NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

RELIABILITY ANALYSIS OF THE 4.5 ROLLER BEARING

by

Cole Muller

June 2003

Thesis Advisor:	David H. Olwell
Second Reader:	Samuel E. Buttrey

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2003	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE: Reliability Analysis of the 4.5 Roller Bearing			**5. FUNDING NUMBERS**
6. AUTHOR(S) Muller, Cole			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.			**12b. DISTRIBUTION CODE**

13. ABSTRACT ***(maximum 200 words)***

The J-52 engine used in the EA-6B Prowler has been found to have a faulty design which has led to in-flight engine failures due to the degradation of the 4.5 roller bearing. Because of cost constraints, the Navy developed a policy of maintaining rather than replacing the faulty engine with a re-designed engine. With an increase in Prowler crashes related to the failure of this bearing, the Navy has begun to re-evaluate this policy. This thesis analyzed the problem using methods in reliability statistics to develop policy recommendations for the Navy. One method analyzed the individual times to failure of the bearings and fit the data to a known distribution. Using this distribution, we estimated lower confidence bounds for the time which 0.0001% of the bearings are expected to fail, finding it was below fifty hours. Such calculations can be used to form maintenance and replacement policies. Another approach analyzed oil samples taken from the J-52 engine. The oil samples contain particles of different metals that compose the 4.5 roller bearing. Linear regression, classification and regression trees, and discriminant analysis were used to determine that molybdenum and vanadium levels are good indicators of when a bearing is near failure.

14. SUBJECT TERMS Reliability Statistics, Data Analysis			**15. NUMBER OF PAGES** 83
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public released; distribution is unlimited.

RELIABILITY ANALYSIS OF THE 4.5 ROLLER BEARING

Cole Muller
Ensign, United States Navy
B.S., United States Naval Academy, 2002

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN APPLIED SCIENCE (OPERATIONS RESEARCH)

from the

NAVAL POSTGRADUATE SCHOOL
June 2003

Author:

Cole Muller

Approved by:

David H. Olwell
Thesis Advisor

Samuel E. Buttrey
Second Reader

James N. Eagle
Chairman, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The J-52 engine used in the EA-6B Prowler has been found to have a faulty design which has led to in-flight engine failures due to the degradation of the 4.5 roller bearing. Because of cost constraints, the Navy developed a policy of maintaining rather than replacing the faulty engine with a re-designed engine. With an increase in Prowler crashes related to the failure of this bearing, the Navy has begun to re-evaluate this policy. This thesis analyzed the problem using methods in reliability statistics to develop policy recommendations for the Navy.

One method analyzed the individual times to failure of the bearings and fit the data to a known distribution. Using this distribution, we estimated lower confidence bounds for the time which 0.0001% of the bearings are expected to fail, finding it was below fifty hours. Such calculations can be used to form maintenance and replacement policies.

Another approach analyzed oil samples taken from the J-52 engine. The oil samples contain particles of different metals that compose the 4.5 roller bearing. Linear regression, classification and regression trees, and discriminant analysis were used to determine that molybdenum and vanadium levels are good indicators of when a bearing is near failure.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

The completion of this thesis signifies a major milestone in my continued education. That being said, without the help of certain individuals, this work would not have been possible.

Foremost, I would like to thank my thesis advisor, Dave Olwell, for all the help he has been to me. I also thank my second reader Sam Buttrey for his assistance, in particular, the *S-Plus* help he has given me. Next, I thank NAVAIR for providing the necessary data to make this thesis work possible. Lastly, I would like to thank Lynn Yates for answering the many questions I had about the data.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. OVERVIEW

The EA-6B Prowler is the only operational electronic attack aircraft flown by the U.S. military today. It has a unique mission that is necessary to the success of most military operations. The Prowler conducts Electronic Attack (EA) missions in support of U.S. and coalition air operations that are vital to our national security interests. It possesses a unique mission capability that is in continual high demand to support worldwide military operations and is the only aircraft in the U.S. inventory that is dedicated to the suppression of enemy air defenses (Pitts, 2002).

The Prowler design is about thirty years old and there is a need to replace it with an airframe that incorporates today's advanced technology. Until a replacement aircraft is built, however, the 122 Prowlers in the U.S. inventory are the only EA-capable aircraft (Pitts, 2002).

Figure 1 - A failed roller bearing.

In the late 1980's, it was found that the J-52 engine, the engine used in the Prowler, had a faulty design. Oil flow was insufficient through the engine, and consequently there were problems within the engine. The 4.5 bearing was the part that most often failed because of the poor engine design. After a few in-flight failures of the bearing, the Navy ordered a study to be done in order to determine how to correct the problem. The study recommended that the engine be redesigned, but due to cost constraints, the Navy decided to replace the bearings whenever the aircraft came in for a major inspection. The cost of a bearing is around two thousand dollars, while a new engine design could be well into the millions of dollars (Barber, 2002).

B. BACKGROUND

In November 2001, two Prowlers crashed within a week, and both crashes were most likely caused by the failure of the 4.5 roller bearing (Selinger, 2002). In the Whidbey Island crash, the 4.5 roller bearing in the "right engine failed, touching off a chain reaction that ultimately destroyed both engines. Other parts in the turbine section in which the bearing was housed 'were liberated' and rocketed into the left engine, severing fuel and hydraulic lines along the way, an examination of the engines revealed" (Barber, 2002).

Immediately following these two crashes, the Navy temporarily grounded its fleet of EA-6B Prowlers, but eventually relaxed this restriction, mainly due to the necessity to have Electronic Attack aircraft available in Operation Enduring Freedom. The Navy has begun to procure a new generation of Electronic Attack aircraft, but until that time, a new replacement policy must be put into place regarding this bearing.

The Prowler is a twin-engine aircraft, and with approximately 122 Prowlers in service for the Navy, there are nearly 250 Pratt and Whitney J-52 engines, not counting spare engines, with the 4.5 bearing operational in the fleet today (Pitts, 2002).

C. PROBLEM & PURPOSE

This thesis looks at two distinct approaches determining the reliability of the 4.5 bearing. We want to model the reliable life of the bearing, as well as identify bearings that are likely to fail shortly. Reliability is defined as the probability that a system will perform its intended function for a specified period of time. In this thesis, reliability is of more interest than availability, which is the fraction of time that a system is available for use (Meeker and Escobar, p.2). Determining the availability of the aircraft is not as important, for this issue, than the reliability. We want to decrease, if not eliminate, in-flight failures, so reliability is the natural way to analyze this situation.

Reliability, in this case, is the probability that the engine bearing will not fail in flight. We are not as concerned with other failures within the engine, but want to look specifically at the 4.5 bearing and the effect it has on the reliability of the aircraft. The life scale that we will use is not actual age, but instead hours in which the engine was operating, or engine hours.

D. APPROACHES

The first of the two approaches examined is a standard life data analysis. A life data analysis is done when the collected data is the time to failure of many identical units, including suspensions for those units not yet failed. This data is fit to a distribution model to get failure rates, quantiles, and probabilities. This type of analysis can produce estimates of the probability of a failure before a specified time, the hazard function at a specified time, and also the proportion of units that will fail in a specified time.

It is important to use the correct distributional model, or results from the model can be greatly different than those that actually take place. Many times extrapolation is necessary to gain useful information from the data. For example, if a test has run for 400 hours and what is desired is the proportion failing at 900 hours, using the wrong distributional model can very well lead to incorrect conclusions. Similarly, if we have data on 250 engines but want to calculate the reliable life associated with 99.9999% reliability, we will extrapolate to a point well before the first observed failure.

The times to failure, given in engine operating hours, that are attributed to the 4.5 bearing were used to fit the data to a distribution. This allows us to predict probability of failure of the bearing over a time interval with a given confidence level. This in turn allows us to evaluate and suggest crude replacement policies.

Life data analysis gives us some good insight into the overall problem, but there are limitations. Since the data is binary (failed or suspended bearings), this type of analysis did not use other available information. This analysis recommends that the Navy should replace each bearing very frequently to maintain a certain reliability level, but this interval is impractical. Therefore, an effective policy based on this type of analysis would be a cost-ineffective policy to adopt.

The other approach used regression techniques from data analysis such as simple linear regression, classification and regression trees (CART), and discriminant analysis. This process explored the levels of metallic materials found in the filter of the engine. The idea was that an increase in certain levels of metals and other elements in the filter could be used to predict failure of a bearing. The elements that make up the bearing that are used in this approach are silver, molybdenum, iron, and vanadium.

The oil filter analysis program is done by "evaluating the residue the filter collects" in order to predict the failure of a component, in this case the 4.5 roller bearing. When the bearing fails, the particles "that are generated are too large to be picked up in a spectroanalysis . . . are picked up in the filter and can be evaluated using filter analysis" (Oil Lab, 2002).

Data on hand that was used for this approach is the percentage of different metals found in the oil filter, along with the cumulative mass of each metal within the filter. The collected data also indicates which oil samples come from engines that have failed bearings so a classification can be made.

The results from these approaches are used to create recommendations to the Navy to influence policy so that the risk of in-flight bearing failures is reduced. The recommendations identify appropriate approaches discussed in this paper to pursue and why.

II. LIFE DATA ANALYSIS

A. APPROACH

The first approach used was a standard life data analysis. A life data analysis is done when the collected data is the time to failure of many identical units, including suspensions for those units not yet failed. The idea of this approach is to develop a model that fits a distribution to the data, which is described below. Once we found an acceptable distribution for our data, we then estimated the parameters of the distribution so that we could generate probability plots and failure-time distribution functions, which are described more thoroughly in Appendix *A*, and also estimate confidence intervals to aid in our analysis.

B. DATA

The data we collected is a mixture of two different types of data. The first type is known as complete data. Complete data is data that has exact values for the time to failure of the bearing. Right-censored data is data that has not failed by a certain time, which is the case for the vast majority of the data that we have. Both types of data were used for the life data analysis (Meeker and Escobar, p.34).

Listed in *Appendix H* is the data we used for this analysis. There are sixty-six data entries, with eleven failures and fifty-five suspensions. This is considered a small sample size because of the number of observed failures is low.

To analyze this type of data properly, we must understand the methods of analysis and parameter estimation appropriate to censored data. These methods are similar to those methods used on complete data, but are modified to fit the needs of censored data. For example, in a complete data set, it is easy to calculate the mean of the data. We simply sum the data entries and divide by the number of entries to get this desired mean. However, when dealing with censored data, we need to adjust for the interval of uncertainty that comes with each data point that is censored. To take the mean of censored data, we must account for the data points that did not fail by the end of the data

collection period. Therefore, the probability of failure, also known as unreliability, needs to be adjusted. These methods are discussed further in Appendices *C* and *D*.

C. ASSUMPTIONS

We expect the 4.5 bearing of the J-52 engine to have an increasing failure rate. As the bearing gets worn down, it should be more susceptible to failure. There is the possibility of infant mortality, in which case the failure rate function would have a 'bathtub-shaped' look. We consider the data to be well past the time of infant mortality, so we can disregard this possibility for the failure rate.

We are primarily concerned with the distribution of the life of the bearing during early life. We considered distributions that have strictly increasing failure rates during early life, but included distributions, such as the lognormal, that fail the strictly increasing failure rate test over the whole domain.

D. DISTRIBUTIONS AND PLOTS

We used a commercial software package, *Weibull++*, to help with the analysis. *Weibull++* is able to take the data and fit distributions to this data and also estimate the parameters of these distributions. It also has a function that tests the goodness of fit of each distribution, and this function then recommends the distributions that are the best fit for the data. The goodness of fit tests are a weighted score of Anderson-Darling, Kolmogorov-Smirnov, and Maximum Likelihood tests.

Properties of common lifetime distributions are given in Appendix *B*. For our data, the first distribution that *Weibull++* recommends to us is the two-parameter exponential distribution. One of the characteristics of the exponential distribution is that it is a memory-less distribution, i.e. it has a constant failure rate. However, one of our assumptions was an increasing failure rate of the bearing, at least during the portion of time that we explored. Because of this, we eliminated the exponential distribution from consideration.

The next best-fit distribution is the three-parameter Weibull distribution. We took a look at the probability plot, seen in Figure 2, and it seemed like a decent fit. Appendix

C describes the idea of probability plotting in greater detail. Looking at the failure rate function in Figure 3, we saw that the PDF is zero before approximately 580 hours. This would make it hard to analyze the data and come up with a one-sided confidence bound given a desired reliability level. Given this, we preferred to look at other distributions that fit the data fairly well, yet allowed these desired one-sided confidence bounds.

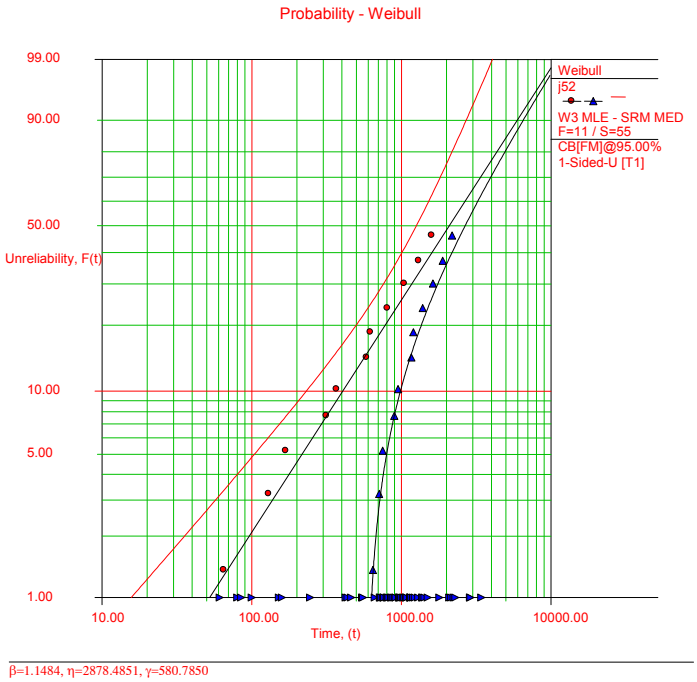

Figure 2 – This is the probability plot using the three-parameter Weibull distribution to fit our data. The curved line on the right is the data fit to the model on the original scale. The straight line is the data fit to the model after the translation parameter has been subtracted. Above the straight line is a one-sided upper confidence bound at the 95% level for the failure probability. The best-fit line is not a great fit to the data, which is not uncommon, as MLE estimates for small samples such as this are known to be biased. Suspensions are noted on the horizontal axis.

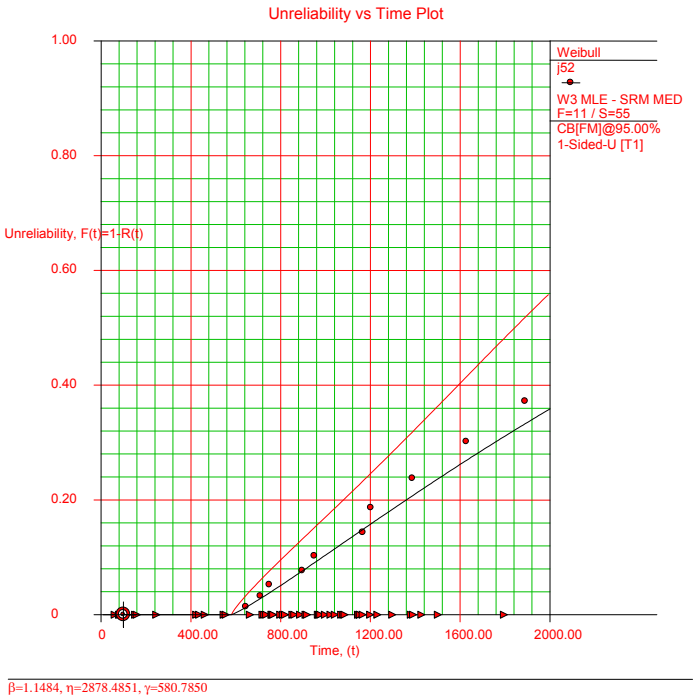

Figure 3 – This is the failure rate using the three-parameter Weibull distribution to fit the data. The failure rate is conditional on the translation parameter, which is fixed. We can see as a result that the function is zero before approximately 580 flight hours. Also drawn on this plot is the upper confidence bound. Because of this, it is hard to determine a one-sided confidence bound with a given reliability level, and thus we want to look at another distribution that would allow us to do so. We also want to at least admit the possibility of early failures, so a translation parameter is not desired.

We then looked at the next recommended distribution, which is the lognormal distribution. The probability plot in Figure 4 looked slightly worse than that from the three-parameter Weibull distribution, although the difference was minor. The failure rate function was increasing, at least in the time interval that we explored. In Figure 5, it can be seen that the failure rate increased until approximately 2800 hours, which is well past the scope of this analysis. We were only interested in what occurs in the first couple hundred hours, so we disregarded the fact that the failure rate will eventually decrease.

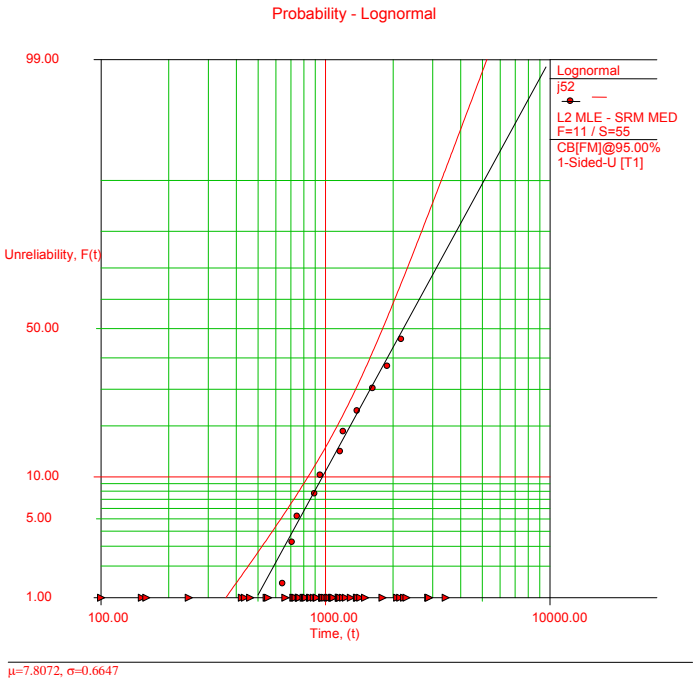

Figure 4 - This is the probability plot with the lognormal distribution modeling the data. This looks to be a better fit than in *Figure 2*. Once again, we also have included the 95% upper confidence bound on unreliability.

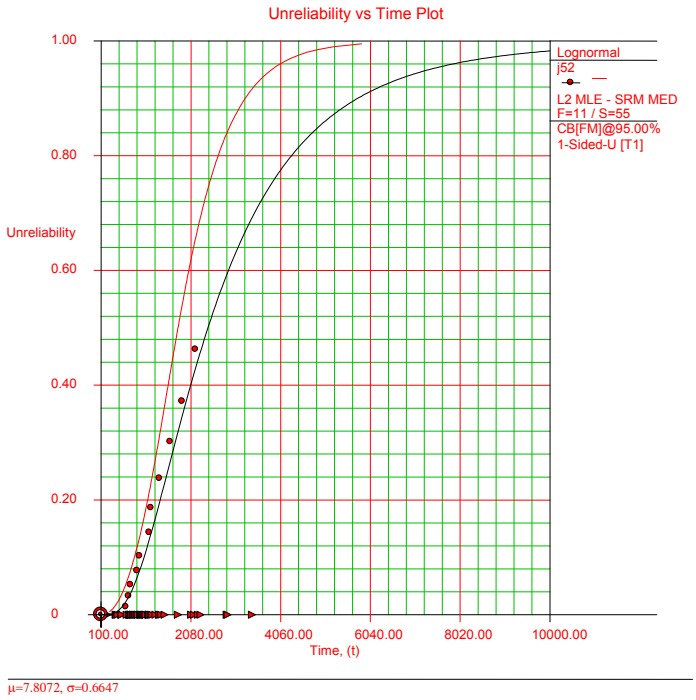

Figure 5 - This is the failure rate of the data when fitted to the lognormal distribution. Despite being a non-monotonic function, it is an increasing function up until a certain point. Our expected time to failure is well before this point, so we can use the lognormal distribution to fit our data.

The last distribution that we explored was the two-parameter Weibull distribution. Even though *Weibull++* ranks this distribution lower than the lognormal, we wanted to take a look at the probability plot of the data fit to this distribution, seen in Figure 6. We can see from this plot that this distribution is a bad fit, especially in the lower tail of interest, and for this reason we omitted this distribution from our analysis.

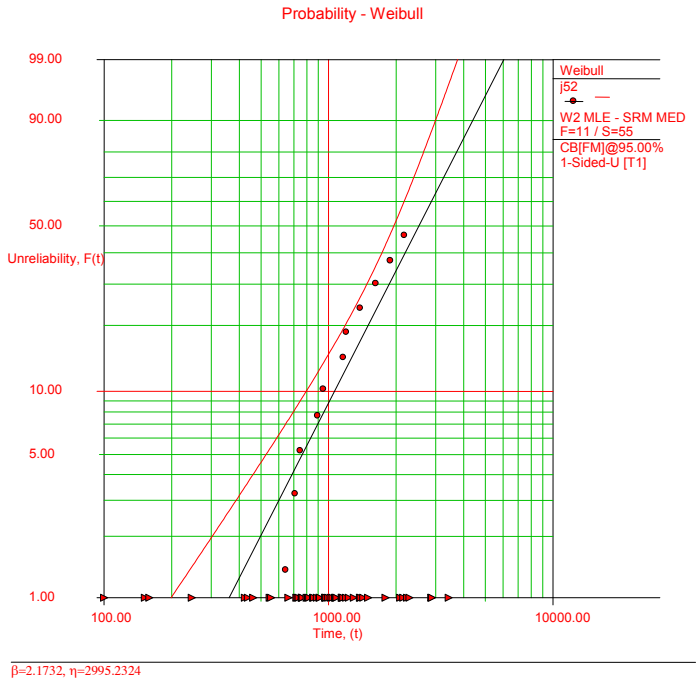

Figure 6 – This is the probability plot modeling the data with the two-parameter Weibull distribution. We again have a 95% one-sided upper confidence bound. The best-fit line is clearly not a good fit.

E. SELECTING A DISTRIBUTION TO MODEL THE DATA

We decided that the exponential distribution and the two-parameter Weibull distribution were not good fits to model our data based on our assumptions. The three-parameter Weibull distribution seemed to represent our data a bit better than those of the lognormal distribution; however it did not allow for easy calculations of the one-sided confidence bound for small probabilities of failure. Also, there is a history of modeling bearing life with the lognormal distribution (Lawless, p.228). Because of this, we chose the lognormal distribution to model our data.

F. PARAMETER ESTIMATES AND CONFIDENCE BOUNDS

Now that we selected a distribution, we needed to estimate the parameters of this distribution. We let *Weibull++* calculate them, using either Rank Regression (RRX) or Maximum Likelihood Estimators (MLE). For more on these two techniques and the differences between them, refer to Appendix *D*. Since we have censored data with a sufficiently large sample size, we used the MLE method to estimate the desired parameters. Using this method, we got the following parameters for our model: $\mu = 7.8072$, $\sigma = 0.6647$.

We set a reliability level that is high enough so that the risk of failure is very small. Due to the catastrophic nature of just one failure, we set the reliability level to be 99.9999%. Once the probability of failure is greater than 0.0001%, we would recommend that the bearing be replaced. Using the probability plot generated from Weibull++, seen in Figure 7, after approximately 100 engine hours, the point estimate of the reliability level has dropped to 99.9999%, our given threshold level.

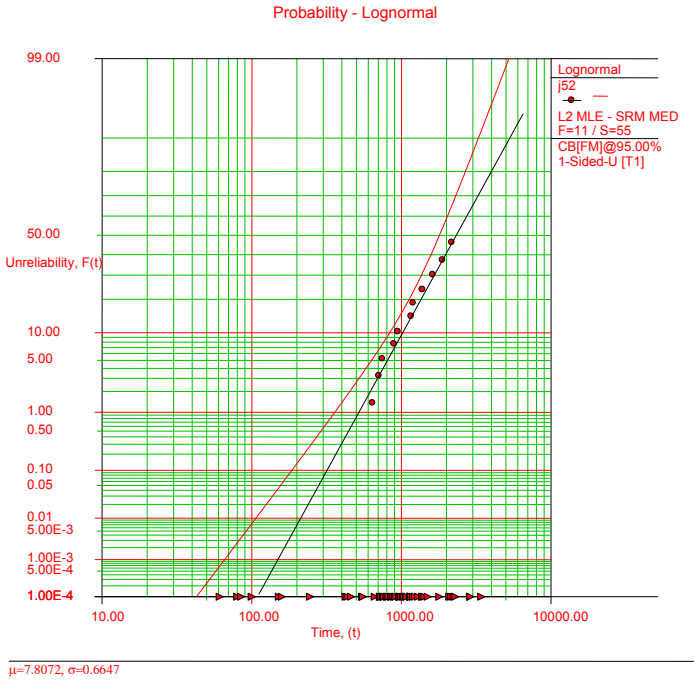

Figure 7 – This is the probability plot of the lognormal distribution with the estimated parameters using MLE. This plot shows the expected time of failure for 0.0001% of all aircraft to be 104.3 hours, with a 95% upper confidence bound (Fisher-Matrix) of 42.8 flight hours. Once again, the circles represent the failures and the triangles represent the suspensions.

We also calculated the exact value of the expected engine hours until 0.0001% of the bearings failed using the MLE method. Using the process described earlier on MLE's, we obtained 104.3 engine hours as the expected time where 0.0001% of the bearings should fail.

We also determined a confidence interval for this time. Since we had a small amount of data, the time 104.3 can vary quite a bit from sample to sample. We sought a confidence interval that is 95% confident that the probability of failure is less than 0.0001%. A discussion on confidence bounds, to gain a better understanding of how they are derived, follows.

Confidence bounds estimate the precision of an estimator. The estimator for us was the number of engine hours by which 0.0001% of the engines had failed. Since we could not get the true reliability value because of the data being only a sample of J-52

engines, we could only estimate this value using different probabilistic methods. Our parameter estimates will change slightly from sample to sample, thus changing the reliability value that we estimated. Confidence bounds produce an interval that should contain the reliability value for a certain percentage of intervals generated from this sampling scheme, in our case, 95%.

Confidence bounds can be one or two-sided. Two-sided bounds have both an upper and lower bound to the interval, while one-sided bounds have either an upper or lower limit, but not both. We can further break one-sided bounds down into two types. We can have either one-sided upper confidence bounds or one-sided lower confidence bounds. One-sided upper confidence bounds have an upper bound on the interval and a lower bound being the start time, in most cases zero. One-sided lower confidence bounds, thus, have a lower bound on the interval and an upper bound that goes out to infinity. In our case, we did not care what happened at later time periods, and therefore had no utility for an upper confidence bound. We gained more useful information if we chose to use the one-sided lower confidence bound to construct our interval for the estimator of the engine hours it takes for 0.0001% of the engines to fail.

There are two different approaches to constructing confidence bounds that we considered for our model. They were Fisher Matrix (FM) confidence bounds and Likelihood Ratio (LR) bounds. Fisher Matrix bounds use the assumption of asymptotically normal MLE estimates of the parameters, and therefore need a sufficiently large sample size. Likelihood Ratio bounds do not depend on asymptotic normality as strongly as doe Fisher Matrix bounds, and thus work better than *FM* when the sample sizes are smaller. We discuss each of these methods in greater detail.

To construct confidence bounds using FM, we need to know the mean and the variance of the function that we are examining. Since we are inquiring about the cdf, or unreliability function, of the lognormal distribution at 0.0001% unreliability, we need the mean and variance of this function in order to use Fisher Matrix bounds.

The mean can be calculated easily enough using MLE's and their invariance properties, and we have already done so earlier. Now we need the variance of the function. The variance of a function depends on the variance of each of the parameters

estimated in the function, as well as the covariance between them. Appendix *E* gives a more in-depth look at this idea.

For our data, the Fisher Matrix method gave us a 95% one-sided lower confidence bound of 42.8 engine hours. Thus, the Fisher Matrix method suggested that a 99.9999% reliability threshold with 95% confidence was 42.8 engine hours.

The Likelihood Ratio bound method is based upon the following equation:

$$-2\ln\left(\frac{L(\theta)}{L(\hat{\theta})}\right) \geq \chi^2_{\alpha,k}, \tag{1.1}$$

where $L(\theta)$ is the likelihood function with unknown parameter(s) θ, $L(\hat{\theta})$ is the likelihood function calculated with our estimated MLE parameters, $\chi^2_{\alpha,k}$ is the chi-squared test statistic with probability α and k degrees of freedom. $L(\hat{\theta})$ is calculated using the MLE method, and since we know the parameter estimates, the only unknown term in the equation is $L(\theta)$. We can then solve for this term (Meeker and Escobar, p.185).

Since we chose the lognormal distribution to model the data, we have two parameters that can vary. Thus, a set of values will satisfy the equation. Numerical methods are then used to find this set. For our data, using the Likelihood Ratio method to calculate our 95% one-sided lower confidence bound, we got a value of 31.1 engine hours for the lower confidence bound. Thus, a 99.9999% reliability threshold with 95% confidence was 31.1 engine hours, according to this method.

A more conservative approach would be to use the method that gets the lower of the two confidence bounds, which would be using the Likelihood Ratio bound method. Both methods, however, suggest replacing the bearing too frequently for practical purposes.

G. RESULTS FROM LIFE DATA ANALYSIS

The Navy currently employs a policy that replaces the bearings at major inspection intervals, which vary but can be more than one thousand engine hours apart, which is drastically high compared with the results from our model (Barber, 2002). The

point estimate of the probability of such a failure by one thousand engine hours is 0.088. This can be seen in Figure 7, although it is difficult to read from the plot. Given this, and that five of the sixty-six bearings failed before this time, we can see that a better policy should be implemented. Replacing the bearings every thirty or forty hours, though, might not be the best policy either. The cost-effectiveness of this policy would not be very high. If we take a look at the data, the first failure occurred after 646 engine hours. This suggests that the bearings more than likely can be used a lot longer than 31.1 hours by accepting slightly more risk. This is a shortcoming of our 99.9999% reliability requirement. Replacing bearings every thirty or forty hours discards a lot of useful life in each bearing without gaining much more reliability. Considering the high cost of both replacing the bearing and maintenance, this does not seem to be a very efficient policy. In light of this, we want to come up with a better model to more accurately determine when a bearing will fail.

III. OIL FILTER ANALYSIS

A. TECHNIQUES

The second approach used techniques from data analysis: simple linear regression models, classification and regression trees (CART), and discriminant analysis. These processes used the levels of metallic materials found in the filter of the engine as a predictor of failure. The idea was that an increase in certain levels of metals and other elements in the filter could be used to predict failure of a bearing. The main elements that make up the 4.5 bearing are molybdenum (4%), iron (90%), and vanadium(1%); the cage that holds the bearing is made of silver. From our analysis, we determined the levels of each of the metals found in the oil filter that are the best indicators of a failed bearing.

B. DATA

As the engine is used, friction between the bearings and the shaft cause particles of the metallic bearings to break off from the bearing. These particles get mixed in with the oil in the oil filter. The data that has been collected is a sample of the debris found in the oil filter. The sample is placed on a patch that is one square centimeter in area and the mass of each metal found in this sample is recorded. The units of the level of each metal found in the sample are grams per square centimeter. The cumulative amount of each metal is the data that is used in the analysis.

Each time the oil filter is changed, the contents of the filter are analyzed for the presence of the aforementioned metallic substances. The data we shall use is updated every time the filter for a specific engine is changed, in order to represent the additional amounts of the different metals found in the single oil filter. We are interested in the cumulative build-up of the metals inside the filter because this could represent the overall life span of the bearing, and therefore could be the best predictor available as to the health of the individual bearing.

There are twenty-one sample engines from which the data was collected. The levels of silver, molybdenum, iron, and vanadium are recorded for each engine. Also

recorded is the health of each bearing. The health of the bearing is broken down into a integer value between zero and five. Zero indicates that the bearing was brand new, while a five indicated that the bearing had completely failed. When the bearing begins to skid and wear on the rollers, it is said that the bearing stage is one. A bearing stage of two indicates there is noticeable wear on the bearing rollers. Stage three is classified as when the cage of the bearing begins to crack. When the cage incurs severe wear, the bearing stage is four. We can interpolate between the stages to come up with a continuous function for bearing stage. Figures 8-11 below present the plots of the level of each metal versus the bearing stage.

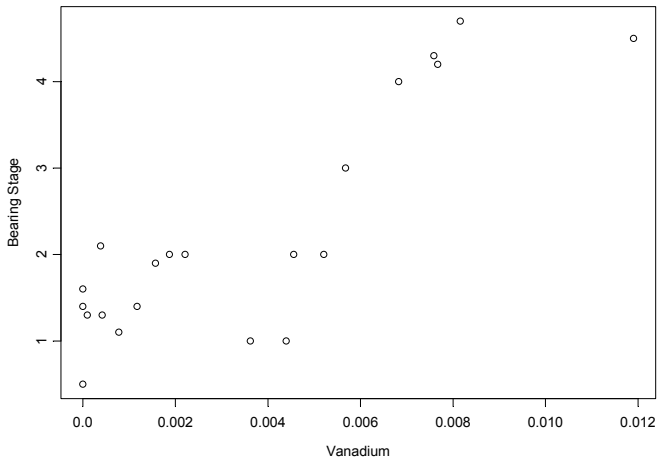

Figure 8 – This is the Bearing Stage vs. Vanadium plot. There seems to be some positive correlation between the two. We found ρ = 0.855.

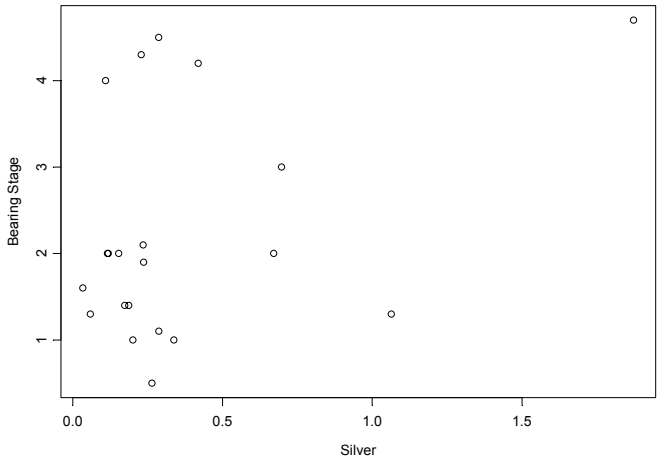

Figure 9 – This is the Bearing Stage vs. Silver plot. It is not clear if there is any correlation between the two. We found ρ = 0.359.

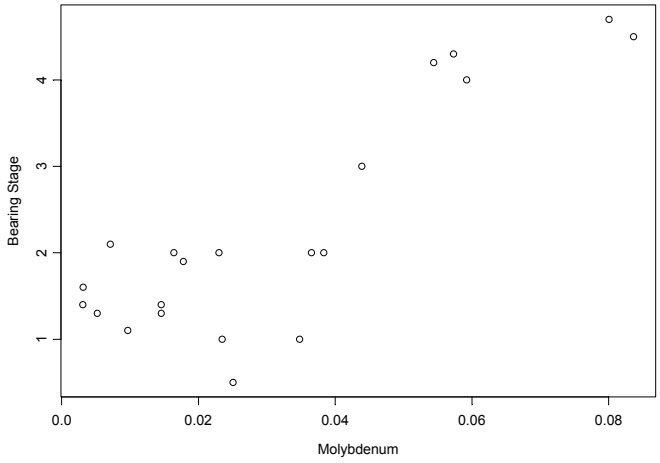

Figure 10 – This is the Bearing Stage vs. Molybdenum plot. Again, it looks like there is some positive correlation between the two. We found ρ = 0.854.

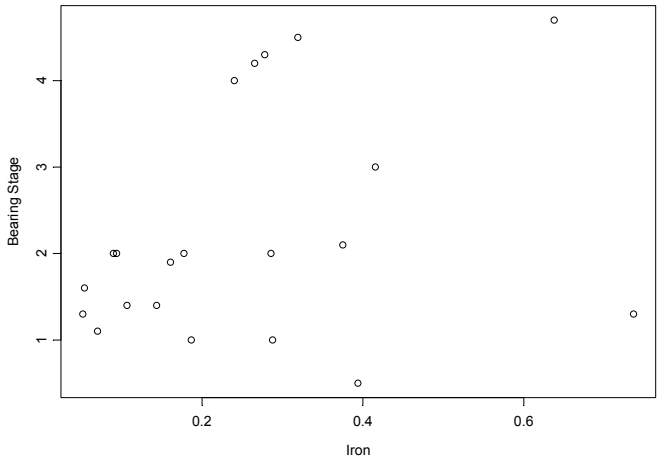

Figure 11 – This is the Bearing Stage vs. Iron plot. There may be a slight positive correlation between the two, but it is not quite clear. We found ρ = 0.304.

Figure 8 shows that the level of vanadium and bearing stage are positively correlated. In general, as the level of vanadium found in the oil filter sample rises, so does the bearing stage. This is also true of molybdenum, seen in Figure 10. Iron and silver do not show this trend. In fact, these two metals do not show any obvious trend that would help predict the stage of the bearing. We can also see in Figures 12 and 13 that the levels of vanadium and molybdenum in this sample can be classified perfectly, meaning that a naïve model by inspection would yield no misclassification of any of the bearings.

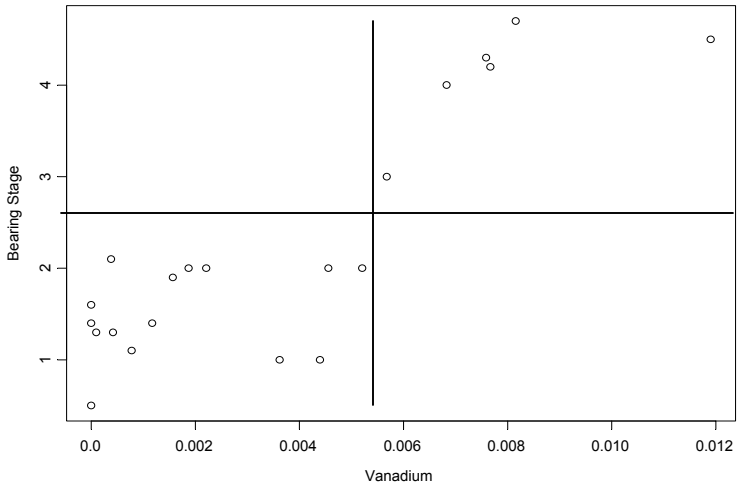

Figure 12 – This is the Vanadium vs. Bearing Stage plot with perfect classification. If we classify bearings with a stage of 2.5 or greater as bearings in need of replacement, then we can say that if the level of vanadium found in the oil filter is greater than approximately 0.0055 g/cm^2, the bearing should be replaced.

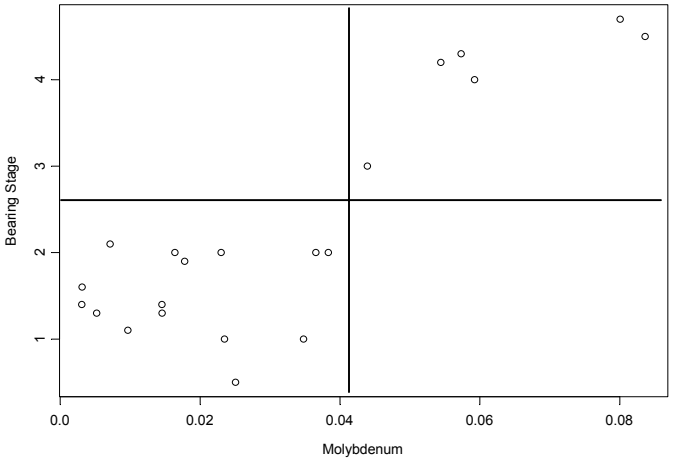

Figure 13 - This is the Molybdenum vs. Bearing Stage plot with perfect classification. If we classify bearings with a stage of 2.5 or greater as bearings in need of replacement, then we can say that if the level of molybdenum found in the oil filter is greater than approximately 0.0041 g/cm^2, the bearing should be replaced.

C. SIMPLE LINEAR REGRESSION MODEL

First, we created a simple linear regression model to predict the bearing stage given the levels of vanadium, iron, molybdenum, and silver found in the oil filter. The regression model equation is of the form

(1.2) $$E[Y] = b_0 + b_1 x_1 + \ldots + b_{k-1} x_{k-1},$$

where k is the number of input variables that will be used to predict the response variable (Hamilton, p.66). To perform this linear regression, we assumed that the expected value of the response variable given the input variables was linear. For our data, we had four input variables. Therefore, the model that we constructed was of the form

(1.3) $$\hat{y} = b_0 + b_V x_V + b_{Fe} x_{Fe} + b_{Mo} x_{Mo} + b_{Ag} x_{Ag},$$

where b_o is the estimate for the intercept and b_k is the estimate for the slope of the k-th input variable.

For each of our data entries, we used the input variables to generate an expected response variable. We then took the difference between this expected value and the observed value. Using all our data entries, we calculated the sum of the squares of this difference, which is known as *RSS*.

(1.4) $$RSS = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2$$

The coefficients that minimized the RSS are the ones we chose for our model (Devore, p.498). We used *S-Plus* to generate these coefficients for our model.

By creating a linear model in *S-Plus*, we got the coefficients seen in Table 1. We see that the p-values of each individual element are high, suggesting that the removal of any one of these elements will not hurt the model. The R^2 of the model is 0.7516.

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	1.0	0.32	3.15	0.0062
V	186.36	211.20	0.8824	0.3906
Fe	-0.71	1.60	-0.4447	0.6625
Mo	19.97	32.13	0.6215	0.5430
Ag	0.41	0.67	0.6152	0.5471

Table 1- Table of coefficients for initial linear model. If the model is accepted, we assume that the errors are normally distributed.

We checked the residuals for normality. Figure 14 shows a plot of the residuals against the standard normal quantiles. The plot does not look too bad, except the lower tail seems a bit skewed, so we decided that the residuals did not look to be normally distributed.

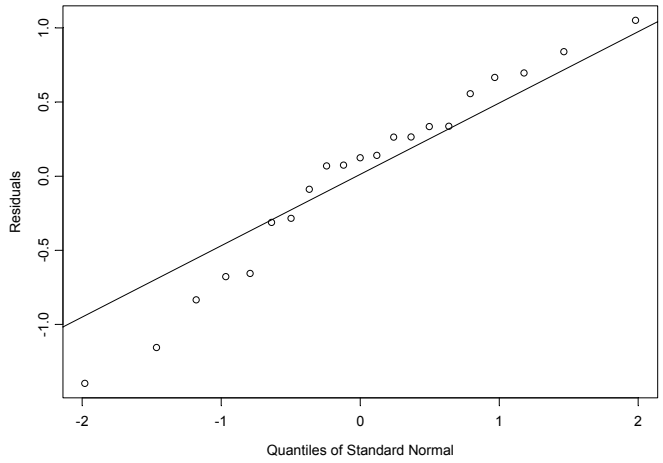

Figure 14 - Residuals of the model plotted against the standard normal quantiles. The fitted line is the line that the residuals should fall on if they are normally distributed.

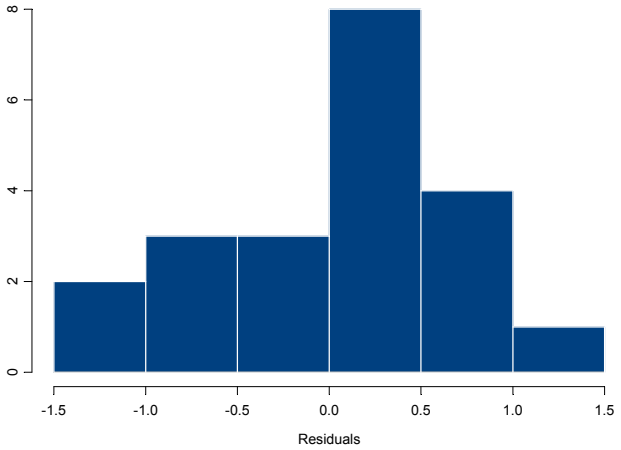

Figure 15 - This is a histogram of the residuals from our simple linear model. The residuals are not quite symmetric, and therefore do not look to be normally distributed.

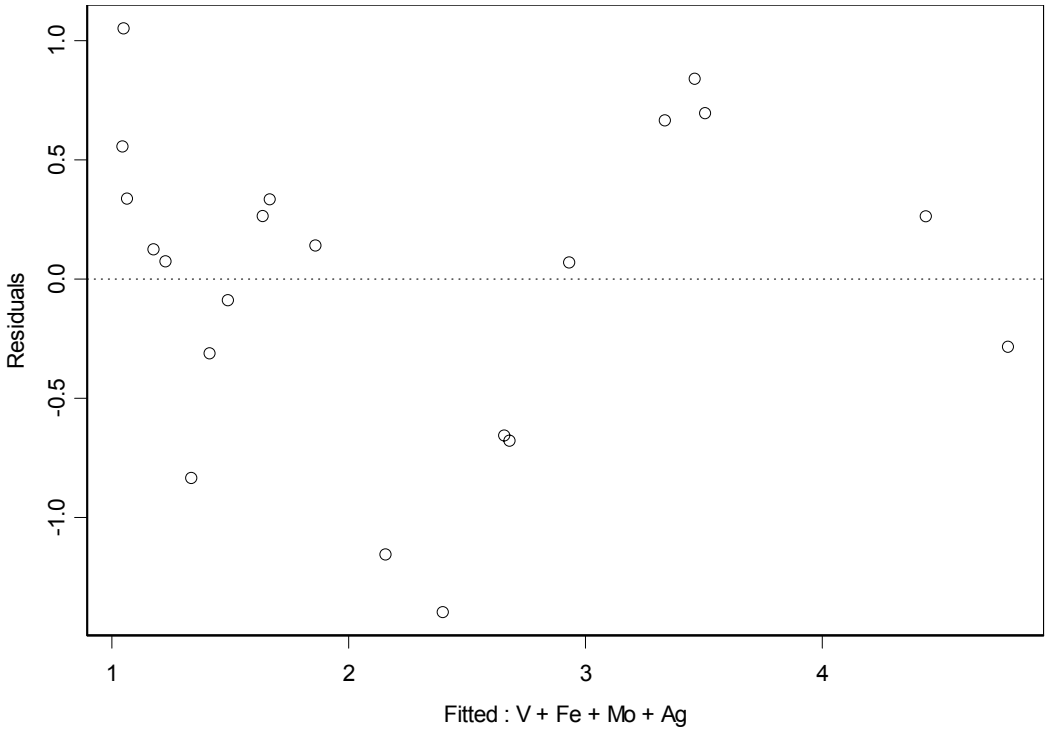

Figure 16 - This is a plot of the fitted values vs. the residuals of our initial model.

Figure 15 is a histogram of the residuals, and from here the residuals also do not look to be from the normal distribution. Figure 16 is a plot of the fitted values against the residuals. This plot does not seem to be suggesting that our model is a good fit. After looking at various transformations of the data, there were no candidates whose residuals looked a lot better than this model, however.

We continued with this model and added two-way interactions into it. After we added the interactions, we ran the *stepAIC* function in S-Plus. The *stepAIC* function automates the stepwise addition of terms from our model that significantly decreases the residual sum of squares (Venables and Ripley, p.186). Table 2 shows the coefficients of the terms that remained in our model. The R^2 of this model is 0.8598, which is significantly higher than other models we looked at.

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	0.64	0.59	1.08	0.3019
V	-1497.62	819.27	-1.83	0.0948
Fe	4.97	3.44	1.44	0.1766
Mo	248.73	115.33	2.16	0.0540
Ag	-4.49	2.52	-1.78	0.1028
V:Fe	4126.47	2222.65	1.86	0.0903
V:Mo	4396.29	2935.63	1.50	0.1624
V:Ag	549.03	377.26	1.46	0.1735
Fe:Mo	-803.90	345.49	-2.33	0.0401
Fe:Ag	8.34	3.78	2.21	0.0493

Table 2 – These are the coefficients of linear model with interactions. Some terms have high *p*-values, suggesting that we can remove that term and the model would not be affected much.

In Figure 17, we see the plot of the residuals of our model with interactions against the standard normal quantiles. Figure 18 is the histogram of our model with interactions. These plots show some improvement from the previous model with respect to the residuals being normally distributed. We removed some of the terms with high *p*-values individually, and although the R^2 value of each of the new models did not drop

significantly, the residuals looked less normal than the residuals from this model. Figure 19 is the fitted values versus residuals plot. This plot suggests to us that this model may be better than our previous one.

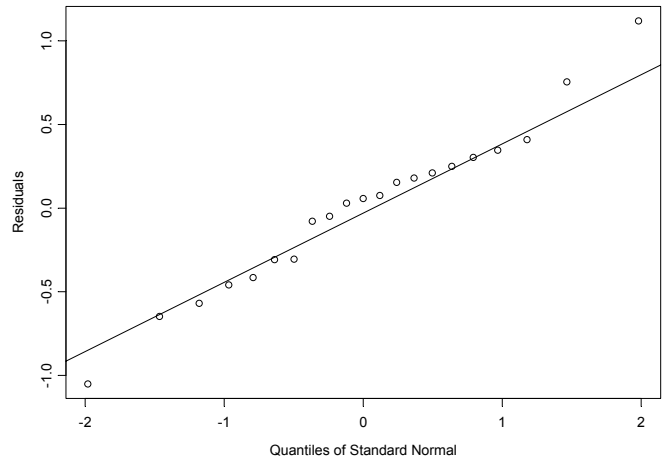

Figure 17 – This is a plot of the residuals of the model with two-way interactions against the standard normal quantiles. Residuals that are perfectly normal would all fall on the line, but the residuals from our model are close to the line in most cases, therefore we conclude the residuals are normally distributed.

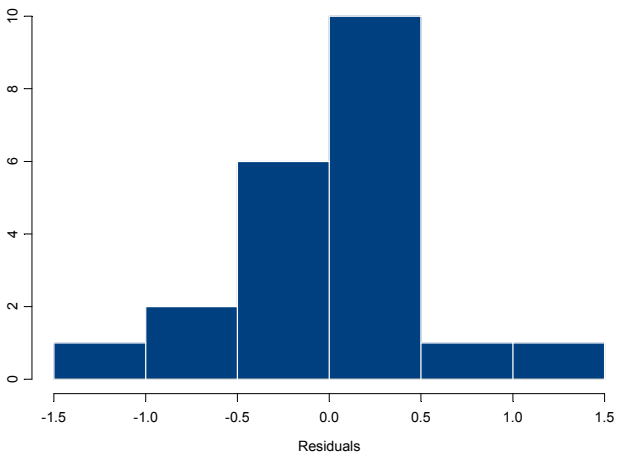

Figure 18 – This is a histogram of the residuals of the model that includes two-way interactions. The shape of this histogram is somewhat close to that of the normal distribution. Given the small sample size, we shall assume that the residuals are normally distributed.

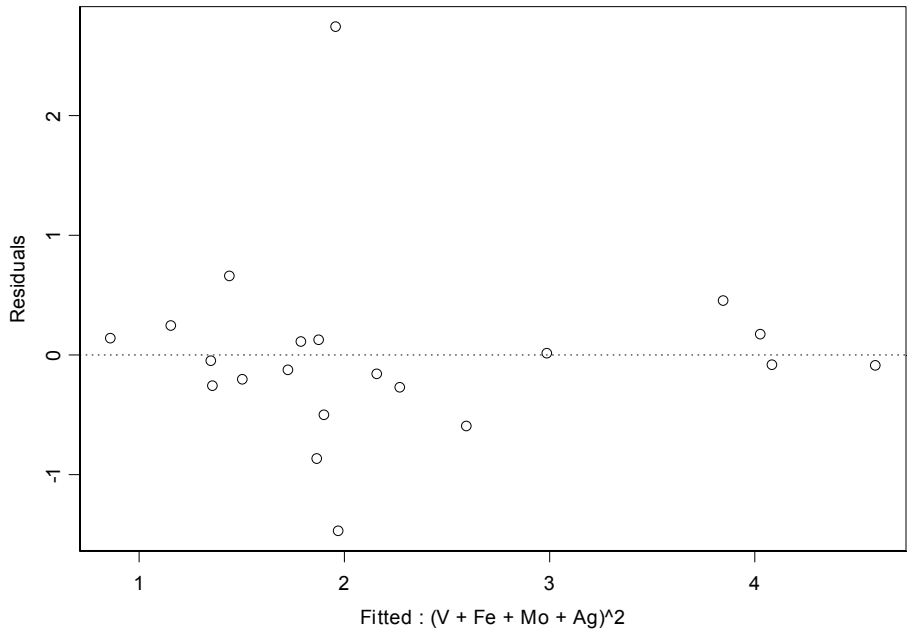

Figure 19 - This is the plot of the fitted values vs. residuals of our model with interactions. There seems to be one extreme outlier, although after looking at the data point, it is not immediately clear why this point is so extreme.

From these models, we saw an emphasis was placed on the presence of vanadium and molybdenum in the oil filter when determining the bearing stage, so we wanted to construct a model that simply modeled the bearing stage as a function of the presence of these two metals together. The coefficients of such a model are seen in Table 3. This model has an R^2 of 0.7674.

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	1.20	0.34	3.55	0.0025
V	71.02	178.10	0.40	0.6950
Mo	11.00	24.34	0.45	0.6569
V:Mo	2483.82	1972.94	1.26	0.2251

Table 3 – This is the table of coefficients of the linear model that models bearing stage by simply the level of vanadium, molybdenum, and the interaction between the two.

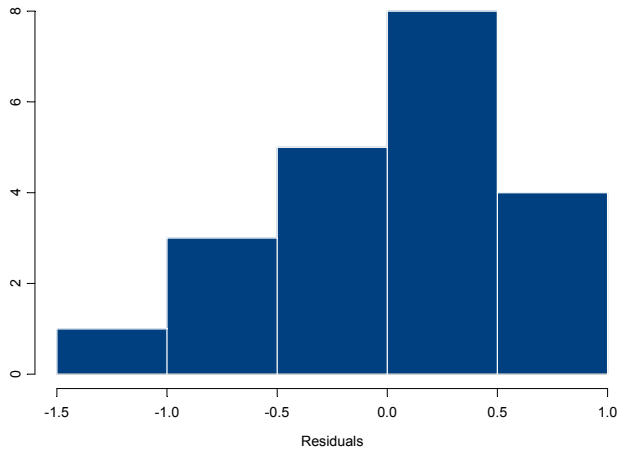

Figure 20- Histogram of Residuals of model with only vanadium, molybdenum and the interaction between the two.

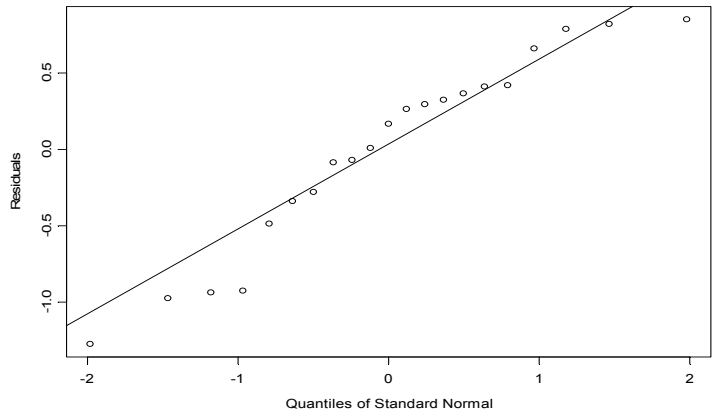

Figure 21 - Residuals of model with vanadium, molybdenum, and the interaction between the two.

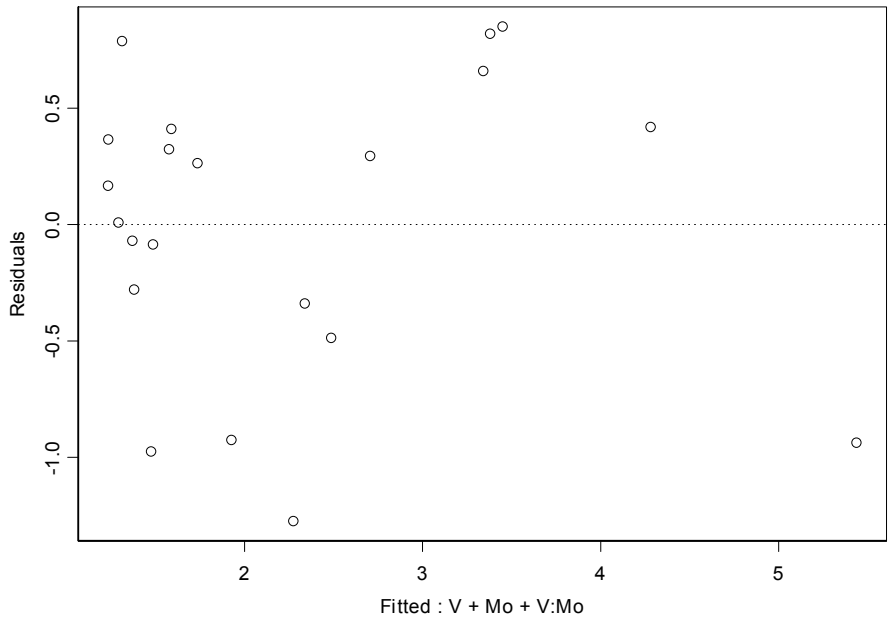

Figure 22 - This is the fitted values versus residuals plot for our model that included just V, Mo, and the interaction between the two.

Figures 20-22 show the plots of the residuals, which do not look as normally distributed as the residuals from the previous model. This last model showed that the levels of vanadium and molybdenum found together in the oil filter are a good indicator of the health of the associated bearing.

Overall, the best model is probably the second one, although there are many terms chosen by *stepAIC* for inclusion in the model that had high *p*-values. However, the plots of the residuals look to be closer to the normal distribution than those of any other model, and the R^2 value of this model is also significantly higher than the other models. We looked at other regression tools to construct some more models in order to validate this result as best we can. Also, Figures 8 and 10 indicate that the response is probably piecewise linear, which makes the use of standard regression models problematic.

D. REGRESSION TREE MODEL

Our second approach to analyze our data was to use regression trees. Regression trees use least squares regression to develop a stepwise tree structure. In regression, there is a response variable, y, and independent variable, x. We use the x-values to construct a predictive model. These models can be used to accurately predict the response variable for future x-values. They also can be used to see the relationships between the x and y variables. Refer to Appendix *F* for more details on regression trees.

We analyzed the oil filter data using regression trees. The data has five ordered variables, one being the bearing stage and the other four being the level of the corresponding metal found in the filter analysis. Since we constructed a model that predicts the bearing stage based on the levels of the different metals found in the filter, our independent variables were the levels of the four metals, and our response variable was the bearing stage.

We used the computer program *S-Plus* to help with the construction of our models. The *tree* function in *S-Plus* was used to create our first regression tree. We declared the response variable to be the bearing stage and all the other variables as our

inputs. *S-Plus* did the calculations of the deviance and constructed a tree that reduced the total deviance of the model. In Figure 23, we see the initial tree that *S-Plus* built based on our data.

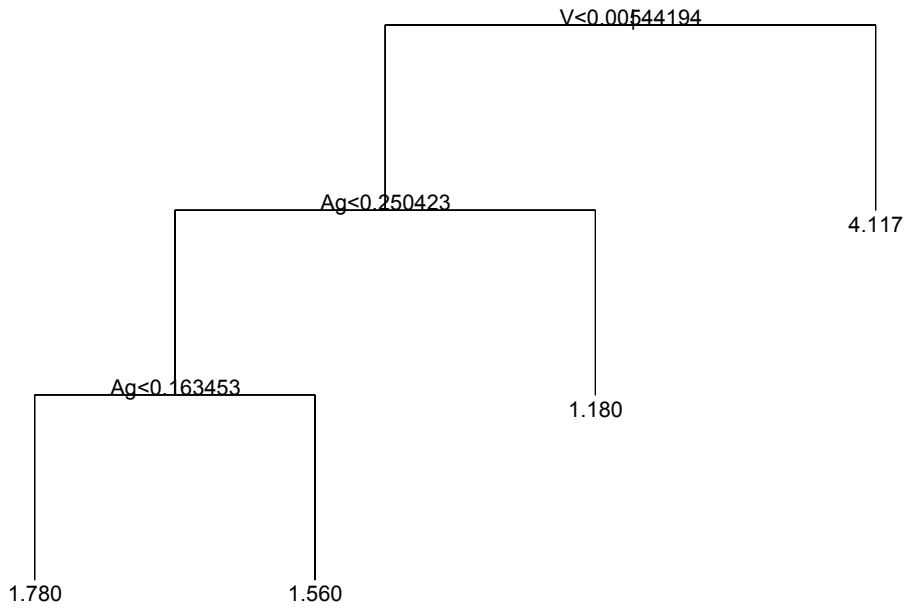

Figure 23 – Initial regression tree with the intermediate nodes labeled with the appropriate split and the terminal nodes labeled with the appropriate predicted bearing stage. This model has four terminal nodes. The 'yes' branch is to the right and the 'no' branch is to the left. Accordingly, the mean bearing stage for an oil sample with V < 0.00544194 and Ag > 0.250423 is 1.180.

From this tree, we see that the first split was determined by the amount of vanadium that is found in the oil filter. This model said that if the total mass of vanadium found in the filter is less than 0.00544194 g/cm^2, then we should follow the branch to the left of the node, which gets us to another intermediate node based on the level of silver in the filter. If the level of vanadium is greater than 0.00544194 g/cm^2, then we branch off to the right of the first node, which leads to a terminal node. The expected bearing stage

at the terminal node is 4.117, which is the average of all the bearing stages from the data entries that have a vanadium level greater than 0.00544194 g/cm^2.

There are four terminal nodes in this model, with a calculated deviance of 0.2445. We see that this model incorporates only the levels of vanadium and silver when determining the predicted bearing stage. This suggests that the levels of molybdenum and iron found in the oil filter have little to no effect on the stage of the bearing or are confounded with other predictors. If we look at the tree, we notice that all of the terminal nodes that descend from the split where the vanadium level is less than 0.00544194 g/cm^2 have predicted bearing stage values less than two. This suggests that the biggest indicator of bearing health is the amount of vanadium that is found in the oil filter.

We now turn our attention to pruning this model. Since we split each node until splitting made no difference, we have run the risk of over-fitting our model to the data. To counteract this, we must prune the tree backwards to get a good balance of descriptive and predictive power. This technique creates a nested sequence of sub-trees, from which the best-sized tree is chosen (Venables and Ripley, p.327).

A tree is determined to be of best size when the deviation is the smallest. In *S-Plus*, we ran a cross-validation function to determine what size tree gives us the smallest deviance. We used cross-validation to suggest to us the best size of our tree without over-fitting the model to the data. Looking at the deviances of the different sizes, we chose the best size to be two. Figure 24 shows us a picture of the deviance compared to the size of the tree.

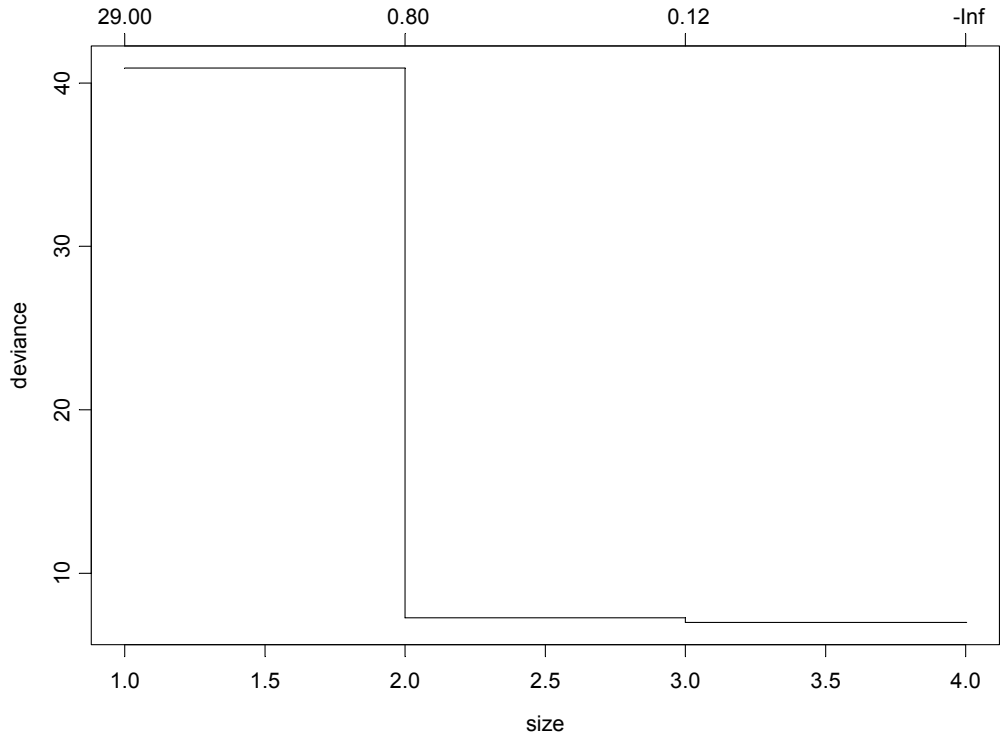

Figure 24 – There is a large drop off in deviance when the size of the model moves from one to two. For sizes larger than two, the deviance does not significantly drop, leading to the best size chosen to be two.

From this picture, we see that the deviance does not decrease noticeably for size values of greater than two. Since we wanted the simplest model possible, we chose the smallest size among all the deviances that are close, which gives us the best size of two for our model.

Then we wanted to prune our tree appropriately, so we took this recommendation and ran the *prune* function, which pruned our model down to a tree with the desired size. Figure 25 shows us our new model, which is a tree of size two.

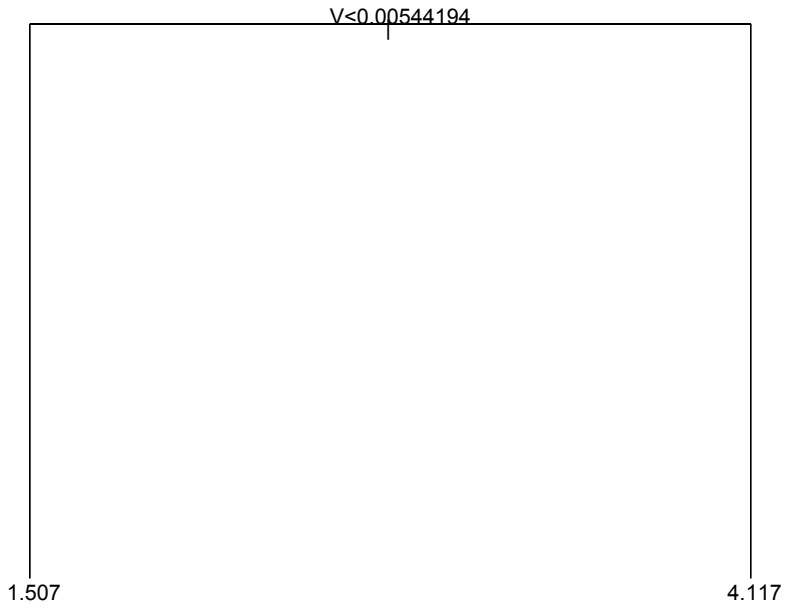

Figure 25 – The pruned regression tree that is of size two. This tree indicates that the only split that really matters is the split on the vanadium levels.

This tree splits the original node with the same variable as our previous tree. This is because all pruning did was remove splits in nodes that are too far down the tree. This new model agrees with what was stated earlier, that the best model simply looks at the level of vanadium in the oil filter. When this level goes above a certain threshold, the predicted stage of the bearing then changes.

The regression tree that we have constructed as our model to predict the stage of the bearing by the level of certain metals found in the oil filter is a very simple one. It suggests that vanadium may be used as a lone indicator of the health of the bearing, as suggested in Figure 12. This model almost seems a bit too simple, so we will take a third approach to analyzing this oil filter data to see if we can gain similar results.

E. CLASSIFICATION TREE MODELS

When analyzing the oil filter data, we did not care as much about what the exact bearing stage was, but instead all we really wanted to know was whether or not the bearing had failed. To do so, we needed to modify the data that we used when we constructed our regression tree model. Instead of an ordered response variable for our bearing stage, we classified the bearing stage into two distinct types: failed bearings and non-failed bearings.

We used another tree-based model to analyze our data: the classification tree. The idea is to select splits in each node so that the subsets created by the split are more pure than the original node. Classification trees are similar to regression trees, except the response variable is categorical instead of a continuous numerical value. Appendix *G* has more on this topic.

Again, we used *S-Plus* to construct our classification trees. We also used many of the same functions in *S-Plus* used to build regression trees. Before we began building our model, though, we first converted the data into a data type that we could use. We decided on a way to label whether a bearing has failed or not. We chose a cutoff value such that all bearings with bearing stage values higher than this cutoff were said to be failed bearings. Based on the levels of perfect classification observed in Figures 12 and 13, we chose the cutoff of the bearing stage to be 2.5, which is after the stage when there is noticeable wear on the bearing and before the cage has cracked. All bearings in our data that have bearing stages that are less than 2.5 were said to have not 'failed', while those with bearing stage values of 2.5 or greater were considered to have failed. With the proper data type, we built our classification model.

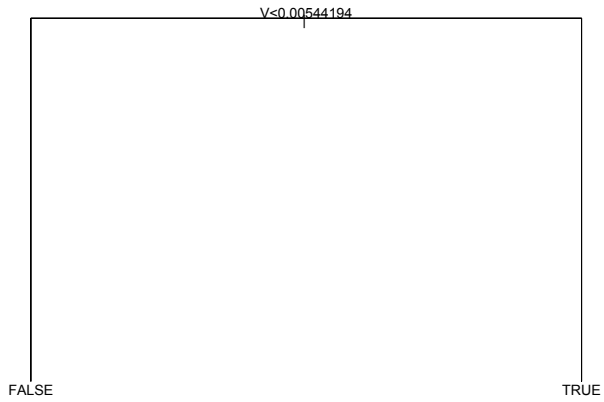

Figure 26 - Classification tree with the intermediate nodes labeled with the appropriate split and the terminal nodes labeled with the predicted response. This model has three terminal nodes.

Figure 26 shows the classification tree that was generated in *S-Plus* from our data set. At the terminal nodes, the response tells whether the bearing is projected to be a failed one. True responses indicate that the bearing has failed, while false indicates otherwise. This model says that the level of vanadium is plays a large part in telling whether a bearing has failed or not. It suggests that the level of vanadium found in the oil filter of a failed bearing is at least 0.00544194 g/cm^2, which is the same cutoff level used in our regression tree model.

With our classification tree, we had a misclassification rate of zero. This follows from Figure 12. The model we constructed had perfect classification of our sample, suggesting that we had a very good model. We did not have a need to prune the model any further, as we were already guaranteed to have minimal misclassification (zero) with a small model size (two terminal nodes).

The results from both the regression and classification tree models are very similar, mainly because the processes used to construct them are similar. It is more intuitive to use the classification tree model because this model gives us a clearer answer. With the regression tree model, the response variable was the expected bearing stage, but with the classification model, the output was simply whether the bearing is projected to have failed or not. This is the question we are trying to answer, so the classification tree

model makes more sense for us to use. Regardless, though, both models give strikingly similar results.

F. DISCRIMINANT ANALYSIS

The last regression technique we explored was discriminant analysis. Discriminant analysis is another classification technique that tries to find a set of coefficients that defines a function that separates groups of variables maximally. This function is known as a Linear Classification Function (LCF). This function can be written $LCF = w_1 V_1 + ... + w_k V_k$, where w is the discriminant coefficient, V is the set of variables, and k is the number of variables.

Discriminant analysis is used in order to find common groupings of variables. A threshold, which is used to classify objects into groups, is determined. If the *LCF* is greater than or equal to the threshold level, the object is classified into one group. If the *LCF* is less than the threshold, then it is in the other group (Groth, p.34-35).

Once again, we used *S-Plus* to aid with constructing the model. After creating a model through the *lda* function, we looked at the coefficients of the linear discriminant generated for our model. These coefficients are those of the *LCF*. We have our *w's* for our function. The input variables, V_k, are used to generate the response variable to determine which group each data entry is classified into. We also obtain the threshold value from *S-Plus*.

We shall omit the precise mathematics behind linear discriminant analysis, but for a more complete overview of the topic, consult Chapter 11 in *Modern Applied Statistics with S-Plus* by W.N. Venables and B.D. Ripley.

The data we used for this approach was the same used in the classification tree approach. We split the stage of the bearings into two groups. A bearing stage of 2.5 or greater suggested that the bearing was in need of replacement. Those bearings less than 2.5 were said to not need replacement. The linear discriminant coefficients for our model are shown in Table 4.

	LD1
V	194.6005467
Fe	0.7540444
Mo	50.0668160
Ag	-0.1222041

Table 4 - Linear discriminant coefficients

Therefore, the *LCF* is the following: $LCF = 194.60V_V + 0.75V_{Fe} + 50.07V_{Mo} - 0.12V_{Ag}$.

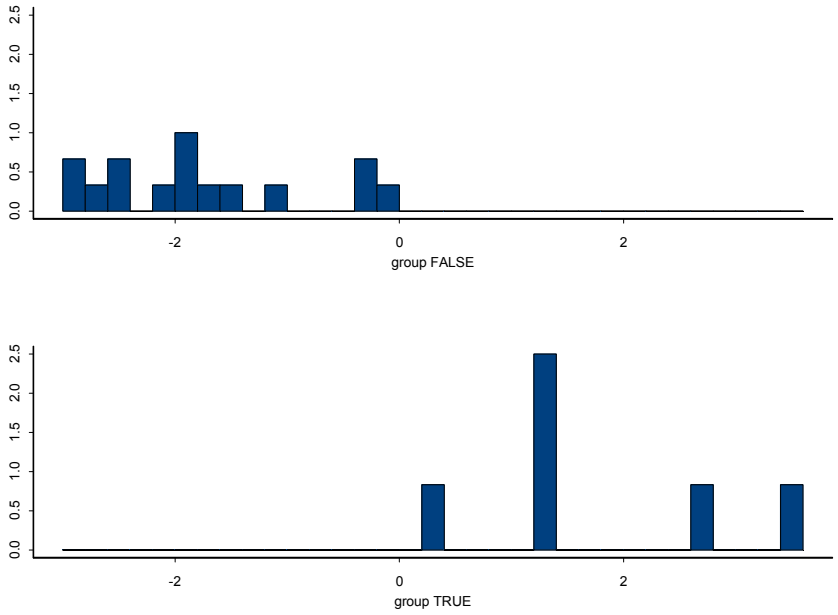

Figure 27 - These histograms show what LCF values each data entry has. The threshold value has been normalized to zero to make it easy to see which side of the threshold each data entry is on. This threshold once again gives us perfect classification.

In Figure 27, we can see the LCF values for each data entry. Each value has been adjusted so that the threshold value is zero. This makes it easy to see which side of the threshold a data entry is predicted to be on. Once again, we have perfect classification of our sample from the model. We can clearly see that the coefficients for vanadium and molybdenum are the dominant terms in the equation, which leads us to conclude that the

amounts of vanadium and molybdenum found in the oil filter are the key to determining whether or not the bearing should be replaced. This result is consistent with the other models that we have constructed.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

The current Navy policy does not give an acceptable reliability level for the bearing (currently 91.2%, according to our model), so we used life data analysis to find a replacement policy that ensures that 99.9999% of the bearings will not fail with 95% confidence. This analysis gave us a lower confidence bound of less than fifty hours, which is not a practical policy either. This policy would cost too much and not be very efficient due to much useful life of each bearing being discarded. Therefore, a policy should be based on some other type of analysis. The analysis that we have used is the oil filter analysis. Every model that we constructed greatly depended on the level of vanadium found in the oil filter sample. Some models also suggested that when molybdenum is present with vanadium, this dependence was even greater. We have therefore determined from this analysis that the level of vanadium found in the oil filter is a key indicator of bearing failure, with molybdenum being a secondary factor when vanadium is already present. The regression models seem to be too crude due to the lack of a larger sample size and non-linear behavior seen in Figures 8-11, so we do not suggest exact levels to use to monitor the health of the bearings. Nonetheless, these models have been helpful in identifying the indicators of a failed bearing.

B. RECOMMENDATIONS

Recommendations for policy changes are to use the results from the life data analysis to obtain an acceptable interval of time in which to gather oil filter samples. To implement an effective policy, we round our result of 42.8 engine hours down to forty engines hours as our time between samples. When the cumulative levels of vanadium get too large in the oil filter, we suggest replacing the degraded bearing. The cutoff that our model suggests is 0.00544194 g/cm^2. Despite this being a result of a small sample size, we recommend a cutoff at this point until a larger sample size can be collected. We also

recommend that data on the 4.5 bearing continue to be collected and the health of the bearings monitored.

C. FUTURE RESEARCH

When future data is collected, it would be helpful to record the age of the bearing as well as the levels of the metals found via the oil filter analysis. With this extra piece of data, we could combine the two approaches that we have done to come up with a more precise model that could be used to determine failed bearings based upon both the levels of the metals found in the oil filter and the age of the bearing. This would allow us to estimate the margin of safety in the oil filter approach.

APPENDIX A. FAILURE-TIME DISTRIBUTION FUNCTIONS

This appendix gives an overview of the failure-time distribution functions used in the analysis.

The probability density function, or PDF, of a distribution completely specifies the probability distribution of a continuous random variable. The PDF is denoted as *f(t)*. The area under the PDF curve is always equal to one. The cumulative density function, *F(t)*, is the total area under the PDF curve up to the point in time, *t*. Thus, we can represent the CDF in *Equation A.1*.

(A.1) $$F(t) = \int_0^t f(s)\,ds$$

The CDF therefore represents the probability of failure in the interval [0,*t*].

The reliability function is simply the probability of a non-failure over the interval [0,*t*]. We know that the CDF is the probability of failure over the same interval, and we now call this the unreliability function. *Equation A.2* shows the reliability function.

(A.2) $$R(t) = 1 - F(t)$$

This is also known as the survival function.

The final function relating to a given distribution that we wish to explore in life data analysis is the hazard rate, or more commonly referred to as the failure rate. The failure rate is the probability of failure at time *t* in the next Δt of time, given that the system has not failed before that time. The failure rate equation is given in *Equation A.3*.

(A.3) $$h(t) = \lim_{\Delta t \to 0} \frac{P(T \in (t, t+\Delta t))}{\Delta t \cdot R(t)}$$

The relation of the failure rate to the PDF and CDF is shown in *Equation A.4* (Meeker and Escobar, p.28).

(A.4) $$h(t) = \frac{f(t)}{R(t)}$$

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. COMMON LIFETIME DISTRIBUTIONS

This appendix gives a brief overview of a few common lifetime distributions.

Weibull Distribution

The Weibull distribution is a reliability distribution commonly used in life data analysis. It tends to be a good model for times-to-failure of both electronic and mechanical equipment. This distribution can have up to three parameters. If the location parameter is assumed to be zero, then the two-parameter Weibull distribution results.

Different behaviors can be modeled by the Weibull distribution, depending on the values of these parameters. The shape parameter, denoted as β, can affect the characteristics of the shape of the PDF curve, reliability, and failure rate. The shape parameter is also called the slope parameter, since it gives the slope of the CDF when plotted on probability paper. Changing the values of β can have distinctively different effects on the distribution properties. The property that we shall explore in detail is the hazard function.

It can be shown that when $0 < \beta < 1$, the failure rate is a monotonic, decreasing function. When $\beta = 1$, the failure rate is constant for all values of *t*. This is because when $\beta = 1$, the Weibull distribution reduces to the memory-less distribution, the exponential distribution. When $\beta > 1$, we can show that the failure rate is a monotonic, increasing function (Devore, p.179-180). *Equations B.1-B.4* show the reliability function, CDF, PDF, and hazard function explicitly for the Weibull distribution.

(B.1) $$R(t) = e^{-\left(\frac{t}{\eta}\right)^{\beta}}$$

(B.2) $$F(t) = 1 - e^{-\left(\frac{t}{\eta}\right)^{\beta}}$$

(B.3) $$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}}$$

(B.4) $$h(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}$$

Exponential Distribution

The exponential distribution has a location parameter and may have a translation parameter. It is usually considered to be the easiest distribution to work with. Because it is so easy to manipulate, it is often misused. It is used to represent systems that are assumed to have a constant failure rate, as seen in the special case of the Weibull distribution with $\beta = 1$. The failure rate for the exponential distribution is λ, which is a constant. The two-parameter exponential distribution has scale parameter, $1/\lambda$, or η, and location parameter, γ. The effect of γ is that the distribution is simply shifted along the *x*-axis. For positive values of γ, this implies that failures cannot occur before time $t = \gamma$ (Devore, p.174-175). The CDF, PDF, reliability function, and hazard function can all be obtained from those of the Weibull distribution with $\beta = 1$.

Normal Distribution

The normal distribution is the most used distribution and is occasionally used for reliability analysis and times-to-failure of electronic and mechanical systems. There are two parameters in the normal distribution that need to be estimated, namely, the mean, μ, of the normal times to failure and the standard deviation, σ, of the times to failure. It should be noted that to use the normal distribution with life data, we must be careful to only consider it when the mean is relatively high and the standard deviation small in comparison. This is due to the PDF of the normal distribution extending to negative infinity, leading to negative times-to-failure, which usually does not make much sense. Since the normal distribution has much utility in the modeling of life data, we can justify this with a high mean compared to the standard deviation. The hazard function of the normal distribution is a monotonically increasing function (Devore, p.158-162). *Equations B.5-B.7* give the CDF, PDF and hazard function for the normal distribution.

(B.5) $$F(t) = \int_{-\infty}^{t} f(s)ds$$

(B.6) $$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$$

(B.7) $$h(t) = \frac{f(t)}{1 - F(t)}$$

Lognormal Distribution

The lognormal distribution can be often used to model reliability data, cycles-to-failure in fatigue, loading variables in probabilistic design and material strengths. This distribution is used when the natural logarithms of the times-to-failure are normally distributed. Like the normal distribution, the lognormal distribution has two parameters; $\mu = E[\ln(t)]$, which is the location parameter, and $\sigma = \sqrt{Var[\ln(t)]}$, which is the scale parameter. This distribution is similar to the normal distribution in many ways, although we must explore the hazard function a bit to see that it is increasing monotonically only for a while, where it then begins to monotonically decrease out towards infinity (Devore, p.181-182). The CDF, PDF, and hazard function for the lognormal distribution are similar to those from the normal distribution and are shown in *Equations B.8-B.10*. The CDF and PDF of the normal distribution are here denoted $F_n(t)$ *and* $f_n(t)$.

(B.8) $$F(t) = F_n \cdot \ln(t)$$

(B.9) $$f(t) = \frac{1}{\sigma} f_n \cdot \ln(t)$$

(B.10) $$h(t) = \frac{f(t)}{1 - F(t)}$$

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. PROBABILITY PLOTTING

This appendix gives a more detailed explanation of probability plotting.

The essence of probability plotting is that if the plot is based on the correct distribution, then plotting sample points should result in a nearly-straight line (Devore, p.186). Probability plotting tries to linearize the cumulative density function (CDF) of the distribution. For the Weibull distributions, we use the logarithms of the times to failure as the inputs, or our x-values, and we must have some method to obtain our y-values, or median ranks value. This method is the median ranks method. From complex calculations, the plots can then be drawn. Refer to *Statistical Methods for Reliability Data* written by W. Meeker and L. Escobar for further detail.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. METHODS OF PARAMETER ESTIMATION

This appendix gives the details of both RRX and MLE methods of parameter estimation.

To fit a line to the data to estimate the parameters of the distribution, we could use a least squares method, known as rank regression on X (RRX), to perform our linear regression. This method takes the sum of the squares of the horizontal difference between the actual value, x, and the x-value that lies on the regression line at point y. The regression line that minimizes this sum is considered to fit best and therefore is the line that we shall choose.

To develop the probability of failure on complete data, we simply rank the failure times in order of occurrence relative to each other. Since all data points are known, this can be done. But with censored data, we cannot always be certain which event will happen next. If a data point was suspended at a time before another data point failed, there is ambiguity as to which data point was the first to fail. We must modify our way of ranking the failures. To do so, we use a weighting scheme. We find out how many different scenarios can occur with the suspension occurring first, and then the same with the failure occurring first. Then we calculate the mean order number (MON), which is represented in *Equation D.1*.

$$MON = \frac{(ca + db)}{c + d} \tag{D.1}$$

The variable a is the position of the failure if the suspension happened first, c is the number of different scenarios that can occur if the suspension occurs first, b is the position of the failure if the failure happens first, and d is the number of different scenarios that can occur from this position. This mean order number is then the value assigned to the failure. After doing this method to each failure, we have a mean order number for every failure entry. We now can proceed as we would with complete data, generating our y-values and knowing our x-values exactly to come up with our probability plots to obtain the parameters of the chosen distribution (ReliaSoft, p.39, 53-55).

While RRX is a popular method of analysis of censored life data, there is a problem with it. Since the mean order number is simply the position of the failed data points relative to other failures, there is no compensation for how spread-out the failures are from each other. To rectify this, we explore the method of Maximum Likelihood Estimators, commonly referred to as MLE's.

First, we shall explain the method of MLE's for complete data in order to get a better grasp at how MLE's will handle censored data better. The underlying idea of MLE's is to get the most likely value of the parameters of the chosen distribution that best describes the data. Say we have a probability density function (PDF) that is as follows:

(D.2) $$f(x;\theta_1,...,\theta_k),$$

where x is a continuous random variable and there are k unknown parameters to be estimated. The likelihood function seen in Equation D.3 is the product of each $f(x_i)$, where i is an element of the set of all x-values, or failure times in our case.

(D.3) $$L = \prod_i f(x_i;\theta_1,...,\theta_k)$$

Taking the partial derivatives of the natural log of the likelihood function with respect to the parameters and setting these equal to zero allows us to solve the system of k equations simultaneously to obtain our estimated θ values for the k parameters to be estimated.

Under regularity conditions, MLE's converge to the correct values as the sample size increases, making MLE's a very efficient and accurate method of parameter estimation for large sample sizes. These regularity conditions, however, do not hold when using a threshold parameter (Meeker and Escobar, p.622). By the Central Limit Theorem, the large sample size also allows us to assume the distribution of the estimates to be normal, allowing us to use the Fisher Matrix confidence bounds that will be explained later. MLE's also deal much better with right-censored data, and we now shall explore the method of using MLE's with censored data (Devore, p.268).

The likelihood function for MLE analysis of data with censored data needs to account for not just the failures, but also the suspensions as well. We use the same

technique described above, but we now add another term to the equation to account for each suspension. Thus, we get a likelihood function as follows:

$$L = \prod_{i \in I} f(x_i, \theta_1, ..., \theta_k) \cdot \prod_{j \in J} [1 - F(x_j; \theta_1, ..., \theta_k)], \tag{D.4}$$

where I is the set of complete data points (i.e. failure times) and J is the set of suspended data points (Meeker and Escobar, p.174-176).

RRX and MLE methods both assume that a distribution is already known, but life data does not usually tell the analyst what distribution it follows, if any, so the analyst must use a variety of factors in order to decide on a distribution and estimate its parameters. Some of the aspects of distributions to consider include the probability density function, cumulative density function, reliability and unreliability functions, the mean life function, commonly referred to as the mean-time-to-failure (MTTF), and the hazard function, which is also known as the failure rate.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX E. FISHER MATRIX BOUNDS

This appendix gives a few more details about Fisher Matrix Bounds.

Using the log-likelihood function, Λ, obtained in the MLE method, we can construct the Hessian matrix of the function, given in *Equation E.1* (ReliaSoft, p.68).

(E.1)
$$F = \begin{bmatrix} \dfrac{\partial^2 \Lambda}{\partial \theta_1^2} & \dfrac{\partial^2 \Lambda}{\partial \theta_1 \partial \theta_2} \\ \dfrac{\partial^2 \Lambda}{\partial \theta_2 \partial \theta_1} & \dfrac{\partial^2 \Lambda}{\partial \theta_2^2} \end{bmatrix}$$

Using the estimated values of the parameters, we can invert F to come up with the covariance matrix, where we can then get the variance of each of the parameters and the covariance between them, seen in *Equation E.2*.

(E.2)
$$\begin{bmatrix} \hat{Var}(\hat{\theta}_1) & \hat{Cov}(\hat{\theta}_1, \hat{\theta}_2) \\ \hat{Cov}(\hat{\theta}_1, \hat{\theta}_2) & \hat{Var}(\hat{\theta}_2) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial^2 \Lambda}{\partial \theta_1^2} & \dfrac{\partial^2 \Lambda}{\partial \theta_1 \partial \theta_2} \\ \dfrac{\partial^2 \Lambda}{\partial \theta_2 \partial \theta_1} & \dfrac{\partial^2 \Lambda}{\partial \theta_2^2} \end{bmatrix}^{-1}$$

We then can go back and use these values to obtain the variance of the function using the delta method. Now that we have the expected value and the variance, we can go ahead and use the assumption of normality to construct our one-sided lower confidence bound (Devore, p.270-271).

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX F. REGRESSION TREES

This appendix gives a more detailed account of regression trees.

Regression trees are models that take input variables and use them to predict the response variable. To use the x-values as a predictor, we need to define how to measure the accuracy of this predictor. One way to do so is to take the average error,

$$\frac{1}{N}\sum_{n=1}^{N}|\, y_n - d(x_n)\,|\,. \tag{F.1}$$

y_n is the response value for the n-th data entry, and $d(x_n)$ is the predicted y_n value from the model. This is known as least absolute deviation regression. Of course, we also have the more traditional measure of accuracy in regression, which is the average squared error,

$$\frac{1}{N}\sum_{n=1}^{N}(y_n - d(x_n))^2\,. \tag{F.2}$$

This is commonly called least squares regression. We shall use the method of least squares to define our measure of accuracy (Breiman, et al, p.222).

The model consists of a sequence of binary splits, where each split results in two more nodes in the model. There are two types of nodes, terminal and intermediate. Intermediate nodes are nodes that branch further down, depending on certain criteria described later. Terminal nodes are nodes where the predicted response variable has been determined, and this value is constant. At each intermediate node, one of the input variables determines which branch of the tree to follow. The predicted y-value, or output, of the model at a terminal node is simply the average of all the y-values of the data entries at that specific node. Each intermediate node has a selected input variable associated with it and the split depends solely on the value of a single variable. A cutoff level is determined for this input variable, and all data entries that have values less than the cutoff level are branched on one side of the tree, and the values greater than the cutoff level branch off along the other side, which creates two new nodes further down the tree. The tree continues to branch at each intermediate node until all branches reach a terminal node.

At each node we are really asking a question with a binary response. The question is whether $x_i \leq c$, where c is defined as the cutoff value and x_i corresponds to the i-th variable, which is where the node split is based. If the answer to the question is yes, then we follow the branch to the left. If the response is no, the branch to the right is followed.

There are three necessary elements needed when determining what the tree should look like in our model. The first is a way to select the split at intermediate nodes. Within each node, the error is calculated and the split that reduces the overall error the greatest is selected. Therefore, the regression tree simply looks to maximize the decrease in the error by splitting nodes as necessary.

The second element needed is a rule for determining when a node is terminal. Since the model is seeking to minimize the error, splitting at a node occurs when the error is significantly decreased. Thus, if splitting a node does not result in a significant decrease in the error, then no split occurs and we say that the node is terminal. There are other criteria used by *S-Plus* that we do not get into here.

The last element is a rule to assign the y-values to each terminal node. We have already stated that the y-value for each terminal node is the average of all the y-values for all data points at that specific node, which yields a constant value. This value is the value that minimizes the within node squared error.

The resulting tree of nodes forms the model that we use to make future predictions. We take the input data and run it down the tree, following the branches that our data point satisfies. When we reach a terminal node, we take the expected y-value for that node to be the predicted y-value for our data point (Breiman, et al, p.228-232).

APPENDIX G. CLASSIFICATION TREES

This appendix gives a more detailed account of classification trees.

As it was with regression trees, the construction of classification trees is based on the same three principles. Instead of an expected *y*-value, though, the output is the predicted class within the response variable to which the data point most likely belongs. The probability associated with each class is calculated and the class with the highest probability is selected.

The process of selecting node splits is different only because we wish to minimize a different impurity function. Terminal nodes are determined when the impurity does not decrease with a split in that node, just as we labeled nodes as terminal in regression trees when the error ceased to decrease.

In a classification tree, we assume the responses to be multinomial. A multinomial response is one where each observation has a probability associated with it of resulting in each of the outcomes. These probabilities are denoted p_i for the i^{th} outcome. Thus, for n trials, the probability of seeing n_i outcomes of type i is proportional to $\prod_i p_i^{n_i}$. This is the likelihood function and taking the logarithm of this function gives us the log-likelihood, the quantity to be minimized in the classification tree. Again, we go through the tree and ask a question at each node. If the variable associated with the node is ordered, then the question remains the same as with regression trees: is $x_i \leq c$? If the node split is determined by a categorical variable, then the question to be asked is whether or not x_i is an element of a determined subset of the responses to that variable. We follow the appropriate branch depending on the answer to the question at that node. When we reach a terminal node, we look at the assigned categorical response. This is the predicted response for our data point (Breiman, et al, p. 27-36).

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX H. DATA USED FOR LIFE DATA ANALYSIS

Life	Condition
1085	S
100	S
1890	F
1390	S
759	S
1380	S
971	S
861	S
1165	S
997	S
1079	S
1152	S
977	S
424	S
3428	S
2087	S
1297	S
727	S
820	S
1388	F
663	S
810	S
2892	S
951	F
1167	F
853	S
546	S
1203	F
2181	F
917	S
1070	S
799	S
1231	S

Life	Condition
1795	S
1500	S
1628	F
1145	S
152	S
246	S
61	S
966	S
462	S
437	S
887	S
1199	S
159	S
1022	S
763	S
555	S
646	F
2238	S
2294	S
897	F
1153	S
1427	S
80	S
2153	S
767	S
711	F
911	S
736	S
85	S
1042	S
2871	S
719	S
750	F

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX I. DATA COLLECTED FOR OIL FILTER ANALYSIS

Bearing Stage	V Mass G/CM2	Fe Mass G/CM2	Mo Mass G/CM2	Ag Mass G/CM2
2.1	0.0004	0.3750	0.0071	0.2353
1.3	0.0001	0.7365	0.0146	1.0639
2.0	0.0052	0.1775	0.0383	0.1534
0.5	0.0000	0.3938	0.0251	0.2644
1.4	0.0000	0.1067	0.0031	0.1735
4.2	0.0077	0.2655	0.0544	0.4189
1.0	0.0044	0.2877	0.0348	0.2011
4.5	0.0119	0.3192	0.0836	0.2866
1.3	0.0004	0.0518	0.0052	0.0591
2.0	0.0046	0.2858	0.0365	0.6711
2.0	0.0019	0.0898	0.0164	0.1157
4.7	0.0082	0.6378	0.0800	1.8727
1.9	0.0016	0.1607	0.0178	0.2364
2.0	0.0022	0.0937	0.0230	0.1186
1.0	0.0036	0.1865	0.0235	0.3373
1.1	0.0008	0.0702	0.0097	0.2874
1.6	0.0000	0.0540	0.0032	0.0337
4.0	0.0068	0.2401	0.0592	0.1096
1.4	0.0012	0.1437	0.0146	0.1866
3.0	0.0057	0.4156	0.0439	0.6974
4.3	0.0076	0.2780	0.0573	0.2288

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Barber, M., "Navy Knew of Faulty Engine Part Years Ago," *Seattle Post-Intelligencer*, 03 Aug 2002.

Breiman, L., Freidman, J., Olshen, R. and Stone, C., *Classification and Regression Trees*, Wadsworth International Group, 1984.

Devore, J., *Probability and Statistics for Engineering and the Sciences*, Duxbury, 1995.

Groth, Robert, *Data Mining: A Hands-On Approach for Business Professionals*, Prentice Hall, 1998.

Hamilton, Lawrence C., *Regression with Graphics*, Duxbury Press, 1992.

Lawless, J.F., *Statistical Models and Methods for Lifetime Data*, Wiley, 1982.

Meeker, W. and Escobar, L., *Statistical Methods for Reliability*, Wiley, 1998.

Oil Lab Analysis, Inc., "Filter Analysis", Oct 2002.
<http://www.oillab.com/index.htm>, Jun 2003.

Pitts, J., Larsen, R. and Kirk, M., "Letter to Appropriations Committee on EWWG Priorities", 25 Apr 2002.
< http://www.house.gov/pitts/initiatives/ew/020425ew-appropspriorities.htm>, Jun 2003.

ReliaSoft Corporation, *Life Data Analysis Reference*, ReliaSoft Publishing, 2000.

Selinger, M., "Prowler Engine Problems Seen Underscoring Need For New Aircraft," *Aerospace Daily*, 2002.

Venables, W.N., and Ripley, B.D., *Modern Applied Statistics with S-Plus*, Springer, 1999.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Fort Belvoir, Virginia

2. Dudley Knox Library
Naval Postgraduate School
Monterey, California

3. Dr. David Olwell
Operations Research Department (Code OR/OL)
Naval Postgraduate School
Monterey, California

4. Dr. Samuel Buttrey
Operations Research Department (Code OR)
Naval Postgraduate School
Monterey, California

5. Lynnwood Yates
Orange Park, FL

6. Craig Paylor
RCM Program Division Manager
Naval Air Depot
Cherry Point, North Carolina

7. ENS Cole Muller
Hatboro, Pennsylvania