**Project 8: PCBs (Polychlorinated Biphenyl Congeners) in fish**

**Author: Zhen Jiang**

**Introduction and background**

Polychlorinated Biphenyl Congeners (PCBs) are compounds that has different numbers of chlorine [1]. There are 209 distinct PCB compounds (known as congeners), which generally occur as mixtures of congeners. PCBs are man-made organic chemicals that were widely manufactured and used in a variety of industries. They are very stable compounds that affect human and environmental health. The harmful effects of PCBs include affecting nervous system, immune system, endocrine systems, as well as reproductive systems. PCBs are also defined as probable human cancer-causing chemicals [2]. For these reasons, PCBs have been prohibited from use in the United States since 1979.

However, there is still much PCBs left in the environment. Since they are stable, they still have effects on human and environmental health. According to the published data, the primary exposure to PCBs is through fish consumption [3]. PCBs could enter the food chain when there was leaking or releasing of PCBs from electrical transformers, or illegal / improper dumping of PCBs wastes that would enter soil and rivers. These PCBs will be absorbed by plants or other creatures which will then be eaten by fish. Since PCBs are very stable and can pass along the food chain, larger predatory fish that are at the top of the food chain should have higher levels of PCBs. PCBs are also stored in fatty tissue, thusso it is expected that fish with higher content of fat tissue should have higher PCBs level.

The object of interest for this study is PCBs’ concentration. However, PCBs’ concentration based on lipid (fatty) tissue (Lipid PCB) will provide different results than those based on fish body weight (Total PCB). Therefore, in this study, researchers calculated PCBs’ concentration per unit of lipid (Lipid PCB) as well as per unit of fish body weight (Total PCB) so as to use both Lipid PCB and Total PCB as objects of interest.

In purpose of evaluating human health risk, the researchers quantified PCBs’ presence in fish which have been collected from different rivers in North Carolina. The goals of this project include the following aims. First, evaluate whether fish species and waterbody affect PCBs’ presence in fish or not. Second, create a model (equation) to predict PCBs’ concentration by using other environmental factors, such as species, waterbody, trophic level, fish weight, lipid content, etc.. Third, three quantification methods for quantifying PCBs will be compared to determine if there is any difference between them.

The significance of this study is that by creating a model to predict fish PCBs’ toxicity, there will be a more cost effective way to accurately estimate PCBs for evaluating human health risk in regarding to PCBs.

**Data collection**

**1. Sample collection**

This project seeks to evaluate the PCBs’ contamination in the Yadkin-Pee Dee River Basin (YPD). Researchers from North Carolina Department of Environmental Quality (NCDEQ) and the United States Environmental Protection Agency (EPA) selected four lakes in YPD for them to collect fish samples for this study. The four lakes are: Badin Lake, High Rock Lake, Lake Tillery, and Falls Reservoir. In 2008, NCDEQ collected 70 fish using electro-shocking techniques from Badin Lake. These fish were used to prepare 30 tissue samples. In 2011, NCDEQ and EPA collected 119 fish from High Rock Lake to prepare 42 samples. They also collected 110 fish from Lake Tillery to prepare 42 samples. 47 fish from Falls Reservoir were collected to prepare 15 samples in 2012.

For each fish, researchers recorded the fish’s species, waterbody, position in lake, trophic status (level within the food chain a fish species takes: top predator, middle predator, bottom feeder), TL (trophic level, numerical representation of trophic status), guild (type of diet the fish species relies on: omnivore, insectivore, piscivore), body length, body weight, lipid content (measures the ratio of lipid tissue weight to body weight). The information for all of the 131 fish were used to create the original dataset.

**2. PCBs measurement and data management**

For measuring PCBs’ content (209 PCB congeners), samples from 2008 were analyzed by SGS Laboratories in Wilmington NC and samples from 2011 and 2012 were conducted by U.S. EPA Region 4 Science and Ecosystem Support Division (SESD) Laboratory in Athens, GA. PCBs’ content measurement results were then being sent back to the researchers.

According to the measurement results for all of the 131 samples, researchers did quantifications to calculate the total amount of 209 PCB congeners in each sample by applying the EPA method 1668A, which is the gold standard for quantifying PCB.

The objects of interest for this study are both Lipid PCB and Total PCB because PCB concentration based on lipid (fatty) tissue will provide different results than those based on total fish body weight. Therefore, researchers then calculated Lipid PCB and Total PCB by using the total amount of 209 PCB congeners that have been calculated above. The Total PCB was calculated as the ratio of the total amount of PCBs to the total fish body weight for each sample. Lipid PCB was calculated as the ratio of the total amount of PCBs to the lipid fish tissue weight for each sample. The Total PCB and Lipid PCB data were then input into the original dataset.

The preceding calculation of Total PCB and Lipid PCB were based on the quantification results of the total amount of 209 PCB congeners using the standard quantification method (EPA method 1668A). In order to compare this standard method with alternative quantification methods, researchers also applied two alternative quantification methods as well as the standard method to quantify the total amount of a subset of only 33 PCB congeners for each sample. These values were also input into the original dataset. The labels for these three methods are: HRMS (Standard Measurement of a Subset of 33 Congeners), LRMS (Alternate Measurement 1 of a Subset of 33 Congeners), and LRMSMS (Alternate Measurement 2 of a Subset of 33 Congeners). The quantified values of PCBs from the three methods will be compared to determine if there are differences between the two alternative methods and the standard method.

**Statistical analysis**

The purpose of this project is to evaluate PCBs’ contamination in the Yadkin-Pee Dee River Basin by analyzing the fish PCBs’ content. The objects of interest are the Total PCB and Lipid PCB.

**1. Evaluate if fish species and waterbody affect PCBs’ concentration in fish.**

Linear regression models followed by several ANOVA analysises were chosen to answer this question. Since both species and waterbody are categorical variables, each level of them will be treated as dummy variable. Here, it may not be appropriate to use ANOVA to test if the interaction term is needed or not because not all of the species are present in all four of the waterbodies, which means that some interactions are not existing. Thus, linear regression model was applied here based on these two factors, and the interaction term will be kept. The full model, which is Model 1 (table 1), includes species, waterbody and the interaction term between them as explanatory variables. Model 2 (table 1) does not include waterbody as explanatory variable, and Model 3 (table 1) does not include species as explanatory variable. The response variable for these models is Total PCB or Lipid PCB because both of them are objects of interest. These three models were first fitted by using Total PCB as response variable, and then these three models were fitted again by using Lipid PCB as response variable. To satisfy the assumptions for applying linear regression, residual plot and qq-plot were generated to access equal variance and normality assumptions. Then transformation of the response variables was made when necessary.

**Table 1: Detailed model description for evaluating**

**if fish species and waterbody affect PCBs concentration**

|  |  |  |
| --- | --- | --- |
|  | **Explanatory variable** | **Response variable** |
| **Model 1 (full model)** | species, waterbody, interaction term | total PCB or lipid PCB |
| **Model 2** | species | total PCB or lipid PCB |
| **Model 3** | waterbody | total PCB or lipid PCB |

To determine if fish species and waterbody affect PCBs’ concentration, two hypothesis tests (HT) (table 2) were set up for both response variables (Total PCB, Lipid PCB). For the first HT, the null hypothesis is that fish species does not affect PCBs’ concentration. To test this hypothesis, ANOVA was applied to compare Model 1 (table 1) with Model 3 (table 1). If p-value < 0.05, then the conclusion is that fish species does affect PCBs’ concentration. Similarly, for the second HT, the null hypothesis is that waterbody does not affect PCBs’ concentration. To test this hypothesis, again, ANOVA was used to compare Model 1 (table 1) with Model 2 (table 1). If p-value < 0.05, then the conclusion is that waterbody does affect PCBs’ concentration.

**Table 2: Hypothesis test for determining**

**if fish species and waterbody affect PCBs concentration.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **H0** | **HA** | **Method** |
| **HT #1** | Fish species does not affect PCBs concentration. | Fish species does affect PCBs concentration. | Use ANOVA to compare model 1 with model 3. |
| **HT #2** | Waterbody does not affect PCBs concentration. | Waterbody does affect PCBs concentration. | Use ANOVA to compare model 1 with model 2. |

**2. Create a model (equation) for predicting PCBs’ concentration.**

Linear regression model was employed here to create the equation for predicting PCBs’ concentration because this is a general method for predicting response by using multiple explanatory variables. Factors that have been considered in building this model includes species, waterbody, trophic level, trophic guild, fish length, fish weight, and lipid content. Table 3 has summary information of these variables. Species, trophic level, and trophic guild are kind of related, if we include all of these three variables in the regression model, it will be redundant. Therefore, in order to determine if larger predatory fish that are at the top of the food chain will have higher PCBs level, the trophic level among these three variables was kept in the regression model.

**Table 3: Summary of Variables**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Name** | **Variable Description** | **Variable Type** | **Levels or Range** |
| Waterbody | Body of water fish sampled | Categorical | 4 |
| Species | Fish species of sample | Categorical | 15 |
| Trophic level | Species level in food chain | Categorical | 3 |
| Trophic guild | Type of fish diet | Categorical | 3 |
| Fish length (mm) | Length of fish | Continuous | [142.0, 867.5] |
| Fish weight (g) | Weight of fish | Continuous | [52.0, 8669.0] |
| Lipid content | Percent of lipid tissue | Continuous | [0.1, 7.9] |
| Total PCB | Total amount of 209 PCB congeners quantified by standard method / fish body weight | Continuous | [106, 127,150] |
| Lipid PCB | Total amount of 209 PCB congeners quantified by standard method / lipid tissue weight | Continuous | [0, 43.8] |
| HRMS | Total amount of 33 PCB congeners quantified by standard method | Continuous | [0, 65,298] |
| LRMS | Total amount of 33 PCB congeners quantified by alternate method 1 | Continuous | [0, 79,620] |
| LRMSMS | Total amount of 33 PCB congeners quantified by alternate method 2 | Continuous | [0, 65,298] |

Again, the response variables are Total PCB and Lipid PCB. So, two linear regression models were fitted for predicting Total PCB and Lipid PCB, separately.

To satisfy the assumptions for applying linear regression, residual plot and qq-plot were generated to access equal variance and normality assumptions. Then transformation of the response variables was made when necessary.

**3. Determine if the three different methods for quantifying PCBs’ concentration provide different results.**

In the previous analysis, we’ve used Total PCB and Lipid PCB values. These values were calculated by using the total amount of 209 PCB congeners, which were quantified by using the standard method (labeled as HRMS). The researchers also wanted to compare the standard method with the other two alternative quantification methods (labeled as LRMS and LRMSMS). Therefore, they applied each of the three different methods to quantify the total amount of a subset of 33 PCB congeners.

To compare these two alternative methods with the standard method (HRMS vs LRMS, and HRMS vs LRMSMS), we applied HT for testing paired difference since each sample has several PCB amount values that come from using different methods (table 4). The first HT is to test if HRMS and LRMS methods have difference in quantifying PCBs. The null hypothesis is that there is no difference between HRMS and LRMS. This means that the mean difference is zero. To test this hypothesis, for each sample, the difference of PCB values that been quantified by using the two methods were calculated. After acquiring mean and standard deviation of all differences, t-value was calculated using this formula: t = mean difference / (sd. of difference / sqrt(sample size)). By knowing the df=130, t-values can be evaluated using the table “percentage points of Students’ t distribution” in order to get the p-values accordingly. If p-value < 0.05, we then conclude that there is difference between the two methods that have been compared. Otherwise, there is no difference. The second HT is to test if HRMS and LRMSMS methods have difference in quantifying PCB. The null hypothesis is that there is no difference between HRMS and LRMSMS. The method for testing this hypothesis is the same as the above.

**Table 4: Hypothesis test for comparing different quantification methods.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **H0** | **HA** | **Method** |
| **HT #1** | No difference between  HRMS and LRMS | There is difference between the two methods | Test paired difference |
| **HT #2** | No difference between  HRMS and LRMSMS | There is difference between the two methods | Test paired difference |

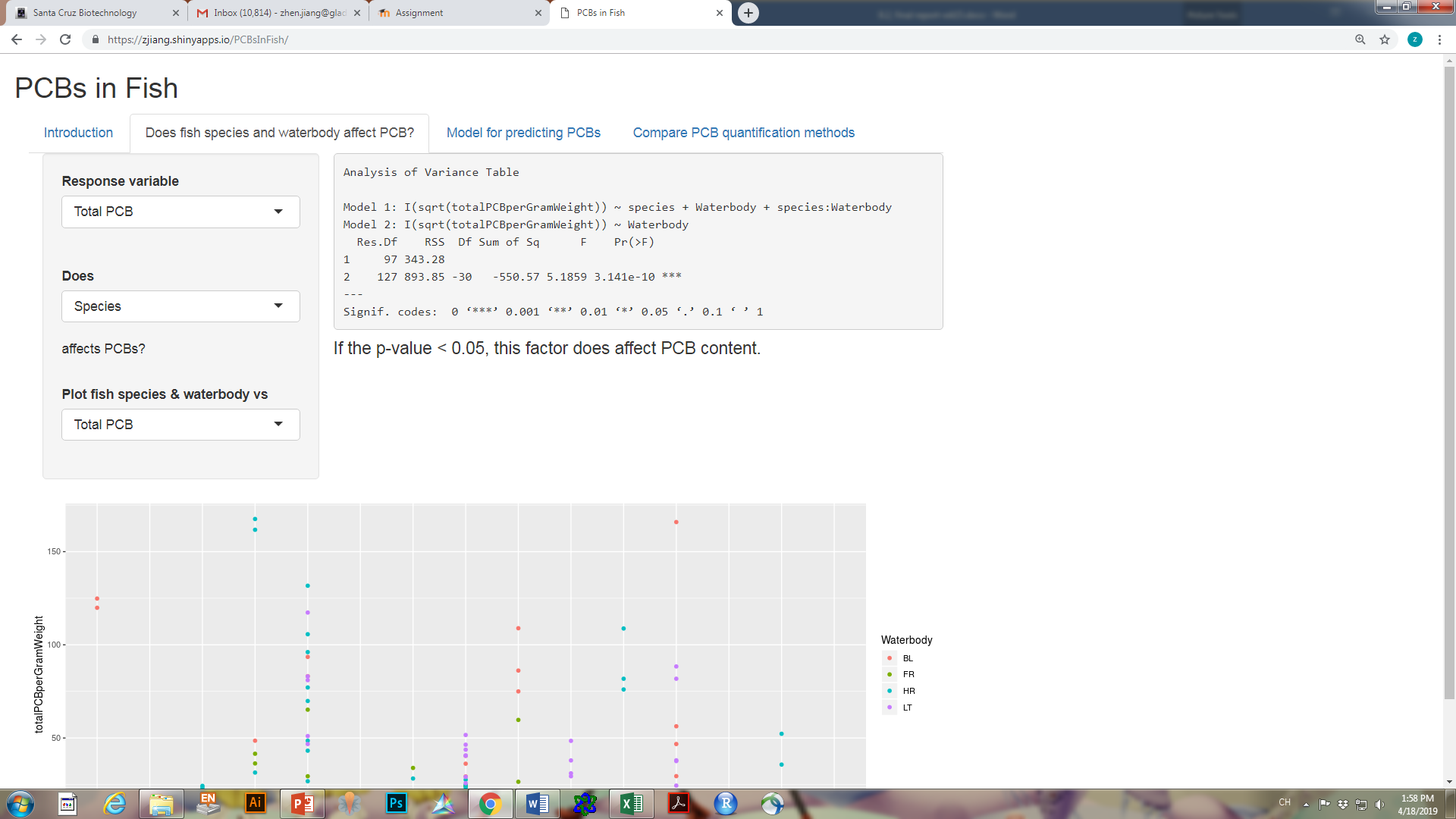
**Results**

**1. Relationship between PCBs’ concentration in fish and fish species / waterbody.**

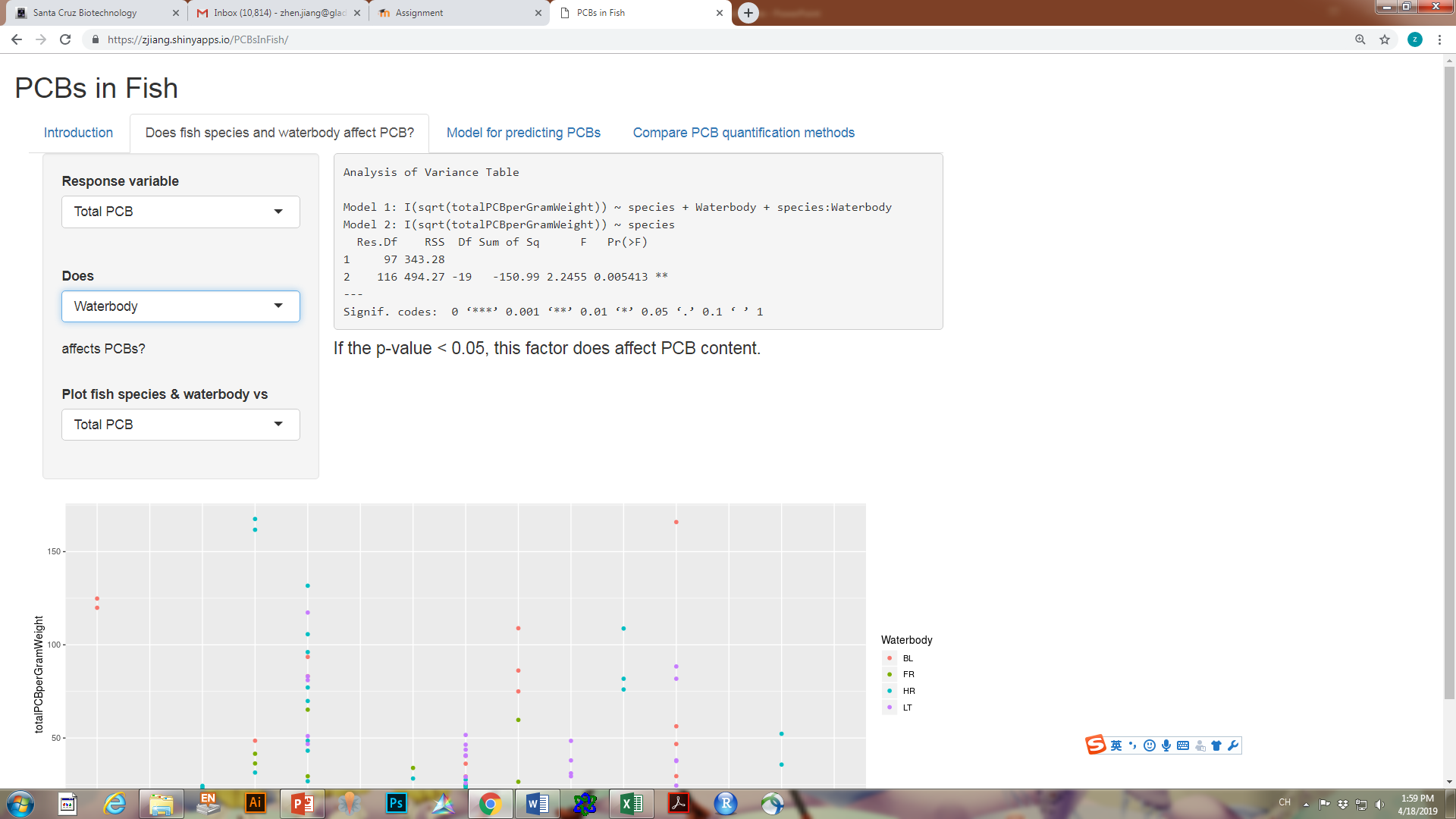
To apply the linear regression model, equal variance and normality assumptions were assessed, the response variable in the model was then transformed into square root of the Total PCB or Lipid PCB content accordingly.

For Total PCB, the hypothesis that fish species does not affect Total PCBs’ concentration (Table 2) was tested by comparing Model 1 (table 1) with Model 3 (table 1) using ANOVA. Figure 1 is the ANOVA output for this hypothesis test. This analysis showed that the p-value is 3.141 x 10^(-10), which is much smaller than 0.05. This suggested that fish species does affect Total PCB’s concentration. Similarly, by comparing Model 1 (table 1) with Model 2 (table 1) using ANOVA, the hypothesis that waterbody does not affect Total PCBs’ concentration(Table 2) was tested. Figure 2 is the output of this analysis. The p-value is 0.0054, which is also really small compare to 0.05. Therefore, waterbody also affects Total PCB’s concentration.

**Figure 1: Fish species does affect Total PCB**

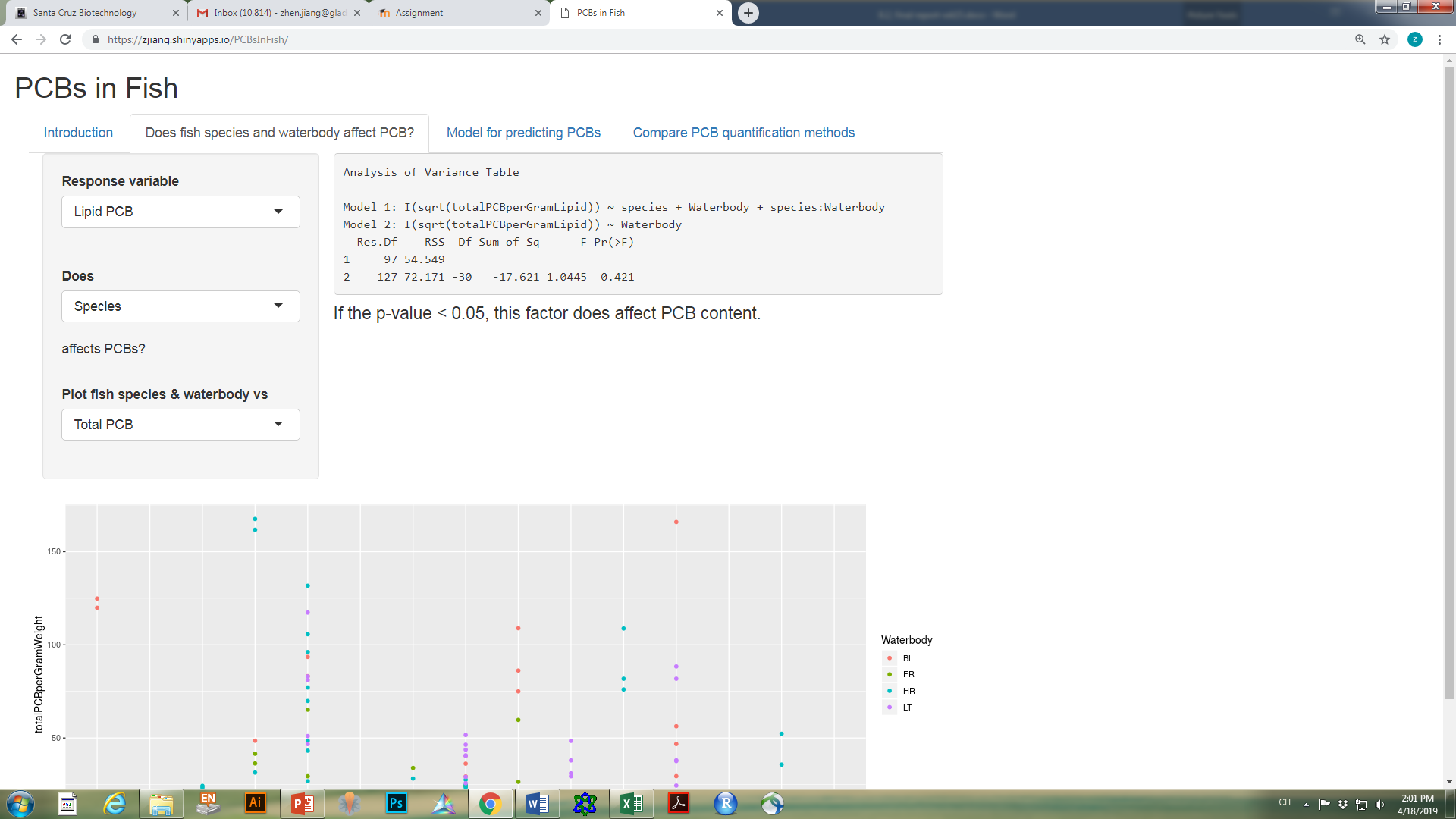


**Figure 2: Waterbody does affect Total PCB**

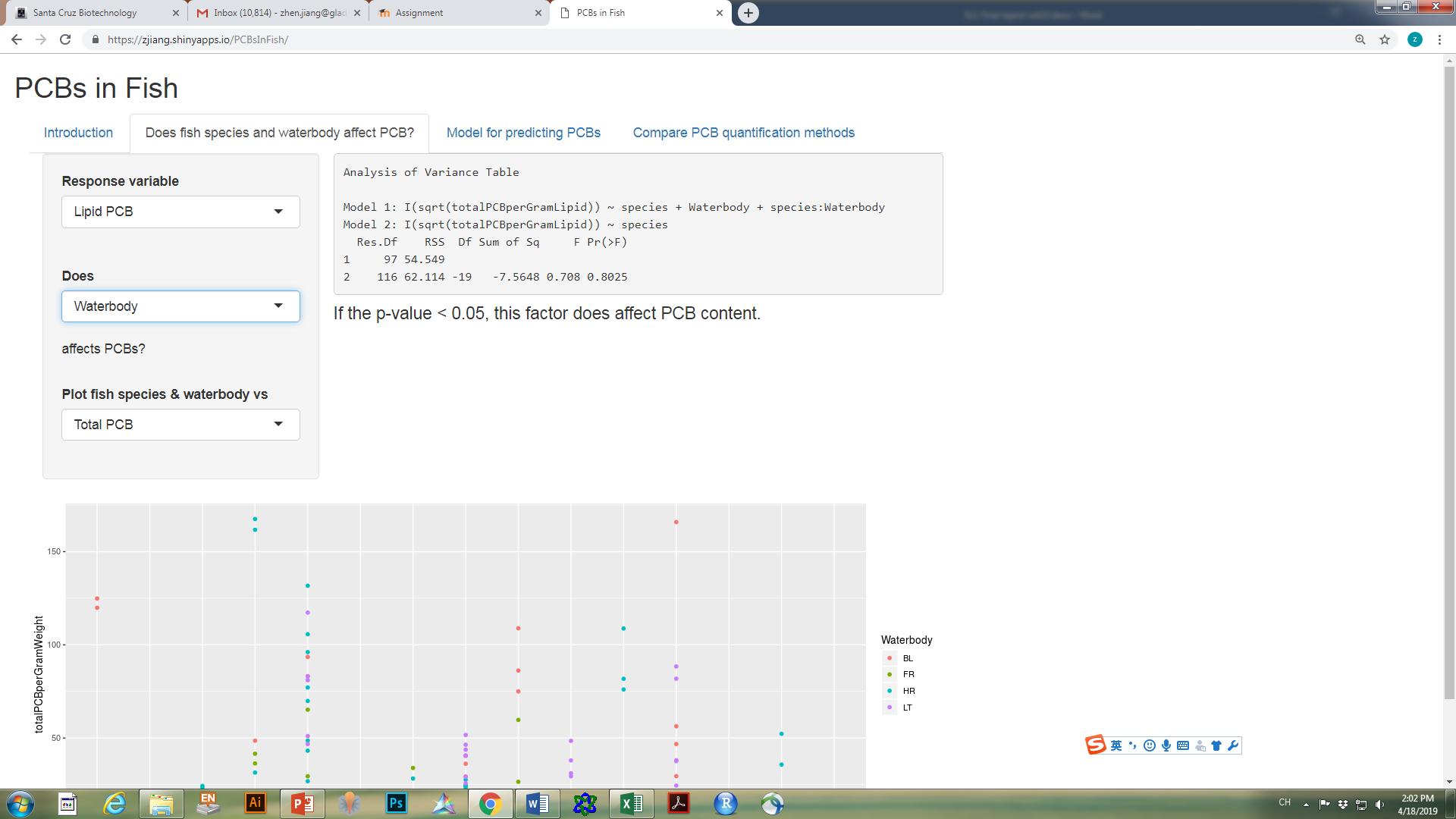


For Lipid PCB, after comparing Model 1 (table 1) with Model 3 (table 1) using ANOVA, the hypothesis that fish species does not affect Lipid PCBs’ concentration (Table 2) was tested. Figure 3 is the output of this analysis. P-value for the hypothesis test is 0.42, which is much bigger than 0.05. This indicates that fish species does not affect Lipid PCB’s concentration. The hypothesis that waterbody does not affect Lipid PCBs’ concentration (Table 2) was then tested by comparing Model 1 (table 1) with Model 2 (table 1) using ANOVA. Figure 4 is the output for this hypothesis test. P-value for this analysis is 0.8, which is also bigger than 0.05. Therefore, waterbody does not affect Lipid PCB’s concentration.

**Figure 3: Fish species does not affect Lipid PCB**



**Figure 4: Waterbody does not affect Lipid PCB**



**2. Model for predicting PCBs’ concentration.**

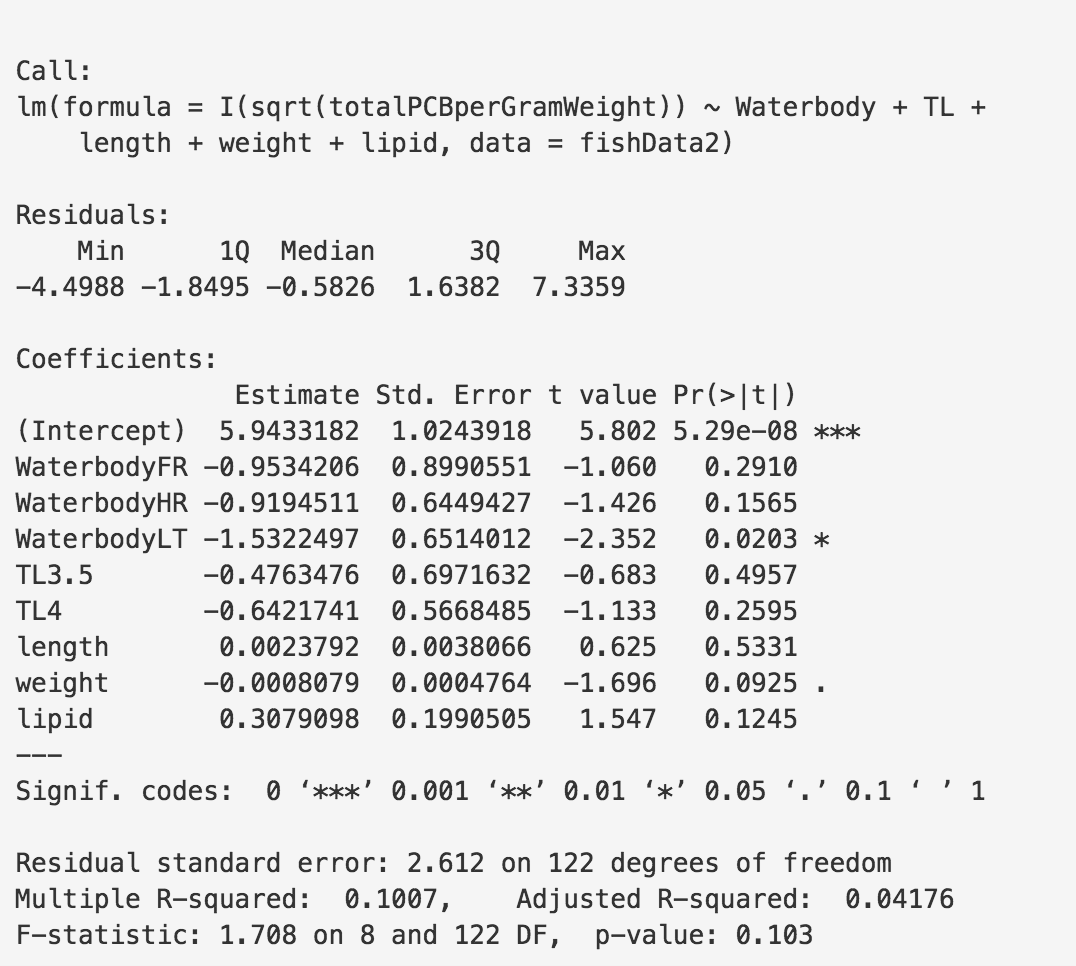
 Linear regression model was applied to build the model / equation for predicting PCBs’. Again, response variable in the model was transformed into square root of the Total PCB or Lipid PCB content to satisfy the assumptions for applying linear regression model.

To predict Total PCB’s concentration, linear regression model was fitted. Figure 5 is the output of the summary of the regression model. Since we have both categorical and numerical variables in this model, interaction terms were not included to avoid making the model too complicated.

According to this regression output, an equation for predicting Total PCB’s concentration was generated:

Total PCB’s concentration = 5.94 - 0.95 x WaterbodyFR – 0.92 x WaterbodyHR -1.53 x WaterbodyLT – 0.48 x TL3.5 – 0.64 x TL4 + 0.0024 x length – 0.00081 x weight + 0.31 x lipid

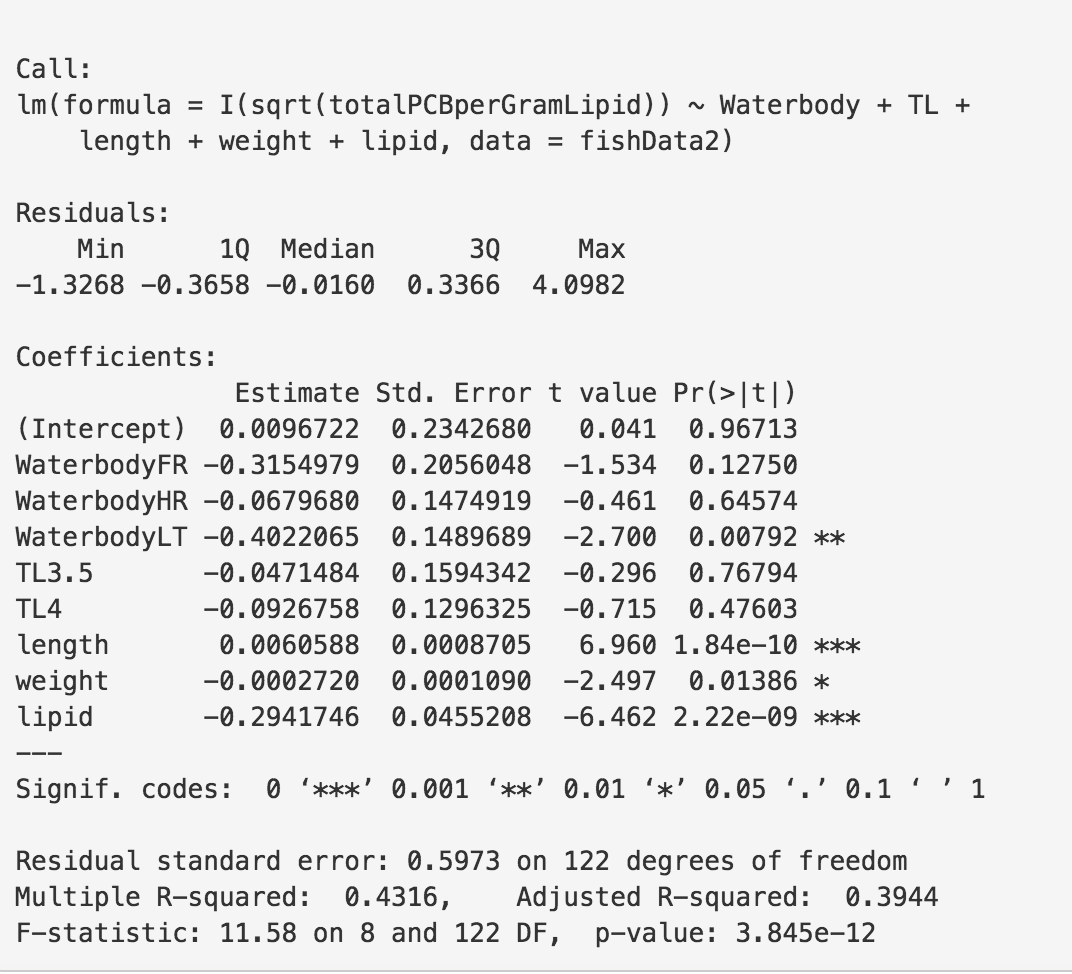
**Figure 5: Regression analysis for predicting Total PCB’s concentration.**



For predicting Lipid PCB’s concentration, another linear regression model was also fitted. According to regression summary output (Figure 6), an equation for predicting Lipid PCB’s concentration was generated:

Lipid PCB’s concentration = 0.0097 - 0.32 x WaterbodyFR – 0.068 x WaterbodyHR -0.402 x WaterbodyLT – 0.047 x TL3.5 – 0.093 x TL4 + 0.0061 x length – 0.00027 x weight - 0.29 x lipid

**Figure 6: Regression analysis for predicting Lipid PCB’s concentration.**



**3. Compare the standard quantification method for PCB with two alternative methods.**

Standard PCB quantification method (HRMS) was applied for previous analysis for quantifying 209 PCB congeners. To compare HRMS with the other two alternative quantification methods (LRMS and LRMSMS), each of the three methods were used to quantify the total amount of a subset of 33 PCB congeners. These quantifications were used for comparison. Here, only the total PCB of a subset of 33 PCB congeners is the interest of study.

To compare HRMS with LRMS, hypothesis test for testing paired difference showed that t-value = 5.36, which means that p-value is less than 0.001. This is much smaller than 0.05. Therefore, there is difference between methods HRMS and LRMS.

To compare HRMS with LRMSMS, hypothesis test for testing paired difference showed that t-value = 1.46. The according p-value is between 0.1 and 0.2, which is bigger than 0.05. Thus, there is no obvious difference between methods HRMS and LRMSMS.

**Discussion**

PCBs can build up in animals over time and along the food chain. PCBs are found in the fatty tissues of animals living in water or on land, particularly those at the top of the food chain [4]. For this study, fish that have higher content of fat tissue, usually these fish species are at the top of the food chain, should have higher PCB’s level.

The analysis results showed that both fish species and waterbody affect Total PCB. This is because of that different fish species belong to different levels of the food chain, thus, they have different lipid content (which will cause different amount of stored PCB). Different waterbody could also have different amount of PCB’s pollution, which will result in different accumulation amount of PCB in fish that live in that specific waterbody. The equation from the regression model for predicting Total PCB also showed that higher lipid content, larger fish, and higher level in the food chain would result in higher content of Total PCB in fish.

However, for the Lipid PCB, neither fish species nor waterbody affect Lipid PCB. The reason should be that PCB is mostly stored in lipid tissue. For a certain amount of lipid tissue, the PCB it can store is also a certain amount. That is to say that PCB’s concentration in lipid tissue should be within a certain range. Therefore, for fish that belongs to different species or waterbody, as long as the PCB’s weight is normalized by lipid tissue weight, the ratio (Lipid PCB) should be similar no matter how heavy the fish is or which waterbody it belongs to. The equation from the regression model for predicting Lipid PCB also showed that neither fish weight nor waterbody would affect the actual content of Lipid PCB in fish.

When comparing the alternative quantification method for PCB, the analysis showed that method LRMSMS provided similar quantification results as that from using HRMS (the standard method). However, in order to conclude that these two methods are equivalent, an equivalence test is still needed to claim that they are equivalent [5].

**References:**

[1] [Jerrold J.Heindel](https://www.sciencedirect.com/science/article/pii/B9780323189071001530" \l "!)[R. ThomasZoeller](https://www.sciencedirect.com/science/article/pii/B9780323189071001530" \l "!). [Endocrinology: Adult and Pediatric (Seventh Edition)](https://www.sciencedirect.com/science/book/9780323189071). 2016, Pages 2640-2652.e3.

[2] [V J Cogliano](https://www.ncbi.nlm.nih.gov/pubmed/?term=Cogliano%20VJ%5BAuthor%5D&cauthor=true&cauthor_uid=9618347). Assessing the cancer risk from environmental PCBs. [Environ Health Perspect](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1532993/). 1998 Jun; 106(6): 317–323.

[3] Results of the monitoring of non dioxin-like PCBs in food and feed. EFSA journal. 2010, 8(7):1701.

[4] Scientific Facts on PCBs Polychlorinated biphenyls – GreenFacts. Sep 15, 2006. https://www.greenfacts.org/en/pcbs/

[5] Equivalence Testing. http://homepage.stat.uiowa.edu/~rdecook/stat6220/Class\_notes/equivalence\_testing.pdf