

Using Statistical Methods to Determine the Effects of Oysters on NC Coast Plants

University of North Carolina at Chapel Hill

Statistics & Operations Research

Client: Lauren Glaze

Yifei Wu PID: 730265525

Abstract

This report aims at investigating how three oyster treatments and different sites affect the growth of a certain type of NC coast plant. According the analysis, different treatments and sites have significant impacts on the stalks' length, but more metrics need to be examined in order to explain more information in the data; Besides, stalks under Live Oyster condition shows a greater probability to have flowers than the other two treatments. To reach the conclusion, this study majorly implemented mixed effect nested ANOVA, nested logistic regression and corrected the result of the logistic model by bootstrap.

Contents

Abstract.....	1
1. Introduction.....	3
2. Data	3
2.1 Experimental design.....	3
2.2 Data Preprocessing.....	4
3. Analysis.....	4
3.1 Shoot_length as response.....	4
3.1.1 Exploratory Data Analysis	4
3.1.2 Mixed Effect Nested ANOVA	7
3.1.3 Model fitting.....	8
3.1.4 Model Evaluation	9
3.1.5 Results.....	10
3.2 Flowering or not as response	13
3.2.1 Multileveled nested logistic regression	13
3.2.2 Model fitting	14
3.2.3 Results.....	14
3.2.4 Prediction	16
4. Conclusion	17
References.....	19

1. Introduction

This study was designed to assist Lauren Glaze, an undergraduate student from UNC Environmental Science Department, with her recent project. The project interested in whether different oysters' treatment affect the stalks' length and flowering of plants from NC coast. The treatments were "Live Oysters", "Oysters Shells" and "No Oysters" and nine sites were randomly selected within each of three treatments. For each treatment, there were two responses set to be investigated, individual stalk's length and whether flowering or not. According to these responses, nested ANOVA and nested logistic were applied to tackle the problem.

The remainder of this report is structured as follows. Section 2 describes the data used in this experiment, including experimental design and preprocessing. Section 3 introduces the analyses performed in terms of different independent variables, and section 4 concludes the final results.

2. Data

2.1 Experimental design

As shown in figure 1 below, the experiment was designed in which it was intended to measure a response ('average stalks length', 'number of stalks' and 'flowering stalk or not') to one of treatments (three levels: 'live oysters (LO)', 'oyster shells (OS)' and 'no pot (NO)'). Initially, seeds were uniformly planted in each of the three treatment, and after a period, nine sites were randomly picked within each treatment, and living stalks were recorded grouped by each site. It should be noted that "opa" means "out of planting area", which was also regarded as a site.

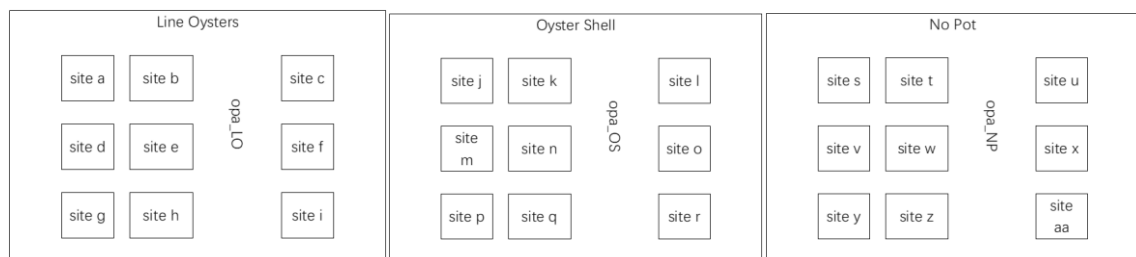


Figure 1. Experimental design

There are 999 observations in total, and for each observation (say stalk), there are 5 features. 2 of them are regarded as the predictors and the rests are responses. Table 1 and Table 2 show the descriptions of 2 predictors and 3 responses. The three responses in this study are "stalk_length", "flowering" and "flowering_number".

Table 1. Description of the predictors

Predictors	Description
treatment	3 different oysters' conditions
site	10 randomly picked sites (including opa) within each treatment, 30 sites in total

Table 2. Description of responses

Responses	Description
stalk_length	The length of each individual stalk
flowering	0 represents this stalk has no flower and 1 indicates it's a flowering stalk
Flowering_number	Number of flowering stalks in that site

2.2 Data Preprocessing

Data cleaning and preprocessing were applied in the project at the first place. Since the study aims at living stalks, those with shoot length zero were not initially recorded. Before actual data analysis, exploratory data analysis was used in the beginning of each section to help select the regression methods. For section 3.1, the normality and homogeneity of variance were also tested to satisfy the assumptions of linear regression models, and Box-Cox transformation was put into practice.

3. Analysis

3.1 Shoot_length as response

3.1.1 Exploratory Data Analysis

The main concern of this section is to investigate how treatment and site influence the variable “shoot_length”. Therefore, an intuitive way to tackle the problem is to use Analysis of Variance (ANOVA). Before jumping into the real analysis, some exploratory data analyses were presented. Figure 2 below shows the boxplots of “shoot_length” in each of the 30 sites. From the plot, the variances of each sites seem to be quite uniform, due to the approximately even length of each box, indicating the variation of data in each site have no obvious difference. However, there are some outliers, which are represented by the dots in the picture. Since the effects of the outliers would not be revealed until the model is fitted, a deeper look into the outliers will be discussed in section 3.1.3.

Intuitively, one would use two-way ANOVA to exam what influences two predictors have on one response. Nevertheless, in this case, prior experience suggested that the sites sampled under each treatment were very uneven and patchy with respect to the

response, it was hence unlikely that these single units would adequately represent the populations. A nested design could be thought of as a hierarchical arrangement of factors whereby a treatment is progressively sub-replicated (Keough, 2002). Therefore, “treatment” is the main fixed effect, and “site” is nested within “treatment”, which should also be treated as a random effect variable. More details about Mixed Effect Nested ANOVA will be revealed in next section.

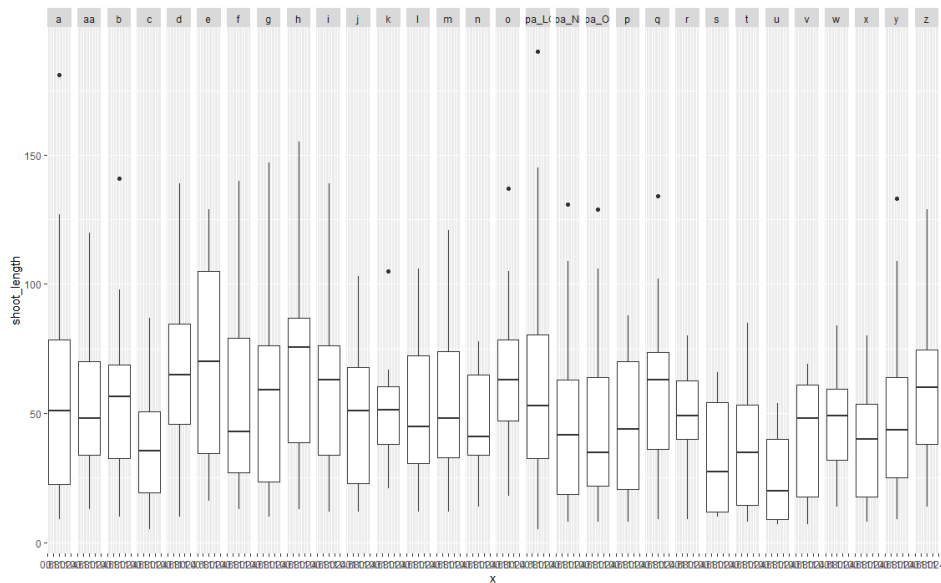


Figure 2. Boxplots of 30 sites

Normality and homogeneity of variance of residuals within each treatment are two essential assumptions for ANOVA. In the nested design the observations are collected at a sub-replicate rather than the replicate level. That is, the actual replicates of the main treatment effect are the sites not individual stalks. Therefore, when assessing assumptions pertaining to the residuals, it is necessary to first aggregate the observations to the sites. Here, the average shoot length of stalks within each site was used as the replicates for the sites.

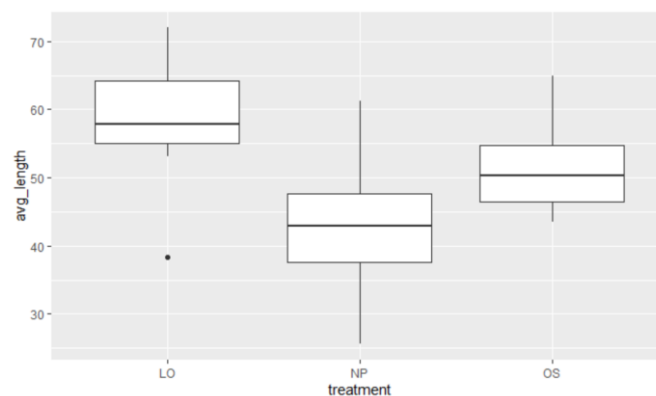


Figure 3. Boxplots of average shoot length under each treatment

No obvious violations of non-normality or homogeneity of variance are shown in figure 3. Note that assessing normality can be a little difficult from such small numbers

of replicates (9 sites per treatment), the distributions of individual stalks are also critical, which needed to be examined as well.

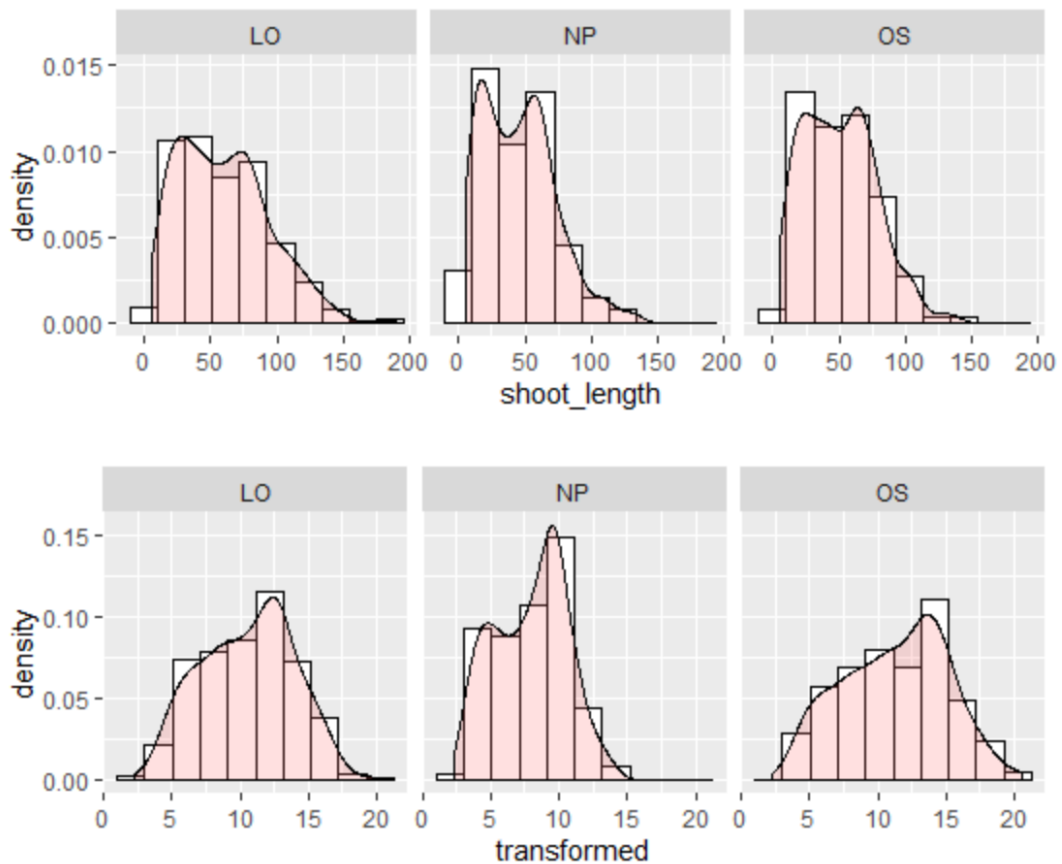


Figure 4. Raw data and transformed data

A Box Cox transformation is a way to transform non-normal dependent variables into a normal shape. Lambda (λ) is a core exponent of Box-Cox transformation, the optimal value of which results in the best approximation of a normal curve.

The top 3 plots of figure 4 shows the distribution of the raw “shoot_length” data grouped by treatments, and the bottom 3 represents the data after Box-Cox transformation. The raw data were far from normally distributed from the plots, and the Box-Cox transformation was hence introduced. The package in R showed that the three optimal λ s were 0.41, 0.45 and 0.43. After the transformation, the right-skewed trends were weakened, and the data were more visually normal.

Table 3. P-values of normality tests results

		Shapiro-Wilk test	Lilliefors test
LO	Raw data	1.635e-6	3.304e-6
	Transformed data	0.0548	0.102
NP	Raw data	0.424e-8	2.583e-8
	Transformed data	0.0125	0.0327
OS	Raw data	2.121e-5	3.315e-5
	Transformed data	0.446	0.290

Shapiro-Wilk and Lilliefors test are the most common ones for examining normality. The null hypothesis of the test was that the data subjects to normal distribution. The significant level used here was 0.01. Therefore, according table 3, p-values improved significantly after the Box-Cox transformation, and they are all larger than 0.01, which means that one can accept the null hypothesis under 0.01 significance. The rest part of the ANOVA section used the transformed data.

Also, the distribution of data in subgroups (sites) is worthwhile to be concerned in advance. Figure 5 demonstrates the transformed data in terms of sites. The distributions are approximate normal, and outliers will be diagnosed after model fitting.

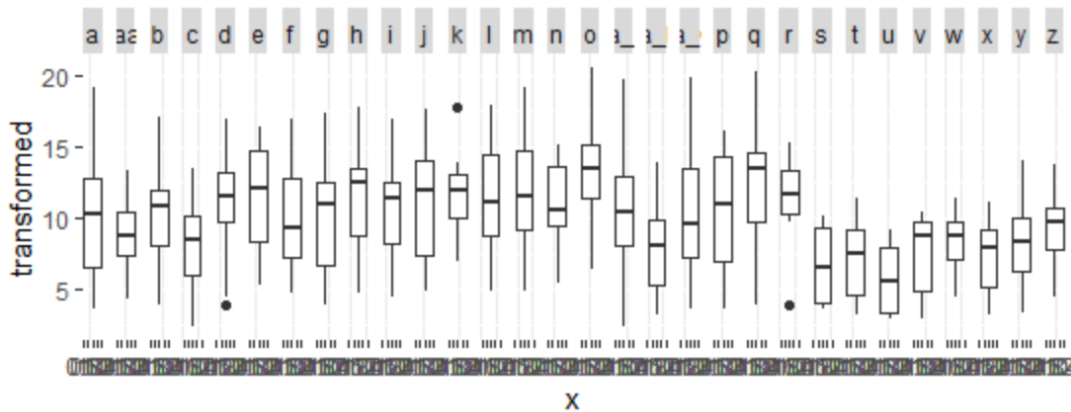


Figure 5. Transformed data of 30 sites

3.1.2 Mixed Effect Nested ANOVA

In this case, ANOVA allows one to determine whether the differences between the shoots' lengths are simply due to sampling errors or whether there are systematic treatment effects that causes the mean in one group to differ from the mean in another.

As mentioned in last section, the predictor "site" is nested within the predictor "treatment". Namely, sites picked under each treatment are different, which makes the normal two-way ANOVA invalid. For a nested ANOVA, 2 F statistics will be calculated to compare the mean among groups at each level. In addition to testing the equality of the means at each level, a nested ANOVA also partitions the variance into different levels.

Mixed effect model provides an approach to solve nested ANOVA. Although the term “mixed-effects” can be used to refer to any design that incorporates both fixed and random predictors, its use is more commonly restricted to designs in which factors are nested or grouped within other factors (sites nested within treatments). Furthermore, rather than basing parameter estimations on observed and expected mean squares, mixed-effects models estimate parameters via maximum likelihood (ML) or residual maximum likelihood (REML). In so doing, mixed-effects models more appropriately handle estimation of parameters, effects and variance components of unbalanced designs (Nakagawa, 2013).

A nested ANOVA has one null hypothesis for each level. In a two-level nested ANOVA, one null hypothesis is that the groups have the same mean. In this study, this null would be that the average shoot length among treatments are the same, and sites have the same means as well.

3.1.3 Model fitting

The models were fitted in R by using “nlme” package. Rather than fitting just a single, simple random intercepts model, it is common to fit other related alternative models and explore which model fits the data best. In this study, a random intercepts and slope model was also fitted as comparison. The random effect only affects the intercept of the first model but has impact on both the intercept and slope of the second model.

The mixed effect models use residual maximum likelihood method to estimate the parameters, so the test comparing two models should be likelihood ratio test, more specifically REML, which assesses the goodness of fit of two competing statistical models based on the ratio of their likelihoods. Table 4 shows the results of likelihood test ratio, the corresponding p value, and AIC/BIC criterion.

Table 4. Likelihood ratio test of two mixed effect models

	AIC	BIC	L.ratio	P value
Simple random intercept	5247.576	5272.095		
Intercept/slope	5256.894	5305.931	0.6820335	0.9839

AIC and BIC are criterions majorly used to compare among models.

The AIC and BIC values are quite similar, which were not informative. The null hypothesis of the test is that the two likelihood scores have no differences. According to the relatively large p value, the null hypothesis was accepted, indicating the more complex random intercepts and slopes model did not fit the data significantly better than the simpler random intercepts model and thus the latter model would be used.

3.1.4 Model Evaluation

3.1.4.1 Residuals

As always, exploring the residuals can reveal issues of heteroscedasticity, non-linearity and violations of basic assumptions. Figure 6 shown below are three plots about residuals of the model. A is a Q-Q plot of the residual, in which the straight line is generated by the theoretical normally distributed data. Since the dots were approximately plotted along the line, there was no obvious violation of normality; B is a visualization of the distribution of residual as well, the blue curve is the theoretical normal curve, and the result is quite similar to A; For C, is a plot examining homoscedasticity of variance, the horizontal blue line and even spreading of dots scattered around the line indicates that variance of residuals is constant. Because there were no apparent patterns of the residuals, the variance could be treated as constant.

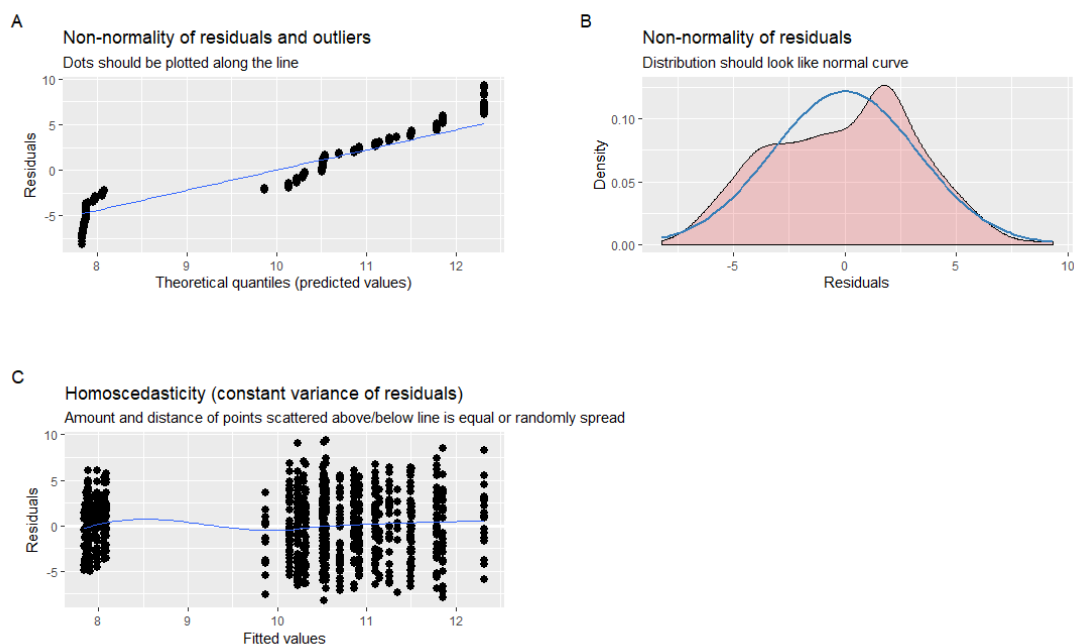


Figure 6. Residuals evaluation

3.1.4.2 Influential points

However, figure 6 didn't show the outliers, which could cause serious deviation of the prediction. Residuals vs Leverage plot was introduced as below. This plot helps researchers to find influential cases if any. Not all outliers are influential in linear regression analysis. Even though data have extreme values, they might not be influential to determine a regression line. That means, the results wouldn't be much different if they are either included or excluded from analysis.

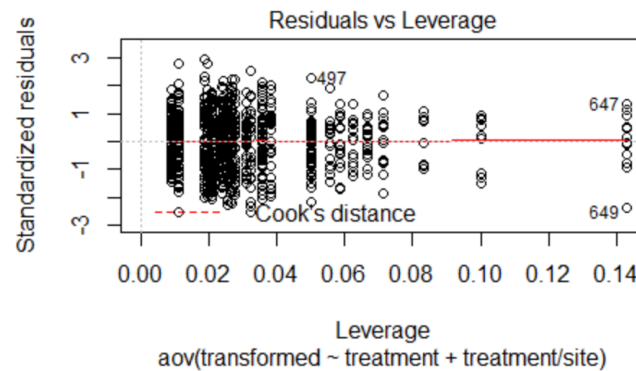


Figure 7. Residuals vs leverage

One watches out for outlying values at the upper right corner or at the lower right corner (like 647 and 649). Those spots are the places where cases can be influential against a regression line. One should also focus on cases outside of a dashed line, threshold of Cook's distance. When cases are outside of the threshold, the cases are influential to the regression results. The regression results will be altered if excluded.

According to figure 7, the dashed line is barely seen in the plot, because all points were well within the threshold. This result shows that there were no influential points and the fitted model was accepted.

3.1.5 Results

3.1.5.1 Parameter estimation

Tables below represent the results of the nested ANOVA model, including estimated values of parameters and the p values.

Table 5. Results of random effect

Groups	Variance	Standard deviation
sites	22.57	4.751
Residual	905.97	30.099

Table 6. Results of fixed effect

	Estimated value	P value
Intercept	59.453	3.15e-14
treatmentNP	-14.731	0.000305
treatmentOS	-8.446	0.001169

Table 5 shows the variance of random effect. The variance and standard deviation of random effect "sites" were obviously larger than 0, which suggests that "sites" indeed accounted for a substantial amount of variability in the data. Next section will show how much variation it contributes to.

Table 6 is the parameter estimation of the fixed effect “treatment”. To interpret the table, all the p values were significantly smaller than 0.01, indicating that all three levels of “treatment” were statistically significant in the model under 0.01 level. That means, one has 99% confidence to say that all three treatments do have influence on the stalks’ length. The Bonferroni adjusted p values showed the same result.

To be more specific, the stalks’ length of level NP and OS were both significantly lower than level LO, due to the negative estimated values. It can also be shown by an effect plot in figure 8.

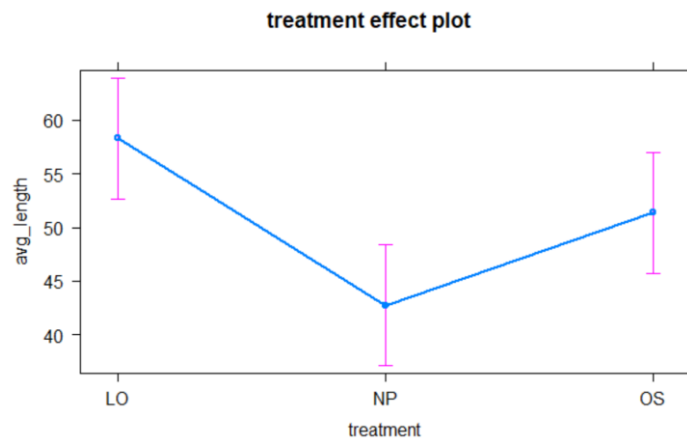


Figure 8. Treatment effect plot

3.1.5.2 Likelihood ratio test

As mentioned before, likelihood ratio test was used to compare two models. In this case, the overall significance of “treatment” was examined by comparing the model with and without the fixed effect. To be note, as this was contrasting a fixed component, the models needed to be fitted with ML rather than REML.

Table 7. Likelihood ratio test of treatment

	AIC	BIC	P value
Without treatment	9673.2	9687.9	
With treatment	9660.9	9685.5	0.0000294

Clearly, the p value was significantly small, indicating that the “treatment” factor was statistically significant in the model.

3.1.5.3 Pairwise comparison

Table 6 only shows the case where treatment NP and treatment OS were compared with treatment LO respectively, the client was interested in the comparisons among each pair of the treatments as well. Tukey multiple comparison was hence applied as Table 8, and 95% confidence intervals were presented by figure 9.

Table 8. Tukey pairwise comparison

	p-value
NP-LO	0.00000
OS-LO	0.08365
OS-NP	0.07564

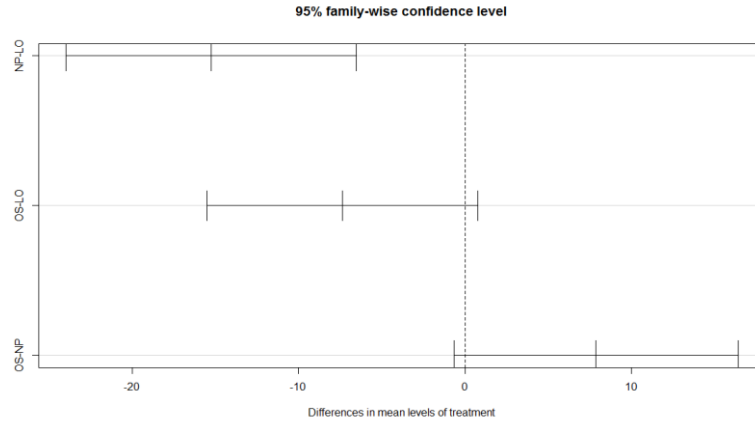


Figure 9. Confidence intervals of differences among groups

Since the p value of NP-LO is strictly smaller than 0.05, the average stalks' length of NP and LO have significant difference, while no obvious difference among the other two pairs. From the 95% confidence intervals (figure 9), 0 was included in the intervals of OS-LO and OS-NP, indicating the same conclusion.

3.1.5.4 R^2 approximation

R^2 is a popular goodness of fit metric in simple linear models, but its use is rarely extended to (generalized) linear mixed effects models.

The journal 'A general and simple method for obtaining R^2 from generalized linear mixed - effects models' discusses the issues associated with R^2 calculations and suggests a series of simple calculations to yield sensible R^2 values from mixed effects models.

An R^2 value quantifies the proportion of variance explained by a model (or by terms in a model) - the higher the value, the better the model (or term) fit. There are two R^2 for mixed effects models (Nakagawa, 2013):

- Marginal R^2 - the proportion of total variance explained by the fixed effects.

$$\text{Marginal } R^2 = \frac{\sigma_f^2}{\sigma_f^2 + \sum_l^z \sigma_l^2 + \sigma_d^2 + \sigma_e^2}$$

where σ_f^2 is the variance of the fitted values, $\sum_l^z \sigma_l^2$ is the sum of the z random effects (including the residuals) and σ_d^2 and σ_e^2 are additional variance components appropriate when using non-Gaussian distributions.

- Conditional R^2 - the proportion of the total variance collectively explained by

the fixed and random factors.

$$\text{Conditional } R^2 = \frac{\sigma_f^2 + \sum_l^z \sigma_l^2}{\sigma_f^2 + \sum_l^z \sigma_l^2 + \sigma_d^2 + \sigma_e^2}$$

Since the linear mixed effect model was concerned in this study, σ_d^2 and σ_e^2 could be ignored due to the normality of the residuals.

Table 9 shows the marginal and conditional R^2 of the model fitted in section 3.1.3.

Table 9. R^2 approximation of the model

	Marginal R^2	Conditional R^2
Fitted model	0.15628	0.18315

The fixed effect “treatment” accounted for approximately 15.63% of the total variation in stalks’ length. The random effect of “sites” accounted for approximately 2.7% of the total variation in stalks’ length and collectively, the hierarchical level of sites (containing the fixed effect) explained approximately 18.32% of the total variation in stalk’s length.

Although the R^2 was relatively small, the “treatment” and “sites” were significant in the model. These two factors did have significant effects on stalks’ length, but there were supposed to be other contributors that were not included in the model influencing the response, since “treatment” and “sites” could not explain the most variation in the data.

3.2 Flowering or not as response

In this part, the client was mainly interested in whether different treatments have significant influence on the flowering of stalks or not. The response is a binary variable, where 1 represents for the stalk having flowers and 0 stands for the ones do not. A multileveled Logistic Regression model was fitted, since individual stalks were nested within sites, and sites were nested within three treatments.

3.2.1 Multileveled nested logistic regression

Logistic regression is often used when the dependent variable is categorical. For this study, the independent variable is whether the stalk has flower or not, hence binary logistic regression was introduced.

The central mathematical concept that underlies logistic regression is the logit—the natural logarithm of an odds ratio, which is defined as

$$\text{logit}(p) = \log \left(\frac{P(y_i=1)}{1-P(y_i=1)} \right) = \beta_0 + \sum_{j=1}^k \beta_j x_{ij},$$

who turns a non-linear problem into a linear regression model.

Therefore, the three-level nested logistic model is written as

$$y_{ijk} = \beta_0 + \beta_1 x_{1ijk} + \beta_2 x_{2jk} + \beta_3 x_{3k} + v_k + u_{jk} + e_{ijk}$$

$$v_k \sim N(0, \sigma_v^2)$$

$$u_{jk} \sim N(0, \sigma_u^2)$$

$$e_{ijk} \sim N(0, \sigma_e^2)$$

where $\beta_0 + \beta_1 x_{1ijk} + \beta_2 x_{2jk} + \beta_3 x_{3k}$ is termed the fixed part of the model and the rest is termed the random part of the model. In this case, fixed part only contained “treatment”, and random part included “site” and “stalk”. The fixed part of the model specifies the overall mean relationship between the flowering and the predictor “treatment”; The random part of the model specifies how the stalks and sites specific relationships differ from this overall mean relationship.

3.2.2 Model fitting

The model was fitted by using “glmer” function in R. One good thing about glmer is that there is no need to specify whether the groups are nested or cross classified, R can figure it out based on the data.

Like in section 3.1.3, a simple random intercept logistic regression model and a model with random intercept and slope were fitted. They were compared based on likelihood ratio test, which is shown in table 10.

Table 10. Likelihood ratio test for two models

	AIC	BIC	P value
Simple random intercept	622.2	641.8	
Intercept/slope	631.9	676.1	0.9983

According the result, the p value is relatively large, indicating that the more complicated intercept and slope model didn’t make the fitting significantly better. Therefore, the simple random intercept was accepted.

3.2.3 Results

3.2.3.1 Laplace approach

The model was firstly fitted by the default Laplace approximation in R

Table 11. Results of random effect

Groups	Variance	Standard deviation
sites	0.06712	0.2591

Table 12. Results of fixed effect

	Estimated value	P value
Intercept	-1.7794	<2e-16
treatmentNP	-1.1849	0.000432
treatmentOS	-0.8774	0.003631

For fixed effect, the p values for both “NP” and “OS” treatment were rather small than 0.01, which means that these two treatments were statistically significant in the model under 0.01 level. For random effect, the standard deviation was obviously different from 0. Therefore, “site” did have significant effects on whether the stalk has flowers or not, but affected less than it did to “stalks’ length”.

And to interpret the coefficients, the coefficient of treatmentNP was -1.1894, which means that, when site remains the same, stalks under treatment LO is more likely to have flowering stalks than those under NP. Specifically, $e^{-1.1894}=0.306$ times compared with LO. These can be more straightforward by the following plot.

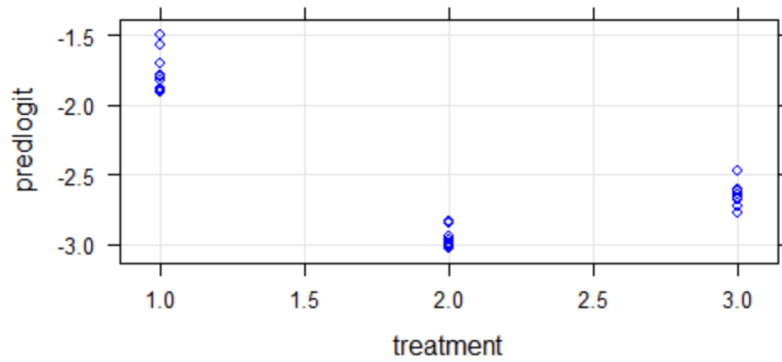


Figure 10. Scatter plot of predicted logit

Figure 10 shows the predicted logit against treatment plot. Each blue point represents for an individual stalk. For example, for a stalk planted under Live Oyster condition, the log-odds of flowering ranges from about -1.95 to -1.5 depending on which site it was planted. This translates to a range in probabilities of $\frac{e^{-1.95}}{1+e^{-1.95}} = 0.125$ to $\frac{e^{-1.5}}{1+e^{-1.5}} = 0.182$, so there were little site effects.

3.2.3.2 Bootstrap for parameter estimation

To assess the model, parametric tests rely on assumptions about the underlying distributions of the residuals to be normally distributed. However, the Shapiro-Wilk test showed that the residuals were far from a normal distribution. In this case, a bootstrap approach was introduced.

Bootstrap is used to enable inference on the statistic of interest when the true distribution of this statistic is unknown. In R, “bootMer” function provides users an easy way to apply bootstrap method in mixed effect models.

The bootstrap estimated values were based on 100 times replications and took the means of 200 samples. Table 13 shows the estimated value of fixed effects and variance of random effect, compared with bootstrap means of each variable.

Table 13. A comparison between Laplace and Bootstrap

	Estimated value	Boot mean
Intercept	-1.7794	-1.002
treatmentNP	-1.1849	-1.2403
treatmentOS	-0.8774	-0.6256
Variance of site	0.067	0.263

From the table 13, there were not too much improvement after applying bootstrap. Therefore, treatment had significant impact on the flowering of stalks, while sites only had little influence.

3.2.4 Prediction

3.2.4.1 ROC curve

2 Receiver Operating Characteristic (ROC) curves for predicting the nested logistic regression model is plotted below. The Area under the ROC curve (AUC) is a widely used measure of accuracy for classification models. Its meaning can be interpreted as follows:

- When $AUC=1.00$, the model assigns all observation to their true class with perfect accuracy.
- When $AUC=0.50$, the ROC curve is equivalent to the 45-degree line. It indicates the model is as accurate as guessing at random.
- When $AUC<0.50$, the model accuracy is worse than guessing at random.

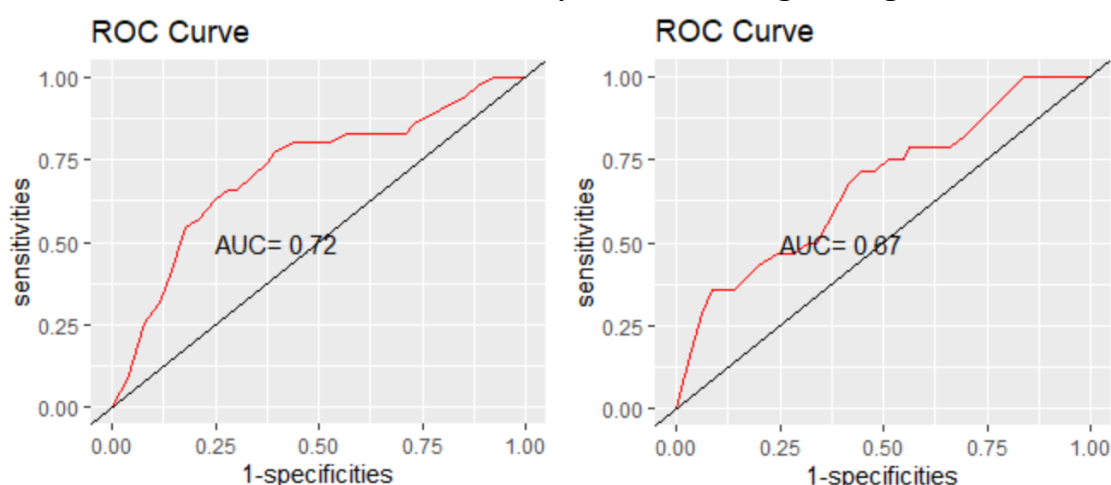


Figure 11. ROC curves of original data

The left panel of figure 11 is the AUC calculated from full original data set (apparent “in-sample” AUC), the right one is an adjusted AUC, which was derived from splitting the data set into training set and validation set. It is not surprisingly that AUC from full data was relatively high, since the predictions were made from the same

dataset on which the model was fitted. The adjusted one seemed to be more reasonable, but it only used a subset of the whole information. Another approach introduced by ‘Regression Modeling Strategies’ (Harrell, 1996) is ‘bootstrap for optimism’.

3.2.4.2 Bootstrap for optimism

Bootstrap approach for AUC correction is as follow:

- Step 1: Fit model to original data, and estimate AUC using original data based on fitted model. Denote this estimate of AUC as AUC_{app} (apparent “in-sample AUC”).
- Step 2: For $b = 1, \dots, 100$:
 1. take a bootstrap (with replacement) sample from the original data
 2. fit the model to the bootstrap dataset, and estimate AUC using this fitted model and this bootstrap dataset. Denote the estimate by $AUC_{b,boot}$ (second column in table 14).
 3. Estimate AUC by applying the fitted model from the bootstrap dataset to the original dataset. Let $AUC_{b,origin}$ (first column in table 14) denote the estimate
- Step 3: Calculate $O = \frac{1}{100} \sum_{b=1}^{100} (AUC_{b,boot} - AUC_{b,origin})$.
- Step 4: Calculate the optimism adjusted measure of AUC as $AUC_{app} - O$.

Table 14. Bootstrap AUC

	$AUC_{b,origin}$	$AUC_{b,boot}$	Optimism
Bootstrap sample 1	0.695	0.718	0.023
Bootstrap sample 2	0.707	0.724	0.017
...			
Bootstrap sample 100	0.712	0.722	0.010
Average	0.704	0.719	0.015

“Optimism” is the amount by which the apparent “in-sample” AUC overestimates the true AUC. The overall estimate of optimism is the average of the differences between the first and second columns of table 14, a value of 0.015 in this example.

Once an estimate of optimism was obtained, it was subtracted from the apparent AUC to obtain an improved estimate of prediction accuracy. Here the bootstrap corrected AUC is $0.72 - 0.015 = 0.705$.

4. Conclusion

In this report, mixed effect nested ANOVA and nested logistic regression were performed, in order to examine how different treatments and sites influence the growth of a certain type of plant. To recap, there are following conclusions:

The effects of treatments and site over stalk’s length were statistically significant under

0.05 level, and plants under treatment “Live Oyster” tend to grow taller than the other two treatments. However, treatment and sites could only account for little amount of variation in data. More features were needed to be included in the model; In terms of the flowering of stalks, treatments had significant impacts on whether the stalks have flowers or not, while “site” didn’t contribute considerably. To be more specific, plants under “Live Oyster” had more flowers than the other two treatments, indicating “Live Oyster” condition significantly influenced the both metrics of the recorded data.

References

- Agresti, A. (2013). *Categorical Data Analysis (3rd Ed.)*.
- Cameron, A. C. (2009). *Advances in Count Data Regression Talk for the Applied Statistics Workshop*.
- Cameron, T. (2009). *Microeconometrics : Methods and Applications*.
- Efron, B. T. (1993). *An Introduction to the Bootstrap*.
- Harrell, F. (1996). *Regression Modeling Strategies*.
- Kabacoff, R. I. (2010). *R in Action*.
- Lu Ma, X. Y. (2014). A Quasi-Poisson Approach on Modeling Accident Hazard Index for Urban Road Segments.
- Nakagawa, S. a. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects.
- Nitipong Homwong, A. D. (2016). Three-Level Mixed-Effects Logistic Regression Analysis Reveals Complex Epidemiology of Swine Rotaviruses in Diagnostic Samples from North America.