

**INVESTIGATING THE IRONWOOD TREE (*CASUARINA  
EQUISETIFOLIA*) DECLINE ON GUAM USING APPLIED  
MULTINOMIAL MODELING**

A Thesis

Submitted to the Graduate Faculty of the  
Louisiana State University and  
Agricultural and Mechanical College  
in partial fulfillment of the  
requirements for the degree of  
Master of Applied Statistics

in

The Department of Experimental Statistics

by

Karl Anthony Schlub  
B.S., University of Utah, 2007  
December 2010

## Acknowledgments

I would like to give my sincere thanks to my extended family: Richard and Judy Schlub for their support during the writing process, to my grandparents Carl and Arvella Schlub who have made contributions towards my education, and to my parents Robert and Joanne Schlub.

I give thanks to my major professor Dr. Brian Marx for motivating me to take on the research my father has done on the ironwood tree. I would also like to thank Dr. Bin Li and Dr. Jing Wang for their time and advice given in regards to my thesis.

Lastly I want to thank the faculty at the Department of Experimental Statistics for helping me in the initial moving in process when I first arrived to Baton Rouge.

## Preface

Guam is a small island in a remote part of the Pacific Ocean known as Micronesia. It is located 13°28' north and 144°45' east. Guam's location is 1500 miles from Tokyo, Japan, and close to 5800 miles from California, and 3200 miles from Sydney, Australia. The island is located on the west side of the International Date Line which places the time of day one day ahead of the United States; it is also in the same time zone as Sydney, Australia.

This island was formed by a large volcano that grew out of the deepest part of the ocean. Over millions of years tectonic plates at the bottom of the ocean shifted forcing the island to rise upward above the sea level. Guam today consists of two distinct halves. The first half, which is now the southern portion of Guam, was formed by the volcano and now consists of narrow valleys and steep hills. The other half, the northern part of the island, was at one time a large coral reef and is now a smooth plateau. Because Guam is surrounded by a natural coral reef, there are many nice beaches along the shoreline.

The native people of Guam are known as Chamorros. The 2000 U.S.A. census estimated Guam's population to be 154,805. Basic racial demographics would be 37% Chamorro, 26% Filipino, 6% Caucasian; the remaining miscellaneous island or Asian ethnicities are not listed. Guam's main language is English. There is an officially recognized local language known as Chamorro. Modern day Chamorro adopts some of its vocabulary from the Spanish and English languages; this is primarily a result of being an occupied land by both Spain and the United States. Guam is now a territory of the United States government and citizens born on Guam are granted United States citizenship. Guam has an established local government consisting of a governor, a legislature, and a supreme court.

The primary industry of Guam is tourism. The tropical weather and its close proximity to Asia makes Guam an ideal weekend escape for people arriving from Japan and South Korea. Outdoor recreation such as scuba diving, fishing, and swimming are some of the many activities of tourists and Guamanians.

Guam's native wild life consists primarily of birds and marine life. The official bird of Guam, the Guam Rail, is a flightless animal that feeds off small insects and seeds. Guam has a species of bird, the Marianas Crow, which resembles its North American relative. The reefs and ocean around Guam contain a variety of marine life from colorful fish to dolphins and sharks.

Guam hosts many invasive species. The most infamous example of an introduced species is the Brown Tree Snake. Arriving possibly from a cargo ship, the Brown Tree Snake has devastated many of the native bird populations on Guam. For instance the Micronesian Kingfisher no longer exists in the wild primarily due to the snake. The rhinoceros beetle, which was introduced to the island in the last decade, could have a profound effect on Guam's coconut groves. The rhinoceros beetle's method of boring holes for food can fatally injure a coconut tree. Currently, efforts are being made to control the rhinoceros beetle from spreading (Moore and Smith 2008). Pigs, introduced by Europeans, have created problems for native vegetation, farmers, and golf courses because of the pigs' rooting behavior in search for food.

Shoreline erosion is another ecological threat to the island. The most important factor preventing erosion of the shoreline is the natural coral reef that surrounds a large portion of the island. The reefs are under threat from pollution due to various human activities such as urban construction. With the recent expansion of construction activity on Guam, soil from recently cleared areas of land has washed off into the many bays. This soil then can suffocate much of the coral killing off the reef. Harmful fishing practices that involve pouring toxic poisons into the waters can harm both fish and the coral reef. Climate change could also pose a threat to the reef. The rising of ocean temperatures can affect coral's ability to grow. If sea levels rise then much of the shoreline could disappear along with the coral reef (Coral Reef Targeted Research Program 2007).

Guam is host to rare species. Today the Guam Rail cannot live on the island because of a host of threats from snakes, feral cats, and wild dogs. The Fanihi, also known as Marianas Fruit Bat, has been under constant threat from a host of factors from poaching for its meat to the loss of its habitat due to urban development. Today recent counts estimate the number of bats to be less than 100. It is not just animals on Guam that are under the threat of extinction; the island also has plants which are vulnerable. A tree species known as Ifit (*Intsia bijuga*), once desired for furniture and building construction, has been overly harvested to a fraction of its original population.

Today many of Guam's common plants have encountered new threats. The betel nut, which is common in many parts of the Pacific, has experienced a decline due to bud rot. The coconut tree which is common on Guam has experienced a decline caused by a disease known as Coconut Tinangaja Viroid. The ironwood tree (*Casuarina equisetifolia*), commonly found in Australia and in other parts of the Pacific has experienced an unidentified threat that is slowly reducing their numbers on Guam.

The ironwood tree is a natural part of Guam's wildlife. The tree grows quickly in loose soil, preventing erosion, and serves as a perch for predatory birds. It is important to determine the causes of the ironwood tree's decline because it could help prevent its decline from happening on neighboring islands such as Saipan and Rota. It could also help in finding a possible solution to removing the ironwood tree from areas where the tree was introduced. Identifying the underlying causes for the ironwood's decline on Guam can be difficult because decline can be attributed to one or many factors, both biotic and abiotic. The use of modern day statistics can help reduce the expense, time, and guesswork in finding the most likely causes for the ironwood tree's decline.

# Table of Contents

|  |      |
|--|------|
| Acknowledgments . . . . .                      | ii   |
| Preface . . . . .                              | iii  |
| List of Tables . . . . .                       | vi   |
| List of Figures . . . . .                      | vii  |
| Abstract . . . . .                             | viii |
| Chapter 1. The Ironwood Tree . . . . .         | 1    |
| Chapter 2. Data Collection . . . . .           | 3    |
| Chapter 3. Statistical Modeling . . . . .      | 6    |
| 3.1 Logistic Regression . . . . .              | 6    |
| 3.2 Cumulative Logit Regression . . . . .      | 7    |
| 3.3 Generalized Estimating Equations . . . . . | 8    |
| 3.4 Model Diagnostics . . . . .                | 8    |
| Chapter 4. Results . . . . .                   | 11   |
| 4.1 Logistic Results . . . . .                 | 11   |
| 4.2 Cumulative Logit Results . . . . .         | 13   |
| Chapter 5. Discussion . . . . .                | 15   |
| References . . . . .                           | 16   |
| Vita . . . . .                                 | 17   |

## List of Tables

|                |  |           |
|----------------|--|-----------|
| <b>Table 1</b> | <b>Measured Variables Grouped by Type . . . . .</b>      | <b>3</b>  |
| <b>Table 2</b> | <b>Logistic Modeling Results . . . . .</b>               | <b>11</b> |
| <b>Table 3</b> | <b>Logistic Model Clustering Trees By Site . . . . .</b> | <b>12</b> |
| <b>Table 4</b> | <b>Cumulative Logit Model Results . . . . .</b>          | <b>13</b> |

## List of Figures

|          |   |    |
|----------|---|----|
| Figure 1 | Locations of Sampled Trees . . . . .                                      | 3  |
| Figure 2 | Visual Estimation of Decline Severity . . . . .                           | 4  |
| Figure 3 | Logistic Pearson Residuals Mapped Over Two Dimensional Grid . . . . .     | 12 |
| Figure 4 | Pearson Residual $P(Y \leq 4)$ Mapped Over Two Dimensional Grid . . . . . | 14 |

## Abstract

The ironwood tree (*Casuarina equisetifolia*), a protector of coastlines of the sub-tropical and tropical Western Pacific, is in decline on the island of Guam where aggressive data collection and efforts to mitigate the problem are underway. For each sampled tree the level of decline was measured on an ordinal scale consisting of five categories ranging from healthy to near dead. Several predictors were also measured including tree diameter, fire damage, typhoon damage, presence or absence of termites, presence or absence of basidiocarps, and various geographical or cultural factors. The five decline response levels can be viewed as categories of a multinomial distribution where the multinomial probability profile depends on the levels of these various predictors. Such data structure is well suited to a proportional odds model thereby leading to odds ratios involving cumulative probabilities which can be estimated and summarized using information from the predictor coefficient. Various modeling techniques were applied to address data set issues: reduced logistic models, spatial relationships of residuals using latitude and longitude coordinates, and correlation structure induced by the fact that trees were sampled in clusters at various sites. Among our findings, factors related to ironwood decline were found to be basidiocarps, termites, and level of human management.



# Chapter 1. The Ironwood Tree

The ironwood tree known by the scientific name *Casuarina equisetifolia* is sometimes referred to as the Australian Pine. The ironwood tree shares some attributes with North American pines in that both have narrow needles and both produce cones. The differences between ironwood and pines are that the needles from ironwood trees have visible segments which give them a soft hair like appearance in contrast to the rigid spiky appearance of American pine needles. The ironwood tree is native to Australia and has a natural range along the South Pacific islands from Malaysia to Micronesia. Ironwood has also been introduced by man to the Hawaiian Islands and to the Florida coast. In Florida the ironwood tree poses a threat to native vegetation as it competes for sunlight and nutrients.

The ironwood tree migrated naturally to Guam thousands of years before Chamorros first settled on the island (Athens and Ward 2004). Because of this tree's ability to grow rapidly in loose or sandy soil, the ironwood tree has been found growing along the sandy shorelines of Guam making it a crucial plant in preventing erosion of the beaches. An ironwood tree during its life creates nitrogen fixings in its roots helping to replenish the soil with nutrients once the tree has died. Nitrogen fixing can help other plants grow in place of a previously living ironwood tree.

The native wild life on Guam has also benefited from the ironwood tree. A newly discovered species of eulophid wasp was discovered on Guam's ironwood trees (Mersha and Schlub 2009). The wasp lays eggs in the ironwood needles. Mariana Fruit Bats have been found to colonize on ironwood. Birds have used the ironwood for both roosting and perching. Roosting is a behavior where birds rest or sleep either as a group or individually. The ironwood tree is a place for birds such as the White-Collared Kingfisher and the Mariana Fruit-Dove to roost. The extension of roosting is nesting which is when birds build nests in an effort to reproduce. The White Tern on Guam has been found to lay eggs in ironwood trees (Mersha and Schlub 2010). The ironwood tree's branches also serve as a perch. An elevated perch offers predatory birds a vantage point from which they can hunt for food. Ironwood trees can grow up to 80 feet tall and with extending branches offering birds a good location for sighting prey. Ironwood trees offer birds a safe place to conserve energy.

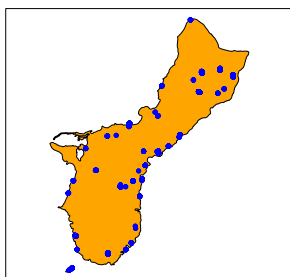
Humans have benefited from the ironwood tree in domestic and industrial uses. The wood has a high density which makes it ideal firewood. The ironwood has been popular as an ornamental plant for homes. The tree has been used in bonsai which means it can be manipulated to grow small. It also can be formed into large hedges which offer homeowners privacy, and for farmers, a natural wind break for their crops. Ironwood frequently sheds needles which can be used as mulch, a substance which is used to control the growth of weeds. Many of the uses for ironwood can be found on Guam where the tree has been a part of the island's native vegetation.

It was in the year 2002, when a local Guam farmer noticed his ironwood trees were showing symptoms of decline. By 2005, ironwood tree decline was widespread across the island. A panel of tree experts from the United States and Australia organized by the University of Guam met in January of 2009. Their objective was to find the possible causes for the decline in Guam's ironwood tree population. The panel had difficulty in assessing significant factors for this decline. Factors that were discussed to have been possible reasons for the ironwood's decline were extensive damage from typhoons, brush fires, termites, fungi, the age of the tree, and human management (Mersha and Schlub

2010). Many other factors were discussed, but none of them were proven or estimated to have been likely causes. Data collection and statistical modeling would need to be performed in order to find the most probable factors for ironwood tree decline.

## Chapter 2. Data Collection

Data collection on Guam began in August 2009, and ended in December of 2009. Approximately 1427 ironwood trees were surveyed across 44 sites. Most of the trees were surveyed due to their proximity to roads. Trees on government controlled land, such as the military bases, were not sampled for the study because of the difficulty to gain clearance.



**Figure 1: Locations of Sampled Trees**

For each sampled tree a set of 15 measurements were taken. However, there were a few variables that could not be measured for the 2009 survey. The age of the tree was one such variable that could not be measured because current methods such as counting tree rings and diameter and height of ironwood trees were not precise indicators for age. The gender of the tree also was not used. Ironwood trees can be either monecious or dioecious. A plant that is monecious contains both male and female reproductive parts allowing the plant to reproduce asexually. A plant that is dioecious contains reproductive parts for only one sex and only reproduces when it is near another plant with reproductive parts of the opposite sex. The gender of specific ironwood trees could be determined by finding male or female flowers on each tree, but most of the surveyed trees did not produce flowers at the time the survey was performed; therefore, gender was a discarded variable. During the survey a newly discovered species of eulophid wasp was found on the needle tips of the ironwood tree. It was decided that the wasp did not cause fatal injury to the tree and was not factored in the survey.

Of the set of measurements that were taken, these 15 measurements are:

**Table 1: Measured Variables Grouped by Type**

| Variable Type | Variable(s) name(s)                 |
|---------------|-------------------------------------|
| Response      | Dec_sev, Dieback                    |
| Structure     | Num_stem, CBH, Density              |
| Stress        | Fire, Conk, Typhoon, Termites       |
| Geographic    | Latitude, Longitude, Altitude, Site |
| Miscellaneous | Human_mgmt, Planted_natural         |

The response variable Dec\_sev, which stands for decline severity, is a visual categorization of an ironwood tree's health based on the fullness of a sampled tree's foliage. Dec\_sev takes discrete values from 0 to 4. A perfectly healthy

tree will have full foliage and will be given a Dec\_sev value of 0. A tree that is sick will have less needles on its branches and will have a Dec\_sev higher than 0 depending on the severity of its loss in foliage(see fig. 2).



**Figure 2: Visual Estimation of Decline Severity**

The variable Dieback is derived from the variable Dec\_sev. If a tree is perfectly healthy and has a Dec\_sev equal to 0, the tree then will be assigned a Dieback value of 0. Otherwise a tree having a Dec\_sev greater than 0, will be assigned a Dieback value of 1.

The structure variables relate to the physical attributes of a sampled ironwood tree. The variable num\_stem is a discrete count of the number of main branches or scaffolds that grow out of the trunk of a sampled tree. The variable CBH stands for the circumference at breast height which is a measurement in centimeters of a tree's circumference at 1.3 meters from the ground. The variable density is the number of ironwood trees found on a given site per square meter.

Stress variables are observations of the damage a sampled tree has suffered. All four variables in the stress group are binary variables 0 and 1; 0 if no signs of stress are present for a given variable and 1 if there are signs of stress. For instance the variables fire and typhoon are values 0 or 1 for a given tree based on signs of damage. If a sampled tree had indications that it was damaged at one time by fire, then the fire variable was recorded as 1; if there were no signs of a fire then the given tree's fire measurement was assigned a 0. Likewise 1 or 0 was assigned to a given tree dependent on damage by a typhoon. The variables termites and conk are similarly 0 and 1. If a sampled tree has conk, a fungus that is visible on the trunk of a tree, then the given tree has a conk value of 1; if no conk is visible then the value is 0. If termites are present on the tree then the termite value for the given tree is assigned a 1; if no termites are present the value is 0.

Geographic variables define basic information regarding the area where a sampled tree is located. The latitude and longitude are the global coordinates for a sampled tree measured using a GPS device. The latitude and longitude were measured because it was theorized that there was some importance in the specific location of the tree. The altitude is the meters above sea level of the sampled tree. Altitude was obtained from the same GPS device that measured the longitude and latitude. The site is an arbitrary number assigned to a location where a group of trees were surveyed.

For example all ironwood trees on a golf course would be assigned the same site number. In total there were 44 sites that were surveyed.

It was suggested at the 2009 ironwood tree meeting that humans were unintentionally damaging the ironwood trees. Therefore a variable `human_mgmt`(human management) was measured. It was given a discrete ordinal value from 0 to 2 to describe the level of management trees at a given site were likely to have received. If a site was untouched by people, such as a nature park, then the ironwood trees on that site were all assigned a `human_mgmt` level of 0. The more intensely managed the site the higher the number. The last variable, `planted_natural` was assigned simple nominal value of  $p$  for planted or  $n$  for natural. It was suggested at the 2009 meeting that the ironwood trees might be dying due to being planted by people.

## Chapter 3. Statistical Modeling

Because the ironwood tree survey contains information on 15 variables (regressors), knowing which of these regressors is significant can be a difficult problem. The goal is to identify important variables through the use of statistical modeling and variable selection. The basic method for creating a statistical model is to fit a response variable to a set of regressors. For instance, the variable *dieback* can be treated as the response variable to a logistic regression model because *dieback* has a natural binary outcome of 0 and 1. The initial reason for using a logistic regression is to explore the variables relationship to the binary outcome: sick and non-sick trees.

It is theorized that ironwood tree decline has more than two stages. A binary response is inadequate to compare the gradual levels of tree decline because there are more than two stages in the decline. Therefore, the variable *dec\_sev* can be treated as a response variable for a multinomial model using a cumulative link function. It is necessary to utilize such a model (cumulative logit model) for two reasons. The first reason is that a cumulative logit model will utilize all five stages of tree decline. The second reason is that the ordinal nature of the response will be used. The variables *dieback* and *dec\_sev* will be used as responses for two types of models to determine if at least one of the models can explain what factors are significant towards ironwood tree decline.

### 3.1 Logistic Regression

Logistic regression requires the response variable *dieback* to have a natural binomial distribution. The values for the response is *dieback* = 1 if a tree is sick, and *dieback* = 0 otherwise. If the data contains a total count of  $n$  trees, then there exists some probability  $\pi$  that  $y$  trees will have a *dieback* = 1. The distribution of  $y$  is known as the binomial distribution which is

$$P(Y = y) = \binom{n}{y} \cdot \pi^y \cdot (1 - \pi)^{n-y} \quad .$$

In general the probability  $\pi$  of a sick tree is unknown, but it can be estimated using a set of regressors in a logistic regression setting. Let  $\pi(\mathbf{x})$  be the probability of a sick tree based off a vector of regressors  $\mathbf{x}$ , then the basic form of a logistic regression is

$$\log \left( \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right) = \alpha + \boldsymbol{\beta}' \mathbf{x} \quad .$$

On the right hand side is a standard linear predictor setting where  $\alpha$  is an estimate of the intercept term,  $\beta$  is an estimate vector of  $k$  coefficients, and  $\mathbf{x}$  is a vector of data that corresponds to values for each of the  $k$  regressor variables. The left hand side of the equation is a log odds ratio. The odds ratio, in the context of the ironwood study, measures the weight of the probability of a sick tree  $\pi$  against the probability of a non-sick tree  $1 - \pi$ . Suppose the probability  $\pi$  of a sick tree is known and assume that  $\pi < 1$  then to calculate the odds of a sick tree relative to non sick, we have

$$\frac{\pi}{1 - \pi} = m \quad \text{where} \quad m \geq 0 \quad .$$

In general the value  $\pi$  is unknown but the odds ratio can be calculated from the equation for a logistic regression if there is a set of values  $\mathbf{x}$ . Using  $\mathbf{x}$  the estimated odds ratio is

$$\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} = e^{\alpha + \beta' \mathbf{x}} \quad .$$

Because logistic regression only evaluates a difference between sick and non-sick trees, a cumulative logit model is needed to evaluate across all five stages of decline. We estimate  $\pi$  and the odds ratio using maximum likelihood estimates  $\hat{\alpha}$  and  $\hat{\beta}$  for  $\alpha$  and  $\beta$ , respectively.

### 3.2 Cumulative Logit Regression

A cumulative logit model is an extension of the ideas used in logistic regression. In the logistic regression we worked in terms of  $\pi(\mathbf{x})$  which was the probability of a singular event. A cumulative logit model, in contrast to a logistic regression, works in terms of a cumulative probability which could be one or more events. In the context of the ironwood study, the cumulative probability is such that for a given tree there is a probability that the tree will have a *dec\_sev* value that is at most equal to  $0 \leq j \leq 4$ . The cumulative probability is defined as

$$P(\text{dec\_sev} \leq j) = \pi_0 + \pi_1 + \cdots + \pi_j \quad ,$$

where  $\pi_i$  is the probability that for a given tree  $\text{dec\_sev} = i$ . Cumulative logit regression offers two advantages over logistic regression. The first advantage is that the response variable can have more than two values. The second is that the response variable's natural ordering is utilized in the model. The general form of the cumulative logit model is

$$\log \left( \frac{P(\text{dec\_sev} \leq j)}{1 - P(\text{dec\_sev} \leq j)} \right) = \alpha_j + \beta' \mathbf{x} \quad \text{where } j = 0, 1, \dots, 3 \quad .$$

The right hand side of the equation resembles a multiple regression. However, note that the intercept term  $\alpha_j$  changes for each category  $j$ . The term  $\beta$  is the vector of coefficients and  $\mathbf{x}$  is the vector of regressors. The left hand side of the equation is a cumulative log odds ratio. This odds ratio suggests a ratio of the probability that a tree's state of health is leaning towards the healthy stages versus the probability that the tree is more likely to be in one of the sicker stages.

The cumulative odds ratio is computed as

$$\frac{P(\text{dec\_sev} \leq j)}{1 - P(\text{dec\_sev} \leq j)} \quad .$$

The probability  $P(Dec\_sev \leq j)$  can be estimated from the cumulative logit model by using

$$P(Dec\_sev \leq j) = \frac{e^{\alpha_j + \beta' \mathbf{x}}}{1 + e^{\alpha_j + \beta' \mathbf{x}}} \quad .$$

To find the probability of a tree having  $Dec\_sev$  equal to  $j$ , simply compute

$$P(Dec\_sev = j) = P(Dec\_sev \leq j) - P(Dec\_sev \leq j - 1) \quad .$$

By necessity  $P(Dec\_sev \leq 4) = 1$ . Estimates for  $\alpha_j$  and  $\beta$  are computed using the maximum likelihood estimates  $\hat{\alpha}_j$  and  $\hat{\beta}$ .

### 3.3 Generalized Estimating Equations

A structure that could be imposed on a statistical model would involve the concept of clustering data. A cluster is a set of experimental units that are assumed to be similar in some way. In the context of the ironwood tree study, a cluster could be all ironwood trees found on a particular site. Let  $Y_i$  denote the response for a tree  $i$  in a particular cluster  $c$ . The generalized estimating equation (GEE) method generates estimates for the standard errors of a model's (logistic or cumulative) parameters by working with a correlation structure imposed on the clustered data. The GEE method works when assuming that each  $Y_i$  has a common distribution, and that the working correlation matrix for a cluster  $c$  has an assumed structure. Results from a GEE model yields robust standard errors and less bias in the model parameters if the assumed correlation structure is appropriate.

### 3.4 Model Diagnostics

The primary objective of using statistical models with the ironwood tree data is to find possible factors that could be related to tree decline. In effect to find the parameters that have a positive or negative impact on the odds ratio. Given a coefficient  $\beta_i$ , the sign(+/-) could change the interpretation of the  $i$ th regressor  $x_i$ . For example if the sign were negative, then  $\beta_i < 0$  which implies that if holding all other regressors constant, the odds ratio for the model will decrease as  $x_i$  increases. If the sign for the  $i$ th coefficient is positive this would imply the opposite; that holding all other regressors constant, the odds ratio would increase as  $x_i$  increases. To calculate the multiplicative effect that  $x_i$  has on the change of odds one would use the following formula

$$\frac{\left. \frac{\pi(x)}{1 - \pi(x)} \right|_{x_{i+1}}}{\left. \frac{\pi(x)}{1 - \pi(x)} \right|_{x_i}} = \exp(\beta_i) \quad .$$

The left hand side is a notation that suggests that there is a ratio between two odds ratios. The numerator is an odds ratio where the regressor  $x_i$  is increased by one unit holding all other regressors constant. The denominator is



the odds ratio without increasing the regressor  $x_i$ . The right hand side is the multiplicative factor by which the odds ratio changes when  $x_i$  is increased by one unit. For example, suppose  $\beta_i = 2$  then  $e^2 = 7.39$ , this would suggest that the odds increases 7.39 times when the regressor  $x_i$  increases by one unit. In order to determine the significance of a regressor's effect on the odds ratio hypothesis testing must be done.

The most common approach to testing a single regressor is using the Wald's chi-square test to determine if a specific coefficient  $\beta_i$  is significant from 0. The Wald's chi-square test statistic for testing  $H_o : \beta_i = 0$  is constructed as

$$z = \frac{\hat{\beta}_i}{\text{S.E}(\hat{\beta}_i)} \quad .$$

The maximum likelihood estimate  $\hat{\beta}_j$  and its S.E. (standard error) must be computed with software. By raising the value  $z$  to the second power, the Wald test statistic will have a chi-square distribution with one degree of freedom. In effect the critical value for  $z^2$  only changes with the  $\gamma$  level of significance and does not depend on the number of observations in the data, assuming the sample size is large enough. The null hypothesis for the Wald's chi-square test is that  $\beta_i$  is not significant from zero. If  $z^2$  is less than  $\chi_\gamma^2$  (the critical value of a chi-square distribution with  $\gamma$  level of significance and 1 degree of freedom) then  $\beta_i$  is not significant. Otherwise,  $\beta_i$  is significant when  $z^2 > \chi_\gamma^2$ . Similar to the Wald's test statistic, the confidence interval for a single coefficient  $\beta_i$  can be constructed as

$$\hat{\beta}_i \pm z_{\gamma/2} \cdot \text{S.E}(\hat{\beta}_i) \quad .$$

If the value 0 is not in the interval then it would suggest that  $\beta_i$  has a significant effect on the change in odds. Finding an accurate model that fits the data involves removing insignificant regressors from the model. However removing several regressors based on each regressor's Wald test scores can be problematic due to potential inflation of error.

A better method for determining if the removal of insignificant regressors can yield better accuracy would be to perform a goodness of fit test. A full model is defined as a model that is fitted with all regressors in the study, and a reduced model is defined as one that has fewer regressors in comparison to the full model. A goodness of fit test is where a full model's residual deviance is compared to a reduced model's residual deviance. The test then becomes

$$D_R - D_F \quad ,$$

where  $D_F$  is the residual deviance of the full model with  $n_f$  residual degrees of freedom and  $D_R$  is the residual deviance for the reduced model with  $n_r$  residual degrees of freedom. The difference of the residual deviances has a chi square distribution with  $n_r - n_f$  degrees of freedom. If  $D_R - D_F < \chi_{\gamma, df=n_r-n_f}^2$ , where  $\gamma$  is the level of significance, then the reduced model is a better fit than the full model. Algorithms such as stepwise, forward, and backward selection are designed to produce a finalized model by using goodness of fit tests. After reducing the model down to only the most significant regressors it might be of interest to check the model's residuals to further assess the model's fit.

A residual in basic terms of the ironwood study is the difference between a tree's observed response value and the model's estimated response value. Residuals in their raw form usually will not give much information, but by standardizing the raw residuals they can give better details on a model's ability to fit the data. The Pearson residuals is a standardized form of raw residuals which have a standard normal distribution with mean 0 and standard deviation 1. Thus it is expected that most residuals in a well fit model will fit approximately within 2 to 3 standard deviations away from the mean. If for a model an observation has a Pearson residual value of 5, it would suggest that the observation is an extreme value. Having several extreme values could suggest that the model does not explain the trends in the data very well. Residual diagnostics is the last step in assessing the adequacy of a model's fit before conclusions could be reached about the model's significant factors for tree decline.

## Chapter 4. Results

A logistic model and a cumulative logit model were fitted using *dieback* as the response in the former model and *dec\_sev* as the response in the latter. For each model there were initially ten regressors. The regressors were from the *stress*, *structure* and *miscellaneous* groups and included the regressor *altitude*. Model selection algorithms were performed on both models to come to a reduced model. After model selection was applied, the regressors *longitude* and *latitude* were added into each reduced model as regressors in a finalized model. The regressors *longitude* and *latitude* are added last into the model because we want to explore the spatial effects after adjusting for other important regressors.

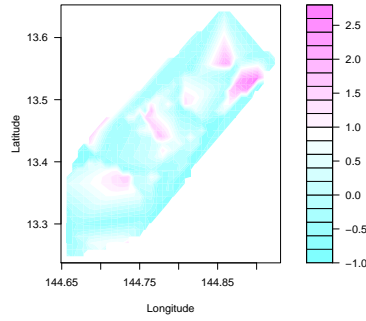
### 4.1 Logistic Results

The following table is output from the finalized logistic model:

**Table 2: Logistic Modeling Results**

|             | Estimate | Std. Error | Wald z-value | $Pr(>  z )$ |
|-------------|----------|------------|--------------|-------------|
| (Intercept) | -417.52  | 225.20     | -1.85        | 0.06        |
| conk        | 3.30     | 0.28       | 11.91        | < 0.01      |
| human_mgmt  | 0.76     | 0.12       | 6.35         | < 0.01      |
| termites    | 0.75     | 0.17       | 4.27         | < 0.01      |
| density     | -1.16    | 1.11       | -1.05        | 0.29        |
| lat         | -4.76    | 1.25       | -3.82        | < 0.01      |
| long        | 3.31     | 1.65       | 2.00         | 0.05        |

Because the response variable *dieback* was assigned 1 for sick and 0 for non-sick, the coefficients that could explain which regressors have a detrimental effect on ironwood tree health are coefficients that are positive and significant from 0. In total three regressors stood out as possible factors relating to the tree's state of health. Among the three regressors *conk* had the largest coefficient value at 3.31. If a tree had the presence of *conk*, and holding all regressors constant, the increase in the odds of a sick tree would be 27.28 times in comparison to a tree that did not have *conk*. Residual diagnostics were performed in order to assess the model's fit. Pearson residuals from the logistic model were analyzed to determine if there were any outliers.



**Figure 3: Logistic Pearson Residuals Mapped Over Two Dimensional Grid**

The Pearson residual plot with interpolation is a way to project observed Pearson residual points onto a two dimensional grid. In addition to observed Pearson residuals, there are predicted Pearson residual points that are estimated between areas of the grid where no observations were made. The plot can then be looked at as a forecast by illustrating the trend of the residuals across the entire island of Guam.

The logistic model’s Pearson residuals were interpolated across a two dimensional grid of the longitude and latitude coordinates of the island. From the graph (Figure 3) most of the residuals fell within the range of -1 to 2.5 which suggests that the logistic model can adequately explain the difference between sick and non-sick trees.

Clustering trees by site changes the estimates for standard error. As displayed in Table 3, the most significant effects after clustering were conk, termites and human management.

**Table 3: Logistic Model Clustering Trees By Site**

|             | Value  | Std. Error | Wald $\chi^2$ value | p-value |
|-------------|--------|------------|---------------------|---------|
| (Intercept) | 417.52 | 612.81     | 0.46                | 0.50    |
| conk        | -3.31  | 0.44       | 57.46               | < 0.01  |
| termites    | -0.75  | 0.27       | 7.56                | < 0.01  |
| long        | -3.31  | 4.51       | 0.53                | 0.47    |
| lat         | 4.76   | 3.57       | 1.80                | 0.18    |
| density     | 1.16   | 2.68       | 0.18                | 0.67    |
| human_mgmt  | -0.76  | 0.28       | 7.29                | < 0.01  |

## 4.2 Cumulative Logit Results

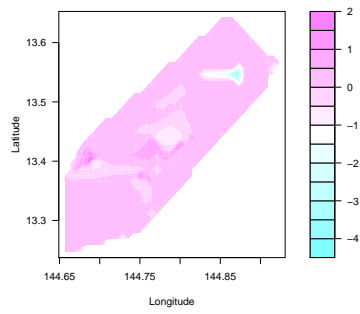
The parameters from the cumulative logit model were estimated to be

**Table 4: Cumulative Logit Model Results**

|               | Value  | Std. Error | Wald $\chi^2$ value | p-value |
|---------------|--------|------------|---------------------|---------|
| (Intercept):1 | 104.35 | 200.96     | 0.27                | 0.60    |
| (Intercept):2 | 105.08 | 200.96     | 0.27                | 0.60    |
| (Intercept):3 | 105.79 | 200.96     | 0.28                | 0.60    |
| (Intercept):4 | 106.50 | 200.96     | 0.28                | 0.60    |
| density       | -2.07  | 0.83       | 6.27                | 0.01    |
| termites      | -0.75  | 0.15       | 24.51               | < 0.01  |
| conk          | -2.64  | 0.17       | 239.76              | < 0.01  |
| typhoon       | -0.35  | 0.18       | 3.97                | 0.05    |
| human_mgmt    | -0.77  | 0.11       | 51.37               | < 0.01  |
| lat           | 3.78   | 1.12       | 11.35               | < 0.01  |
| long          | -1.06  | 1.47       | 0.51                | 0.47    |

As the response variable *dec\_sev* is from 0 to 4, healthy to near dead, determining which regressors were related to tree decline would depend on whether the parameter estimate was negative and if it was significant from zero. The significant regressors that are detrimental to a tree's health from the table are *density*, *termites*, *conk*, *typhoon*, and *human\_mgmt*. The results are similar with the logistic model. It suggests that the regressor *conk* decreases the odds of a tree being in the healthy stages by a multiplicative factor of 0.07 or 92.87%.

Residuals were also analyzed for the cumulative logit regression in order to assess the model's fit. In contrast to the logistic regression, the cumulative logit regression had four sets of residuals; one set of residuals for each category of *dec\_sev* in the model. Each set of residuals were interpolated and then mapped along a two dimensional grid. In evaluating each graph there were signs of outliers.



**Figure 4: Pearson Residual  $P(Y \leq 4)$  Mapped Over Two Dimensional Grid**

For the category of residuals involving  $P(Y \leq 4)$  the range was from -4 to 2. Since some of the data points were outside the 99% range of -3 to 3 this suggests possible outliers in the model. Despite the initial promise that the parameter estimates suggested significant regressors that would relate to ironwood tree decline, the cumulative logit model was not adequate to explain the various stages of decline. At most the conclusions that can be reached is that the logistic model is preferred to the cumulative logit model, and that the logistic model's results suggests that *conk*, *termites*, and *human management* are possibly related to the ironwood tree's state of health on Guam.

## Chapter 5. Discussion

There may be several ways to improve the modeling so that all five categories of tree decline can be utilized. Some alternatives might be to add new regressors or change the model from a cumulative logit model to a multinomial model in which the response is not ordinal. New regressors could be added into the model based on survey data that the United States government has done on the island of Guam. Some of the regressors include soil type, water levels, and vegetation type. Another alternative would be to work with a multinomial model.

One reason to use a multinomial model is that perhaps the cumulative logit model does not meet the proportional odds assumption. With proportional odds the change in odds for a given parameter  $\beta_i$  is always the same regardless of the stage of decline. A multinomial model that is not based on the proportional odds is different from the cumulative logit model. The basic form of the multinomial model is

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j' \mathbf{x}, \quad \text{where } j = 1, \dots, J - 1 \quad .$$

The right hand side of the equation differs from a cumulative logit regression in that for each category  $j$ , the vector of parameters  $\beta_j$  freely varies. In effect the change in category changes the odds. The left hand side of the equation is referred to as a baseline log odds. A category  $J$  is chosen as a baseline when calculating all the parameters in the model. The value  $\pi_J$  is the probability that a given tree has a  $dec\_sev = J$ . Because a multinomial model with baseline odds does not require the response to be ordered, any  $dec\_sev$  category could be chosen to be a baseline. For the purpose of the ironwood tree study the baseline should be  $dec\_sev = 0$  so that the results could state information about the odds between the various states of tree decline versus perfectly healthy trees.

At present the results determined by the logistic regression model have helped the biologists study the possible effects of ironwood tree decline. More study is being done on the various species of conks that are found on the ironwood tree. The conks are being tested for possible pathogenic properties that could explain why they have been prevalent on many of the sick ironwood trees.

## References

- Agresti, Alan. 2007. *An Introduction to Categorical Data Analysis*. 2nd ed. Hoboken, New Jersey: John Wiley & Sons Inc.
- Athens, J. Stephen & Jerome V. Ward. 2004. Holocene Vegetation, Savanna Origins and Human Settlement of Guam. *Records of the Australian Museum, Supplement* 29:15-30.
- “Australian Pine.” *PCA Alien Plant Working Group*. Accessed July 16, 2010. <http://www.nps.gov/plants/alien/fact/caeq1.htm>
- Coral Reef Targeted Research Program. “Global Warming Is Destroying Coral Reefs, Major Study Warns.” *ScienceDaily*, December 14, 2007. Accessed October 8, 2010. <http://www.sciencedaily.com/releases/2007/12/071213152600.htm>.
- D. L. Rockwood, R. F. Fisher, L. F. Conde, and J. B. Huffman. “Casuarina.” *Northeastern State and Private Forestry*. Accessed July 16, 2010. [http://www.na.fs.fed.us/pubs/silvics\\_manual/volume\\_2/casuarina/casuarina.htm](http://www.na.fs.fed.us/pubs/silvics_manual/volume_2/casuarina/casuarina.htm)
- Drenth, A. 2004. Phytophthora Bud Rot In Areca Nut On Guam and The Northern Marianas. *CRC for Tropical Plant Protection*, Indooroopilly Research Centre, Brisbane. 27 pp.
- “Ecology of Casuarina Equisetifolia.” *Global Invasive Species Database*. Accessed July 16, 2010. <http://www.issg.org/database/species/ecology.asp?fr=1&si=365>
- Mersha, Z., Schlub, R. L., and Moore, A. 2009. The State of Ironwood (*Casuarina equisetifolia* subsp. *equisetifolia*) Decline On The Pacific Island of Guam. Poster in 2009 APS annual proceedings Portland, Oregon: Phytopathology 99:S85
- Mersha, Zelalem, Schlub, Robert L. 2010. Pre and Post January 2009 Guam Ironwood (*Casuarina equisetifolia*) Tree Decline Conference. Paper presented at Fourth International Casuarina Meeting, 22-26 March, at Haikou, Hainan, China.
- Moore, Aubrey & Smith, L. Sheri. “Early Detection Pest Risk Assessment Coconut Rhinoceros Beetle.” *US Forestry Service* March 2008. Accessed October 7, 2010. <http://www.fs.fed.us/>
- Wall, G. C. and Randles, J. W. 2003. Coconut Tinangaja Viroid. Chapter 36 in: Hadidi, A. et al. (Eds.) 2003. CSIRO Publishing, Australia, or Science Publishers, Enfield, NH. Pp. 242-245.



## Vita

Karl Schlub was born March 14, 1983, in Covington, Louisiana. He moved to the island of Guam in 1996 where he lived for 8 years. He graduated from George Washington High School which is in Mangilao, Guam. In 2001 Karl first attended the University of Guam with the intentions of earning a computer science degree. He earned a Bachelor of Science degree in mathematics at the University of Utah in 2007. In the course of that time, Karl also attended University of Georgia for one year as a national exchange student. After graduating in 2007, Karl Schlub enrolled in the fall of 2008, at Louisiana State University where he is working on a Master of Applied Statistics.