Statistical Analysis of Salt Flooding on Juvenile Maritime Tree Species

Kyle Anderson, Odalys Barrientos & Abby Eck

MA 321 Statistical Consulting
Monmouth University
3 May, 2021
Dr. Richard Bastian
Client: Dr. Pedram Daneshgar

1 Introduction

Maritime coastal forests' proximity to the ocean makes these trees vulnerable to climate change. Rising sea level and frequent storms have been a relevant topic in the impact on maritime trees. To understand how coastal forests are impacted by rising sea level and the frequency of storms, we will analyze two greenhouse experiments. Both experiments used eight species of juvenile trees such as, Eastern red cedar (JV), pitch pine (PR), red maple (AR), black cherry (PS), black gum (NS), willow oak (QP), Southern red oak (QF), and sweet gum (LS). The first experiment used 16 trees from each species all at different levels of salinity. There were four levels of salinity: the control with 0 parts per thousand (ppt), 5 ppt, 10 ppt, and 15 ppt. Figure 1 provides a visual representation of the salinity experiment.

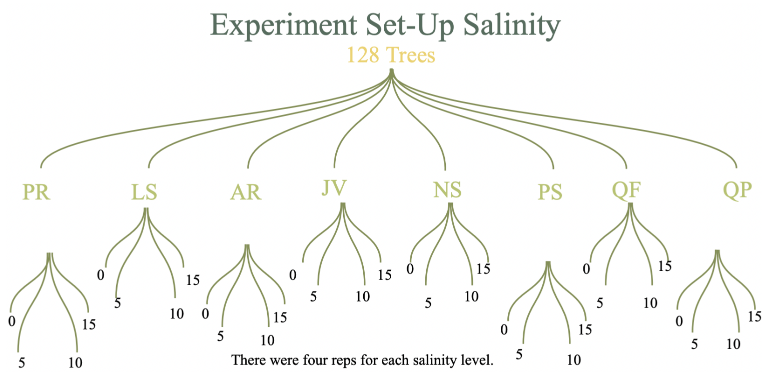


Figure 1: Visual of Salinity Experiment Setup

The second experiment used nine trees from each species all at different frequency levels. Monthly rates of flooding were determined by flooding in the mid-Atlantic coast in 2019 and the predicted flooding rates in 2030 and 2045. The three frequency levels all received the same amount of salt water (10 ppt) at different rates. Treatments 19 received the salt solution once every 30 days, 2030 once a week, and 2045 three times a week. To ensure all trees received the same amount of water, treatments 2019 and 2030 received the control treatment three to two times a week. Figure 2 provides a visual representation of the frequency experiment.

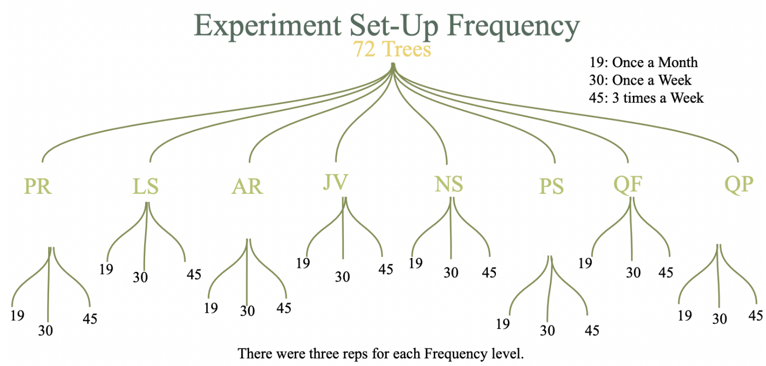


Figure 2: Visual of Frequency Experiment Setup

It is hypothesized that as the levels of salinity and frequency change, there will be a difference in biomass and a difference in the probability of surviving at a given point in time.

2 Methodology

2.1 ANOVA

For both the salinity experiment and the frequency experiment, we used a bootstrapped ANOVA test to test if there was a difference in the total biomass of each tree by the type of treatment they received. A bootstrapped ANOVA test was used because our data did not meet the assumptions of normality and equal variance. Bootstrapping is sampling with replacement from observed data to create simulated samples. These simulated samples estimate the statistic of interest. Since the eight species are different sizes, we ran eight separate bootstrapped ANOVA's for each experiment. This then allowed us to look at the different tests for each species to see which species did yield a statistically significant difference in the total biomass by the treatment they received and which ones did not.

2.2 Survival Analysis

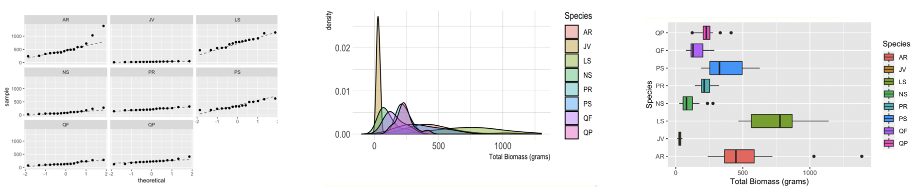
Survival Analysis was also used on the salinity and frequency experiment to analyze the time to the event of the death of the trees. To define death in both experiments, the Julian day in the Julian date calendar was used. If the tree did die during the experiment, it was marked as a one and had their Julian day correspond to when they died. If the tree did not die, then it was marked as a zero since the event of death did not occur. This analysis was done by the treatment they received, as well as by species. Survival Analysis was used to test if there was a difference between the populations in the probability of dying at any point in time. Through the use of this test, the day at which fifty percent of the trees in each category is given as well as the comparison between each category to show which treatments or species were different from each other using the Log Rank Test with bonferroni correction. If the median is not given, this means that the experiment would need to be longer in order to determine when fifty percent of the plants would die.

3 Results

3.1 Salinity Experiment

3.1.1 Descriptives

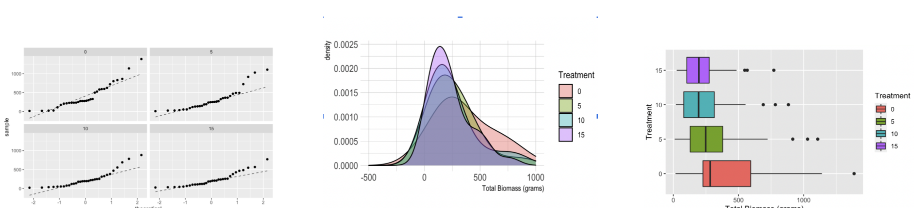
In order to use ANOVA, we need to verify if our data meets the assumption of equal variance and normality. Figure 3 shows the data split by the eight different species in our experiment. From the QQ-plot and the density plot we can see that our data does not meet the assumption of normality. Using the boxplot we can conclude that our data does not meet the assumption of equal variance.



(a) QQ Plot (b) Density Plot (c) Box Plot

Figure 3: Salinity Normality Plots by Species

Figure 4 shows the data split by the four different treatments used in the salinity experiment. From the QQ-plot and the density plot we can see that our data does not meet the assumption of normality. Using the boxplot we can conclude that our data does not meet the assumption of equal variance.



(a) QQ Plot (b) Density Plot (c) Box Plot

Figure 4: Salinity Normality Plots by Treatment

Since our data does not meet the assumptions of an ANOVA we will use a bootstrapped ANOVA. In this instance it would not be a good idea to use the Kruskal-

Wallis non-parametric test since that test assumes the distributions among the different group are similar in shape and shifted over. From the density plots we can see that the distributions are not similar in shape. Thus, we do not meet this assumption of the non-parametric test. Therefore, we will proceed with a bootstrapped ANOVA.

3.1.2 ANOVAs

From Figure 5 we can see that seven out of the eight species decrease in biomass as the levels of salinity increase. Some of these species have larger differences in biomass as the levels of salinity increase. While other species have smaller differences in biomass as the levels of salinity increase. Eastern red cedar (JS) is the only species who has a slight increase in biomass as the levels of salinity increase.

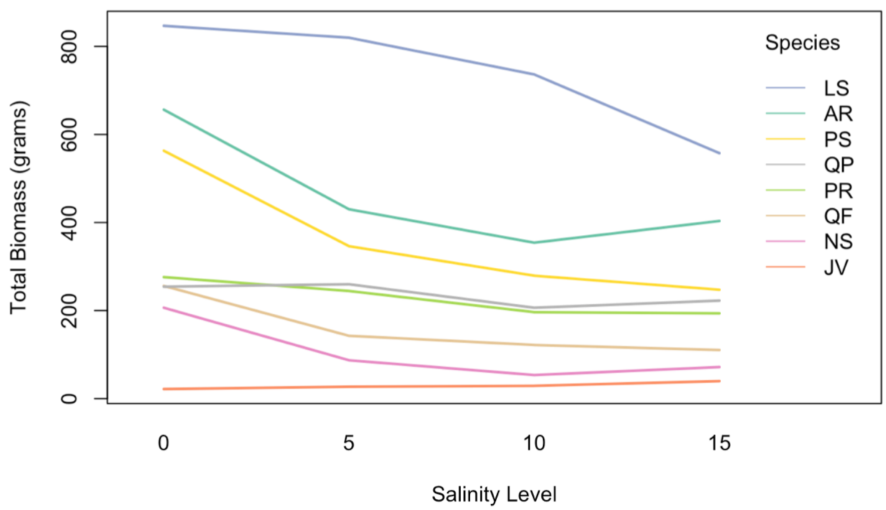
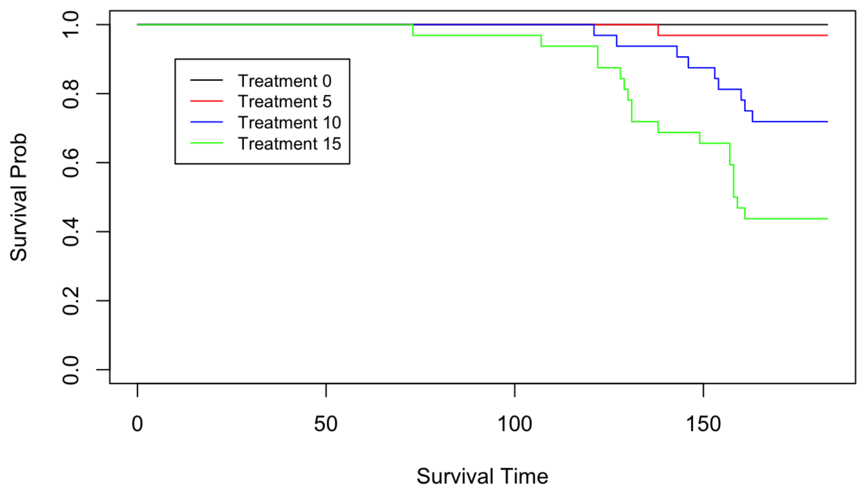


Figure 5: Means Plot for Salinity

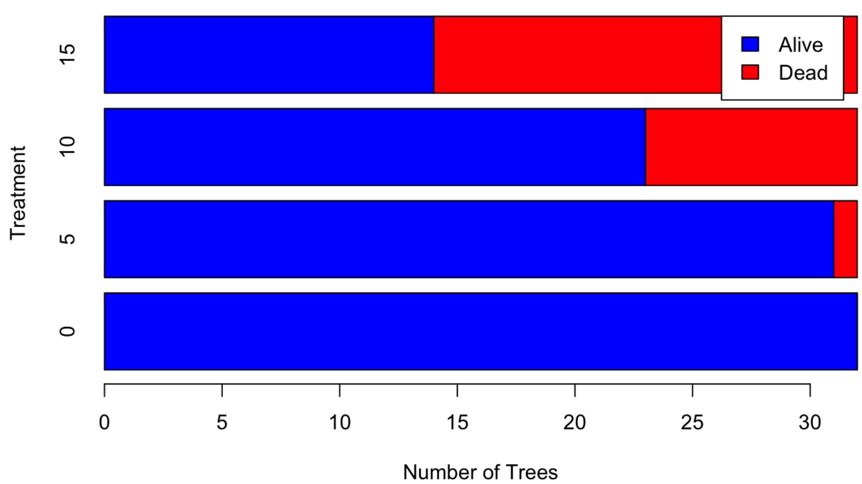
We computed eight different boostrapped ANOVA to test if there is a difference in biomass by the different treatment which were increasing levels of salinity. There were four species that gave us statistically significant p-value $(\alpha = 0.05)$ meaning there is a statistically significant difference in biomass when considering the treatment they were given. These trees were black cherry (PS), Southern red oak (QF), black gum (NS), and pitch pine (PR). All the other species decreased in biomass as the level of salinity increased but by smaller amounts. Eastern red cedar (JV) was the only species whose biomass slightly increased as the levels of salinity increased.

3.1.3 Survival Analysis

The following images and tables below show the results of the Survival Analysis given the treatment administered:



(a) Survival Analysis Results By Treatment



(b) Bar Graph of Survival Analysis by Treatment

	n	events	median
Treatment=0	32	0	NA
Treatment=5	32	1	NA
Treatment=10	32	9	NA
Treatment=15	32	18	158

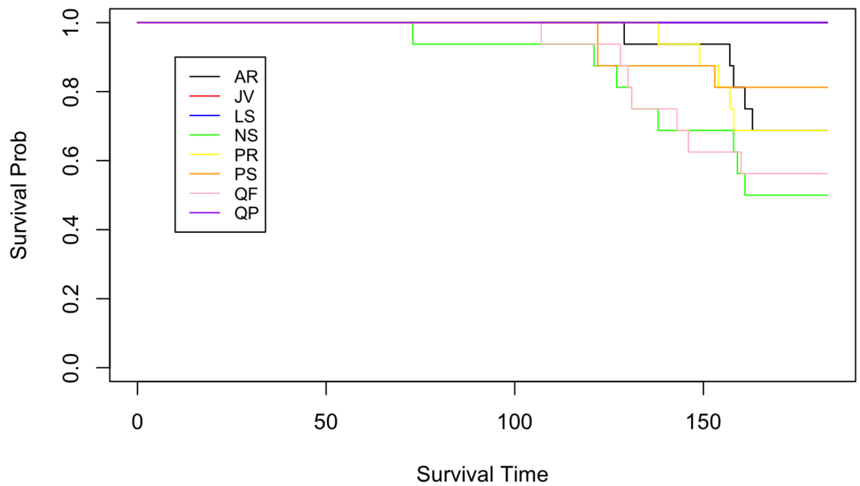
Table 1: Survival Results Given Treatment Given: Salinity Data

	0	5	10
5	1.000	-	-
10	0.0078	0.0408	-
15	$3*10^{-6}$	$1.9*10^{-5}$	0.1005

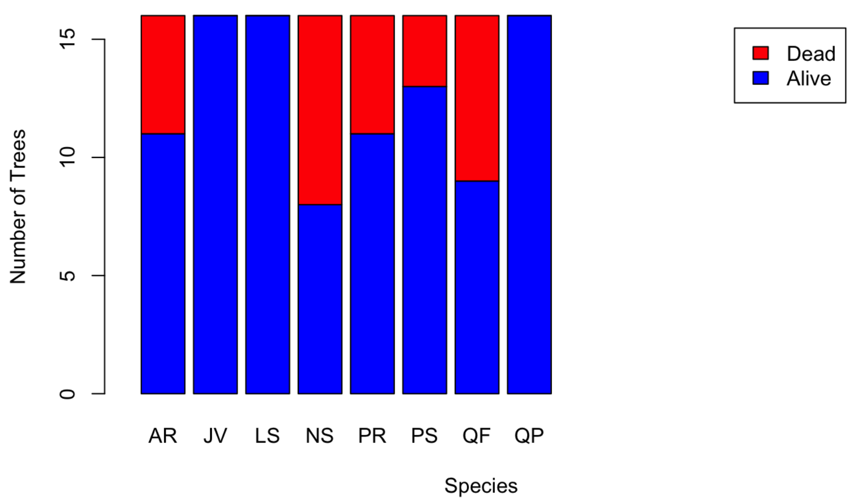
Table 2: Pairwise Comparison Using Log-Rank Test: Salinity Data, Treatment

The results show us that plants from the control group survived the length of the experiment, while one plant given salinity treatment of 5 ppt died, nine plants given salinity treatment of 10 ppt died, and eighteen plants given salinity treatment of 15 ppt died. In addition, plants given salinity treatment of 15 ppt were the only ones to reach the median threshold on the 161 day of the experiment, indicating that at least half of the plants given this treatment died by this point in the experiment. This shows us that as the level of salinity increased, the number of plants that died increased. As for the Pairwise Comparison results, there was statistical significance when comparing each salinity level, with the exception of the control treatment in comparison to the 5 ppt treatment, and when comparing the 10 ppt treatment to the 15 ppt treatment, as these comparisons had p-values greater than $\alpha = 0.05$, indicating that there was not enough evidence to show a significant difference in the number of plants that died between these treatments. Therefore, with the exception of the comparisons previously stated, there is statistical evidence indicating that there is a significant difference in the number of plants that died given the salinity treatment administered to the plant.

The following images and tables below show the results of the Survival Analysis given the type of plant.



(a) Image of Survival Analysis by Species, Salinity Experiment



(b) Bar Graph of Survival Analysis by Species

	n	events	median
Species=AR	16	5	NA
Species=JV	16	0	NA
Species=LS	16	0	NA
Species=NS	16	8	161
Species=PR	16	5	NA
Species=PS	16	3	NA
Species=QF	16	7	NA
Species=QP	16	0	NA

Table 3: Survival Results Given Plant: Salinity Data

	AR	JV	LS	NS	PR	PS	QF
JV	0.460	-	-	-	-	-	-
LS	0.460	1.000	-	-	-	-	-
NS	1.000	0.034	0.034	-	-	-	-
PR	1.000	0.460	0.460	1.000	-	-	-
PS	1.000	1.000	1.000	1.000	1.000	-	-
QF	1.000	0.086	0.086	1.000	1.000	1.000	-
QP	0.460	1.000	1.000	0.034	0.460	1.000	1.000

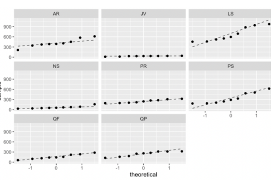
Table 4: Pairwise Comparisons Using Log-Rank Test Given Plant Type: Salinity Data

From the results of the Survival Analysis, it was found that all the plants from species JV, LS, and QP survived the length of the experiment, while plants from species AR, NS, PR,PS, and QF died throughout the experiment, with species NS being the only one to reach the median threshold on the 161 day of the experiment, meaning at least half of the plants from this species died by this point during the experiment. This indicates that all plants from Species JV, LS, and QP survived after receiving varying levels of salinity, while plants from species AR, NS, PR, PS, and QF died after receiving varying levels of salinity. When running Pairwise Comparison of the given species, it was found that the only statistically significant results were found when comparing species NS, which had the most plants die throughout the experiment, with species in which none of the plants died, including JV, LS, and QP, giving p-values less than $\alpha = 0.05$. The other comparisons had p-values higher than 0.05, indicating that there is not enough evidence to say that there is a significant difference in the number of plants that died within the species when compared to other species.

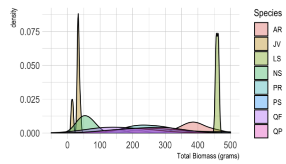
3.2 Frequency Experiment

3.2.1 Descriptives

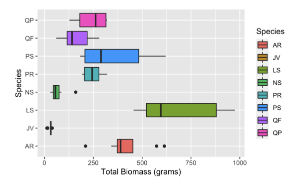
We will first check if we meet the assumptions of an ANOVA such as normality and equal variance. Figure 8 shows the data split by the eight different species in our experiment. From the QQ-plot and the density plot we can see that our data does not meet the assumption of normality. Using the boxplot we can conclude that our data does not meet the assumption of equal variance.



(a) QQ Plot



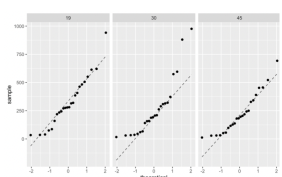
(b) Density Plot



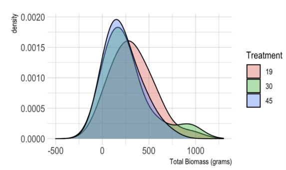
(c) Box Plot

Figure 8: Frequency Normality Plots by Species

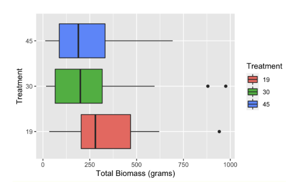
Figure 9 shows the data split by the three different treatments used in the frequency experiment. From the QQ-plot and the density plot we can see that our data does not meet the assumption of normality. Using the boxplot we can conclude that our data does not meet the assumption of equal variance.



(a) QQ Plot



(b) Density Plot



(c) Box Plot

Figure 9: Frequency Normality Plots by Treatment

We will continue with a bootstrapped ANOVA since we do not meet the assumptions of a traditional ANOVA.

3.2.2 ANOVAs

From Figure 10 we can see that two out of the eight species decrease in biomass as the levels of frequency increase, with those species being Willow Oak (Quercus falcata) and Eastern Red Cedar (Juniperus virginiana). Five of the species had changes in biomass that alternated as the level of frequency increased: biomass for the species Sweet Gum (Liquidamar styraciflua) and Pitch Pine (Pinus Rigida) increased as the frequency level went from once every 30 days to once a week, and then the biomass decreased as the frequency level went from once a week to three days a week. By contrast, biomass for the species PS, AR, and NS decreased as the frequency level went from once every 30 days to once a week, and then the biomass increased as the frequency level went from once a week to three days a week. Some of these species have larger differences in biomass as the level of frequency increase, while other species have smaller differences in biomass as the level of frequency increase. Eastern red cedar (Juniperus virginiana) is the only species who has a slight increase in biomass as the level of frequency increase

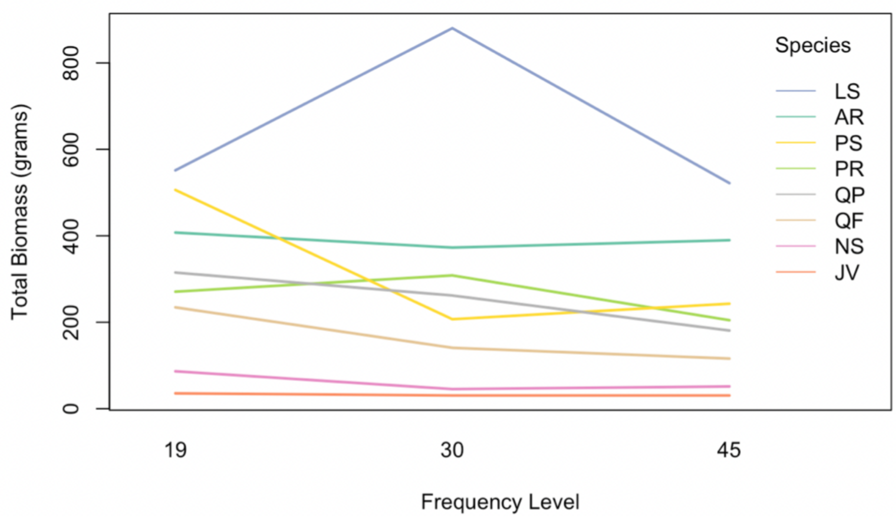
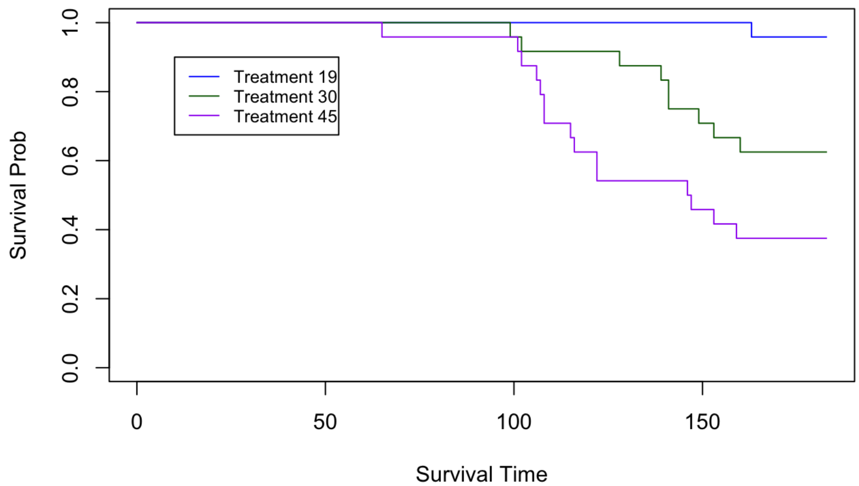


Figure 10: Means Plot for Frequency

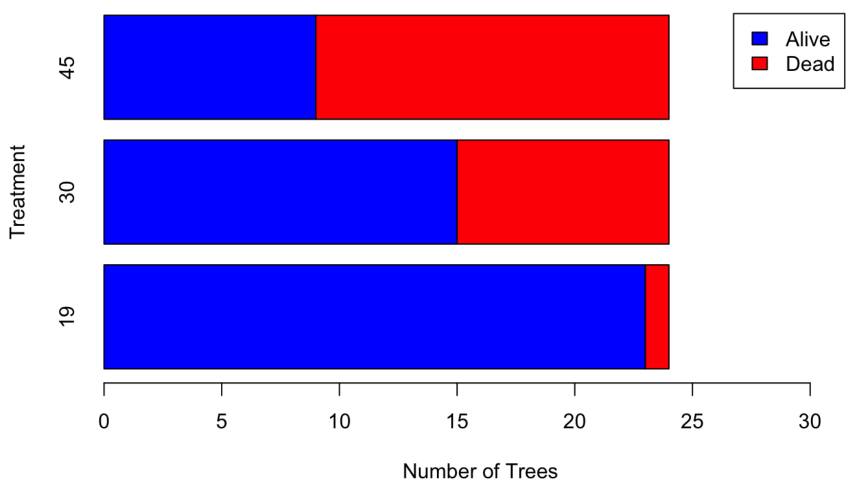
We computed eight different boostrapped ANOVA to test if there is a difference in biomass by the different treatment which were increasing levels of salinity. There were two species that gave us statistically significant p-value ($\alpha = 0.05$) meaning there is a statistically significant difference in biomass when considering the treatment they were given. These trees were black cherry (Prunus serotina) and Southern red oak(Quercus falcata). All the other species, with the exception of the two significant species and red cedar (Juniperus Virginia) whose biomass slightly increased with each higher level of frequency, had varying changes in biomass as the level of salinity increased. This included species increasing and than decreasing in biomass, with other species decreasing and than increasing in biomass.

3.2.3 Survival Analysis

The following images and tables below depict the results of the Survival Analysis of the frequency data given the frequency in which the salinity was administered to the plants:



(a) Image of Survival Analysis by Treatment, Frequency Experiment



(b) Bar Plot of Survival Analysis by Treatment

	n	events	median
Treatment=19	24	1	NA
Treatment=30	24	9	NA
Treatment=45	24	15	146

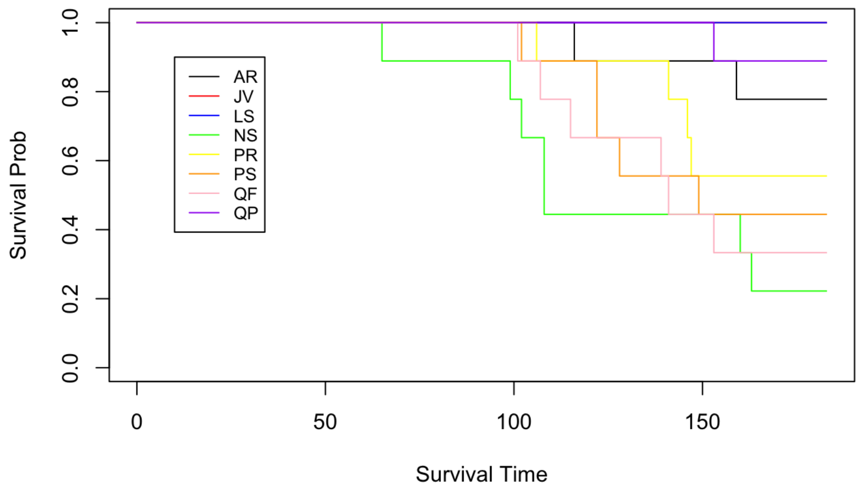
Table 5: Survival Analysis Given Treatment: Frequency Experiment

	19	30
30	0.011	-
45	$2.8 * 10^{-}5$	0.150

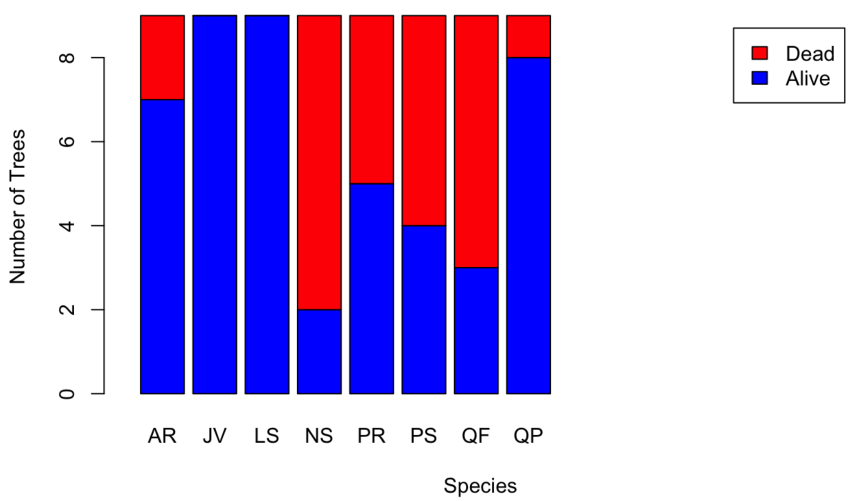
Table 6: Pairwise Comparison Using Log-Rank Test Given Treatment: Frequency Data

From the results of the survival analysis, it was found that one plant died given the salinity treatment once a month, nine plants died given the salinity treatment once a week, and fifteen plants died given the salinity treatment three days a week. In addition, plants given the salinity treatment three days a week were the only ones to reach the median threshold on the 146 day, meaning at least half of the plants from this treatment died at this point during the experiment. This tells us that as the frequency in which the salinity was administered increased, so did the number of plants that died. When running Pairwise Comparisons of the varying levels of frequency, the results indicate that there is a statistically significant difference when comparing the number of plants that died between each of the treatments, with the exception of the plants given the salinity treatment once a week in comparison to the plants given the salinity treatment three days a week, as this comparison produced $geq 0.05$. This indicates that there is not enough evidence to say that there is a significant difference in the number of plants that died between these groups.

The image and tables below display the results of the Survival Analysis in the Frequency experiment given the species of the plant:



(a) Image of Survival Analysis by Species, Frequency Experiment



(b) Bar Plot of Survival Analysis Given Species

	n	events	median
AR	9	2	NA
JV	9	0	NA
LS	9	0	NA
NS	9	7	108
PR	9	4	NA
PS	9	5	149
QF	9	6	141
QP	9	1	NA

Table 7: Survival Results Given Plant Species: Frequency Data

	AR	JV	LS	NS	PR	PS	QF
JV	1.000	-	-	-	-	-	-
LS	1.000	1.000	-	-	-	-	-
NS	0.428	0.022	0.022	-	-	-	-
PR	1.000	0.762	0.762	1.000	-	-	-
PS	1.000	0.283	0.283	1.000	1.000	-	-
QF	1.000	0.088	0.088	1.000	1.000	1.000	-
QP	1.000	1.000	1.000	0.115	1.000	1.000	0.310

Table 8: Pairwise Comparisons Using Log-Rank Test Given Plant Type: Frequency Data

From the results of the Survival Analysis, it was found that all the plants from species JV, and LS survived the length of the experiment, while some plants from species AR, NS, PR,PS, QF, and QP died throughout the experiment. In addition, NS, PS, and QF were the only species to reach the median thresholds during the experiment, meaning at least half of the plants from these species were dead by a certain point during the experiment, with that point being the day of the experiment listed in the . This indicates that all plants from Species JV, LS, and QP survived after receiving varying levels of salinity, while plants from species AR, NS, PR, PS, and QF died after receiving varying levels of salinity. When running Pairwise Comparison of the given species, it was found that the only statistically significant results were found when comparing species NS, which had the most plants die throughout the experiment, with species in which none of the plants died, including JV, LS, and QP, giving p-values less than $\alpha = 0.05$. The other comparisons had p-values higher than 0.05, indicating that there is not enough evidence to say that there is a significant difference in the number of plants that died within the species when compared to other species.

4 Conclusion

Using the output of the eight ANOVA's, out of the eight tree species, four of the them had a statistically significant difference in biomass by treatment. These four species included Black Cherry(PS), Southern Red Oak(QF), Black Gum(NS), and Pitch Pine(PR). The other four trees had smaller differences than the significant four. Using the Survival Analysis, as the salinity level increased, the probability of surviving decreased. The Black Gum(NS) species reached the fifty percent survival mark.

Using the output of the eight ANOVA's, out of the eight tree species, two of the plants had a statistically significant difference in biomass by treatment. These two species included Black Cherry(PS) and Southern Red Oak(QF). The other four trees had smaller differences than the significant two. Using the Survival Analysis, as the frequency level increased, the probability of surviving decreased. The Black Gum(NS), Black Cherry(PS), and Southern Red Oak(QF) species reached the fifty percent survival mark.

Using the analysis of both the salinity and the frequency experiment, Black Cherry(PS), Southern Red Oak(QF), and Black Gum(NS) will have the greatest impact from salt out of the eight total maritime species.

5 Personal Reflections

5.1 Abby

This semester, I was a lot more confident coming into MA 321 because of taking it last semester. For me personally, I learned that practice makes you feel a lot more confident. This semester I did a lot of comparing to how my first semester went taking this class and I saw many improvements. This included being less nervous to speak in front of the classroom and more confident in my statistics knowledge, where I could work more independently on the project rather than always as a team. I was very proud of this because then when I take this course again, I will be prepared to be in a group with students who might have less knowledge than I do. Statistically, I got to learn about Survival Analysis which was really cool. I enjoyed looking at the Kaplan-Meier Curve because I could really understand how the trees were reacting to the salt over time. I also learned that sometimes you can analyze a bunch of information just to get rid of it the next day and start from square one. In class, we always have a procedure of answering questions that always leads us to the answer. Learning that you might not always choose the right way to analyze information and have to start over was very beneficial since this type of project is meant to show us how projects in statistics work in the real world, not just another ordinary math class.

One of my favorite parts of this project was realizing why we can not predict initial biomass. Drawing the experiment on the board of how Dr. Daneshgar set up the experiment and realizing that at the beginning of the experiment, none of the trees had any salt given to them before the experiment. Therefore, how were we supposed to use the trees final biomass to predict the initial when they all had different treatments at the end. I was proud to figure that out because I felt that I had fully understood how the experiment was created to understand why it wasn't possible. I also found it interesting that Pitch Pine had a statistically significant difference in biomass by treatment in the salinity experiment since it was one of the two evergreen trees included experiment. I originally assumed that both of the evergreen trees would not result in a statistically significant result since they are more durable than conifer trees.

5.2 Kyle

This course provided me with unique, hands-on experiences that have made me both a better statistician and a better team player. I was able to learn the ins and outs of data collection regarding specifically from biology perspective, including some practical restrictions that come with designing an experiment in this field. For instance, when working with Dr. Daneshgar, we mentioned that if he were to run this experiment again, he should plant each species without any salinity treatment, and then pull measure them in the beginning of the experiment, that way he would have an initial biomass for each of the plant species to act as a comparison. He mentioned that he actually did consider doing that, but because it cost so much money for each plant, he decided to opt out of it. This made me realize that when designing an effective experiment, you need to also consider practical restrictions like budgeting. As for the actual research topic, I learned that increasing levels of salinity have a negative effect on the growth of most plants, as a majority of the plants in our experiment had a decrease in biomass as salinity exposure increased. What I found to be the most interesting, however, is that some plants actually benefit from increased exposure to salinity, as the Eastern Red Cedar had an increase in biomass with increased exposure to salinity. I was initially under the impression that any exposure would have negative effects on the growth of the plant, so it was interesting to learn that this is not always the case. This class also gave me the opportunity to reinforce my ANOVA and survival analysis skills, as these were topics already covered in previous courses. As for what I learned about myself, I learned that I know more about statistics than I originally thought I did. Even though this was my second time taking this course and I was not as worried as I was the first time, I still get in my head and second guess myself, worrying that I am making the wrong decisions when it comes to analyzing data. However, when working on this project, I would think to myself "oh we did something similar to that in a previous course". Being able to see how my previous courses came together to create ideas and solutions to completing our project was an amazing feeling and helped me realize that I do know more than I think I do, and just need to be more confident in myself.

One of the things I enjoyed about this course was getting to work hands on with a client and walk through the process of how they made their experiment. Even though this is not the first time I have take this course, it is the first time that I actually got to meet with the client in person on a day to day bases since the pandemic was dying down. As we sat down to meet with Dr. Daneshgar every week, he would walk us through the idea that he was going for in his experiment, and based on his responses, would come to a decision on how to move forward with our analysis. I always liked this part of the project because it let me work on both my communication skills by conveying our results to Dr. D in a comprehensible manor, and my decision making skills of listening to what Dr. D wanted in his results and coming up with a solution. In addition, working with Dr D was very enjoyable: meeting with him every week was one of the highlights of the semester, and I will miss working with him. As for what I did not enjoy, I was not a fan of the emotional rollercoaster that our group went through. We were initially a month behind schedule since we made a mistake with our analysis, but than we somehow went to being a month ahead of schedule. This constant uncertainty about where we were in terms of completing the project made me feel uneasy at times, but fortunately, we were able to finish all of our work on schedule.

In order to make this course a little bit better, I would suggest adding more RStudio notes at the beginning of the semester for students who haven't really familiarized themselves with R yet. I would definitely keep the ice breaker questions/ introductions at the beginning of the semester, since it gives the Stats majors an opportunity to introduce themselves to any non stats majors, and vice versa. All in all though, I would not want this course to change really: this is often my most favorite class because the experiences you make are so unique, and I woudn't want to change anything that would take away from said uniqueness.

This course definitely helped me decide about a career and a future revolving around the main attributes that go into consulting: communication, teamwork, and decision-making concerning the approach to statistical analysis. I really enjoyed working with survival analysis again: I always found it interesting how it shows the percent chance of an event happening, as it allows statisticians to essentially predict the future using data. This leaned me in the direction of choosing a career that would allow me to work with survival analysis more often, such as actuarial work or consulting. This also prepared me for a future career as it gave me some insight as to what is expected in occupations that require consulting: you need to be able to work well with others, you have to listen to what your client wants, and you have to be able to think outside of the box. I would never have gotten the opportunity to prepare myself for the outside world without this course, and am thankful to have been given the opportunity to take it more than once.

It's rather difficult for me to compare this course to any other that I have taken here at Monmouth, because consulting is such a unique experience. The only class that comes to mind as being similar is the bio-stat/ MA 327 course that was taught by both Dr. D and Dr. B. The only difference between that course and this one is that we got to create the experiment. I have always had an enjoyable time in this course and have found it to be more beneficial than most of the classes I have taken at Monmouth. This course allows students to come to their own decisions when working with a client, while also being able to ask questions in a safe place outside of the real world. Therefore, as I have said before, there is nothing I would change about this class, and I wish other courses were taught similarly to this one, with actual, real life application.

5.3 Odalys

Every time I take MA321 I learn something new. This semester I learned that we cannot predict initial biomass. I learned that it is okay to question your client's judgment. I learned that we should encourage each other to critique each other's work because that it how we find errors. Finally, I learned that no matter how perfect we believe the experiment was planned out we will always find something wrong. I don't think there exists a perfect experiment. I really believed this experiment was planned out and every possible thing that could go wrong would not go wrong. I beloved this was true until I went through the data and find errors. I liked that I took MA327 last year because I could visualize and understand what the people who designed and performed this experiment went through. I thought this project was relatable and honestly an easy concept to understand. This semester I felt the most comfortable with my topic and with my client. I was able to come out of my shell and talk confidently about my statistical knowledge.

I really enjoyed working with Dr Daneshgar and his project. Dr. Daneshgar is an awesome client and because he is a great client, I really wanted to make this project work. I wanted to find the things he hypothesized would happen. In a way, we did find the things he was looking for but at the same time we did not. Working with real data is hard and it never works out perfectly. This project is a great example of that. While we had great findings a lot of what we found was inconclusive and confusing. I think that can be frustrating when we can't say something is different with confidence because we have A, B, and C that occurred. That is real data!

MA 321 is awesome. I have taken this course 4 times and I come back because I learn something new every time. I find this class very useful for my presenting skills, for my statistical skills, and this class has help build my resume. I have had a couple people tell me I'm a good presenter and I get very surprised when people say that. I don't give myself enough credit for the amount of improvement I've made since I took this class the first time. I get very happy when people say I am good presenter because I know it's because of MA321. I honestly wouldn't change anything about this course. I have only positive things to say, and I would take it again.