STATISTICAL METHODS FOR TURTLE BYCATCH DATA

by

JOHNATHAN E. DEAN

(Under the Direction of Daniel B. Hall)

ABSTRACT

As part of a hook exchange program initiated by the Eastern Pacific Regional Sea Turtle Program (EPRSTP), data on fish and sea turtle catch rates were collected for a large number of fishing lines. The goal of this paper is to find a statistically sound way to analyze these data, comparing two types of fish hooks in terms of their effects on turtle bycatch and catch rates for merchantable fish. The overall goal of this environmental project is to minimize the number of endangered sea turtles being caught on these fishing lines. We suggest an appropriate method for analyzing turtle bycatch data of the type collected by the EPRSTP. In this study, we generate probabilities at a line level under three different patterns simulating possible probabilities of catching a turtle along a line. Once these data were generated four different Mantel-Haenszel methods were compared as methods for analyzing the data. This study suggests that Mantel-Haenszel estimators proposed by Liang (1985) are appropriate for analyzing turtle bycatch data.

INDEX WORDS: Mantel-Haenszel, permutation, odds ratio, risk difference.

STATISTICAL METHODS FOR TURTLE BYCATCH DATA

by

JOHNATHAN E. DEAN

B.S., Western Carolina University, 2003

M.Acc., Western Carolina University, 2005

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2008

STATISTICAL METHODS FOR TURTLE BYCATCH DATA

by

JOHNATHAN E. DEAN

Major Professor:	Daniel B. Hall
Committee:	Cheolwoo Park
	Lynne Seymour

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2008

DEDICATION

To my parents, for all of their love, support and guidance throughout my life and for putting up with me during my extended academic career.

ACKNOWLEDGEMENTS

This work would not have been possible without the help, direction, and patience of my advisor, Dr. Daniel B. Hall, under whose supervision I chose this topic and began the thesis. I owe him many thanks for being my thesis advisor and for sharing his knowledge of nonlinear mixed-effect models and for guiding me through my consulting project.

I would also like to thank my committee members Dr. Cheolwoo Park and Dr. Lynne Seymour for their support and encouragement during my time here at the University of Georgia. I truly believe I would not have made it through the graduate program without their help. I believe their greatest attributed is their heart. They both have made it very clear how much they care for students. The memorable times spent talking over dinner or the times spent on a basketball court; those memories are what I am most grateful.

I also want to express my gratitude to my former officemate and buddy, Jin-Hong Park. While working on his Ph.D. I remember he would be consumed by his work yet he would find the time to ask me how I was doing and to offer his help with my work without me asking. For a first year student such as myself, struggling with his course load, sharing an office with a friend like Jin-Hong is more than I ever could have hoped for. Dr. Park, I am truly grateful for all he as has done.

To each of the above, I extend my deepest appreciation. Thank you!

TABLE OF CONTENTS

Page

CHAPTER

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

BACKGROUND INFORMATION

1.1 NATURE OF THE PROBLEM

Individuals living in many of the poverty–stricken communities along the coast of Central and South America lean on fishing as a major source of income. These professional fishermen are targeting fish such as tuna, billfish, sharks, and mahi-mahi. While they are targeting these marketable fish they are also catching several different species of sea turtles as ``bycatch'', or unintended, non-marketable marine species.

Over the past few years, the numbers of sea turtles along the Pacific beaches have seen decreasing trends. In fact, several sea turtle species are endangered. In hopes of reducing the number of turtles being caught in the longline fishing industry, a 'hook exchange' program was started in 2004 called the Eastern Pacific Regional Sea Turtle Program (EPRSTP). Prior to this program most fishermen were using traditional Japanese-style tuna hooks (or J hooks). It is the belief of the organizers of the EPRSTP that the use of a different type of fish hook, called a circle hook (or C hook), may significantly reduce the number of turtles caught.

This is a completely voluntary program asking fishermen to exchange J hooks for C hooks and adopt an alternating pattern of J and C hooks along a line. The C hooks were provided at no charge to fishermen. As part of the program, an EPRSTP staff member accompanied each fishing crew participating in the hook exchange program to record data on turtles and merchantable fish caught on the two hook types during the fishing trip. The sponsors of the EPRSTP would like to compare catch rates on these J and C hooks for both sea turtles and merchantable fish. The hope is that the use of C hooks will reduce turtle bycatch rates without adversely affecting the merchantable fish harvest. Because a switch to C hooks would be inconvenient and difficult in several respects, such a switch would only be worthwhile if it results in a major reduction in turtle bycatch. Therefore, both whether C hooks reduce bycatch rates and the magnitude of this reduction is of interest.

1.2 NATURE OF THE DESIGN

The EPRSTP conducted this observational study and provided us with a portion of the resulting data. This dataset summarizes information from 539 fishing lines collected from three different countries of interest. There are a number of explanatory variables that were gathered and which are relevant to this turtle bycatch problem. These variables include country of origin, trip number, vessel id, size of boat used, set number, latitude and longitude of the location at which the line was set, elapsed time that the line was in the water, percent of time out in daylight, length of the main line, distance between hooks, bait type used, total number of hooks, number of J hooks, number of C hooks, number of both fish and turtles caught on J hooks, number of both fish and turtles caught on C hooks. While all of these variables are potentially important in determining bycatch rates, the main variables of interest in our analysis are the line identifier, the number of C & J hooks, and the number of turtles caught on each type of hook. While more detailed and intricate analyses are possible, the methods advocated in this thesis have the great advantage of controlling for effects of all between-line covariates (e.g., potential confounders) without specifically accounting for them in the analysis.

A generalized diagram of how a line appears once it is launched in the water is shown in Diagram 1. A typical line would include a mainline and numerous branchlines of which both can be of various lengths. The floatline would adjust the desired depth of the mainline from the water's surface. Then branch lines would be attached to the mainline at various intervals set by the fishermen. At the other end of each of the branch lines would be the fish hook and some type of bait chosen by the fishermen.

For a more detailed example a typical line (also called a "set") would be released from the back of the boat. As the currents would pull the floats out behind the boat a fisherman would attach a shorter branch line onto the main line one after another in which the fisherman would attempt to alternate the type of fish hook used J, C, J, C,…etc. After a certain amount of time in the water, the line would be reeled back in and the hooks would be removed along with any animals caught. The researcher on board would record all of the data at the line level; the number of turtles and the number of marketable fish caught and the type of hook the catch was caught on.

As seen in Diagram 1, lines are layed out in catenaries that suspend the hooks at multiple different depths. Depth could create heterogeneity in the catch probability from hook to hook. For example, turtles are known to be topdwellers. If a line is suspended in such a way that the end of the line is much deeper than the part of line closest to the boat and the depth of the hooks at the end of the line are deeper than where turtles usual navigate then these hooks would have a much lower probability of catching a turtle. Because of the inherent characteristics of the line you wouldn't necessary expect these to be independent identical trials like you would have in a binomial. There are other with-in line factors that may affect hooking events such as distance from the boat, distance from a float, distance between the hooks.

There are other things that could create dependence that are not line specific characteristics. For example, merchantable fish have a tendency to school together and therefore can create hooking encounters in local bunches along a line. This may cause one part of a line to encounter more hooking activity than another part of the line. This sort of thing would potentially violate an assumption of independent identical trials. Failure of that assumption will play a crucial role in our choice of methods and our analysis of the relative performance of those methods. The alternating hook design controls (to a large extent) for some of the within-line sourses of variability.

1.3 NATURE OF THE DATA

Tables 1.1, 1.2, and 1.3 summarize some simple features of the data. From Table 1.1 it is clear that we are dealing with a large number of hooks: 96,443 J hooks and 69,158 C hooks on a total of 539 lines. Also note that J hooks out-number C hooks. Table 1.2 breaks out the number of turtles caught and the number of fish caught on each type of hook and provides a comparison of percentages for J hooks vs. C hooks. For example, 69% (167) of the turtles caught were on J hooks and the other 31% (76) were on C hooks. These figures can be misleading because J hooks and C hooks do not occur in a 50/50 ratio. There are more J hooks than C hooks, so one would expect that more turtles were caught on J hooks. However, even accounting for the greater number of J hooks, the proportion of hooks which caught turtles is higher for J hooks: 0.17% and 0.11% for J and C hooks, respectively (Table 1.3).

In Figure 1 it is clear to see that there were few lines that caught turtles. This figure shows that 74.21% of the lines caught zero turtles and another 16.51% caught only one turtle. Figures 3, 4, 5 and 6 show graphical evidence of sparseness in the data collected. In Figure 2 it is clear to see that there were few lines that caught large numbers of fish. This figure shows that 13.54% of the lines caught zero fish and 69.76% of the lines caught five or fewer fish. More evidence of the sparseness of the data can be seen in Table 1.3. This table shows the ratio of the number of animals caught on a specific type of hook per line, divided by the number of specific type of hooks per line. For example, only 0.17% of J hooks caught a turtle and only 0.11% of C hooks caught a turtle.

Figure 7 is a three-dimensional graph that compares the frequency of turtles caught for J hooks vs. C hooks. There were more lines that caught 0 turtles on C hooks than on J hooks. It is also visible that those lines that caught more than zero turtles are proportionally smaller for C hooks than J hooks. Figure 8 provides the same information for marketable fish. This figure is a bit more difficult to see but those lines that caught larger numbers of fish were on J hooks as opposed to C hooks.

Table 1.1:

Number (Percentage) of Hooks of Each Type

	J Hook	C Hook	Total
Number of Hooks	96,443 (58.2%)	69,158 (41.8%)	165,601 (100%)

Table 1.2:

Number (Percentage) of Turtles/Fish Caught By Hook Type

	J Hook	C Hook	Total
Number of Turtles Caught	167 (68.7%)	76 (31.3%)	243 (100%)
Number of Marketable Fish Caught	1,395 (47.6%)	1,535 (52.4%)	2,930 (100%)

Table 1.3:

Proportion of Hooks that Caught Turtles/Fish By Hook Type

	J Hook	C Hook	Total
Proportion of Hooks That Caught Turtles	0.0017	0.0011	0.0015
Proportion of Hooks That Caught Fish	0.0145	0.0222	0.0177

1.4 GOAL OF THIS PAPER

This paper focuses on coming up with an appropriate method for analyzing turtle bycatch data of the type collected by the EPRSTP. The EPRSTP would like to be able to say that the proportions of turtles caught on C hooks are significantly lower than on J hooks. Implementing this change to a different type of fish hook would hopefully allow turtles a better chance at survival in this area without causing any drastic changes to the fishing industry.

The 'naïve' analysis would be to cross classify all the hooks by hook type and whether a turtle was caught, creating a simple 2×2 table which would be analyzed as if the hooks were not organized onto different lines. This analysis will ignore heterogeneity from line to line and treats all the hooks as independent, identical trials. This simple test is a way of comparing whether proportions are the same for two groups. The odds ratio is used to make this comparison and it is simply the ratio of the odds of a turtle hooking for J and C hooks. This yields an odds ratio of 1.58 for the turtle data. This result suggests that the odds of catching a turtle is 58% higher on a J hook than on a C hook. The same 'naïve' analysis can be done on the fish data collected. It results in an odds ratio of 1.55. This result suggests that the odds of catching a fish is 55% higher on a C hook than on a J hook.

Based on this simple naïve analysis, evidence suggests that C hooks may provide the desired advantages the EPRSTP is looking for. However, given the complexity of the data collection scheme and the multiple sources of potential heterogeneity and dependence among these data, this analysis is clearly over-simplistic, and a more sophisticated analysis with assumptions consistent with the data collection scheme is needed. There are several options to consider when attempting answering the question, what is an appropriate method for analyzing these data? The main goal of this paper is to fix this naïve analyses and see if the resulting 'non-naïve' analyses gives essentially the same answer.

1.5 STATISTICAL METHODS WORKSHOP

In November, 2007, the EPRSTP organized a workshop in Costa Rica to study statistical methods to analyze turtle bycatch data. The participants in this workshop discussed a variety of approaches for analysis of these data including modeling, M-H methods and permutation/randomization methods. A detailed description of the proceedings of the workshop can be found in Christman et al. (2008).

One approach discussed at the workshop was to model all sources of variability in a complex regression model to model the probability that a turtle was caught as a function of hook type and a variety of other covariates, and then compare catch probabilities for J and C hooks through inferences based on this model. Such a model would have to account for many things, including the large proportion of zeros in the data. This approach would be quite challenging, and may be avoidable through the use of simpler, non-model-based statistical methods.

In terms of the question of, is there a difference between the J and C hooks, there is a much simpler way to approach this without building a complex model while simultaneously controlling all of those external sources of variability. In particular, Mantel-Haenszel (M-H) methods are appropriate here. M-H methods are appropriate for testing and quantifying association between two categorical variables like two binary variables such as hook type (J or C) and whether an animal was caught (Yes or No) while controlling for one or more categorical nuisance variables. That is, they are appropriate for analyzing the association within multiple (or stratified) 2×2 tables. Since we are only concerned with making a comparison between hook types we could create a 2×2 table to characterize the data for each line. This would provide for a simpler and more direct approach that gets right to the hypothesis test that we are interested in.

1.6 ORGANIZATION OF THESIS

We want to use a nonparametric test that uses the fewest assumptions as possible. Two of the approaches that we consider are the M-H test and the other is the permutation test, which is closely related to a randomization test. As the first goal of the thesis we will discuss more thoroughly what permutation and M-H tests are and establish that under certain situations the permutation test is the same thing as a M-H test, which is summarized in Chapter 2. We will discuss the applicability of classical M-H methods to the turtle bycatch problem and discuss violations of standard M-H assumptions that are likely to be present in the data collected by the EPRSTP. In Chapter 3, we will present alternative forms of the M-H test and then go over details of our simulation analysis in Chapter 4. Through simulation, we will investigate the effects of such assumption violations on the performance of the classical M-H test as well as generalizations of the M-H test that have been proposed in the statistical literature. We will present several scenarios or patterns used in our simulations which are designed to be realistic given the nature of the data collection scheme used by the EPRSTP, and which represent differing degrees of failure of the M-H assumptions. In Chapter 5, we will draw conclusions from the simulations results and use those to draw conclusions about the real data. Finally, in Chapter 6, we will make some suggestions to possible future work.

CHAPTER 2

MANTEL-HAENSZEL METHODS & PERMUTATION TESTS

Our data can be summarized into multiple 2×2 tables, one table for each line, set up as in Table 2.1 below.

Table 2.1: 2×2 Contingency Table

	Turtle Caught	**No Turtle Caught**	
J hook	x_i	$n_i - x_i$	n_i
C hook	y_i	$m_i - y_i$	m_i
	t_i	$N_i - t_i$	N_i

Here i is the specified line from 1 to 539, x_i is the number of turtles caught on J hooks on the i^{th} line, y_i is the number of turtles caught on C hooks on the i^{th} line, t_i is the total number of turtles caught on the i^{th} line, n_i is the total number of J hooks on the i^{th} line, m_i is the total number of C hooks on the i^{th} line, n_i - x_i is the number of J hooks on the i^{th} line that did not catch a turtle, m_i - y_i is the number of C hooks on the i^{th} line that did not catch a turtle, and N_i is the total number of hooks on the i^{th} line.

We need to find whether there is an association between the row (type of hook) and column (turtle caught or no turtle caught) variables. Using Fisher's Exact test or a Chi Square test one could analyze a single 2×2 table and conclude if the bycatch probabilities on the two types of hooks are significantly different. We need a similar type of analysis that can analyze multiple lines at once because we have data collected from 539 lines. So, we would be dealing with 539 2×2 tables.

In this Chapter we will introduce M-H methods and briefly discuss their relationship with the permutation test.

2.1 MANTEL-HAENSZEL TEST

The Mantel-Haenszel (M-H) test is a test for detecting average partial association between two categorical variables while adjusting for the effects of one or more extraneous variables. The combinations of the levels of these extraneous variables are treated as strata, within which the two-way relationship (or conditional association) is of interest. The M-H methods are appropriate for detecting association within i stratified 2×2 tables. The test statistic is given by

$$T_{\text{MH}} = \frac{\{\sum_i (x_i - n_i t_i / N_i)\}^2}{\sum_i n_i m_i t_i (N_i - t_i)/\{N_i^2 (N_i - 1)\}}, \tag{1}$$

which approximately has a chi-squared distribution with $df = 1$ for a large sample. The M-H method is most powerful for detecting patterns of association across I strata when there is a tendency for the majority of the differences $X_i - \frac{n_i t_i}{N_i}$ among partial tables to consistently have the same sign. If the differences are in opposite directions the M-H method may fail to detect the association.

It is also inappropriate to combine results across multiple partial tables into a single 2×2 marginal table then to be analyzed by an odds ratio (for example). This analysis could lead to inaccurate results. It is possible for the marginal table to exhibit an association that is completely different from the individual partial tables. This is known as Simpson's Paradox. A detailed description of Simpson's Paradox can be found in Agresti (1996).

2.2 MANTEL-HAENSZEL ODDS RATIO ESTIMATOR

The so-called common odds ratio provides a way to estimate the strength of association between J hooks and C hooks rather than doing a simply hypothesis test comparing J hooks and C hooks, which would be a test on the overall or marginal odds ratio collapsing across strata. When this association is stable across partial tables, we can estimate an assumed common value of the I true odds ratios. The M-H estimator of the common odds ratio equals

$$\hat{\Psi}_{MH} = \frac{\sum_i \left[n_i (m_i - y_i) / N_i \right]}{\sum_i \left[(n_i - x_i)(y_i) / N_i \right]} = \frac{\sum_i R_i}{\sum_i S_i}, \tag{2}$$

and the standard error of $\hat{\Psi}_{MH}$ can be consistently estimated by

$$\sqrt{\sum_i (R_i - \hat{\Psi}_{MH}\, S_i)^2 / (\sum_i S_i)^2}\ . \tag{3}$$

The M-H odds ratio estimator is a powerful summary of evidence against the hypothesis of conditional independence, as long as the sample associations fall primarily in a single direction. Further detail of the M-H odds ratio estimator can be found in Kuritz et al. (1988).

2.3 MANTEL-HAENSZEL RISK DIFFERENCE ESTIMATOR

Another tool related to the odds ratio is called risk difference. The common risk difference also provides a way to estimate the strength of association between J hooks and C hooks. M-H estimator of the common risk difference equals

$$\hat{\delta}_{MH} = \sum_i (x_i m_i / N_i - y_i n_i / N_i)/(\sum_i n_i m_i / N_i),$$

and the standard error of $\hat{\delta}_{MH}$ can be consistently estimated by

$$\sqrt{\sum_i L_i / \sum_i (n_i m_i / N_i)^2},$$

where

$$L_i = [x_i(n_i - x_i)m_i^3 + y_i(m_i - y_i)n_i^3]/(n_i m_i N_i^2).$$

The risk difference is often used in epidemiology settings and is preferred when one wants to quantify the "public health impact" of an exposure to a disease. In our case we will use the risk difference to quantify the number of turtles saved by implementing the replacing of J hooks with C hooks. Further description of the risk difference can be found in Greenland et al. (1985).

2.4 EXACT MANTEL-HAENSZEL INFERENCE

For stratified 2×2 tables, conditional on the marginal totals, the M-H test of conditional independence depends on the cell counts through $\sum_i x_i$. Conditional on the table margins, the hypergeometric distribution determines probabilities for $\{x_i, i = 1,..., I\}$ and therefore determines the distribution of their sum.

The null hypothesis of conditional independence states that all conditional odds ratios $\{\theta_{XY(i)}\}$ equal 1. A "positive" conditional association corresponds to the one-sided alternative: $\bar{\theta}_{XY} > 1$, where $\bar{\theta}_{XY} = \frac{1}{I}\sum_i \theta_{XY(i)}$. The M-H P-value for this alternative is hypergeometric probability that $\sum_i x_i$ is at least as large as observed, for the fixed marginal totals. Similarly, for the alternative $\bar{\theta}_{XY} < 1$, the P-value equals the hypergeometric probability that $\sum_i x_i$ is no greater than observed. Two-sided alternatives can use a two-tailed probability of those outcomes at least as unlikely as the observed one. Additional information about the exact version of the M-H test can be found in Agresti (1996).

2.5 PERMUTATION TEST & ITS RELATIONSHIP TO EXACT MANTEL-HAENSZEL

The goal of this section is to clarify the connection between the permutation test and the exact M-H test. The classical definition of a hypergeometric distribution goes as follows. You have an urn with 11 balls, 3 are red, 8 are black and a sample of 6 balls is chosen. What proportion of them is red? This is a textbook example of the hypergeometric distribution. In terms of our turtle problem, you have a line with n hooks, n_1 of J hooks and n_2 of C hooks. We sample k hooks (the hooks that caught a turtle) and we count the number x of those hooks which are of type J. The distribution on x is hypergeometric.

Now, suppose a line has 6 J hooks and 5 C hooks for a total of 11 hooks. Suppose 3 turtles are caught on this line of which 2 were on J hooks and 1 on a C hook. All other hooks either caught no turtle or perhaps caught a fish. So $2/6$ of the J hooks caught turtles and $1/5$ of the C hooks caught turtles. Using this data we could compute the odds ratio as one form of the test statistic to test the hypothesis of no association. If the type of hook doesn't matter then there are many different ways to assign J and C hooks to these 11 positions such that you have 6 J hooks and 5 C hooks. The pattern of hooks along the line could be alternating or it could be that the first 6 hooks are J hooks and the last 5 hooks are C hooks, etc. For every possible layout of J and C hooks along a line one could compute the odds ratio associated with it.

The odds ratio is determined by how many of the J hooks caught turtles. As soon as you know that 2 of these J hooks caught turtles, the number of total J hooks there are, how many C hooks there are, and how many total turtles were caught we know everything about the line. So each distinct value of the odds ratio corresponds to a distinct value of x, the number of J hooks that caught a turtle. The permutation distribution of the odds ratio is the permutation distribution of x. Once x is known the odds ratio is known. Out of all the possible permutations of the assignments of J and C hooks to the line, what proportion of them have an odds ratio (or equivalently have a value of x) at least as extreme as the one we obtained. This is the permutation test p-value. It quantifies how extreme the observed odds ratio is relative to the permutation distribution that assumes the J and C hook labels do not matter, and it is given by a hypergeometric probability.

This setup of the permutation distribution on x fits the definition of a hypergeometric probability distribution. There is equivalence between Fisher's Exact test in terms of this turtle problem (the classic definition of the hypergeometric distribution) and Fisher's Exact test being a permutation test. In the data that we have analyzed there is not a single line, but there are actually 539 lines. However, this does not change the argument. The same process can be done for every line and the resulting permutation p-value is that given by the exact version of the M-H test based on hypergeometric probabilities. If we had two lines then one would look at all the permutations in the first line and all permutations in the second line and compare the actual data observed in the first line to how extreme it is of getting 3 J hooks with turtles in two lines. This can be extended to many lines. Simply take all the permutations possible from each line and compare how extreme the observed case is to getting 3 J hooks with turtles in 539 lines like in the case we have. It is hypergeometric in each line and it is hypergeometric with multiple lines.

CHAPTER 3

LIMITATIONS OF MANTEL-HAENSZEL TESTS & ALTERNATIVES

One of the limitations of the M-H test is an underlying assumption of independence among the experimental units. The source of the concern here is there is potential for correlation in the catch event among the hooks that we are using to form the 2×2 tables. For example, imagine a fish or turtle being caught and flopping around on a hook. Then the two hooks adjacent to this hook may be directly affected. Standard M-H methods are not designed to accommodate such dependence. In addition to independence, the M-H test also assumes a constant success probability within each row of the stratified 2×2 tables. This assumption too is of some concern for the turtle bycatch data. As mentioned above, differences in the depth at which the hooks are set within a line may make for non-constant hooking probabilities, or long lines that pass through waters that have different environmental conditions or habitats may have similar effects. Fortunately there are a number of generalizations of the M-H methods that exist in literature designed for correlated data. Because correlation and heterogeneity among Bernoulli events both lead to a similar phenomenon of over-dispersion in the sum of such events, one can expect these methods also to work well in the presence of non-constant catch probabilities along a line. Below we investigate these methods under both scenarios.

Because the derivation of both the asymptotic and exact sampling null distribution of the M-H test statistic depends strongly on the assumption of independent, identical trials, the appropriateness of the M-H test is in doubt when this assumption is invalid. Liang (1985) proposed two new test statistics that relax the assumptions of the M-H test and allow for possible correlation and/or heterogeneity of the response probabilities. Below we present Liang's tests.

As stated above in Chapter 2, the M-H test, under the hull hypothesis where the odds ratio is equal to 1, can be written as in equation (1). Below Liang uses the same numerator as T_{MH} but replaces the denominator with its sample variance formula $\Sigma (x_i - n_i t_i / N_i)^2$ which results in

$$T_1 = \frac{\{\sum_i (x_i - n_i t_i / N_i)\}^2}{\sum_i (x_i - n_i t_i / N_i)^2}.$$

By taking the numerator of T_{MH}, which has a mean of 0 under the null hypothesis, and dividing by its standard deviation, results in T_1 being the square of an approximately standard normal random variable. Liang's second proposed test statistic is

$$T_2 = \frac{\{\hat{\Psi}_{MH} \log \hat{\Psi}_{MH}\}^2}{\sum_i L_i / \sum_i (n_i m_i / N_i)^2},$$

where the M-H estimator $\hat{\Psi}_{MH}$ can be defined as in equation (2) and the standard error of the $\hat{\Psi}_{MH}$ can be consistently estimated by equation (3). Under the hull hypothesis, where the odds ratio is equal to 1, both T_1 and T_2 have asymptotic chi-squared distributions with one degree of freedom.

CHAPTER 4

SIMULATION STUDY DESIGN

We want to evaluate all the above versions of the M-H test under different conditions and see whether or not they are appropriate methodologies and if so is there an advantage of one over another. In order to do this we needed to do a simulation study. In this study we wanted to generate data that mimic the real data that we have. One problem that we encountered is that we don’t know exactly how a turtle interacts with the hooks along a line. The data that we have were collected at the line level. So we know how many animals were caught on a specific line but what we don’t know is the location of the hook along the line the animal was caught on.

In the study we generated the data at the individual hook level. We then summarized them into stratified two way tables where we then analyzed them using different versions of the M-H Test discussed above.

4.1 OVERVIEW OF THE SIMULATIONS

Since we are not certain of how turtles interact with hooks, we generated data under three different scenarios. The differences between these three patterns come from how the probabilities of catching a turtle along a line were generated. The simulations are meant to account for a variety of factors that may affect catch rates along longlines: heterogeneity in hook probabilities across lines (found in all three patterns), heterogeneity in hook probabilities within a line (found in patterns 2 & 3), correlation among hooks within a line (found in pattern 3), and effect size (found in all three patterns). A comparison of all three patterns can be seen in Table 4.1, located at the end of this chapter.

In addition there were some factors that we simply controlled for in the experiment instead of trying to vary. We felt that it would be unrealistic to assume that there is not heterogeneity from line to line. For example, some lines are being soaked where many fish or turtles are located while other lines are being soaked where no fish or turtles are located.

We controlled the size of the effect to be detected by the M-H tests by manipulating the difference between the J and C hook catch probabilities. Three effect sizes, none, small and large, were examined in each scenario, but the specific settings differed across scenarios so are summarized below.

We felt it would be unrealistic to assume that the level of catch probabilities would be the same across lines. For example, one line is pulled through a school of fish while another line came in contact with a small grouping of turtles and another line never came in contact with any sort of sea creature. This seemed realistic so we included it in all of the simulations.

To make these simulations as real as possible we kept a number of the parameters the same as our real data. Each simulation was composed of 539 lines and each line was the exact length (total number of hooks) as the observed data. We created a constant pattern of difference between the lines to create the type of variability across strata that the M-H test is designed to control for and which are to be expected in the longline fishery.

4.2 SIMULATION PATTERN 1

Pattern 1 is the most simplified scenario. It is when the classical assumptions hold under the M-H test. The probability of catching a turtle remains constant throughout the entire line. There is a P_J (probability of catching a turtle on a J hook) and a P_C (probability of catching a turtle on a C hook) for each line. So within a given line the number of turtles hooked on a J hook and the number of turtles hooked on a C hook will respectively both be binomials because those will be the number of successes out of the number of independent trials, each with a constant success probability. We initialized the P_J by calculating the actual proportion of turtles caught on J hooks.

$$P_J = \frac{167}{96{,}443} = 0.0017,$$

where 167 turtles were caught on 96,443 J hooks.

After establishing an initial value for P_J we imposed three different effect levels where $P_J = P_C$, $P_J = 1.4 \times P_C$, and $P_J = 1.8 \times P_C$ along a line. The effect size is the difference between J and C hook probabilities. The main idea is that in the no effect the probability of catching a turtle on a C hook is equal to that of a J hook. The small effect would have some increased probability of catching a turtle on a C hook equal to that of a J hook. Finally, the large effect would have a larger increased probability of catching a turtle on a C hook equal to that of a J hook. The effect sizes were set up where there were increased levels of difference between J and C hook probabilities and where the rejection rates were comparable. If the effect size is too large then everything gets rejected.

We also created a constant pattern of differences between the lines by setting $P_J = 0.0017$ (as mentioned above) for one-third of the lines $P_J = 1.5 \times 0.0017$ in a second third of the lines, and $P_J = 0$ in the remaining third of the lines. Our reasoning for making these settings is to mimic spatial variability in the prevalence of sea turtles are located, whereas other lines could be finished in unsuitable habitats thus never encounter a turtle.

4.3 SIMULATION PATTERN 2

The probability of catching a turtle was generated using a uniform probability between 0 and 0.05. This scenario allows the probabilities within a line to vary. So whether you catch a turtle on each hook is independent within a line. In this scenario we are violating the constant success probability assumption.

Again, we imposed the three different effect levels where $P_J = P_C$, $P_J = 1.05 \times P_C$, and $P_J = 1.1 \times P_C$ along a line. Here the effect sizes changed from Pattern 1. The effect sizes of 1.4 and 1.8 were much too large resulting in a 100% rejection rate for all tests. The effect sizes were lowered to make the rejection rates comparable.

We also created a constant pattern of differences between the lines by setting the first third of the lines equal to what is generated, second third equal to 1.5 times the value generated, and the last third is equal to zero. Like in Pattern 1, we did this to mimic how some lines are going to encounter more turtles than others and to reflect other between-line sources of variability.

4.4 SIMULATION PATTERN 3

Similar to Pattern 2, Pattern 3 represents violations of the assumptions under the classical M-H test. This pattern violates the constant success probability assumption and the assumption of independence under the M-H test. This pattern is designed to have correlated binary responses. In other words, hooks closer together will have a higher correlation than hooks further apart. For example, the line goes out and some of it goes into water that is unsuitable for turtle migratory condition. Or some may go into a school of fish but you want that probability to be higher in one region of the line and lower in another. We feel this scenario is theoretically reasonable but have no evidence to suggest this is apparent in real data.

We separated each line up into four approximately equal sized subgroups and assigned starting probabilities for each subgroup. The probabilities of each of the four subgroups of a given line were set as (0.01, 0.02, 0.03, 0.04). We generated correlated binary data according to the method of Oman and Zucker (2001). In particular, we generate a continuous variable Z_i with an AR(1) dependency structure using the formula

$$Z_i = U_i Z_{i-1} + (1 - U_i)\varepsilon_i \quad (i \geq 2),$$

where U is Bernoulli(0.5), Z_1 is Normal(0,1) and the ε_is are independent Normal(0,1) random variables. Then correlated binary variables are created via

$$Y_i = 1_{(Z_i \leq \theta_i)},$$

where $\theta_i = \Phi^{-1}(p_i)$, Φ is the standard normal distribution function, and p_i is the desired probability of catching a turtle. The resulting series $\{Y_i\}$ does not have an exact AR(1) structure, but as described by Oman and Zucker (2001), has the desired marginal mean (sequence of success probabilities) and a correlation structure that is close to AR(1), subject to the constraints of the specified marginal and the binary nature of the data. The exact theoretical lag-1 autocorrelation among the Y_i s is a complex function of their marginal mean and the parameters used to generate $\{Z_i\}$, but the estimated sample lag-1 autocorrelation for these binary deviates was about 0.36.

Again, we imposed the three different effect levels where $P_J = P_C$, $P_J = 1.1 \times P_C$, and $P_J = 1.5 \times P_C$ along a line. Here the effect sizes changed once again. The effect sizes of 1.4 and 1.8 were again much too large. All of the hypothesis tests were being rejected. The effect sizes were lowered to make the rejection rates comparable.

We also created a constant pattern of differences between the lines by making the first third of the lines equal to what is generated, second third equal to 1.5 times the value generated, and the last third is equal to zero, thus creating heterogeneity across lines. Like in Patterns 1 and 2, we did this to mimic how some lines are going to encounter more turtles than others.

4.5 DETAILS OF THE ANALYSIS

Once the data had been generated under each scenario explained above they were analyzed using four different test statistics. The first two are the exact version and an asymptotic version of the M-H test. Both of these versions discussed in Chapter 2 are the most familiar and can be found in most statistical programming packages. Also included in the analysis are two generalizations of the M-H test which were explained in Chapter 3.

Each simulation generated the data at the individual hook level and then we summarized that simulated set of 539 fishing lines into a stratified 2×2 table. From there we calculated each of the test statistics and the associated p-value. We simulated 1,000 data sets and calculated the rejection rates, which is the proportion of times the significance level exceeded the p-value. The rejection rate is the empirical size of the test when data are generated under the null, and is the empirical power when the null is false. Significance levels of 0.05 and 0.01 were used in the simulation study.

Table 4.1: Simulation Patterns

	Pattern 1			Pattern 2			Pattern 3		
Type of Pattern Along a line	**Homogeneity of Catch Probabilities and Independence within a Line**			**Catch Probabilities Generated Uniformly within a Line**			**Correlated Catch Probabilities within a Line**		
Proportion of lines	1/3	1/3	1/3	1/3	1/3	1/3	1/3	1/3	1/3
Heterogeneity Across Lines	(P_J, P_C)	$1.5 \times (P_J, P_C)$	$0 \times (P_J, P_C)$	(P_J, P_C)	$1.5 \times (P_J, P_C)$	$0 \times (P_J, P_C)$	(P_J, P_C)	$1.5 \times (P_J, P_C)$	$0 \times (P_J, P_C)$
Simulation	1	2	3	4	5	6	7	8	9
Effect Levels	NONE	SMALL	LARGE	NONE	SMALL	LARGE	NONE	SMALL	LARGE
	$P_J = P_C$	$P_J = 1.4 \times P_C$	$P_J = 1.8 \times P_C$	$P_J = P_C$	$P_J = 1.05 \times P_C$	$P_J = 1.1 \times P_C$	$P_J = P_C$	$P_J = 1.1 \times P_C$	$P_J = 1.5 \times P_C$
Heterogeneity Within Line	P_J is Constant along a line			P_J generated from Uniform(0,0.05)			P_J generated with Correlation: each line split into 4 subgroups		

CHAPTER 5

RESULTS

5.1 SIMULATION RESULTS

In the results from the no effect setting within each of the three patterns yield conclusions regarding whether or not the tests have the correct nominal size (0.05 or 0.01). This can be seen below in Tables 5.1 and 5.2. Table 5.1 contains the proportion of p-values that are less than 0.05. We are also interested in the rejection rates at each increasing level of hook effect, which is called the empirical power. Power can be the determining factor when choosing which test is the best. It can be used for comparing multiple tests of the correct size and the one with the greater power is considered the best test.

The motivation behind this research is to determine if M-H methods provide an appropriate way to analyze real turtle bycatch data. So, if we are generating these data under a realistic scenario and we take that data and analyze it with our M-H methods are we really getting a valid test?

The initial simulations that we ran were simply comparing the M-H test statistics that were built into R which include both asymptotic versions (with and without the continuity correction) and the exact version. These simulations show that using the M-H test to analyze data that violate certain assumptions does really matter. This conclusion is drawn from Tables 5.1 and 5.2. In Table 5.1 for example we are expecting to see 5%

rejection rate (proportion of the simulations where we rejected) in each of the no effect columns of each of the three patterns. Pattern 3 results suggest that we have a 0% rejection rate, which is too conservative. This conservativeness lowers the power of the test and thus, the basic M-H test is not appropriate to analyze data that were correlated as in the simulated Pattern 3. The real data are possibly correlated from one hook to the next. We should be using a method that allows for that and doesn't assume independence.

In addition to the exact version and the asymptotic version of the M-H test statistic built into R we imposed two additional versions of the M-H test. Additional simulations were ran for these two new test statistics which can also be seen in Tables 5.1 and 5.2. These simulation results include Liang's test statistics and compare them to the basic M-H test statistics that are built into R. These new tests are performing roughly the same as the M-H in Patterns 1 and 2. They are doing significantly better in Pattern 3. This suggests the test statistics T_1 and T_2 work well under all three scenarios. We get close to nominal size in all of three patterns using either of Liang's test statistics. So it doesn't matter if the data are correlated or heterogeneous within a line or not. Liang's second test statistic T_2 seems to have slightly stronger power in some case but it appears from the analysis that both of Liang's test statistics perform equally well. Liang's test statistics are definitely the better proposed generalized version of the M-H test because Liang's test statistics are not limit to the assumptions under the basic M-H.

Table 5.1: Rejection Rates

Alpha=0.05

	Pattern 1			Pattern 2			Pattern 3		
	NO	SMALL	BIG	NO	SMALL	BIG	NO	SMALL	BIG
EXACT	**0.041**	**0.306**	**0.671**	**0.057**	**0.734**	**0.897**	**0.000**	**0.321**	**0.789**
TMH	**0.049**	**0.335**	**0.706**	**0.057**	**0.743**	**0.899**	**0.000**	**0.332**	**0.794**
LIANG-T1	**0.052**	**0.333**	**0.705**	**0.059**	**0.735**	**0.901**	**0.055**	**0.768**	**0.974**
LIANG-T2	**0.052**	**0.333**	**0.703**	**0.060**	**0.746**	**0.905**	**0.055**	**0.769**	**0.975**

Table 5.2: Rejection Rates

Alpha=0.01

	Pattern 1			Pattern 2			Pattern 3		
	NO	SMALL	BIG	NO	SMALL	BIG	NO	SMALL	BIG
EXACT	**0.013**	**0.133**	**0.433**	**0.016**	**0.501**	**0.747**	**0.000**	**0.076**	**0.445**
TMH	**0.013**	**0.145**	**0.461**	**0.017**	**0.507**	**0.754**	**0.000**	**0.076**	**0.453**
LIANG-T1	**0.014**	**0.141**	**0.462**	**0.017**	**0.504**	**0.751**	**0.018**	**0.513**	**0.892**
LIANG-T2	**0.014**	**0.140**	**0.451**	**0.017**	**0.516**	**0.758**	**0.018**	**0.522**	**0.895**

5.2 REAL DATA RESULTS

In Tables 5.3 and 5.4 we report some real data analysis for two versions of the M-H test build into R software and both of Liang’s test statistics. Note that there are some differences between the two versions found in the software and Liang’s statistics because the two versions of the M-H test built into the software rely strongly on the assumption of independent, identical trials whereas neither of Liang’s statistics requires such an assumption.

From Table 5.3 it can be seen that using either of Liang’s test statistics we can conclude that the number of turtles caught on J hooks is significantly higher than the number of turtles caught on C hooks. There is evidence to suggest that these C hooks are better at catching few turtles. In fact, from Table 5.4, we see the odds ratio of catching a turtle on a J hook is estimated to be 1.97 meaning the odds of catching a turtle on a J hook is estimated to be 97% higher than on a C hook. The M-H odds ratio $\left({}^{\text{J hook}}\!/_{\text{C hook}}\right)$ is greater than one and the risk difference (J hook – C hook) is 0.000976 with a 95% confidence interval that does not contain zero, as shown in Table 5.4. This implies the difference in hooking risk between J and C hooks is statistically significant.

Table 5.3: Turtle Result – Rejection Rates

	T-Statistic	P-value
EXACT	[*]	0.00000109
TMH	23.68	0.00000114
LIANG-T1	19.36	0.00001083
LIANG-T2	19.85	0.00000837

Table 5.4: Turtle Result – M-H Estimators and Standard Errors

Odds Ratio	1.97
Standard Error	0.301
95% CI OR	(1.50,2.60)
Risk Difference	0.000976
Standard Error	0.000194
95% CI RD	(0.00136,0.00060)

---

[*] No test statistic is given in the exact form of the test, only a p-value.

Analyzing the merchantable fish data gives similar results. From Table 5.5 it can be seen that using either of Liang's test statistics we can conclude that the number of fish caught on C hooks is significantly higher than the number of fish caught on J hooks. There is evidence to suggest that C hooks are better at catching more fish. In fact, from Table 5.6, we see that the odds of catching a fish on a J hook is estimated to be 0.72 meaning the odds of catching a fish on a C hook is estimated to be 38% higher than on a J hook. The M-H odds ratio $\left({}^{\text{J hook}}\!\big/_{\text{C hook}}\right)$ is less than one and the risk difference (J hook – C hook) is -0.00592 with a 95% confidence interval that does not contain zero, as shown in Table 5.6. This implies the difference in hooking risk between J and C hooks is statistically significant.

Table 5.5: Fish Result – Rejection Rates

	T-Statistic	P-value
EXACT	*	0.00000000
TMH	76.4757	0.00000000
LIANG-T1	23.30481	0.00000134
LIANG-T2	29.01486	0.00000007

Table 5.6: Fish Result – M-H Estimators and Standard Errors

Odds Ratio	0.72
Standard Error	0.00189
95% CI OR	(0.672,0.779)
Risk Difference	-0.00592
Standard Error	0.00071
95% CI RD	(-0.00731,-0.00454)

CHAPTER 6

CONCLUSION

We generated data using different scenarios which are undoubtedly much simpler than real conditions, but which were chosen to capture the main complicating features that one might expect to occur in real life. After looking at these different patterns we determined that Patterns 2 and 3 violated assumptions under the M-H test. Which pattern generates the most realistic data? Our intuition tells us that Pattern 3 may be the most realistic because it would seem that hooks closer together would be correlated and would be more strongly correlated than hooks far apart. However, we do not know how turtles actually interact with hooks on a line we cannot be sure that Pattern 3 is the most realistic. In either case, if we continue to use the exact version or either of the asymptotic versions of the M-H test these could lead to incorrect inferences. The poor performance of the exact M-H test was clearly shown in Tables 5.1 and 5.2.

We would like to find a generalization of the M-H test that relaxes some of the assumptions and that can calculate sparse correlated binary data. The best case scenario would be if we could find a generalization of the M-H test that would accurately analyze data from any of those three patterns that we created simply because we don't have the actual hook level data to test to see how the real data behaves. It turns out that Liang's test statistics do just that. Either version of his test statistics shows evidence of good performance analyzing the turtle data. The result shows that C hook is recommended to use. C hooks caught significantly fewer turtles and caught significantly more fish when compared to J hooks.

6.1 FUTURE RESEARCH

One can develop complex regression models including Bayesian methods, Case-control models, and generalized linear mixed-effects models to analyze these data. The comparison of the inferences from such models and the inferences obtained from these MH methods can be suggested as future work.

A limited amount of hook-level data is available and methods to analyze hook level data can be developed and compared with line level approach developed in this thesis. One approach to analyzing hook level data while minimizing the data collection demands is to just collect the location at which a turtle was caught and the result of a neighboring hook. Then such data could potentially be analyzed using methods appropriate for matched-case-control designs.

REFERENCES

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. New York: John Wiley and Sons.

Beverly, S. and Chapman, L. (2007). Interactions between sea turtles and pelagic longline fisheries. Document WCPFC-SC3-EB SWG/IP-01 of the Scientific Committee Meeting of the Western and Central Pacific Fisheries Commission, 13-24 August, 2007, Hawaii, USA.

Christman, M., Hall, D., Hall, M., Kinas, P., Lennert, C., Manly, B., McCracken, M., Minami, M., Sims, M. and Thompson, S. (2008). Workshop on Turtle Bycatch Mitigation for Longline Fisheries: Experimental Design and Data Analysis. Inter-American Tropical Tuna Commission. Special Report 17.

Greenland, S. and Robins, J. (1985). Estimation of a Common Effect Parameter from Sparse Follow-Up Data. *Biometrika*, **41**, 55-68.

Kuritz, S., Landis, J. and Koch G. (1988). A General Overview of Mantel-Haenszel Methods: Applications and Recent Developments. *Annual Review of Public Health*, **9**, 123-60.

Liang, K. Y. (1985). Odds ratio inference with dependent data. *Biometrika*, **72**, 678-82.

Oman, S. and Zucker, D. (2001). Modelling and Generating Correlated Binary Variables. *Biometrika,* **88**, 287-90.

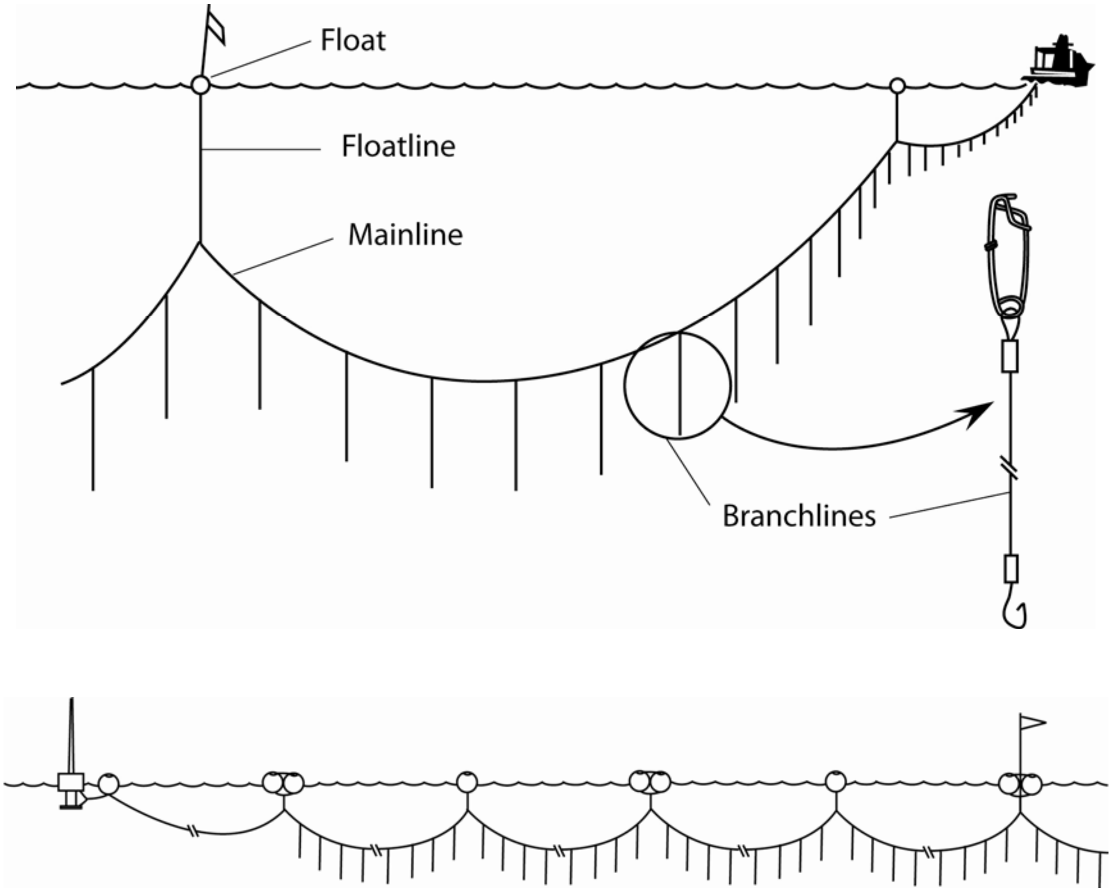
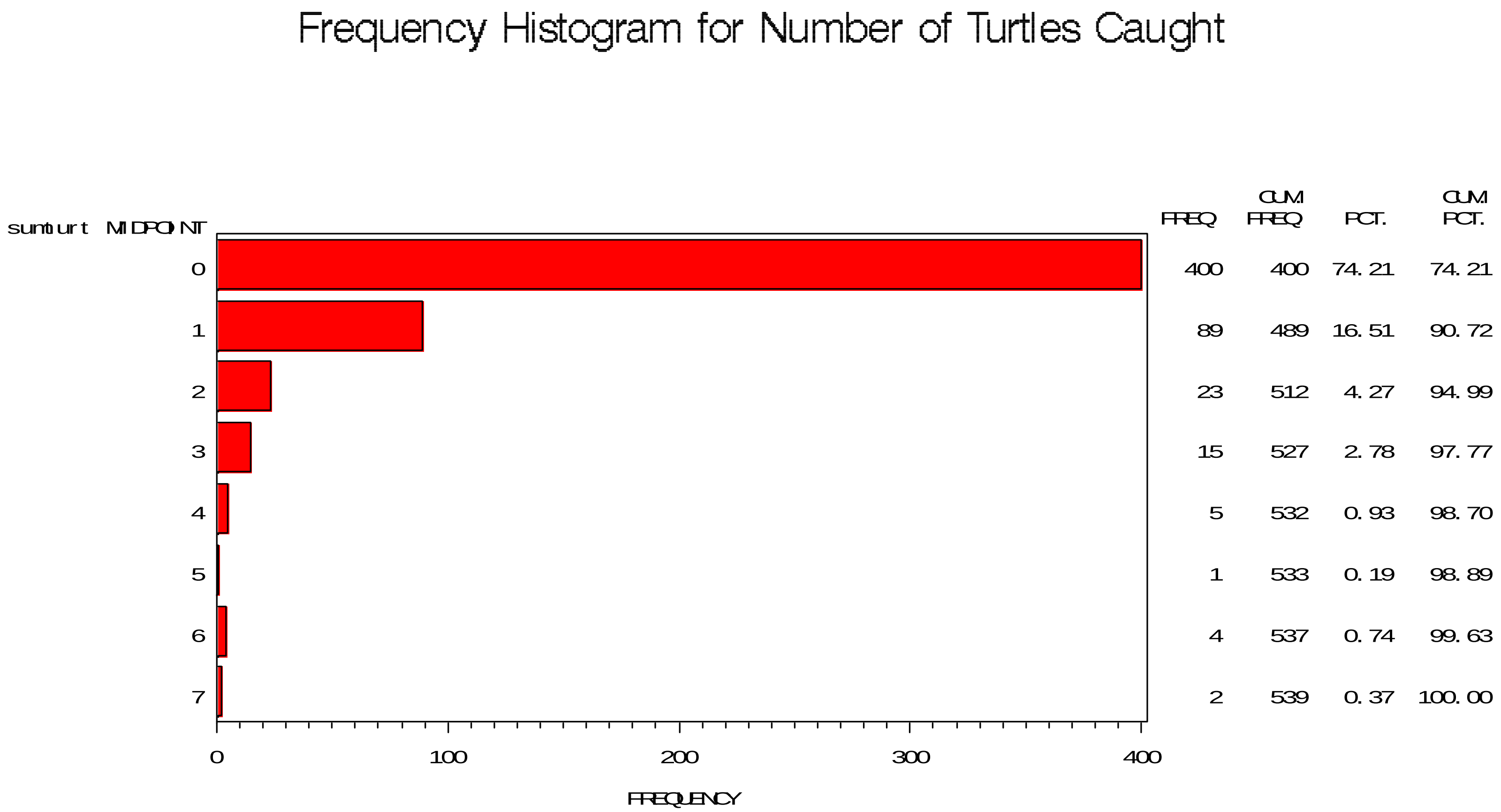
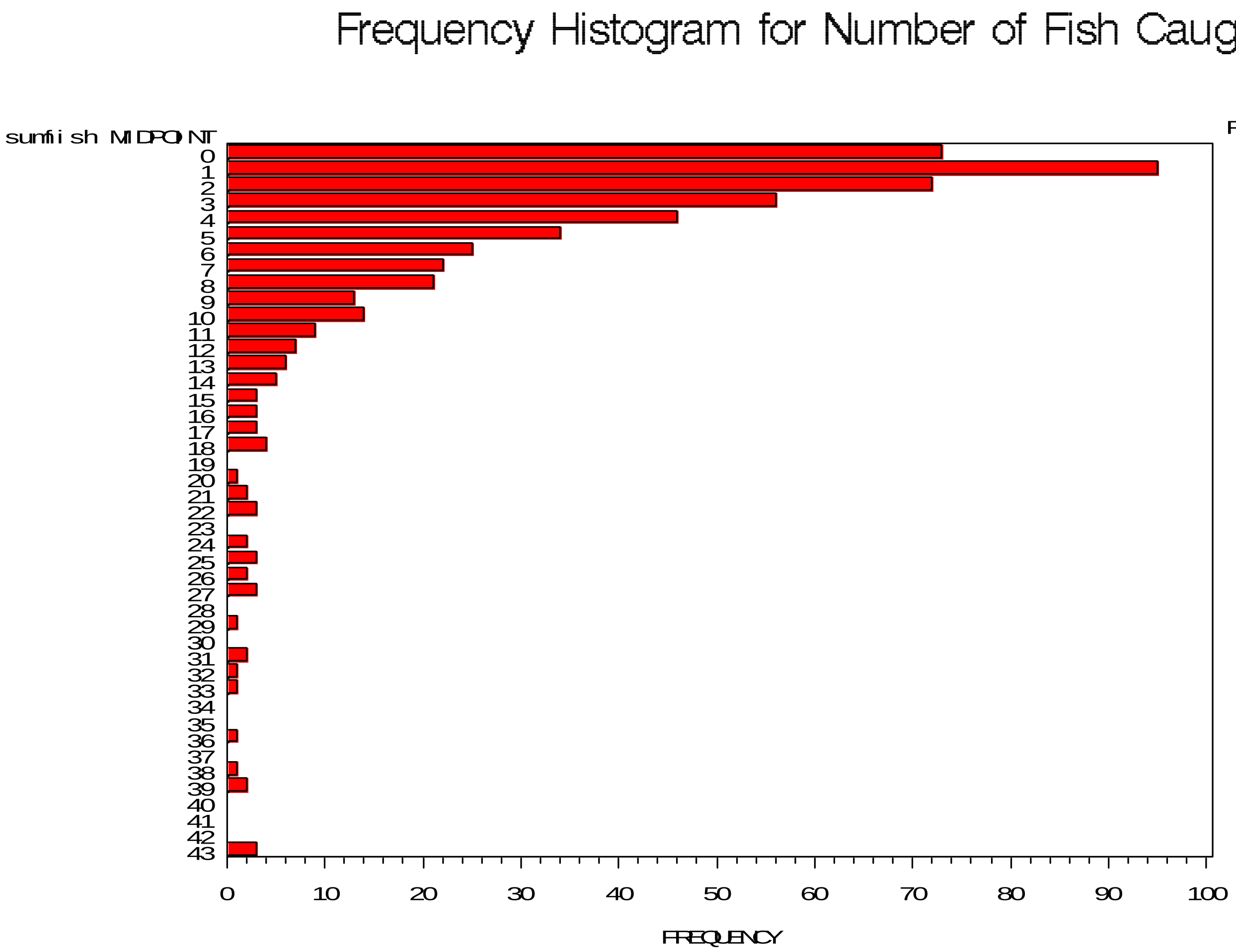


Diagram 1: Example of a Generic Longline from Beverly and Chapman (2007), with permission.



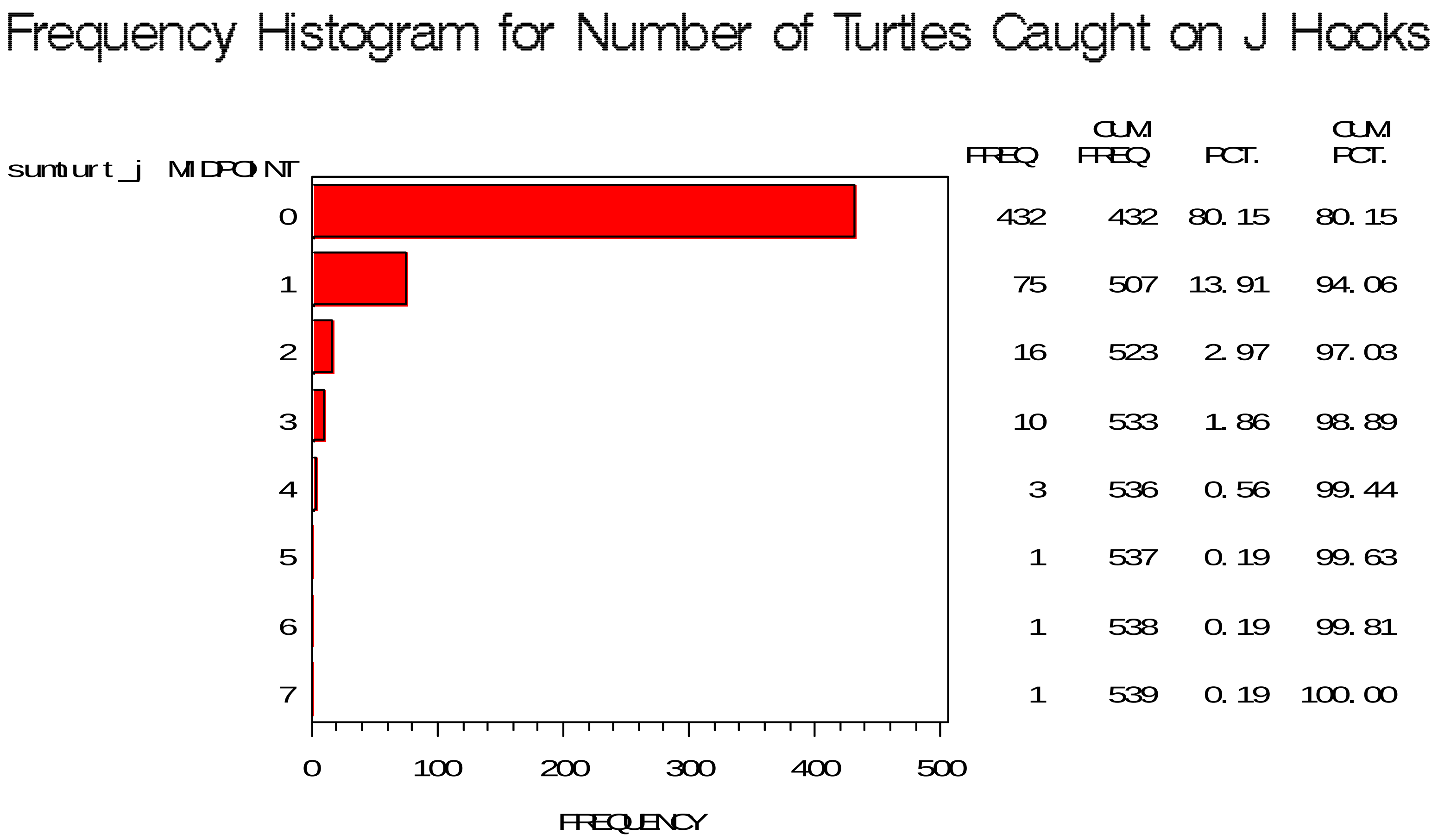
sumturt MIDPOINT	FREQ	CUM FREQ	PCT.	CUM PCT.
0	400	400	74.21	74.21
1	89	489	16.51	90.72
2	23	512	4.27	94.99
3	15	527	2.78	97.77
4	5	532	0.93	98.70
5	1	533	0.19	98.89
6	4	537	0.74	99.63
7	2	539	0.37	100.00

Figure 1: Frequency Histogram for Number of Turtles Caught



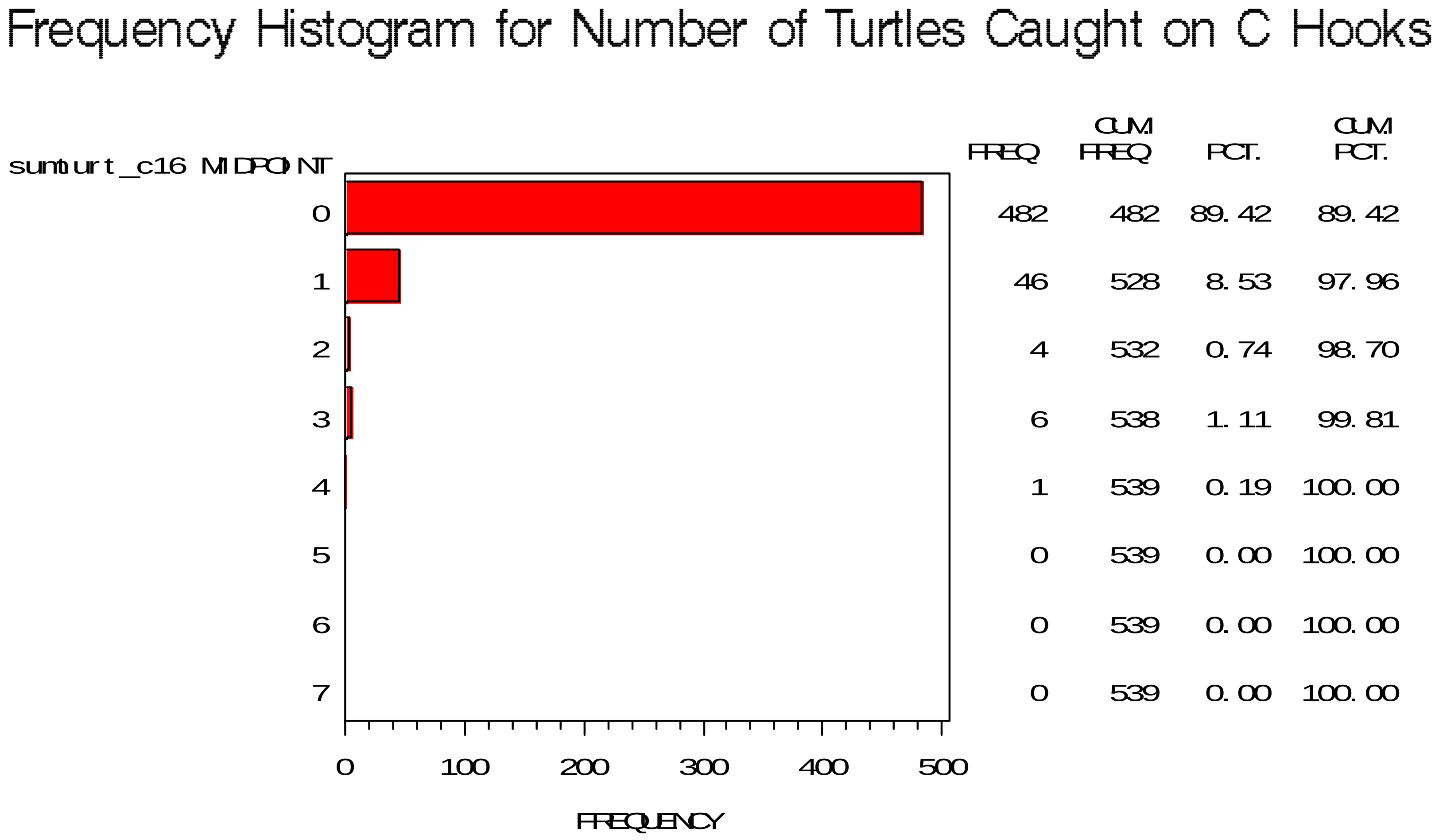
sunfish MIDPOINT	FREQ	CUM. FREQ	PCT.	CUM. PCT.
0	73	73	13 54	13 54
1	95	168	17 63	31 17
2	72	240	13 36	44 53
3	56	296	10 39	54 92
4	46	342	8 53	63 45
5	34	376	6 31	69 76
6	25	401	4 64	74 40
7	22	423	4 08	78 48
8	21	444	3 90	82 37
9	13	457	2 41	84 79
10	14	471	2 60	87 38
11	9	480	1 67	89 05
12	7	487	1 30	90 35
13	6	493	1 11	91 47
14	5	498	0 93	92 39
15	3	501	0 56	92 95
16	3	504	0 56	93 51
17	3	507	0 56	94 06
18	4	511	0 74	94 81
19	0	511	0 00	94 81
20	1	512	0 19	94 99
21	2	514	0 37	95 36
22	3	517	0 56	95 92
23	0	517	0 00	95 92
24	2	519	0 37	96 29
25	3	522	0 56	96 85
26	2	524	0 37	97 22
27	3	527	0 56	97 77
28	0	527	0 .00	97 77
29	1	528	0 19	97 96
30	0	528	0 00	97 96
31	2	530	0 37	98 33
32	1	531	0 19	98 52
33	1	532	0 19	98 70
34	0	532	0 00	98 70
35	0	532	0 00	98 70
36	1	533	0 19	98 89
37	0	533	0 00	98 89
38	1	534	0 19	99 07
39	2	536	0 37	99 44
40	0	536	0 00	99 44
41	0	536	0 00	99 44
42	0	536	0 00	99 44
43	3	539	0 56	100 00

Figure 2: Frequency Histogram for Number of Fish Caught



sumturt_j MIDPOINT	FREQ	CUM FREQ	PCT.	CUM PCT.
0	432	432	80.15	80.15
1	75	507	13.91	94.06
2	16	523	2.97	97.03
3	10	533	1.86	98.89
4	3	536	0.56	99.44
5	1	537	0.19	99.63
6	1	538	0.19	99.81
7	1	539	0.19	100.00

Figure 3: Frequency Histogram for Number of Turtles Caught on J Hooks



sumturt_c16 MIDPOINT	FREQ	CUM. FREQ	PCT.	CUM. PCT.
0	482	482	89.42	89.42
1	46	528	8.53	97.96
2	4	532	0.74	98.70
3	6	538	1.11	99.81
4	1	539	0.19	100.00
5	0	539	0.00	100.00
6	0	539	0.00	100.00
7	0	539	0.00	100.00

Figure 4: Frequency Histogram for Number of Turtles Caught on C Hooks

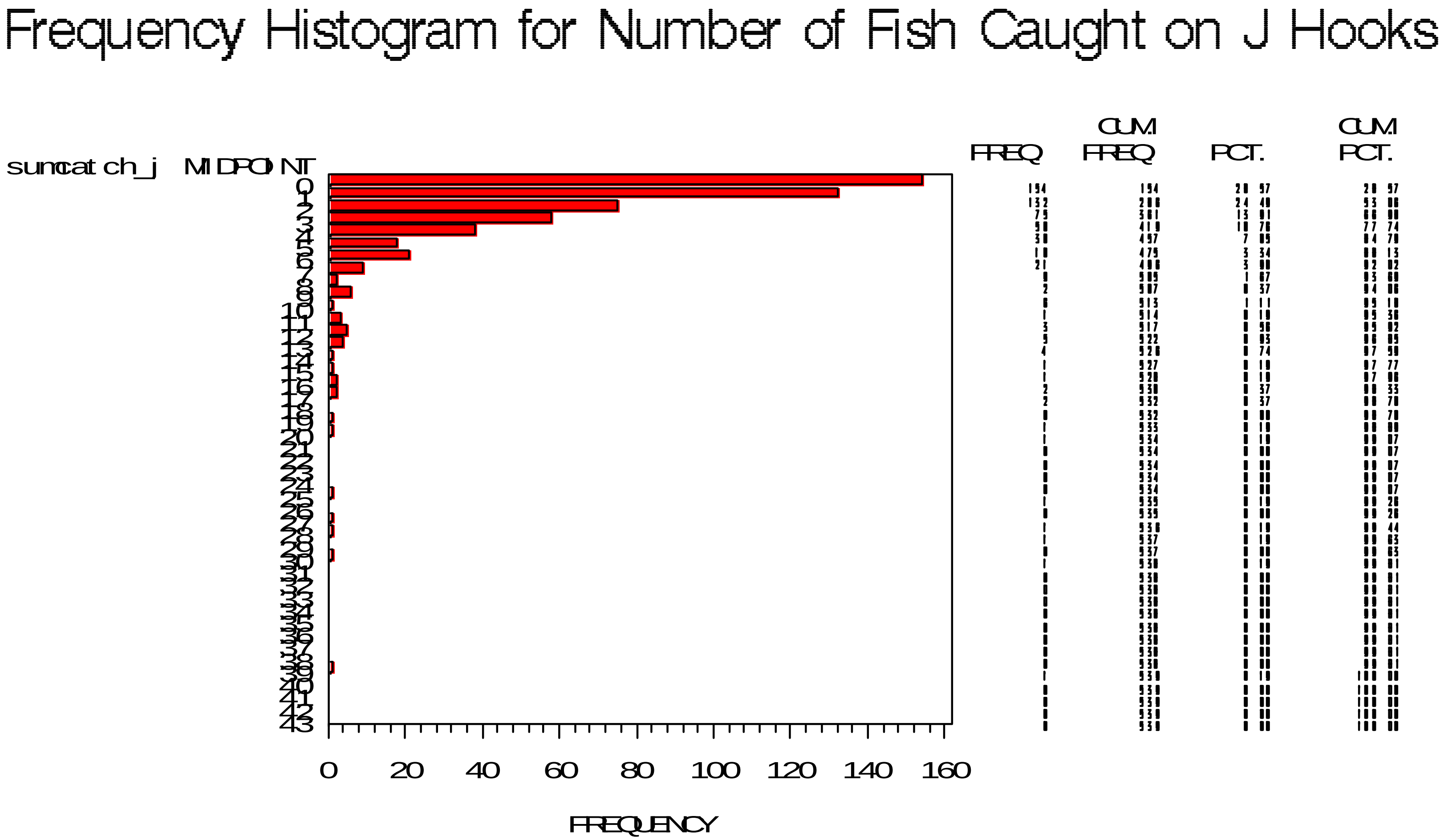


Figure 5: Frequency Histogram for Number of Fish Caught on J Hooks

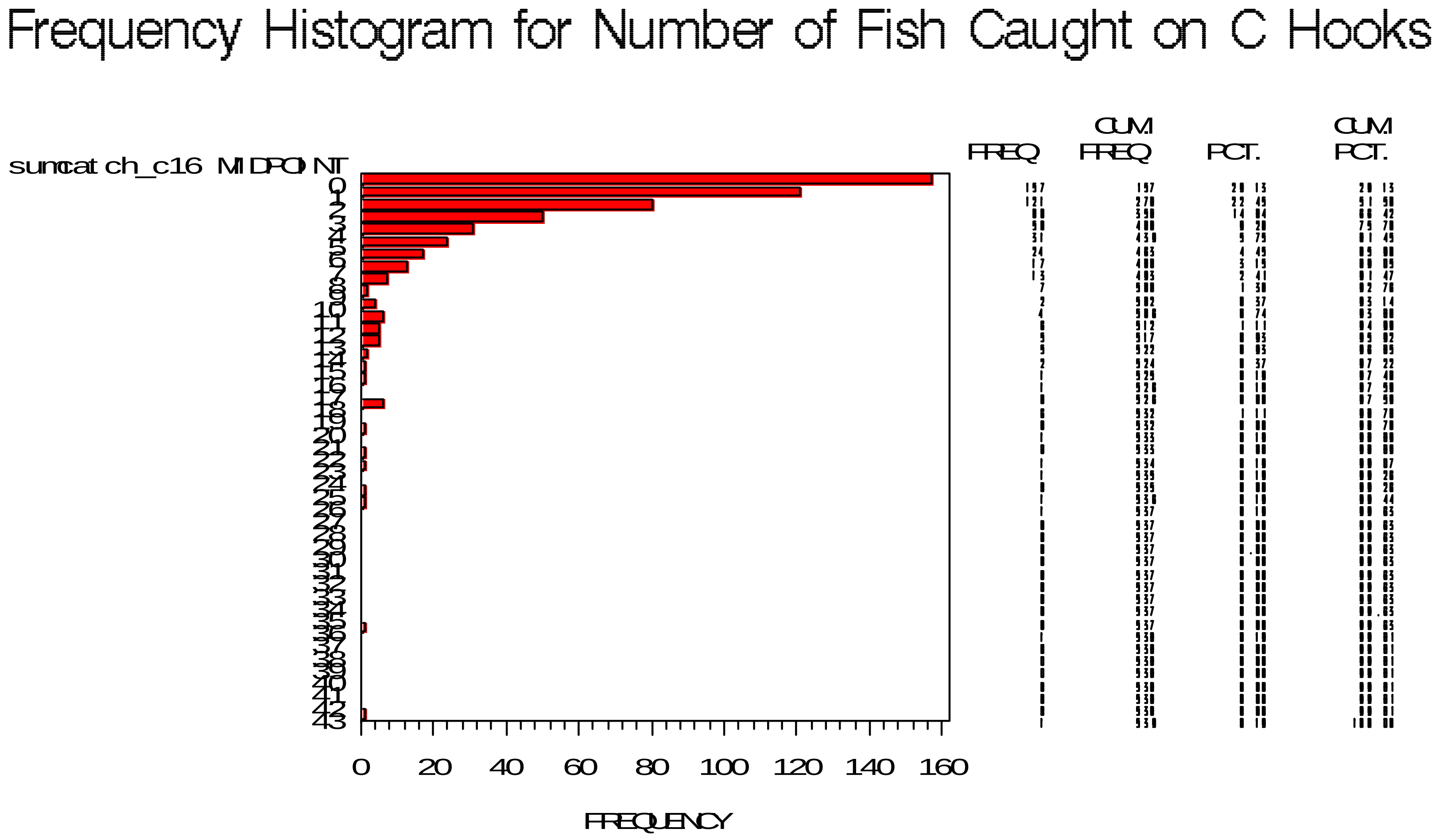


Figure 6: Frequency Histogram for Number of Fish Caught on C Hooks

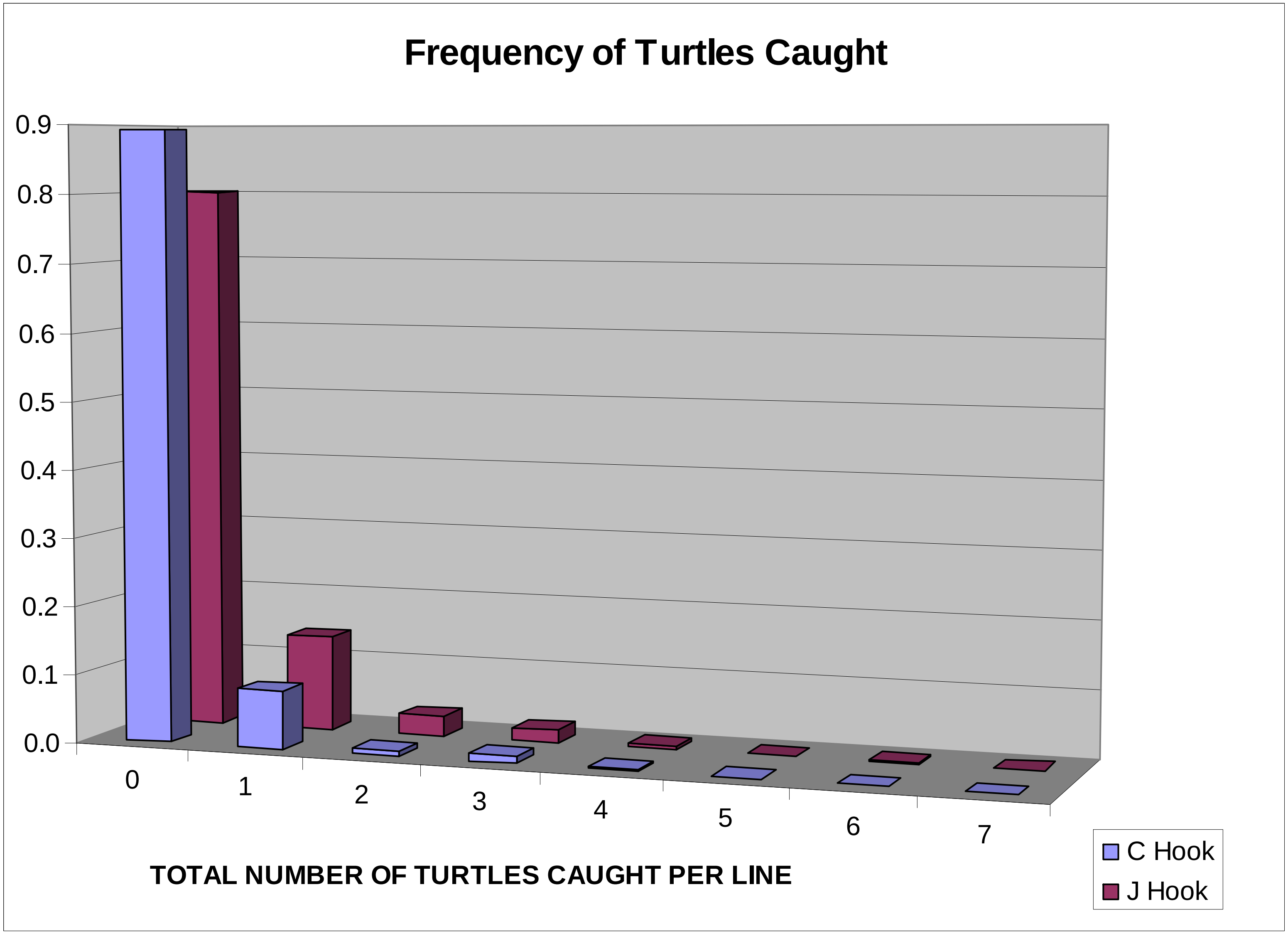


Figure 7: Frequency of Turtles Caught

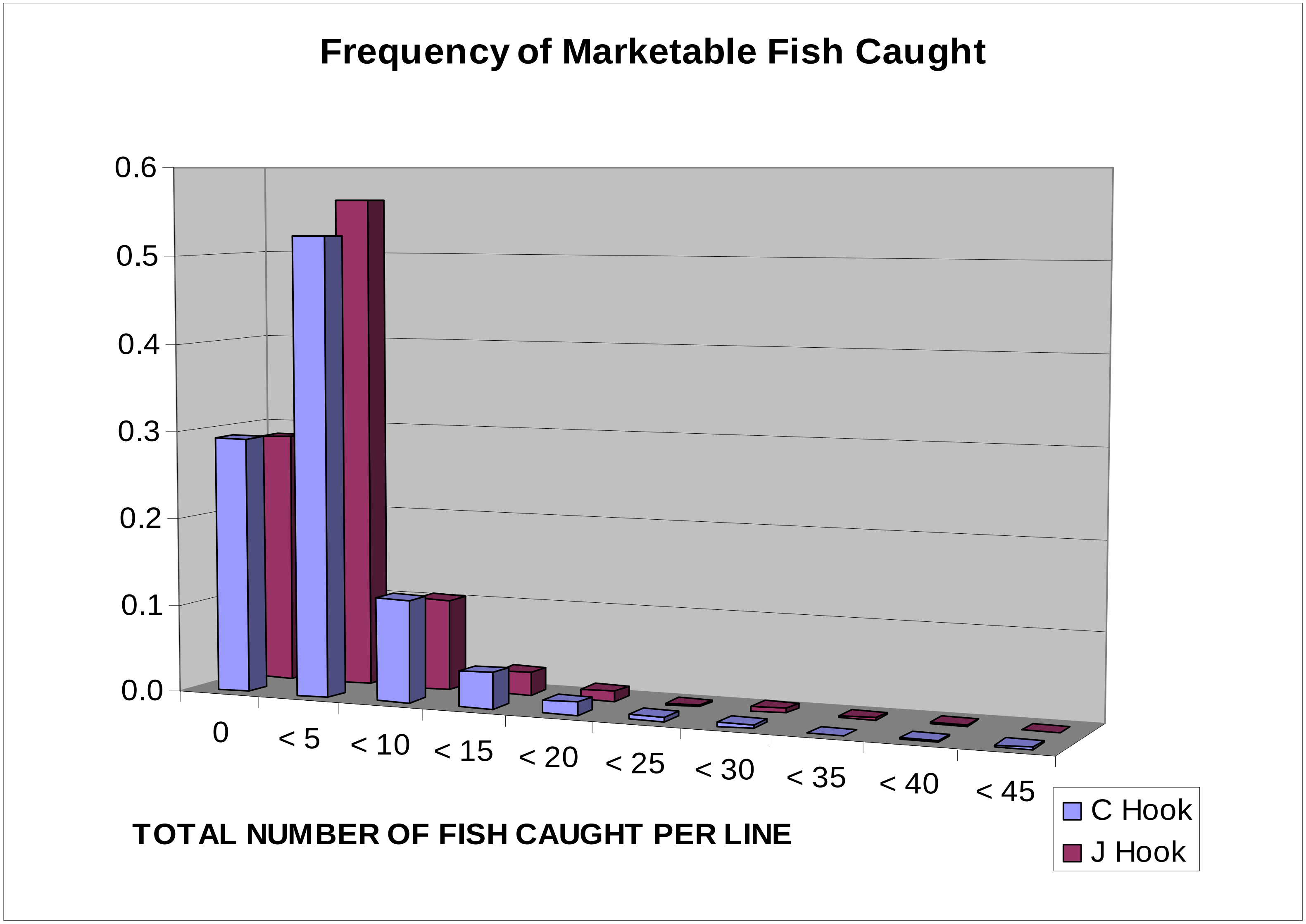


Figure 8: Frequency of Marketable Fish Caught