**Go Fish! The Important Factors in Population Management for Atlantic Cod in the Northeast Continental Shelf**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**MSBA 6120 - Statistics for Data Science**

Yali (Lydia) Li

Dimo Liao

Lauren Moore

Anupam Shandilya

Submitted August 17, 2019

**Introduction:**

Motivation:

As the world begins to grow and change, ensuring sustainability becomes a primary concern. The Food and Agriculture Organization (FAO) of the United Nations monitors the state of the world’s fisheries. The FAO reported that from 1990 to 2007, approximately one-quarter of fish stocks have either been overexploited, depleted, or in a state of recovering from depletion [1].

There are numerous historical examples which show the importance of correctly approximating the amount of fish to ensure a sustainable level. For example, the Atlantic cod is currently classified as depleted. Originally, the cod were an abundant species in the North Atlantic [2]. Specifically, in 1852 the biomass of Atlantic cod was about 1,260,000 metric tons [3]. Unfortunately, due to the high demand of Atlantic cod and an underestimation of the mortality rate led to overfishing which in turn caused the population of Atlantic cod to plummet. As of 2005, the biomass of Atlantic cod has been estimated to be 50,000 metric tons [3]. This is a 96% decrease in the biomass of Atlantic cod.

The goal of this project is to determine which predictors influence the biomass of the Atlantic cod’s population in the Northeast Continental Shelf. By knowing which predictors best estimate the biomass, we are able to loosely determine the condition of the fish-stock. These components could include the biomass of the Atlantic cod’s main prey, the Atlantic herring. Other predictors of interest may include the biomass of the silver hake, the temperature of the water, both at the surface and the bottom of the water. Finally, the depth of trawl may be of interest. Because the data has been collected over multiple seasons, we will only focus on the fall. This will take into account the migration patterns. The data is also divided into five different regions, but we will focus on all regions of the Northeast Continental Shelf.

Fish Overview:

There are three different types of fish species we will focus on this report, the Atlantic cod, Atlantic herring, and silver hake. We chose to focus on the Atlantic cod as our variable of interest due to their state of depletion as mentioned in the previous section. The Atlantic herring are an important species because they are the primary prey for the Atlantic cod. The silver hake are a highly predatory species and consume Atlantic herring. The relationship between herring and hake may affect the biomass of Atlantic cod. Basic understanding of the preditor-prey relationship would tell us that as the prey population increases, the predator population increases. Thus it is necessary for us to describe each of the species in order to accurately build a model to estimate the biomass of the Atlantic cod’s population in the Northeast Continental Shelf. It is important to note that the Atlantic herring will be the only prey taken into consideration for the Atlantic cod.

Atlantic Cod:

The National Oceanic and Atmospheric Administration (NOAA) Fisheries reported that Atlantic cod are overfished according to their 2017 stock assessment. Atlantic cod in the Northeast continental shelf live in the Georges Bank and the Gulf of Maine regions [2].

Atlantic cod often stay close to the ocean floor, and large cod are often closer to the ground than the smaller ones. Atlantic cod form compact schools during the day and will scatter at night [4]. They migrate twice per year, prior to and after spawning [5]. Atlantic Cods migrate north to spawning grounds in March and return south in June [4]. Larger Atlantic cod travel very little outside of the spawning season. Like most predators, if the Atlantic cod exhaust their food supply in one spot, they must move to fresh foraging grounds.

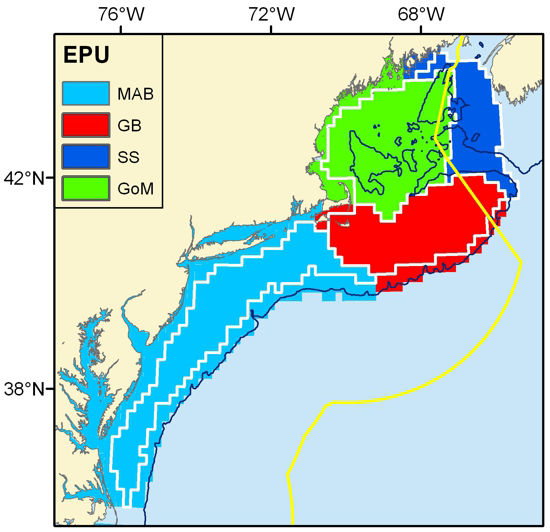
Atlantic Herring:

NOAA Fisheries reported that the Atlantic herring are not overfished according to their 2018 stock assessment [6]. Additionally, NOAA estimates the Atlantic herring population is at about 517,930 metric tons, which is above their target population of 157,000 metric tons. Atlantic herring are found from Labrador to North Carolina [6]. This means Atlantic herring are found throughout the Northeast Continental Shelf.

Atlantic herring travel in schools of up to hundreds or thousands of fish. Herring are not frightened by boats, which could cause the data collected from the bottom trawl survey (BTS) to be skewed in comparison to other fish species. The Atlantic herring migrate twice per year, north in the spring and south in the fall. The spring migration occurs between May and June and the fall migration occurs between November and December. Spawning occurs between July and October.

Silver Hake:

NOAA Fisheries reported that the silver hake are not overfished according to their 2017 stock assessment [7]. Hake are known to be voracious predators that consume Atlantic herring and other fish in the region. Spawning season varies by region, but on average spawning occurs earlier in the south (May to June) than in the north (July to August). In the summer months, silver hake migrate to shallow waters in the Northwest Atlantic to spawn. Silver hake are found from Newfoundland to South Carolina [7].

Data Used: 

The data used was obtained through a previous Research Experience for Undergraduates with the Northeast Fisheries Science Center (NEFSC) [8]. The NEFSC provided survey data from 1968 to 2013. This data is collected twice a year (spring and fall) at 640 different survey stations in four different regions. These regions consists of the Mid-Atlantic Bight (MAB), Georges Bank (GB), Scotian Shelf (SS), and the Gulf of Maine (GoM), each of which is shown in the figure to the right.

Each region of the Northeast Continental Shelf is then further subdivided into strata. Data is collected via a BTS at randomly selected strata, where each strata has an equally likely choice of being selected. In these surveys, a trawling boat goes to the preselected strata, drops their net, and drags it for 20 minutes along the ocean floor. When the time is up, the net is pulled up to the surface and the fish are sorted by species [9]. Data is then collected for each species. Data such as abundance (count), total mass of the catch, surface and bottom temperature, and depth of trawl are recorded for each species.

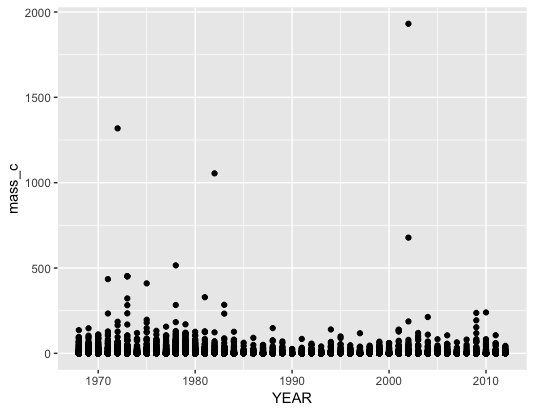
For this paper, we will be utilizing the following data:

|  |  |  |
| --- | --- | --- |
|  | **Description** | **Units** |
| ***mass\_c*** | Total biomass of the Atlantic cod collected in the survey | Kilograms |
| ***mass\_h*** | Total biomass of the Atlantic herring collected in the survey | Kilograms |
| ***mass\_ha*** | Total biomass of the Silver Hake collected in the survey | Kilograms |
| ***b\_temp*** | Temperature of the trawl depth when the survey was taken. | Celsius |
| ***s\_temp*** | Temperature of the ocean surface when the survey was taken. | Celsius |
| ***depth*** | Depth between the trawling boat and the depth when the survey was taken. | Meters |

The surveys are conducted every spring between March and May, and every fall between September and November [9], but we do not know the exact month in which each survey was conducted. Additionally, we do not have information on fishing patterns for either the Atlantic cod or the Atlantic herring in the Northeast continental shelf. This may create some limitations as we begin to analyze the data.

Data Observations:

After filtering the data to fit our criteria, we have 18,546 total survey observations of either Atlantic cod, Atlantic herring, or silver hake in the Northeast Continental Shelf during the fall. On average, the biomass for the Atlantic cod is 3.598kg in a single trawl. It is important to note that the maximum biomass for the Atlantic cod is 1930.930kg in a single trawl, the minimum biomass for the Atlantic cod is 0.0kg in a single trawl, and the median biomass for the Atlantic cod 0.0kg.

Through this analysis of the data, we know that the distribution of Atlantic cod is likely skewed right. Because of this, it may be better to look at the general plot of the species biomass over the years. This is shown in the figure to the right. 

There is a clear uniform distribution of the biomass of the Atlantic cod. We can also see that the maximum biomass for the Atlantic cod (1930.930kg) in a single trawl is likely an outlier. This provides more reassurance as we enter the analysis of our question

Knowing this information will allow us to better understand the construction and output of our model.

**Analyses:**

Full Model:

With the goal of this report to determine which predictors influence the biomass of the Atlantic cod’s population in the Northeast Continental Shelf during the fall, the first step is to run a full model. This model will contain every predictor we believe has a chance of influencing the biomass of the Atlantic cod. A detailed analysis of this model will appear in the appendix, this will just be an overview. When we fit our data with the predictors in question, we obtained the following model:

18.184646 - 0.617038\*(*s temp*) - 0.227014\*(*b temp*) - 0.011407\*(*depth*)

+ 0.366996\*(*mass h*) - 0.093778 \*(*mass ha*)

After confirming that none of the predictors have a large relationship with one another, we can interpret the coefficients of the model as follows:

* For every increase in degree Celsius of surface temperature, on average there is a decrease of 0.617038 kilograms in the biomass of Atlantic cod in the Northeast Continental Shelf during the fall, keeping all other variables fixed.
* For every increase in degree Celsius of bottom temperature, on average there is a decrease of 0.227014 kilograms in the biomass of Atlantic cod in the Northeast Continental Shelf during the fall, keeping all other variables fixed.
* For every increase in meter of depth of trawl, on average there is a decrease of 0.011407 kilograms in the biomass of Atlantic cod in the Northeast Continental Shelf Bank during the fall, keeping all other variables fixed.
* For every increase in kilogram in the biomass of Atlantic herring in the Northeast Continental Shelf during the fall, on average there is an increase of 0.366996 kilograms in the biomass of Atlantic cod in the Northeast Continental Shelf during the fall, keeping all other variables fixed.
* For every increase in kilogram in the biomass of silver hake in the Northeast Continental Shelf during the fall, on average there is a decrease of 0.093778 kilograms in the biomass of Atlantic cod in the Northeast Continental Shelf during the fall, keeping all other variables fixed.

According to some of the output of our full regression model, we have determined that our model does not appear to fit the data in a practical sense. In this current model, only 9.581% of the variation in the biomass of the Atlantic cod in the Northeast Continental Shelf during the fall in our sample is explained by the model. Meanwhile in a statistical sense, the current model does show a relationship between the predictors and the biomass of the Atlantic cod in the Northeast Continental Shelf during the fall. Specifically, there is a chance of obtaining a test statistic more extreme than the actual sample value given there is no relationship between the biomass of Atlantic cod in the Northeast Continental Shelf during the fall and the predictors in the model.

Through the output of our regression model, we can determine which predictors have a relationship with the biomass of Atlantic cod in the Northeast Continental Shelf during the fall. What we have learned from this information is stated below:

* There is a chance of obtaining a test statistic more extreme than the actual sample value given there is no relationship between the surface temperature and the biomass of Atlantic cod in the Northeast Continental Shelf during the fall.
* There is a chance of obtaining a test statistic more extreme than the actual sample value given there is no relationship between the bottom temperature and the biomass of Atlantic cod in the Northeast Continental Shelf during the fall.
* There is a chance of obtaining a test statistic more extreme than the actual sample value given there is no relationship between depth of trawl and the biomass of Atlantic cod in the Northeast Continental Shelf during the fall.
* There is a chance of obtaining a test statistic more extreme than the actual sample value given there is no relationship between the biomass of Atlantic herring and the biomass of Atlantic cod in the Northeast Continental Shelf during the fall.
* There is a chance of obtaining a test statistic more extreme than the actual sample value given there is no relationship between the biomass of silver hake and the biomass of Atlantic cod in the Northeast Continental Shelf during the fall.

From this, we can determine that the biomass of Atlantic herring is indeed relevant in predicting the biomass of the Atlantic cod. This follows with the biological analysis of the Atlantic cod. Because Atlantic cod eat herring, we would expect there to be a strong relationship between the two species. Other notable predictors include surface temperature, depth, and biomass of silver hake.

(*A thorough analysis of the assumptions will be conducted in the appendix.)*

Simplified Model:

After a thorough investigation, we were unable to find a model ‘better’ than the full model described above. Instead, we were able to simplify the model and obtain results that are equivalent to the full model. In this simplified model, two predictors were removed: *s\_temp* and *mass\_ha*.

Surface temperature was removed for two reasons. First, it was related with depth and bottom temperature. This relationship was not significant enough that it was crucial to remove surface temperature as a predictor. The reason we chose to was due to logic. The species of fish being observed, the Atlantic cod, prefer to be in deeper waters. Thus they are very unlikely to be near the surface to begin with. Surface temperature appears to be impractical considering the biological activities of the Atlantic cod.

The biomass of the silver hake was removed because they rarely affect the Atlantic cod. While both consumer Atlantic herring, the are above their target population. There are plenty of herring for both the silver hake and Atlantic cod. If the Atlantic herring was in a state of depletion or recovering from a state of depletion, this would not be the case.

When we fit our data with the new set of predictors in question, we obtained the following model:

14.369809 - 0.671580\*(*b temp*) - 0.024012\*(*depth*) + 0.368004\*(*mass h*)

After confirming that none of the predictors have a large relationship with one another, we can interpret the coefficients of the model as follows:

* For every increase in degree Celsius of bottom temperature, on average there is a decrease of 0.671580 kilograms in the biomass of Atlantic cod in the Northeast Continental Shelf during the fall, keeping all other variables fixed.
* For every increase in meter of depth of trawl, on average there is a decrease of 0.024012 kilograms in the biomass of Atlantic cod in the Northeast Continental Shelf Bank during the fall, keeping all other variables fixed.
* For every increase in kilogram in the biomass of Atlantic herring in the Northeast Continental Shelf during the fall, on average there is an increase of 0.368004 kilograms in the biomass of Atlantic cod in the Northeast Continental Shelf during the fall, keeping all other variables fixed.

Additionally, none of the signs on the coefficients changed. The signs continue to be appropriate for the situation.

According to some of the output of our regression model, we have determined that our model still does not appear to fit the data in a practical sense. In this current model, only 8.706% of the variation in the biomass of the Atlantic cod in the Northeast Continental Shelf during the fall in our sample is explained by the model. While this is lower than the previous model, it is not significantly lower. This model is no less practical than the full model. Meanwhile in a statistical sense, the current model does show a relationship between the predictors and the biomass of the Atlantic cod in the Northeast Continental Shelf during the fall. Specifically, there is a chance of obtaining a test statistic more extreme than the actual sample value given there is no relationship between the biomass of Atlantic cod in the Northeast Continental Shelf during the fall and the predictors in the model. This percentage is identical to that of the previous model.

Through the output of our regression model, we can determine which predictors have a relationship with the biomass of Atlantic cod in the Northeast Continental Shelf during the fall. What we have learned from this information is stated below:

* There is a chance of obtaining a test statistic more extreme than the actual sample value given there is no relationship between the bottom temperature and the biomass of Atlantic cod in the Northeast Continental Shelf during the fall.
* There is a chance of obtaining a test statistic more extreme than the actual sample value given there is no relationship between depth of trawl and the biomass of Atlantic cod in the Northeast Continental Shelf during the fall.
* There is a chance of obtaining a test statistic more extreme than the actual sample value given there is no relationship between the biomass of Atlantic herring and the biomass of Atlantic cod in the Northeast Continental Shelf during the fall.

From this, we can determine that the biomass of Atlantic herring is still relevant in predicting the biomass of the Atlantic cod. This follows with the biological analysis of the Atlantic cod. Because Atlantic cod eat herring, we would expect there to be a strong relationship between the two species. Other notable predictors include bottom temperature and depth.

(*A thorough analysis of the assumptions will be conducted in the appendix.)*

**Discussions:**

Summary:

The goal of this project was to determine which predictors influence the biomass of the Atlantic cod’s population in the Northeast Continental Shelf during the fall. The components we observed include the biomass of the Atlantic cod’s main prey, the Atlantic herring, biomass of the silver hake, the temperature of the water, both at the surface and the bottom of the water, and the depth of trawl. To answer this question, we composed two models: the full model and the simplified model. Below is a table comparing the coefficients of both models side-by-side:

|  |  |  |
| --- | --- | --- |
|  | **Full Model** | **Simplified Model** |
| *mass\_h* | 0.366996 | 0.368004 |
| *b\_temp* | -0.227014 | -0.671580 |
| *depth* | -0.011407 | -0.024012 |

From this table, we can see more clearly that the signs of the coefficients have not changed. Additionally, the magnitude of the coefficients is unchanged. Both of these observations in addition to statistical reasoning are noted in the appendix, we can comfortably say that both models are equivalent in predicting the biomass of the Atlantic cod. Neither model is worse than the other.

Key Takeaways:

Both models provide evidence of a relationship between the biomass of Atlantic cod and Atlantic herring, bottom temperature at time of trawl, and the depth of trawl in the Northeast Continental Shelf in the fall. Atlantic herring are the main food source of Atlantic cod, so there will be a direct relationship between the biomass of the two species. But, in both models certain assumptions do not hold which indicate that the data is not normal. Refer appendix for the test of assumptions.

Further transformations can be done to the data to make this model more useful. Now we know which predictors best estimate the biomass, and if we were given the criteria of classifying the condition of fish stock, we may be able to loosely determine it. Other models such as Lotka-Volterra (predator-prey model) can be used in conjunction to the above model to get higher predictive accuracy of the biomass of Atlantic cod and determine the condition of the fish stock. This model would more accurately take into account the birth rate, mortality rate, and carrying capacity of the environment.

**Appendix:**

Technical Analysis of the Full Model:

As mentioned in the introduction, there are four predictors that could lead to an estimation of the biomass in Atlantic cod. These predictors were *mass\_h*, *mass\_ha, s\_temp, b\_temp*, and *depth*. The first model we ran included all three predictors:

After running this regression in R, we obtained the following fitted model:

18.184646 - 0.617038\*(*s temp*) - 0.227014\*(*b temp*) - 0.011407\*(*depth*)

+ 0.366996\*(*mass h*) - 0.093778 \*(*mass ha*)

The exact R code and output is shown below:

> fit1 <- lm(mass\_c ~ s\_temp + b\_temp + depth + mass\_ha + mass\_h, data = data\_final)

> summary(fit1)

Call:

lm(formula = mass\_c ~ s\_temp + b\_temp + depth + mass\_ha + mass\_h,

data = data\_final)

Residuals:

Min 1Q Median 3Q Max

-382.95 -5.10 -2.18 0.90 1487.20

Coefficients:

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

(Intercept) 18.184646 0.821651 22.132 < 2e-16 \*\*\*

s\_temp -0.617038 0.047538 -12.980 < 2e-16 \*\*\*

b\_temp -0.227014 0.054880 -4.137 3.54e-05 \*\*\*

depth -0.011407 0.003075 -3.709 0.000209 \*\*\*

mass\_ha -0.093778 0.021042 -4.457 8.37e-06 \*\*\*

mass\_h 0.366996 0.009758 37.610 < 2e-16 \*\*\*

---

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

Residual standard error: 24.12 on 18540 degrees of freedom

Multiple R-squared: 0.09581, Adjusted R-squared: 0.09557

F-statistic: 392.9 on 5 and 18540 DF, p-value: < 2.2e-16

Before interpreting any of the coefficients, we checked the collinearity of the predictors. There appears to be a relationship between *b\_temp* and *s\_temp* with a correlation coefficient of 0.63. With a correlation coefficient of this size, we can continue to interpret the predicted coefficients of each of these predictors, but cautiously. Additionally, there appears to be a relationship between *b\_temp* and *depth* with a correlation coefficient of -0.61. With a correlation coefficient of this size, we can continue to interpret the predicted coefficients of each of these predictors, but cautiously. The correlation coefficient was obtained from the output below:

> mydata <- data\_final%>%select("mass\_c", "mass\_h","mass\_ha","s\_temp","b\_temp","depth")

> cormat <- round(cor(mydata),2)

> cormat

mass\_c mass\_h mass\_ha s\_temp b\_temp depth

mass\_c 1.00 0.27 0.02 -0.16 -0.12 0.02

mass\_h 0.27 1.00 0.07 -0.07 -0.08 0.03

mass\_ha 0.02 0.07 1.00 -0.21 -0.20 0.17

s\_temp -0.16 -0.07 -0.21 1.00 0.63 -0.20

b\_temp -0.12 -0.08 -0.20 0.63 1.00 -0.61

depth 0.02 0.03 0.17 -0.20 -0.61 1.00

We should note that all of the signs of the predictors in the model appear to be appropriate. Atlantic cod prefer to stay in cooler water, thus as the surface and bottom temperature increases it is logical to assume that the biomass of Atlantic cod will decrease. Additionally, as Atlantic cod mainly consume Atlantic herring, it is logical to assume that as the biomass of Atlantic herring increases, the biomass of the Atlantic cod will also increase. As the silver hake eat more Atlantic herring, there are fewer for the Atlantic cod to consume. This causes a decrease in Atlantic cod biomass. Depth is a variable that is dependent upon multiple things. The depth varies depending on the proximity to the coast and thus the location of the cod. If we assume the depth increases, on average, with respect to an increase in distance from shore, the sign of the coefficient makes sense. As the depth increases, the biomass of Atlantic cod will decrease. The intercept is unintelligible. If there happens to be a bottom and surface temperature of 0 degrees Celsius, depth of 0 meters, and a biomass of 0 herring and hake, there will be no Atlantic cod in the Northeast Continental Shelf during the fall. Meanwhile, our intercept says there should be a biomass of 18.18kg of Atlantic cod.

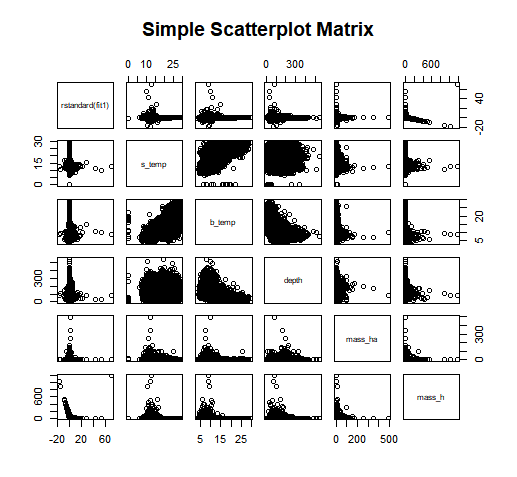
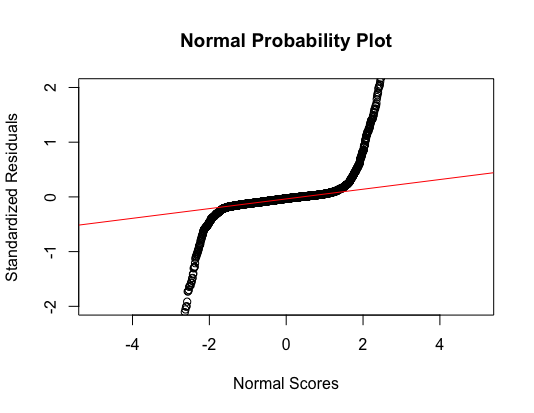
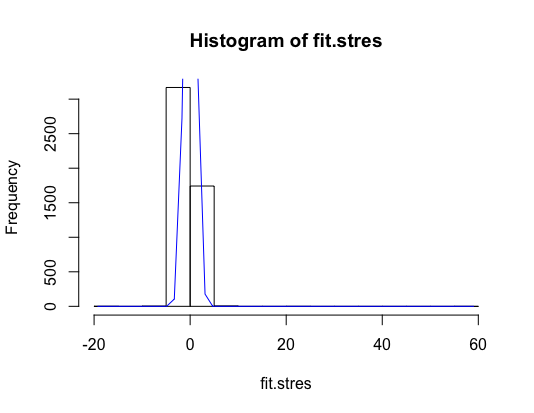
The p-value of each predictor is testing the following relationship:

The biggest takeaway is that our first model indicates a relationship between the *mass\_h* and the *mass\_*c. Essentially, there is a chance of obtaining a test statistic more extreme than the actual sample value given there is no relationship between the biomass of Atlantic herring and the biomass of Atlantic cod in the Northeast Continental Shelf during the fall. It is important to note that all predictors are statistically significant and thus have a relationship with the biomass of Atlantic cod.

According to the , our model does not appear to fit the data in a practical sense. In this model, 9.581% of the variation in the biomass of the Atlantic cod in the Northeast Continental Shelf during the fall in our sample is explained by the model. Meanwhile in a statistical sense, our model does show a relationship between the predictors and the biomass of the Atlantic cod in the Northeast Continental Shelf during the fall. This is valid because the p-value of the model is . This p-value is the result of the following hypothesis test:

Finally, we need to check all assumptions required of the model. First, we know the data is collected with random sampling. We also know the data was collected using the same process, thus there is stability over time. Additionally, we need to test whether ~ N(0, σ) constant over all predictors ( *i* = 1, 2, 3, 4, 5).

The histogram and the QQ-plot below show the distribution of standardized residuals, which is not normally distributed. This violates the assumption of the residuals are normally distributed with a mean centered around zero.

****

The code for this output is shown below:

fit1.stres <- rstandard(fit1)

plot(fit1$fitted.values, fit1.stres, pch=16,

main="Standardized Residual Plot", xlab="Fitted Cod biomass",

ylab = "Standardized Residuals")

abline(0, 0, lty=2, col="red")

h <- hist(fit1.stres)

x <- fit1.stres

xfit <- seq(min(x), max(x), length = 50 )

yfit <- dnorm(xfit, mean=mean(x), sd = sd(x))

yfit <- yfit\*diff(h$mids[1:2])\*length(x)

lines(xfit, yfit, col= "blue")

qqnorm(fit1.stres, main = "Normal Probability Plot", xlab = "Normal Scores", ylab = "Standardized Residuals", xlim = c(-2, 2), ylim = c(-2,2))

qqline(fit1.stres, col = "red")

pairs(~ rstandard(fit1) + s\_temp + b\_temp + depth + mass\_ha + mass\_h, data = data\_final, main="Simple Scatterplot Matrix")

Technical Analysis of the Simplified Model:

As mentioned in the introduction, there are four predictors that could lead to an estimation of the biomass in Atlantic cod. These predictors were *mass\_h*, *mass\_ha, s\_temp, b\_temp*, and *depth*. The first model we ran included all three predictors:

After running this regression in R, we obtained the following fitted model:

14.369809 - 0.671580\*(*b temp*) - 0.024012\*(*depth*) + 0.368004\*(*mass h*)

The exact R code and output is shown below:

> fit2 <- lm(mass\_c ~ b\_temp + depth + mass\_h, data = data\_final)

> summary(fit2)

Call:

lm(formula = mass\_c ~ b\_temp + depth + mass\_h, data = data\_final)

Residuals:

Min 1Q Median 3Q Max

-381.73 -4.70 -2.65 0.13 1488.72

Coefficients:

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

(Intercept) 14.369809 0.772056 18.612 < 2e-16 \*\*\*

b\_temp -0.671580 0.041833 -16.054 < 2e-16 \*\*\*

depth -0.024012 0.002939 -8.171 3.26e-16 \*\*\*

mass\_h 0.366512 0.009789 37.442 < 2e-16 \*\*\*

---

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

Residual standard error: 24.23 on 18542 degrees of freedom

Multiple R-squared: 0.08721, Adjusted R-squared: 0.08706

F-statistic: 590.5 on 3 and 18542 DF, p-value: < 2.2e-16

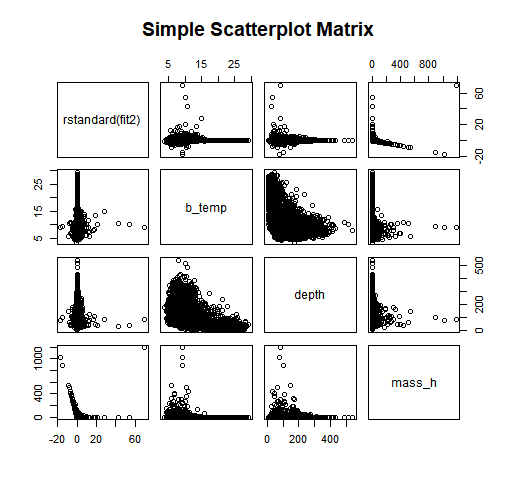
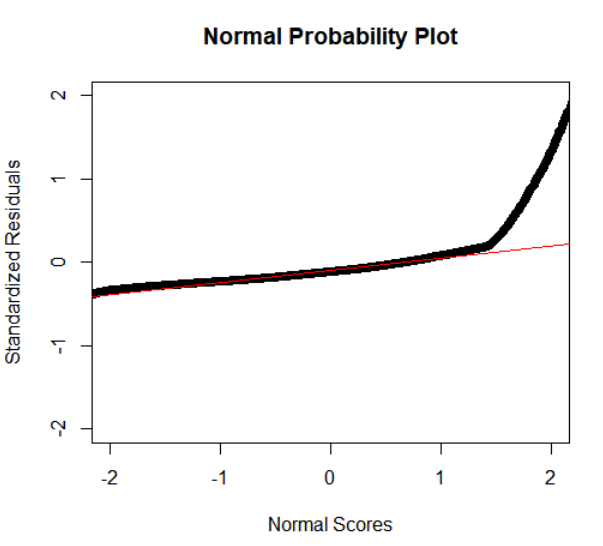
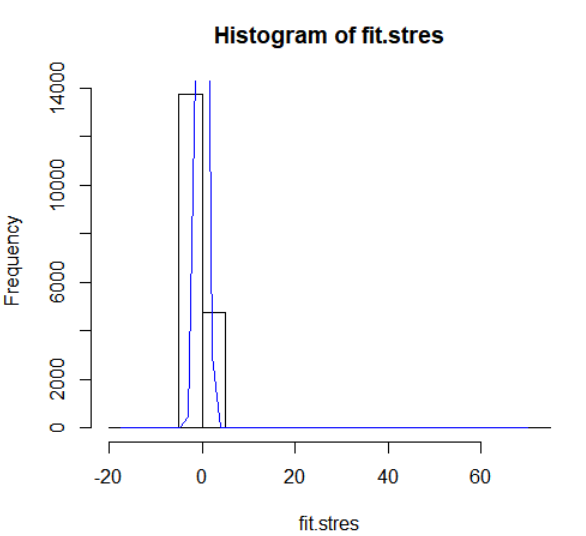
Before interpreting any of the coefficients, we checked the collinearity of the predictors. This was done in the previous section. We should also note that all of the signs of the predictors in the model continue to be appropriate. The intercept is still unintelligible. If there happens to be a bottom temperature of 0 degrees Celsius, depth of 0 meters, and a biomass of 0 herring, there will be no Atlantic cod in the Northeast Continental Shelf during the fall. Meanwhile, our intercept says there should be a biomass of 14.37kg of Atlantic cod.

The p-value of each predictor is testing the following relationship:

The biggest takeaway is that our simplified model indicates a relationship between the *mass\_h* and the *mass\_*c. Essentially, there is a chance of obtaining a test statistic more extreme than the actual sample value given there is no relationship between the biomass of Atlantic herring and the biomass of Atlantic cod in the Northeast Continental Shelf during the fall. It is important to note that all predictors are statistically significant and thus have a relationship with the biomass of Atlantic cod.

According to the , our model does not appear to fit the data in a practical sense. In this model, 8.721% of the variation in the biomass of the Atlantic cod in the Northeast Continental Shelf during the fall in our sample is explained by the model. Meanwhile in a statistical sense, our model does show a relationship between the predictors and the biomass of the Atlantic cod in the Northeast Continental Shelf during the fall. This is valid because the p-value of the model is . This p-value is the result of the following hypothesis test:

Again, like the histogram and the QQ-plot test we run in the first model. The three plots of the second model below show almost the same result as the first one, showing the distribution of standardized residuals are not normally distributed. This violates the assumption of the residuals are normally distributed with a mean centered around zero.



The code for this output is shown below:

fit2.stres <- rstandard(fit2)

plot(fit2$fitted.values, fit2.stres, pch=16,

main="Standardized Residual Plot", xlab="Fitted Cod biomass",

ylab = "Standardized Residuals")

abline(0, 0, lty=2, col="red")

h <- hist(fit2.stres)

x <- fit2.stres

xfit <- seq(min(x), max(x), length = 50 )

yfit <- dnorm(xfit, mean=mean(x), sd = sd(x))

yfit <- yfit\*diff(h$mids[1:2])\*length(x)

lines(xfit, yfit, col= "blue")

qqnorm(fit2.stres, main = "Normal Probability Plot", xlab = "Normal Scores", ylab = "Standardized Residuals", xlim = c(-2, 2), ylim = c(-2,2))

qqline(fit2.stres, col = "red")

pairs(~ rstandard(fit2) + b\_temp + depth + mass\_h, data = data\_final, main="Simple Scatterplot Matrix")

**References:** If appropriate, be sure to cite any references used, including your data sources. Any format is fine as long as the essentials are included: e.g., author, title, source information.

* [1] J. Beddington, D. Agnew, and C. Clark, “Current Problems in the Management of Marine Fisheries,” Science, vol. 316, no. 5832, pp. 1713 – 1716, 2007.
* [2] “Atlantic Cod,” https://www.fisheries.noaa.gov/species/atlantic-cod, 2018.
* [3] A. Myers, N. Hutchings, and N. Barrowman, “Hypotheses for the decline of cod in the North Atlantic,” Marine Ecology Progress Series, vol. 138, no. 293-308, 1996.
* [4] H. Bigelow and W. Schroeder, Fishes of the Gulf of Maine. US Government Printing Office, 1953, vol. 53.
* [5] M. Morin, “Movement of Atlantic Cod in And Among the Western Gulf of Maine Rolling Closures as Determined Through Mark and Recapture,” Master’s thesis, University of New Hampshire, 2000.
* [6] “Atlantic Herring,” https://www.fisheries.noaa.gov/species/atlantic-herring, 2019
* [7] “Silver Hake,” https://www.fisheries.noaa.gov/species/silver-hake, 2019
* [8] S. Amato, L. Moore, K. Ragosta, and S. Stowe, “Modeling the Spatio-Temporal Dynamics of Interacting Fish Species in the Northeast Continental Shelf,” pp. 1 - 7, 2018. *(In process of being published)*.
* [9] P. Politics, J. Galbraith, P. Kostovick, and R. Brown, “Northeast Fisheries Science Center Bottom Trawl Survey Protocols for the NOAA Ship Henry B. Bigelow,” 2014, Northeast Fisheries Science Center Reference Document.