2 |  |  |

REGRESSION FOR FISHERMEN GROUP (AFTER REMOVING THE OULIERS) IN R

a. TotHg

Call:

lm(formula = fish2$TotHg ~ fish2$fishmlwk, data = fish2)

Residuals:

Min 1Q Median 3Q Max

-6.2417 -1.3436 -0.1867 1.0874 11.5213

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.10074 0.51924 4.046 0.000107 \*\*\*

fish2$fishmlwk 0.19838 0.05253 3.777 0.000278 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.495 on 94 degrees of freedom

Multiple R-squared: 0.1317, Adjusted R-squared: 0.1225

F-statistic: 14.26 on 1 and 94 DF, p-value: 0.0002784

b. MeHg

Call:

lm(formula = fish2$MeHg ~ fish2$fishmlwk, data = fish2)

Residuals:

Min 1Q Median 3Q Max

-5.9816 -1.3456 -0.1835 0.9915 11.2134

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.03743 0.50621 4.025 0.000115 \*\*\*

fish2$fishmlwk 0.18872 0.05121 3.685 0.000382 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.433 on 94 degrees of freedom

Multiple R-squared: 0.1262, Adjusted R-squared: 0.117

F-statistic: 13.58 on 1 and 94 DF, p-value: 0.0003817

REGRESSION FOR CONTROL GROUP IN R

a. Total Hg

Call:

lm(formula = control$TotHg ~ control$fishmlwk, data = control)

Residuals:

Min 1Q Median 3Q Max

-1.26692 -0.23034 -0.02892 0.24081 0.96908

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.8926 0.1393 6.406 2.93e-07 \*\*\*

control$fishmlwk 1.6762 0.1082 15.485 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4957 on 33 degrees of freedom

Multiple R-squared: 0.879, Adjusted R-squared: 0.8754

F-statistic: 239.8 on 1 and 33 DF, p-value: < 2.2e-16

b. MeHg

Call:

lm(formula = control$MeHg ~ control$fishmlwk, data = control)

Residuals:

Min 1Q Median 3Q Max

-1.21459 -0.24273 -0.04187 0.26000 0.93441

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.8461 0.1365 6.197 5.41e-07 \*\*\*

control$fishmlwk 1.6627 0.1061 15.676 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4857 on 33 degrees of freedom

Multiple R-squared: 0.8816, Adjusted R-squared: 0.878

F-statistic: 245.7 on 1 and 33 DF, p-value: < 2.2e-1

**References**

N.B. Al-Majed and M.R. Preston (2000)

Lipfert et al. (1995)

ROPME (1988)

Al-Majed and Rajab (1998)

IPCS (1990)