Multivariate Analysis of Organic Contaminants in Farmed and Wild Salmon

Math 257 Final Project

Fall, 2020

Dr. Andrea Gottlieb

Team Members:
Kaylyn Vo
Sagar Shahi
Nicole Pereira

Acknowledgements

Thank you, Dr. Gottlieb for teaching us multivariate statistical analysis, for your patience and always willing to help and always offering more office hours than we could ever asked for.

Thank you, Dr. Ronald Hites for kindly provided us with an interesting dataset that we could apply our understanding of multivariate statistics and get to analyze toxins in one of our favorite seafoods.

Tables of Contents

1. Summary

Global production of farmed Atlantic salmon is estimated 2.6 million tons in 2019[1]. Five countries made up 95.6% of the production of farm salmon with Norway, Scotland, Faroe Islands account for 55.3%, 7.6%, 3.3% respectively in Europe, Chile (25.4%) in South America, and Canada (6%) in North America[2]. Although the health benefits of eating fish such as salmon have been well studied; however, there is only small number of studies on human health risks of consuming salmon as they bioaccumulate toxic contaminants through their feeds and environment[4]. We used multivariate analysis techniques to analyze concentrations of contaminants and fatty acids in farmed and wild salmon. The first analysis focuses on comparing concentrations of contaminants and healthy fats between farmed and wild salmon as a baseline. The second analysis focuses on comparing concentrations of contaminants from the top farmed salmon production countries – Norway, Scotland, Faroe Islands, Chile, and Canada - as they account for the more than ninety five percent of total salmon production.

Our analysis indicates that farmed salmon have substantially higher contaminants, fatty acids and percent lipid than wild salmon. Further analysis shows that farmed salmon raised in Europe have higher concentrations of contaminants and lipid percentage and lower in total Omega-6 fatty acids than farmed salmon in North America while the total Omega-3 are the same between the two continents. The best choice for consumption is wild salmon but they are relatively more expensive than farmed salmon; thus, the next best choice would be farmed salmon from the west coast of North America as they have the lowest contaminant profiles. Although our analysis shown that there are statistically significant differences of contaminants between Europe and North America, further studies are needed to access whether the differences are practical significant and how much of concentrations of contaminants would adversely affect our health. The analyses were inspired by a research “Global assessment of organic contaminants in farmed salmon” by Ronald Hites and et al. For consumption advisories, please refer to a related study “Consumption advisories for salmon based on risk of cancer and non-cancer health effects” by Ronald Hites and et al.

2. Materials and Sampling Methods

We analyzed four organochlorine contaminants and two healthy fat concentrations along with the lipid percentage in farmed and wild salmon collected around the world. A total of 594 individual whole farmed Atlantic salmon were purchased from wholesalers to assure that they were appropriate size and in the same sampling period. The farmed salmon were purchased from 51 farms in eight farming regions in six countries (Scotland, Norway, Faroe Islands, Eastern Canada, Maine, Western Canada, Washington state, and Chile). Additionally, 144 farmed Atlantic salmon fillets were purchased at supermarkets in 16 major cities (Boston, Chicago, Denver, Edinburgh, Frankfurt, London, Los Angeles, New Orleans, New York, Oslo, Paris, San Francisco, Seattle, Toronto, Vancouver, and Washington D.C.)[5]. For comparisons, samples of five Pacific wild salmon species - chum, coho, chinook, pink, sockeye - were collected from Alaska, British Columbia, and Oregon[4].

A final total of 459 whole farmed salmon, 144 farmed supermarket fillets, and 135 wild salmon fillets were used in the study. Each composite sample consisted of fillets from three salmon per location or three fillets per supermarket. The three fillets were ground and reground together to make a homogeneous composite and then analyzed by gas chromatographic high-resolution mass spectrometry[5]. One farmed sample did not pass the QA/QC criteria and was dropped from the dataset. We have a total of 200 farmed salmon samples and 45 wild salmon samples. Of the 200 farmed salmon samples, 11 observations from the supermarket samples did not have known place of origins; therefore, only 189 samples were used when analyzed for differences of contaminants and healthy fats between Europe, North America, and South America.

The contaminants that we especially looked in details are dieldrin, dioxin equivalent, total polychlorinated byphenyls (PCBs), and total toxaphene equivalent because they are highly toxic industrial compounds, and they pose serious health risks from prolonged to small amount of these compounds[6]. These four inorganic contaminants were used as insecticide or pesticide in the 1970s and 1980s or by-products of industrial processes[6,7,8,9,10]. Although they are banned in the United States, they persist in the environment as they don’t dissolve in water and are slow to break down and accumulate in the sediments at the bottom of lakes, streams and coastal areas[6,7,8,9,10].

The total PCBs is the sum of polychlorinated byphenyls-like compounds. Similarly, toxaphene is a sum of related inorganic compounds. Dioxin and dioxin-like compounds were used to calculate dioxin equivalent. It is calculated by multiplying the actual gram weight of each dioxin and dioxin-like compound by its corresponding toxic equivalent factor (TEF) and then summing the results[10]. Dieldrin was measured directly. Although, some of the contaminants were derived from other chemicals, they are unrelated and independent of each other. All units were in ng/g except for Dioxin which was reported as pg TEQ/g.

Lipid analysis was performed by a method based on one recommendation from the Association of Official Analytical Chemists[5]. Fatty acid concentrations are reported as total n-6 fatty acids and total n-3 fatty acids. The total of n-6 fatty acids is the sum of the concentrations of linoleic, γ-linolenic, eicosadienoic, homo- γ-linolenic, arachidonic, and docosatetraneoic acids and the total of n-3 fatty acids is the sum of the concentrations of α-linolenic, eicosatrienoic, eicosapentaenoic, docosapentaenoic, and docosahaexanoic acids. The two fatty acids were measured in unit of mg/g[5]. Total lipids determinations were performed by gravimetric analysis in conjunction with the contaminant analysis. The result was expressed in percentage of tissue weight[5].

3. Statistical Analysis

Contaminants, fatty acids and lipid percentage in farmed and wild salmon were analyzed using multivariate analysis of variance (MANOVA). We decided to use multivariate statistical approach for three reasons: 1) there are correlations between the variables, 2) there will be multiple statistical inferences in the study, and we want to control the overall rate of false positives, 3) univariate statistical approach ignores correlations between the variables, and it is possible that univariate tests fail to differentiate real differences between the populations. In comparing farmed versus wild salmon, the farmed salmon from wholesalers and from supermarkets were considered one group and wild salmon as the other group. Additionally, we also analyzed the concentrations of contaminants and fatty acids in farmed salmon from Europe, North America and South America.

3.A. Principal Component Analysis

We used principal component analysis (PCA) to explore the data by reducing the number of dimensions while minimizing the loss of information of the real structure of the data as possible. The recording units for contaminants and total fatty acids are different; thus, we scaled each variable by its own average and standard deviation before we applied PCA to the data. About 94% of total variation in the data accounted for by the first three principal components, with the first two components accounted for 86% of the total variation. The first principal component is a weighted average with the weights are about equally distributed across the seven variables. The second component can be interpreted as a contrast between the contaminants and the fatty acids with all negative coefficients for the contaminants and all positive coefficients for the fatty acids and lipid percentage. We also computed principal component scores for the three principal components. Figure 1 shows three plots of these scores. The plot of scores of principal component (PC) one against the scores of PC two reveals that large positive scores computed from the first component and large negative score from the second component are from salmon with high concentrations in contaminants but relatively low concentrations in fatty acids. The interpretation for PC three is not as direct as the first two PCs but we can still infer that farmed salmon have large principal component scores than wild salmon. This means that measurements of contaminants and healthy fats are lower in wild salmon than farmed salmon. Furthermore, we can also see that Atlantic farmed salmon from Europe have the highest concentration of contaminants, on average, in comparison to farmed salmon from North and South America. An unusual pattern that we observed is that there seems to be two groups of farmed salmon from South America and North America. By keeping the total amount of fatty acids and lipid percentage fixed, we see that one cluster of salmon has higher concentrations of contaminants than the other. We also observe potential outlying observations from North America and Europe from all three plots. We will investigate these observations in further analyses. Overall, principal component analysis gives us insights and intuitions of the data that we couldn't possibly achieve otherwise with a dataset of seven dimensions.

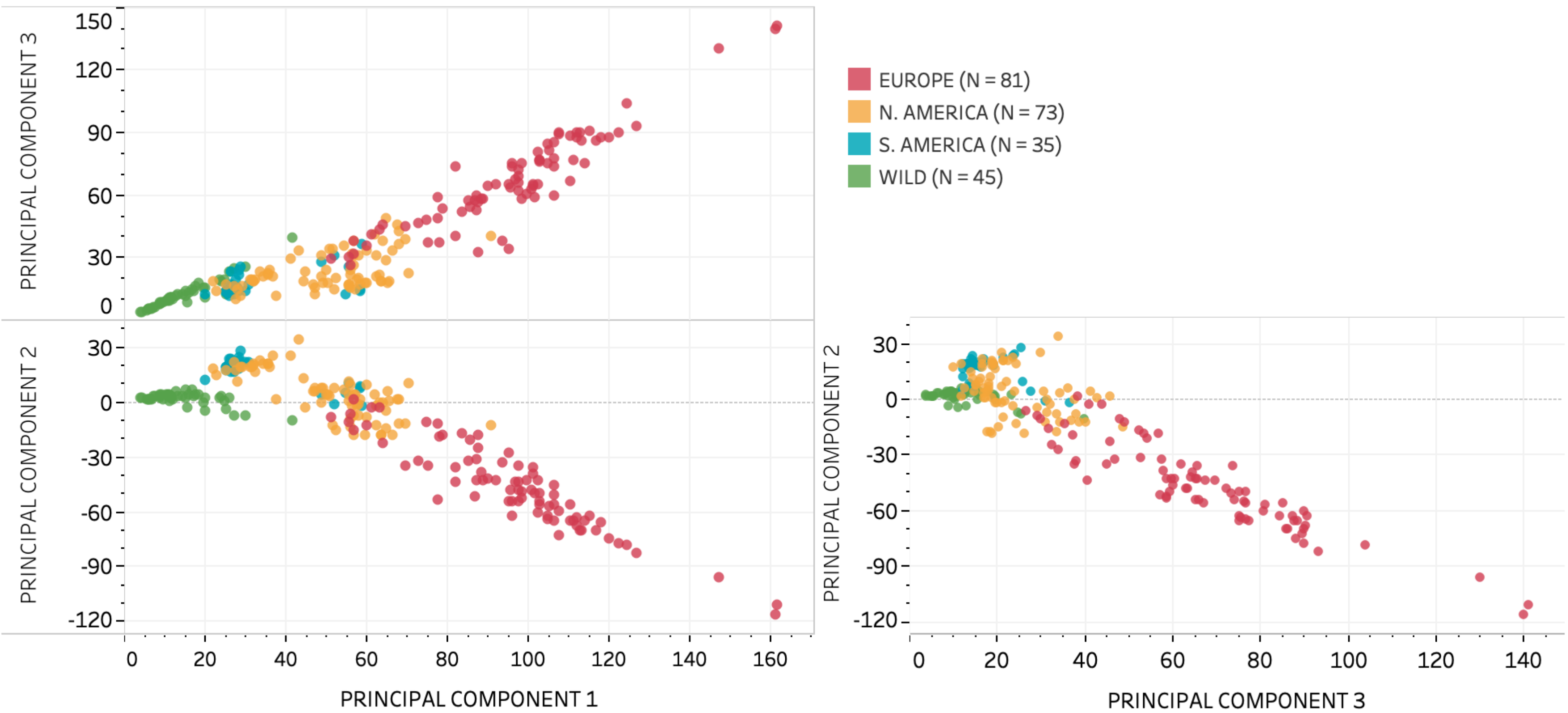

Figure 1: Principal component scores of first three components.

3.B. Means Comparison of Contaminants and Fatty Acids Between Farmed and Wild Salmon

We check for the multivariate normality by examining the distribution of the seven variables from the farmed salmon and wild salmon samples. The original distributions of the four contaminants and the total Omega-6 fatty acids of both samples are skewed. We decided to fit logarithmic transformation to the five variables of both samples. After applying the logarithmic transformation, the distributions of all variables from the wild salmon sample are approximately Normal distributed. Additionally, we used a Chi-square distribution with the degree of freedom of seven to further access the normality assumption. Both methods suggest that the seven variables from the wild salmon sample are individually Normal distributed. Although, the two approaches do not guarantee multivariate normality; however, for practical purpose these two steps are often enough for checking normality.

Even after the transformation, the distribution of the contaminant Dieldrin still shows a serious deviation from normality in farmed salmon. We looked closely at a scatterplot and a boxplot of Dieldrin as seen in Figure 2, there is a large variation in the logarithmic scale ranging from -0.456 to 0.996 with two distinct clusters of observations. The first cluster of 46 observations consists only farmed salmon from the coast of North America and South America with observations in the range of -0.456 to 0.033. This phenomenon also reflects in the boxplots of the variable Dieldrin of North and South America in Figure 4. The second cluster of 139 observations contains a majority of farmed salmon from east coast of North America and Europe with observations between 0.274 to 1. We also looked at scatterplots of the remaining three contaminants to check for the similar pattern. The scatterplots suggest that farmed salmon from Europe have higher concentration contaminants than from western side of North America and South America on average, but there were not any distinct clusters as seen in the distribution of Dieldrin in Figure 2.

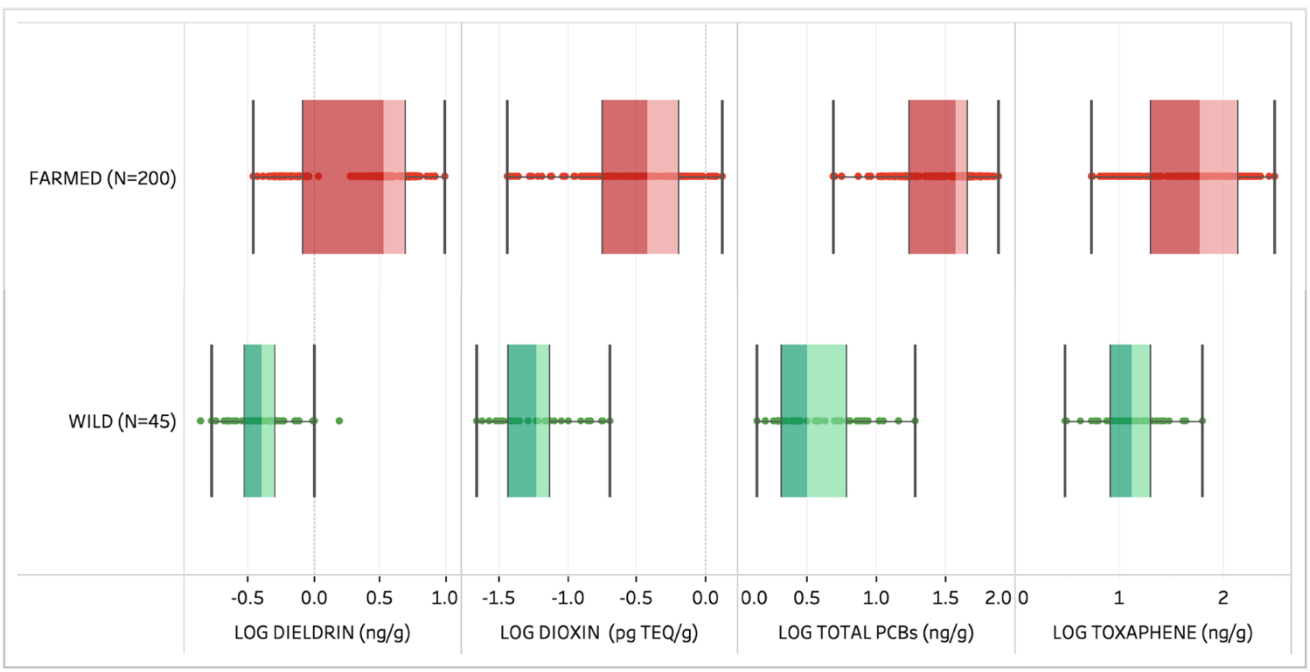

Figure 2: Distributions of logarithmic transformation of four contaminants from farmed and wild samples

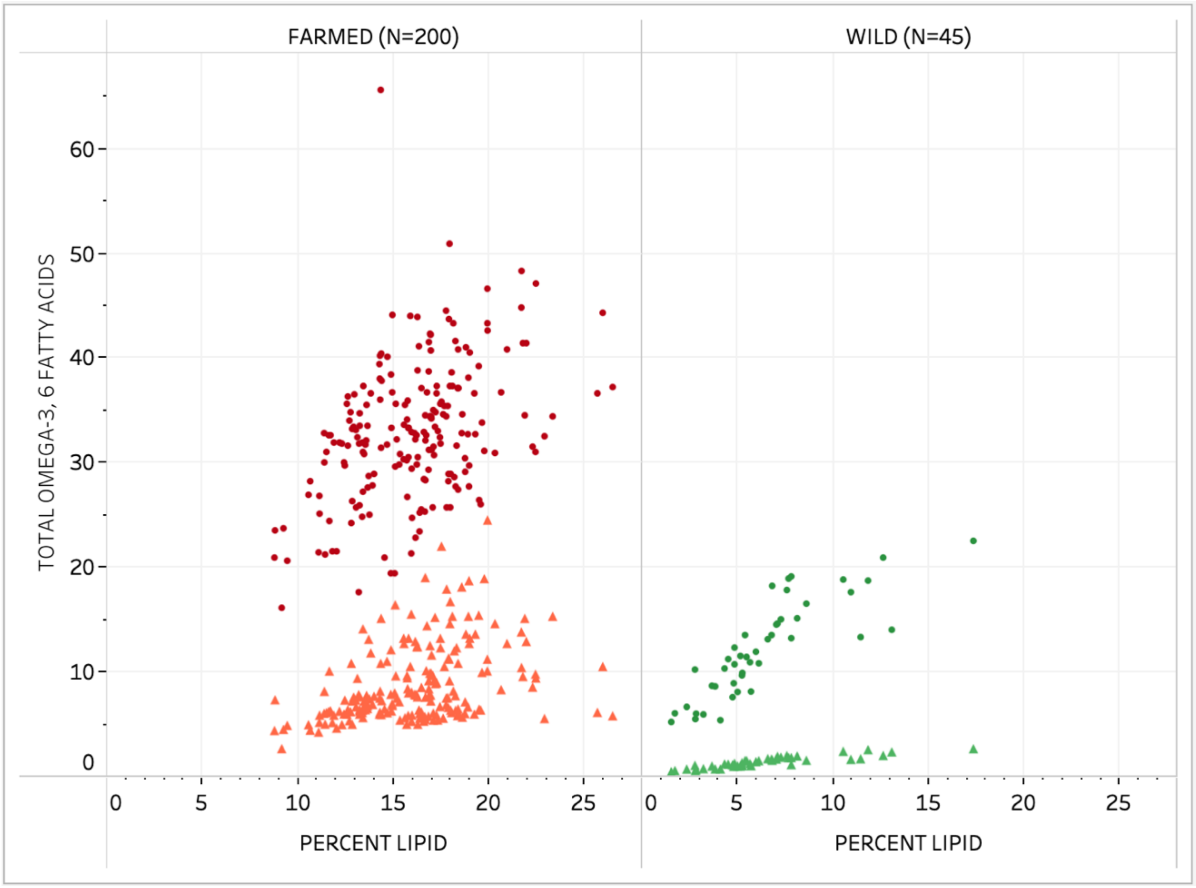

Figure 3: Total Omega 3 and 6 and lipid percentage of farmed and wild salmon

We did not transform the total Omega-3 fatty acids and lipid percentage as the variables were already approximately Normal distributed. Figure 3 shows the total fatty acids with lipid percentage in farmed and wild salmon. The plots show that the three averages of total fatty acids and lipid percentage are higher in farmed than wild salmon. Additionally, there are more variabilities in the measurements of fatty acids in farmed salmon than wild salmon.

We also examined potential outlying observations. These six observations came from Europe and the east side of North America. Their concentrations of contaminants are higher than average across all four contaminants, and this behavior is expected because high concentration in one contaminant would likely lead to high concentrations in the remaining contaminants. We decided to keep them in the analysis.

We have a sample of 200 observations and six of the seven variables are approximately Normal distributed. The deviation from normality of the contaminant Dieldrin doesn't seriously inflate Type I error as the analysis method is relatively robust to departures from normality. We decided to proceed with the multivariate analysis of the means comparison of two populations. After logarithmic transformation, the sample variances of the wild and farmed salmon are still quite different. Thus, we decided to proceed with the analysis using the Chi-square test instead of the exact F test. The approximate test statistics is 1634 with $\chi^2(0.05, 7) = 14.07$. We reject the null hypothesis and conclude that the differences of contaminants and fatty acids between farmed and wild salmon are statistically significance. The result of the test only tells us that there is at least one significance difference, so we computed confidence intervals to know exactly how many statistically significant differences between farmed and wild salmon. As expected, the average differences of all seven variables are positive. This means that the averages of contaminants, fatty acids, and lipid percentage are higher in farmed than wild salmon.

3.C. Means Comparison Farmed Salmon from Europe, North America, and South America

Figure 1 suggests that there might be differences in contaminants of farmed salmon sampled from Europe, North America, and South America. We analyzed the averages of concentrations of contaminants and fatty acids in farmed salmon sampled from the three continents. The original distributions of the four contaminants and the total Omega-6 fatty acids were skewed; therefore, we decided to fit a logarithmic transformation to the five variables. Figure 4 shows the logarithmic transformation distributions of the four variables of farmed salmon from the three continents.

After the transformation, the measurements from Europe are approximately Normal distributed. Additionally, we also used a Chi-square distribution to further access the normality assumption. Both methods suggest that the measurements of seven variables from Europe are individually Normal distributed. The distributions of the observations from North and South America need further inspections. The distribution of the contaminant Dieldrin shows an unusual pattern for both North America and South America. We further inspected the pattern of North America sample using a scatterplot and we found that two clusters of Dieldrin concentration can be separate into two regions – east and west side - of North America. Deviation from normality still persists even if we divide North America into two regions. Thus, we decided to analyze the observations from two regions of North America as one sample. There were not serious deviations from normality for the rest of the variables. We also looked at a potential outlying observation from North America. This observation has higher than average concentrations for all four contaminants, we decided to keep it in our analysis because it was not due to sampling error. Although there is deviation from Normality for the contaminant Dieldrin; however, with a relatively large sample size of 73 observations and the remaining six variables are approximately Normal distributed, we decided to proceed with the analysis with the assumption of Normality approximately satisfied.

The four boxplots of the contaminants sampled from South America as seen in Figure 4 reveals that there seems to be two distinct groups with one group of six observations has higher concentrations of contaminants than the other. We also check for the average weight and length measurements of five of the six observations, they have similar measurements with the rest of observations. We concluded that these five observations are not errors because the pattern of concentrations are similar across the four variables. With only 35 observations and the separation into two distinct groups across the four contaminants, we cannot justify the normality assumption of the data sampled from South America. We decided to exclude South America out of our analysis. Figure 5 shows the total fatty acids with the percent lipid concentration in farmed salmon of the three continents. The scatterplots of total fatty acids of farmed salmon also suggest that the averages are about the same between the two samples.

Since there was a large different of variances of the contaminants between the two samples, so we proceed with the multivariate analysis of means comparison of the two populations using the approximate Chi-square test. The approximate test statistics is 413 with $\chi^2(0.05, 7) = 14.07$. We reject the null hypothesis and conclude there is at least one significant difference between Europe and North America samples. We calculated simultaneous confidence intervals to determine which variables are different. The four contaminants and the lipid percentage are higher in farmed salmon from Europe than North America. On the other hand, the average of total Omega-6 from Europe is lower than North America while the average concentrations of Omega-3 are the same between the two continents.

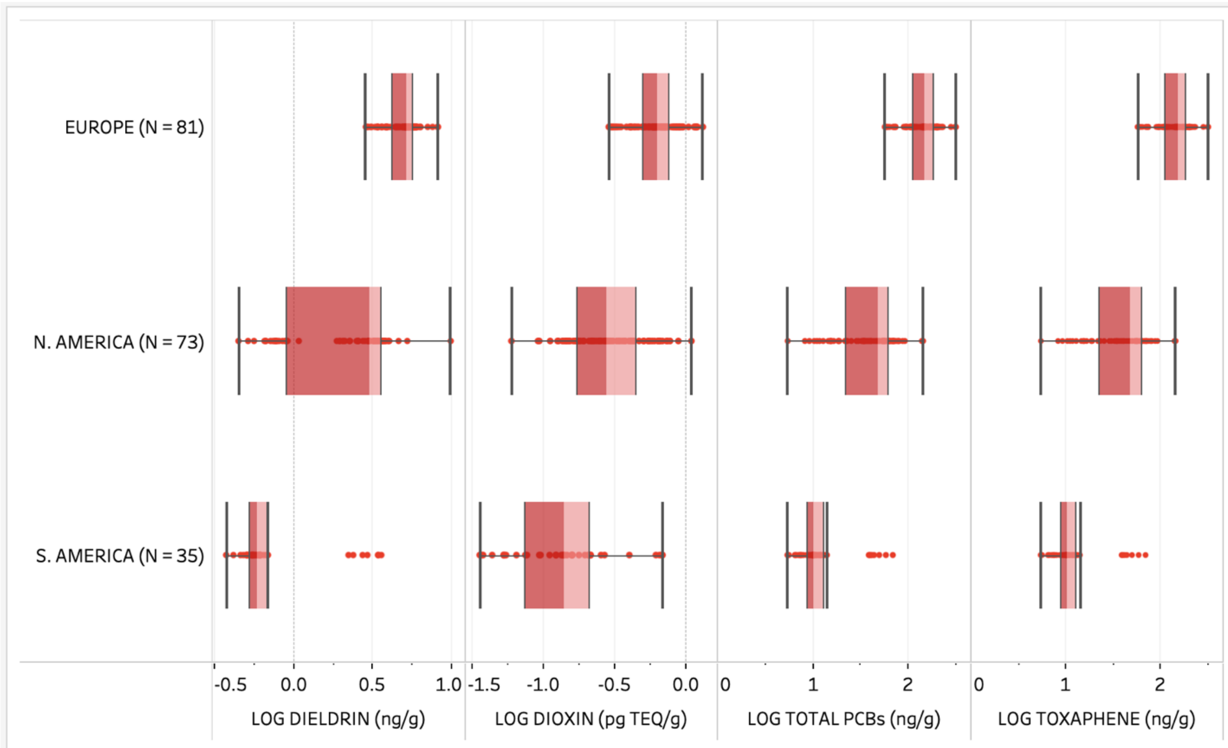

Figure 4. Distributions of logarithmic transformation of four contaminants from the three continents

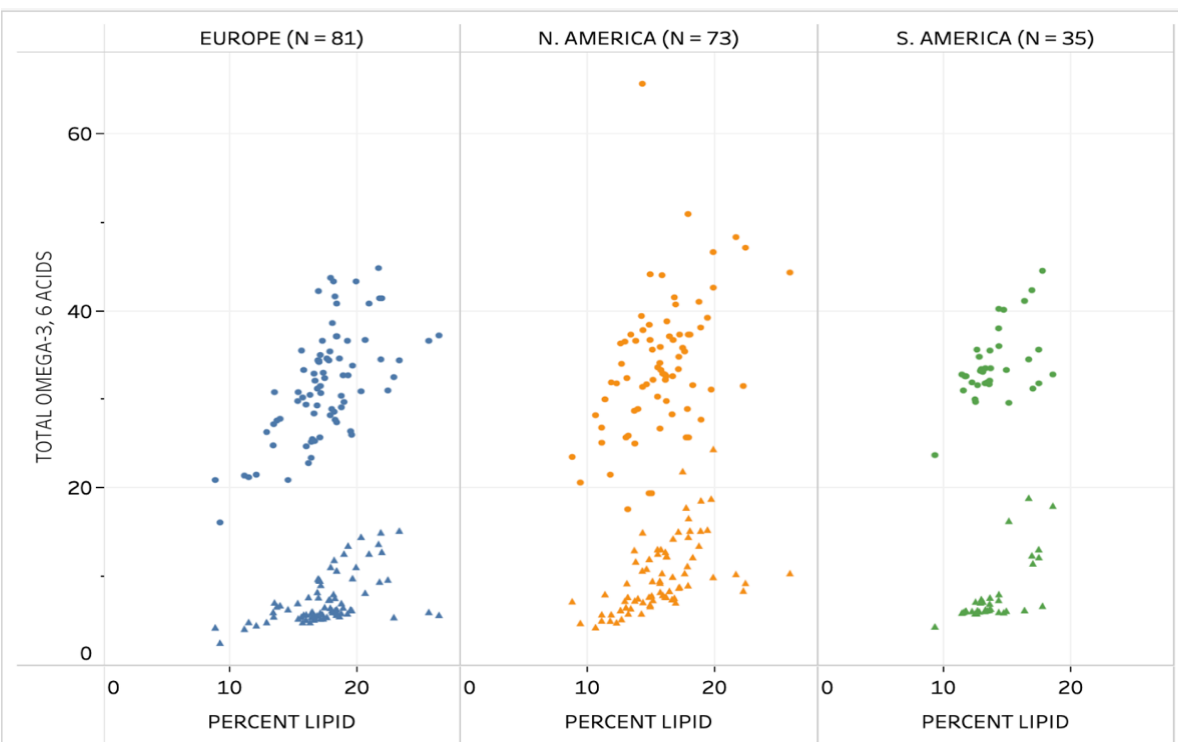

Figure 5: Total Omega 3 and 6 and lipid percentage from Europe, North America and South America.

4. Results and Discussion

Salmon are relatively fatty fish that provide essential omega-3 and omega-6 fatty acids that the human body cannot produce, however they also bioaccumulate toxins through their environment. Inspired by a research work that analyzed the concentrations of contaminants in farmed salmon by Ronald Hites and et al, we also analyzed the concentrations of contaminants using a multivariate approach. Our analysis indicate that farmed salmon have substantially higher contaminants, fatty acids and percent lipid than wild salmon. Further analysis shows that farmed salmon from Europe have the highest concentrations of contaminants while farmed salmon from the west side of North America have the lowest concentrations of contaminants. Unfortunately, we were not able to analyze samples from South America; however, we examined the available data and patterns suggest that the averages of contaminants are lower than farmed salmon from Europe and east side of North America. However, we still need more observations to make a stronger argument and conclude definitely. Although the study focused on contaminants, as seafood enthusiasts who are also mindful of antibiotics in our food sources, we would also like to analyze the concentrations of antibiotics in farmed salmon from Norway, Scotland, Chile, and Canada as they are responsible for total 95% of farmed salmon market, so that we can make the best decision for our health and our budget. Finally, our analyses depended on a set of assumptions and without proper justifications of those assumptions, the results would be meaningless. We agree that the assumption of normality for the variable Dieldrin in both analyses are not exactly satisfied and one could perform non-parametric univariate analyses of the variable separately. However, we would still arrive at similar conclusions and results because of the overwhelming message from the remaining contaminants.

5. Citations

1. “FAO.org.” *Salmon | GLOBEFISH - Information and Analysis on World Fish Trade | Food and Agriculture Organization of the United Nations Market Reports | GLOBEFISH | Food and Agriculture Organization of the United Nations*, www.fao.org/in-action/globefish/market-reports/salmon/en/.
2. Iversen, Audun, et al. “Production Cost and Competitiveness in Major Salmon Farming Countries 2003–2018.” *Aquaculture*, Elsevier, 12 Feb. 2020,
3. Omega-3 Fatty Acids: Harvard T.H. Chan, School of Public Health
4. Global assessment of organic contaminants in farmed salmon, *Science*, 303, 226-229 (2004); with J. A. Foran, D. O. Carpenter, M. C. Hamilton, B. A. Knuth, S. J. Schwager. Responses to several comments, 305, 475-478 (2004) and on-line.
5. Lipid composition and contaminants in farmed and wild salmon, *Environmental Science & Technology*, 39, 8622-8629 (2005); with, M. C. Hamilton, S. J. Schwager, J. A. Foran, B. A. Knuth, and D. O. Carpenter.
6. “PCBs in Fish and Shellfish.” *Seafood Selector*, 19 Feb. 2013, seafood.edf.org/pcbs-fish-and-shellfish.
7. Agency for Toxic Substances and Diseases Registry: Toxaphene
8. Agency for Toxic Substances and Diseases Registry: Dieldrin
9. EDF Seafood Selector: PCBs
10. Dioxin: United States Environmental Production Agency
11. World Health Organization: Dioxin and their effects on human health