12-2005

A Reliability Case Study on Estimating Extremely Small Percentiles of Strength Data for the Continuous Improvement of Medium Density Fiberboard Product Quality

Weiwei Chen
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a thesis written by Weiwei Chen entitled "A Reliability Case Study on Estimating Extremely Small Percentiles of Strength Data for the Continuous Improvement of Medium Density Fiberboard Product Quality." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Statistics.

Timothy M. Young, Major Professor

We have read this thesis and recommend its acceptance:

Frank M. Guess, Ramón V. León

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a thesis written by Weiwei Chen entitled "A Reliability Case Study on Estimating Extremely Small Percentiles of Strength Data for the Continuous Improvement of Medium Density Fiberboard Product Quality". I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Statistics.

Timothy M. Young
Major Professor

We have read this thesis
and recommend its acceptance:

Frank M. Guess

Ramón V. León

Accepted for the Council:

Anne Mayhew
Vice Chancellor and
Dean of Graduate Studies

(Original signatures are on file with official student records.)

A Reliability Case Study on Estimating Extremely Small Percentiles of Strength Data for the Continuous Improvement of Medium Density Fiberboard Product Quality

A Thesis
Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Weiwei Chen
December 2005

ACKNOWLEDGEMENT

This work was supported by the United States Department of Agriculture, Special Wood Utilization Grants Program as administered by The University of Tennessee Institute of Agriculture Experiment Station, contract R11-2219-109, and the Tennessee Agricultural Experiment Station McIntire Stennis project TEN00MS-89. A special thanks for funding support goes to the University of Tennessee Forest Products Center, Dr. Timothy Rials, Director, and former Department Head Dr. George Hopper, currently Dean, Mississippi State University.

I would like to express my gratitude to the Committee Co-chairs and Professors Timothy Young, Dr. Frank Guess, and Dr. Ramón León, for their guidance throughout this research project. It has been great pleasure and a privilege to work with such an excellent team that nurtures and keeps developing young talents. I also want to thank Dr. Bill Seaver and Dr. Halima Bensmail for their valuable advice on my research and for reviewing the thesis.

I am thankful of the honorarium awarded by SAS and the unique opportunity of presenting part of my thesis work at the 2nd Annual JMP® Conference, June 7-8, 2005, SAS World Headquarters, Cary, North Carolina, U.S.A.

ABSTRACT

The objective of this thesis is to better estimate extremely small percentiles of strength distributions for measuring failure process in continuous improvement initiatives. These percentiles are of great interest for companies, oversight organizations, and consumers concerned with product safety and reliability. The thesis investigates the lower percentiles for the quality of medium density fiberboard (MDF). The international industrial standard for measuring quality for MDF is internal bond (IB, a tensile strength test). The results of the thesis indicated that the smaller percentiles are crucial, especially the first percentile and lower ones.

The thesis starts by introducing the background, study objectives, and previous work done in the area of MDF reliability. The thesis also reviews key components of total quality management (TQM) principles, strategies for reliability data analysis and modeling, information and data quality philosophy, and data preparation steps that were used in the research study.

Like many real world cases, the internal bond data in material failure analysis do not follow perfectly the normal distribution. There was evidence from the study to suggest that MDF has potentially different failure modes for early failures. Forcing of the normality assumption may lead to inaccurate predictions and poor product quality. We introduce a novel, forced censoring technique that closer fits the lower tails of strength distributions, where these smaller percentiles are impacted most. In this thesis, such a forced censoring technique is implemented as a software module, using JMP® Scripting Language (JSL) to expedite data processing which is key for real-time manufacturing

settings.

Results show that the Weibull distribution models the data best and provides percentile estimates that are neither too conservative nor risky. Further analyses are performed to build an accelerated common-shaped Weibull model for these two product types using the JMP® Survival and Reliability platform. The use of the JMP® Scripting Language helps to automate the task of fitting an accelerated Weibull model and test model homogeneity in the shape parameter. At the end of modeling stage, a package script is written to readily provide the field engineers customized reporting for model visualization, parameter estimation, and percentile forecasting.

Furthermore, using the powerful tools of Splida and S Plus, bootstrap estimates of the small percentiles demonstrate improved intervals by our forced censoring approach and the fitted model, including the common shape assumption. Additionally, relatively more advanced Bayesian methods are employed to predict the low percentiles of this particular product type, which has a rather limited number of observations. Model interpretability, cross-validation strategy, result comparisons, and habitual assessment of practical significance are particularly stressed and exercised throughout the thesis.

Overall, the approach in the thesis is parsimonious and suitable for real time manufacturing settings. The approach follows a consistent strategy in statistical analysis which leads to more accuracy for product conformance evaluation. Such an approach may also potentially reduce the cost of destructive testing and data management due to reduced frequency of testing. If adopted, the approach may prevent field failures and improve product safety. The philosophy and analytical methods presented in the thesis also apply to other strength distributions and lifetime data.

TABLE OF CONTENT

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I. INTRODUCTION

Medium Density Fiberboard (MDF) is a superior engineered wood product of high reliability with desirable machining capabilities. MDF provides enhanced qualities of a consistent surface, uniform core density, and freedom from irregularities found in naturally grown wood products. MDF is a non-structural wood composite which is used primarily in furniture, cabinets, shelving, flooring, molding, etc. Reliability of products made from MDF is important to the end-user.

Product "life" for MDF can be measured in terms of the strength to failure, as opposed to the time to failure. The strength to failure is a crucial reliability parameter of the product. Estimation of the strength allows the producer to make assurances to customers about the safe, useful "strength" range of the product. The key measure of reliability for MDF is internal bond (IB) which is a tensile strength destructive test (units of measure are p.s.i. - pounds per square inch; or metric units of kilograms per square centimeter). See Guess, Walker, and Gallant (1992), Guess and Proschan (1988), and Guess, Hollander and Proschan (1986) for other measures of reliability.

The lower percentiles may be of particular interest for companies, oversight organizations, and consumers in specifying the product reliability of MDF. Compare Kim and Kuo (2003), Kuo, Chien, Kim (1998), and Kuo, Prasad, Tillman, and Hwang (2000) for more on percentiles. Also, see Walker and Guess (2003) for strengths of container bottles using Kaplan and Meier graphs and nonparametric approaches. Guess, Edwards, Pickrell and Young (2003) explored and viewed graphically MDF data, but did not provide confidence intervals for percentiles. We compute and discuss in later chapters of this thesis for a new MDF IB data set such interval estimates for lower percentiles using accepted statistical methods such as parametric modeling, nonparametric, bootstrapping, and Bayesian prediction.

In this research, we investigate two important MDF product types defined as "Type 1" and "Type 5". The physical difference between the two product types is density. Type 1 is more demanded which requires higher production volume, while Type 5 provides more value for a consumer niche requiring higher density MDF. Both types are of great commercial interest to both the manufacturer and consumer. The production costs of Type 5 is higher than Type 1 given its higher density, i.e., higher density requires higher raw material inputs of wood and resin which requires slower pressing speed. Different sample sizes of Type 1 and Type 5 products existed given the differences in the production volume for each product (Table 1).

Table 1. Key Specifications of Type 1 and 5 Products

Type	Density	Thickness	Width	Tests	Note
1	A	same	same	396	Standard density
5	B	same	same	74	High density

The objectives of the research are:

- Estimate the first percentile of internal bond for both product types;
- Investigate the failure mode based on statistical evidence;
- Demonstrate a complete case study of sound analytical strategy;
- Develop new statistical methods for data preparation and analysis.

We introduce a novel technique called median censoring to weight lower observations. Results of the analyses for the complete data and forced censoring at the median for product Type I are discussed in Chapter III of the thesis. There is evidence from the forced censoring analyses to suggest that MDF has potentially different failure modes for early failures. Probability plots illustrate that expected failure distributions like the Weibull, do not fit the raw data satisfactorily. Even the distribution of overall best fit assuming the normal distribution provides poor estimates of the smaller percentiles. After applying this technique, a better goodness of fit in the lower tails is obtained where the smaller percentiles are impacted the most.

The exploratory results discussed in more detail in Chapter III show that the Weibull distribution fits the lower strength MDF tests better, while the overall strength appears to be best fit by the normal distribution. This conclusion supports Weibull's theory of a "weakest link model" for early failures (Weibull 1939, 1951); and assuming overall failures are normally distributed by the use of the Central Limit Theorem (CLT) is more appropriate . The CLT normality may be a result of the physical properties making up the overall strength which is typically the sum of many individual fiber strengths. Chapter III also presents results of percentile estimates using a simple model.

Chapter IV explains in greater detail the mechanism of median censoring and the extended forced censoring technique at any percentile in any censoring type (left, right, interval). Another practical example for the application of this technique is discussed in Chapter IV as an extension of the results presented in Chapter III.

In Chapter V, both modeling methods and the median censoring technique are cross-validated by the bootstrapping method. The confidence intervals for various parametric models for both the complete and the forced censoring cases are included. The Weibull distribution is the best model for the strength of Type 1 product. Bootstrap estimates of the small percentiles improve the consistency of the fitted model's percentile confidence intervals and support the use of the forced censoring technique. Both percentile bootstraps and t bootstrap intervals algorithms are described in detail in Chapter V.

With confidence in the Weibull model and given the uniqueness of the Type 5 product, we illustrate graphically and parametrically both product Types 1 and 5 in Chapter VI. JMP® is used extensively for its simplicity, interactivity and graphic discovery capabilities (SAS Institute, Inc. 2004). An interesting discovery presented in this chapter is the similar shape that both product types demonstrate on the probability plot.

Chapter VII starts by reexamining the graphical significance of the common shape location-scale model. A rigorous statistical test is performed to prove the common-shape or homogeneity hypothesis. The automation of this customized test in JSL is also introduced with the interpretation of hypothesis test results. Chapter VII concludes with an additional examination of the common-shape model through the residual plot.

Chapter VIII starts by discussing the sample size issue pertaining to the Type 5 product and other potential types. Bayesian methods are introduced to help solve the problem. The roots of Bayesian philosophy are reviewed and the difference of Bayesian interpretation of results from the classical approach is stressed. Chapter VIII also generalizes and critiques the results of low percentile estimates in all previous sections. Finally, Chapter IX is a summary of the overall strategy, methods and results of the thesis.

The statistical software S+ (http://www.insightful.com/products/default.asp) and a free add-on called Splida (http://www.public.iastate.edu/~splida/) are used with some Matlab (http://www.mathworks.com) in the analysis for the thesis. JMP® (http://www.jmp.com, a SAS® division), a statistical discovery software platform with scripting, is also used in the analysis for the thesis. Tutorials on the use of both software for reliability applications can be found at Professor Ramón V. León's course webpage at http://web.utk.edu/~leon/.

CHAPTER II. LITERATURE REVIEW

Chapter I outlined the objectives and methodologies of the thesis. A review of the fundamental principles and scholarly work that are the basis of the thesis is presented in this chapter.

Quality Management Principles

What is quality? In a popular sense, quality seems purely a judgment call. Each person may have his/her own perception of the quality of something, product or service. Yet, this truly reveals the nature of "quality" because quality judgment is the response of customers. Quality is *not* meeting written specifications and nothing more, as some writers on quality control may have suggested. Quality must be judged in terms of customer satisfaction. When Crosby (1979) defines quality as "conformance to requirements", he does not just mean conformance to specifications. Deming (1986) warns against the "fallacy of zero defects" and that "the supposition that everything is all right inside the specifications and all wrong outside does not correspond to this world". Also see Taguchi's customer loss function (1986). Mendenhall and Sincich (1995) have discussed the operating characteristic curve that describes both the customer's and the producer's risk given an acceptable quality level (AQL). English (1998) summarizes that quality is "consistently meeting customer's expectations", and "not necessarily exceeding them".

Quality is not intangible. It can be measured with the most fundamental business measures, e.g., bottom line figures, cost of non-quality products and services, lost profit due to customer dissatisfaction, or created revenue because of return customers, goodwill, etc. Quality is not a by-product; rather, it should be treated as genetic part of the asset, just like employees, working capital, and other resources. Therefore, quality is manageable. There are established principles by quality pioneers such as Deming, Juran, Crosby, Ishikawa, Shewhart, Imai, English and others (Deming 1986, Juran 1988, Crosby 1979, Ishikawa 1994, Shewhart 1986, Imai 1989 & 1997, English 1998). The key components of these principles can be summarized as follows:

Customer focus: listen to the customer; understand the market; learn the customer's needs; establish a partnership mind set and relationship with the customer; educate and help your business partner be successful.

Continuous process improvement: or Kaizen; "the art of continuous and incremental improvement" in Japanese; always be the best and get ahead of the curve in knowing the customer's needs; improve everything in the organization by encouraging everyone to take responsibility for the process.

Scientific methods: statistical methodologies and techniques such as statistical process control (SPC), Shewhart Cycle, Six Sigma (Snee and Hoerl, 2003) including Design for Six Sigma (DFSS) and DMAIC (Define-Measure-Analyze-Improve-Control).

The integration of process thinking, understanding of variation, and data-based decision making is often referred to as statistical thinking (Hoerl and Snee, 2002).

Strategy for Reliability Data Analysis and Modeling

Technically, reliability is defined as the probability that a product or subject will perform its intended function under operating conditions, for a specific period of time (Meeker and Escobar, 1998). Condra (1993) emphasizes that "reliability is quality over time". In today's world, customers expect the product to be reliable and safe; on the other hand, the global marketplace forces the manufacturers to compete in multiple fronts, such as brands, price, quality, innovation, etc. One successful business strategy is to build competitive advantage on the quality of products and services, enhanced by advanced technology and well-trained personnel, instead of relying purely on low price and cheap labor.

To implement this strategy, it is essential to hire and train qualified employees with quantitative knowledge and skills for running designed experiments or tests, collecting quality data, assessing various facets of the data, and making accurate forecasts. Meeker and Escobar (1998) provide a useful general strategy for data analysis and modeling:

1. Start the analysis by visually examining the data without any distributional or strong model assumptions. The primary tool for these initial steps is graphical analysis;

2. It is useful to fit one or more parametric models to the data for the purpose of description, estimation, or prediction. Sometimes, one can combine prior knowledge or data into the current analysis. Many software packages provide these model fitting functions or modules;

3. Examine appropriate diagnostics and assess the adequacy of model assumption. Graphical tools, analytical measures, simulations, and validation techniques are useful at this stage;

4. Once the assumed model is adequate, generally proceed to estimating parameters and predicting desired statistics. However, state the results, with caution, which should include information that reflects uncertainty, variability, or conditions of model assumptions;

5. Display the results graphically; pay attention to the importance of model interpretation.

Information and Data Quality

There are still a few more issues that may go beyond the scope of Meeker and Escobar's (1998) book but must be addressed in the practical world of reliability engineering. First, data collection in the real world is often not simple. In some applications, a massive amount of data or information need to be simultaneously collected and stored as the production process is running. Field engineers or operators need prompt analysis results in order to monitor and manage the process. See Young and

Guess (2002) for how such data is stored and used in a real time data base with regression modeling to predict strength.

English (1999) proposed 14 points of information quality, expanding upon Deming's well-known 14 points for management transformation. He relates data integrity to the information quality in a TQM setting and stresses that low-quality data scrap and rework is essentially the same as the physical product defects in the industrial age. In the Total Information Quality Management (TIQM®) methodology developed by English (1999), low-quality data cost can and should be quantified, and eventually, businesses should design quality into the collection of data rather than depending on data inspection. An alternative approach to English's work is called Total Data Quality Management by Huang, Lee, and Wang (1999), which develops different metrics in the evaluation of data quality. Also compare Redman (2001) for data accuracy, clear definition of terms, and the relevancy of data.

Even though the data may not be severely contaminated, data cleansing and reengineering are often useful in preparing for better statistical analysis. Sometimes, carefully devised data preparation can guard the analyst against mis-specified model assumptions and consequently erroneous estimates. For example, the normal distribution is often assumed for many applications during the quality improvement process (Meeker and Escobar, 1998). However, there are often many practical cases where a better fit of the data are from non-normal distributions. Where normality is not appropriate, forcing the normal distribution model can lead to inaccurate prediction of key process parameters and result in poor product quality. Stanard and Osborn (2002) have discussed general strategies for handling non-normality in a "Six Sigma Quality" context. Guess, León, Chen, and Young (2004) have presented a case study, in which the internal bond or strength of medium density fiberboard (MDF) does not follow perfectly a normal process. The estimation of crucial lower percentiles can be poor when incorrectly assuming the normal distribution and such analytical errors can be very costly for manufacturers; compare Guess, Edwards, Pickrell and Young (2003).

CHAPTER III. GRAPHICAL EXPLORATION AND PRELIMINARY STUDIES OF TYPE 1 PRODUCT DATA

The complete data set of 396 failures for the Type 1 product is initially fitted to several popular distributions of lifetime data. The qualities of the model fits are examined graphically on the respective probability plot in Figure 1. It is highly recommended by the author to implement this exploratory step before making any further statistical inferences. By plotting the data, one can quickly identify underlying issues and proceed with the most appropriate strategies including median censoring. Recall that the IB used throughout the thesis analysis is measured in pounds per square inches (psi) and is pressure to failure data compared to typical life to failure data.

Figure 1 displays that observed early failures deviated from the straight lines of parametric Maximum Likelihood (ML) estimates. There are a few data points on the lower tail and mostly the upper tail that were not well captured by any of the distribution models, evidenced by both tails stretching outside the coverage of pointwise 95% confidence interval of ML estimated models. Later, it is important for quality goals that need both a specification number and pointwise confidence interval on the reliability of the product. We notice that the amount of sampling variability at the extreme observations can be rather large, as suggested by the simultaneous confidence bands in Figure 2. See, for example, Section 3.8 of Meeker and Escobar (1998) for more details.

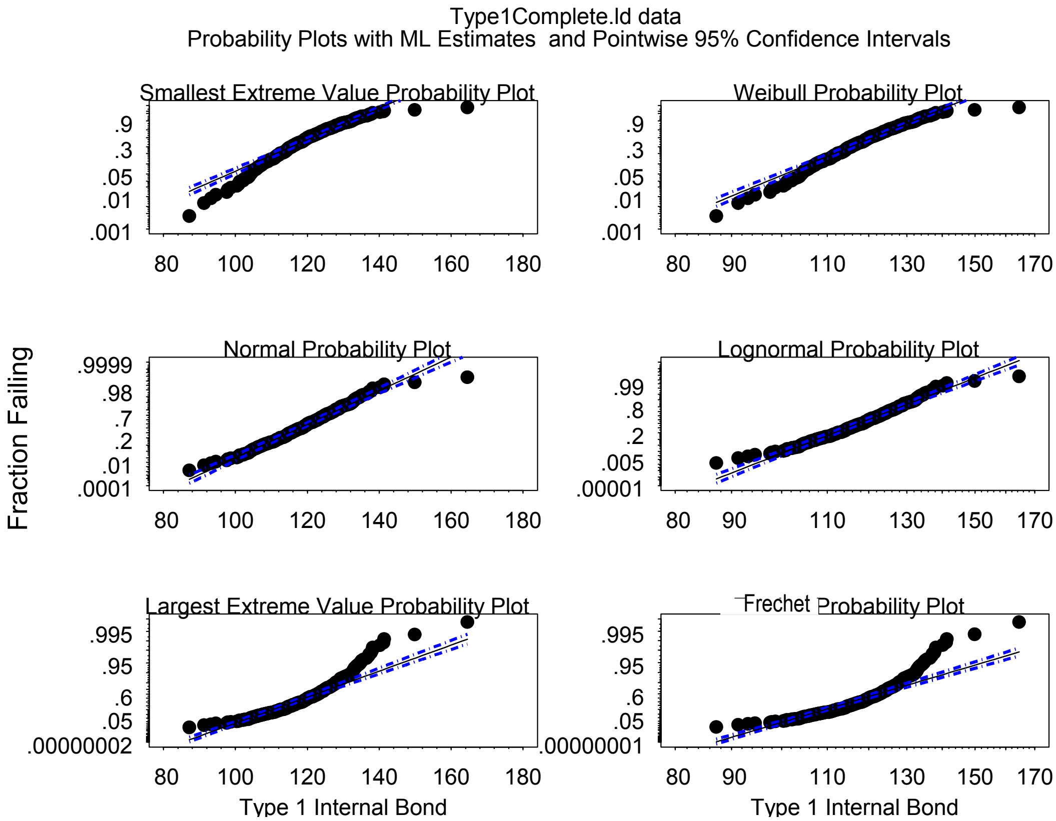


Figure 1. Complete Data Probability Plots with ML Estimates

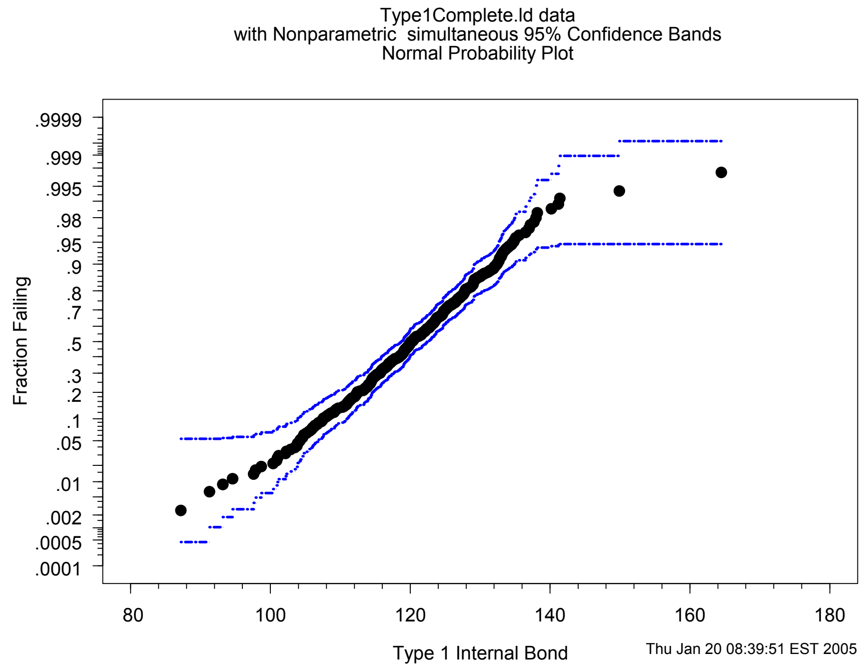


Figure 2. Normal Probability Plot for the Complete Data and Simultaneous Approximate 95% Confidence Bands

The illustrations in Figures 1 and 2 suggest that the ML estimated normal distribution model seems to be the best fit to the entire data, and that some curvature change exists no matter which model is fitted. The existence of such behavior in the data might be signs of potential different failure modes, or mixture of subpopulations at the extremes, or of outliers during the breakdown, or measurement error (section 6.6, Meeker and Escobar, 1998). In these cases, a certain model, for example the normal distribution, may fit the majority of the data better than the other, but this is merely achieved by compromising the local approximation of failure modes toward extreme values, lower or upper. Or, the shape of an empirical failure model, such as Weibull, happens to be largely determined by the upper part of data, while the desired lower percentiles deviate from the observed data which are less influential.

We will further present quantitative evidences in later chapters that the first percentile (and lower percentiles) estimates using the complete data naively were generally unreliable, either too optimistic or overly conservative. This may lead to higher costs of manufacturing when product reliability is misjudged, e.g., “over-engineering” the product with higher raw material inputs than necessary. With the existing data set that has included sufficient information for the small percentile estimates, it is a cost-efficient and statistically sound solution to reengineer and cleanse the data of potential outliers and reassess the pragmatic information quality for the lower percentiles. See English (1999).

Because the goodness of a global model fit sacrifices the more important lower percentile estimates, we may use a forced median-censoring technique to increase the model dependence on the lower tail information. In a traditional reliability context, censoring refers to an observed subject’s true failure time recorded as being either before or after the time of inspection, if the subject does not fail at that exact time. In the proposed forced censoring technique, we retain all the observations no larger than the median intact as exact failures. Observations beyond the median are censored at a forced value slightly larger than the median, but less than the next true observed failure above the median. Essentially, such a technique reengineers the data set so that the upper half of the complete data set is regarded as being censored at the median. Hereby, these large observations are not as informative as the smaller observations in that their breakdown strengths are only known to be larger than the median. In other words, more weights are put on the observations of smaller values in fitting a model. None of the data integrity is violated.

This weighted data of Type 1 product is fitted by select models in Figure 3. We retain 198 observations on the lower tail while censoring the upper half of data (198 observations).

Upon censoring the upper half of the data, the fitted ML estimated lines of the Weibull and Smallest Extreme Value distributions (Figure 3) are able to capture the pattern of small extreme values more “closely” and more importantly, the data on the lower tail, than other models. The lowest data, which would be considered incorrectly as outliers if it were without median censoring (compare both Figures 1 and 2), now falls completely within the 95% confidence interval of a Weibull or S.E.V. model.

For additional specific numbers, say 90 psi or a previous first percentile, for example, with improved, continuous quality goals, we really want and prefer to have pointwise confidence intervals for their new, improved reliability to report to management. When the interval fails to enclose the observed data, however, it is an appropriate conclusion that the data is not as consistent with the model hypothesis (Section 7.3.2, Meeker and Escobar, 1998). Thus, we suspect different underlying failure modes over the whole range of observed failures. The early failures are similar to the

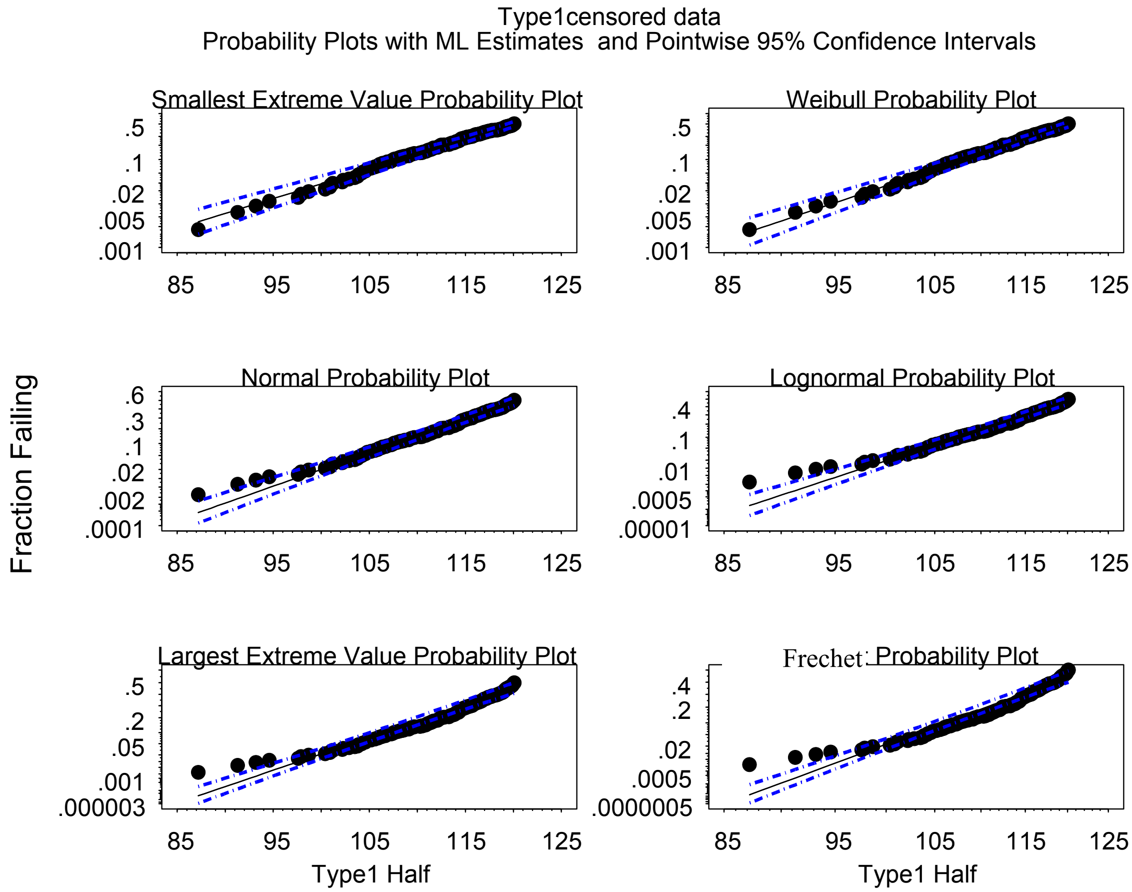


Figure 3. Probability Plots of Median Censored Data with ML Estimates

"infant mortality" for many manufacturing settings. Recall the Weibull model (Weibull 1939, 1951), which governs the "weakest link" of many competing failure processes for the catastrophic effect of even a very small external force upon a certain portion of inferior products, here mostly the lower percentiles. However, the breakdown of the majority of MDF products is determined by a combined strength of individual fibers and bonding between the fibers, i.e., the Central Limit Theorem appears to be suitable.

Table 2 illustrates the loglikelihood and AIC scores of select models as quantitative evidence for a different early failure mode than the normal model. The Akaike's Information Criterion (AIC) for model selection (Akaike, 1973, 1974, 1987; Bozdogan, 2004) favors the model that minimizes AIC score based on the same information (median censoring or not). Therefore, the Weibull ML fit, also seen in Figure 4, is the best approximating model to the censored data set.

Table 2. Select Model Scores for the Complete and Censored Data

ML fit	With median censoring		W/O median censoring	
	Log likelihood	AIC	Log Likelihood	AIC
Weibull	-868.8	**1741.6**	-1518	3040
S.E.V.*	-869.4	1742.8	-1527	3058
normal	-871.5	1747	-1469	2942
lognormal	-874.3	1752.6	-1471	2946
exponential	-1277.2	2558.4	-2293	4590
logistic	-869.2	1742.4	-1461	2926
logLogistic	-869.6	1743.2	-1463	2930
L.E.V.*	-885.1	1774.2	-1512	3028
frechet	-892.6	1789.2	-1539	3082

* S.E.V. denotes the Smallest Extreme Value model where applicable in this thesis; L.E.V. for the Largest Extreme Value model.

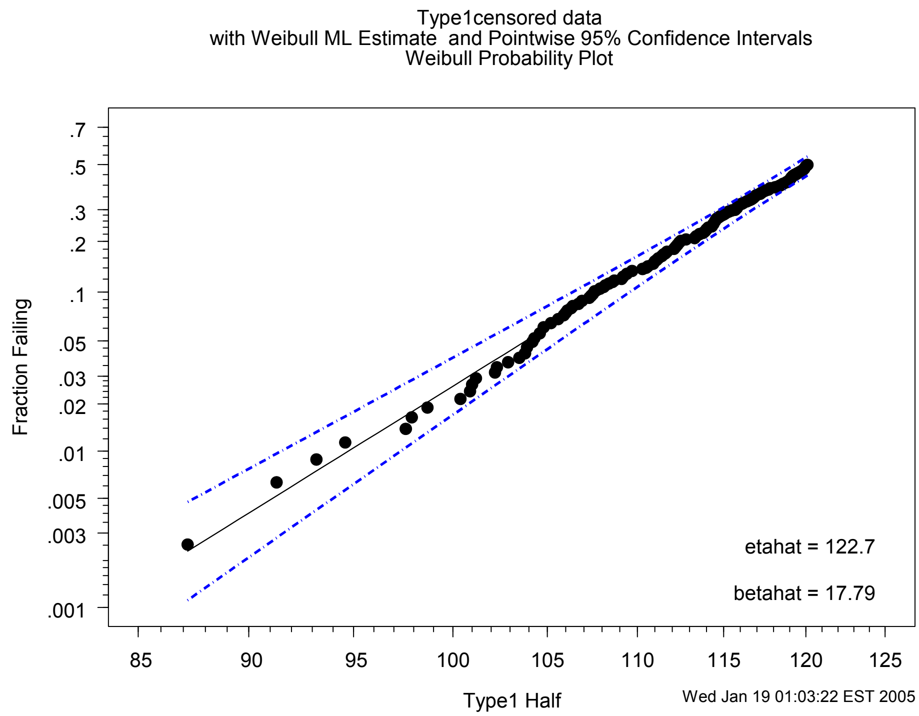


Figure 4. Median Censored Data on the Weibull Probability Plot

Figure 5 shows the Weibull probability plot and how the first percentile estimates are obtained from all three models in Table 3. The solid straight line and the corresponding 95% pointwise confidence bands show the Weibull ML fit, while the curve of normal ML fit deviates the most severely from the lower tail of observed failures. The difference between the Weibull and S.E.V. model on the first percentile is trivial. The S.E.V. model may be of interest if a conservative estimate is preferred in the practical context of reliability evaluation. It is noticeable that the S.E.V tends to produce overly underestimated results as the percentage (quantile) becomes smaller than 1% (0.01).

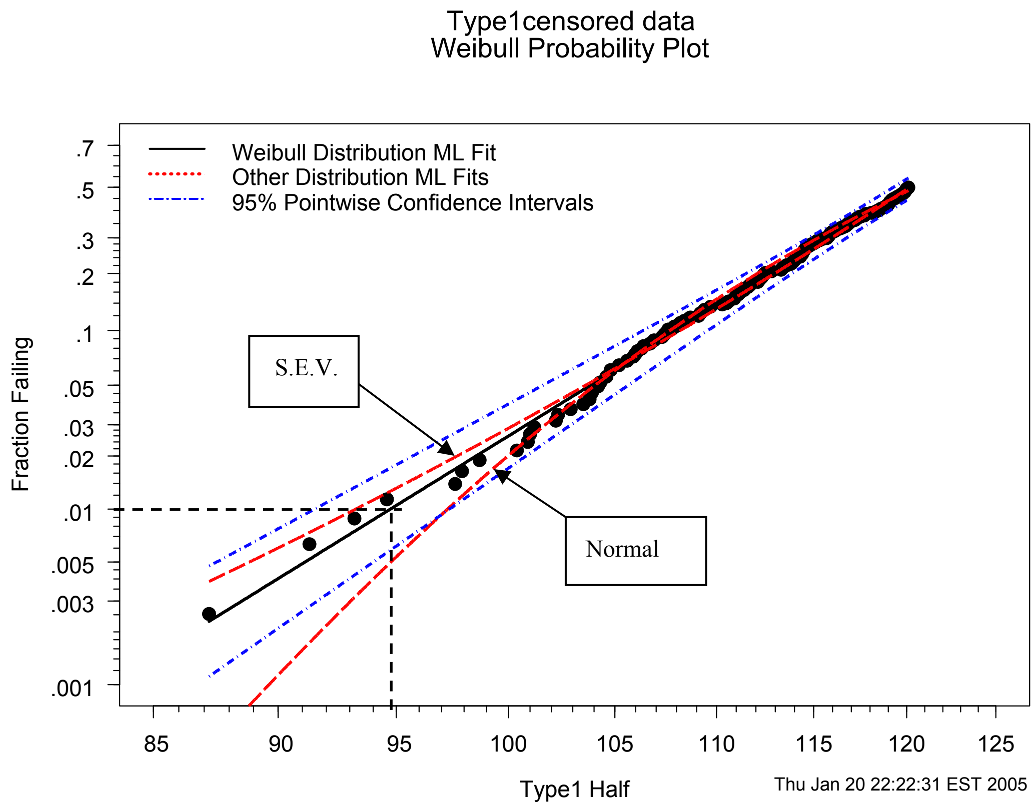


Figure 5. Estimating the First Percentiles from Select Models

Table 3. The First Percentile Normal-approximation Estimates of Select Models for the Censored Data

ML fit	p	Percentile	Std_Err_	95%_Lower	95%_Upper
Weibull	0.01	**94.746**	1.47018	91.908	97.672
S.E.V.	0.01	93.255	1.75203	89.821	96.689
Normal	0.01	97.262	1.16150	94.986	99.539

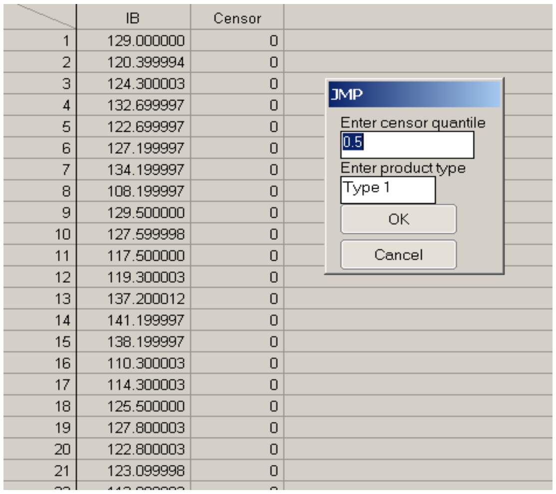
Splida also computes the asymptotic normal-approximation confidence intervals while generating the "probability plot with parametric ML fit", which is a macro in the Splida menu. Table 3 presents the 95% confidence intervals generated based on the Weibull, S.E.V., and normal ML fits. The S.E.V. model gives the most conservative estimate, while the normal model is too optimistic because the data is unduly fitted. See Section 7.3.3 and 8.4 of Meeker and Escobar (1998) for more details on the normal assumption of log-percentile in this estimation method. Meeker and Escobar (1998) comment, that "with moderate-to-large samples (the normal approximation) are useful for preliminary confidence intervals" and "quick, useful, and adequate for exploratory work". Other alternatives of estimating confidence intervals, including a simple nonparametric estimation, various bootstrap and Bayesian methods, are discussed in the later chapters of the thesis. There will be an overall assessment of the first percentile confidence interval estimates presented at the end of Chapter VIII.

CHAPTER IV. FORCED CENSORING TECHNIQUE AND JSL IMPLEMENTATION

In the last chapter, we have introduced a median censoring technique described as: "all the observations no larger than the median are retained intact as exact failures, while observations beyond the median are censored at a forced value slightly larger than the median but less than the next true observed failure above the median." After applying this censoring technique a better goodness of fit is found in the lower tails, where the smaller percentiles are impacted the most.

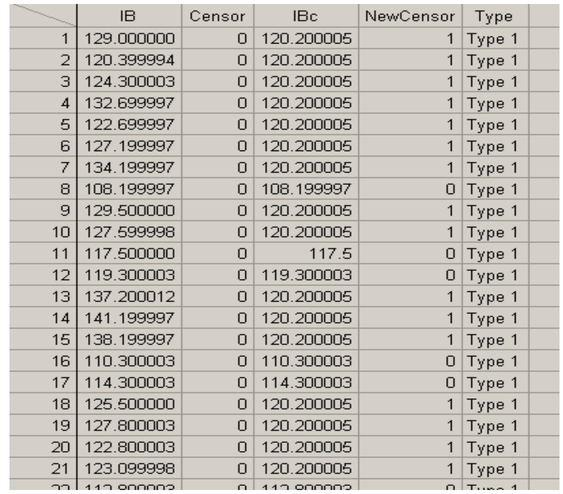
We can further extend the median censoring technique to any portion of a data set. Employing the power of JMP® Scripting Language (JSL), we scripted a module in JMP® that automatically "force-censors" the data from any percentile point of interest. More specifically, the implementation of this JMP® script is to replace the observations larger than a specified percentile value with this new percentile value, and label the replaced observation as "censored". Note that in JMP® by default, censor label values of zero indicate the event (e.g., 0: failure) and a non-zero (e.g., 1: able to customize) code is a censored value; whereas Splida uses 1 for exact failures and 2 as censors.

The script of JSL-implemented forced censoring can be found in the Appendix. Figure 6 includes the illustrations of interactive JMP® dialog before censoring and an example data table readily useful for further modeling.



	IB	Censor
1	129.000000	0
2	120.399994	0
3	124.300003	0
4	132.699997	0
5	122.699997	0
6	127.199997	0
7	134.199997	0
8	108.199997	0
9	129.500000	0
10	127.599998	0
11	117.500000	0
12	119.300003	0
13	137.200012	0
14	141.199997	0
15	138.199997	0
16	110.300003	0
17	114.300003	0
18	125.500000	0
19	127.800003	0
20	122.800003	0
21	123.099998	0

a.) JMP® dialog asking for customized censor quantile



	IB	Censor	IBc	NewCensor	Type
1	129.000000	0	120.200005	1	Type 1
2	120.399994	0	120.200005	1	Type 1
3	124.300003	0	120.200005	1	Type 1
4	132.699997	0	120.200005	1	Type 1
5	122.699997	0	120.200005	1	Type 1
6	127.199997	0	120.200005	1	Type 1
7	134.199997	0	120.200005	1	Type 1
8	108.199997	0	108.199997	0	Type 1
9	129.500000	0	120.200005	1	Type 1
10	127.599998	0	120.200005	1	Type 1
11	117.500000	0	117.5	0	Type 1
12	119.300003	0	119.300003	0	Type 1
13	137.200012	0	120.200005	1	Type 1
14	141.199997	0	120.200005	1	Type 1
15	138.199997	0	120.200005	1	Type 1
16	110.300003	0	110.300003	0	Type 1
17	114.300003	0	114.300003	0	Type 1
18	125.500000	0	120.200005	1	Type 1
19	127.800003	0	120.200005	1	Type 1
20	122.800003	0	120.200005	1	Type 1
21	123.099998	0	120.200005	1	Type 1

b.) Data prepared for further analysis

Figure 6. Screen Illustrations of Forced Censoring Implemented in JMP®

The right-censoring mechanism is sufficient in our case study of extremely small percentiles. Other product applications may require modeling the upper part or an intermittent portion of data. For example, a process engineer may want to estimate the number of particles on a silicon wafer which leads to defective computer chips. Both the small and large percentiles of the distribution of particle numbers per wafer would be key indicators of the quality of the production run. The normal probability plot may show a severe departure from the straight line on both the lower and upper parts of the distribution. Further analysis may reveal inherently non-normal data with no known simple distribution function yielding satisfactory estimates to the key percentiles on either end of the distribution. Different portions of the distribution would need to be examined by themselves in such a complex case. Observations may be treated as right-censored, left-censored, interval-censored, or remain entirely uncensored dependent on analytical needs. The example script (Appendix) should be able to implement a modified all-purpose forced censoring mechanism in JSL. All three types of censoring mechanism, right, left, or interval can be customized in one uniform format of interval censoring, also called “arbitrary censoring”. See JMP® Manual: Statistics and Graphics Guide, section “Interval Censoring” in the topic titled “Survival and Reliability Analysis”.

The central philosophy of the forced censoring technique is to preserve as much useful information as possible in the raw data and to extract desired local information from leveraged data. This is a very useful technique when data is complex in nature and the data collection is expensive. The forced censoring technique is different from other known strategies such as truncation, Box-Cox transformation, or segmentation, when working with non-normal data. The complexity of data structure, like multiple failure modes, is well respected and captured as a whole even when estimating a local parameter.

The forced censoring technique can be used for many other applications beside strengths of materials and their lower percentiles. For example, it can be employed successfully for warranty or lifetime data analysis when estimates of new warranties are based on smaller percentiles.

CHAPTER V. USING THE BOOTSTRAPPING METHOD FOR MODEL VALIDATION AND PERCENTILE ESTIMATION

The novel technique of forced median censoring shows its capability in helping detect possibly different failure modes and improving the model fit, as well as percentile estimates on the lower tail. However, there are some potential weaknesses, both theoretic and practical, in the approach thus far. Figure 2 has suggested that the sampling variability at the extremes can be rather large so that the ML fit plots may give the false impression in model comparisons (Section 6.4.1 Meeker and Escobar, 1998). The entropic information model selection criterion such as AIC affirms our conclusions drawn from probability plotting; yet, the normal-approximation confidence interval still has its theoretic shortcomings. For example, the normal assumption of transformed data may not be the case especially when the sample size is not large. In this section, we rely on the bootstrap method to further demonstrate the estimation improvements from applying the forced median censoring technique, which will provide more accurate confidence intervals. This may help practitioners' work and improve the decision-making capabilities of management. Table 4 presents the 95% confidence intervals of the first percentile for both complete and median censored data, using the approximate and bootstrap nonparametric and parametric methods.

Table 4. 95% Confidence Intervals of the First Percentile Computed Under Various Model Assumptions and With/Without Median Censoring Technique

Model assumption	With median censoring		W/O median censoring		Interval Method
	95%_Lower	95%_Upper	95%_Lower	95%_Upper	
Nonparametric	87.2	98.7	87.2	98.7	Normal-Approximation
Nonparametric	86.647	100.035	86.151	101.242	Bootstrap-t
Nonparametric	87.200	100.630	87.200	100.676	Bootstrap-Percentile
Weibull	91.908	97.672	87.969	91.601	Normal-Approximation
Weibull	91.834	97.392	88.085	97.164	Bootstrap-t
Weibull	91.836	97.711	78.134	92.051	Bootstrap-Percentile
S.E.V.	89.821	96.689	81.305	86.346	Normal-Approximation
S.E.V.	89.878	96.358	80.456	94.572	Bootstrap-t
S.E.V.	89.808	96.347	64.647	87.956	Bootstrap-Percentile
Normal	94.986	99.539	95.402	99.147	Normal-Approximation
Normal	94.363	99.672	94.552	99.607	Bootstrap-t
Normal	94.175	99.771	94.741	99.739	Bootstrap-Percentile

The main idea of the bootstrap method is to simulate the repeated sampling process, reduce the sampling variations in the data, and compute intervals from the simulated distribution of needed statistics without having to making assumptions about the appropriate sampling distribution. The following are three standard steps: 1.) generate a resampled data set, called bootstrap sample, repeatedly for a large number of times, 2.) compute the desired statistic for each bootstrap sample, and 3.) extract information from the distribution of the statistics obtained in 2.), which is the simulated sampling distribution of the population statistic.

For step 1.), the resampling method can be either parametric or nonparametric. See Section 9.2.2 of Meeker and Escobar (1998). We choose the nonparametric bootstrap sampling scheme for all of our bootstrap samples. There are B = 2000 bootstrap samples, each consisting of 396 failures resampled with replacement from the actual data cases, bound with their respective original censoring information. For step 2.), the statistic (first percentile here) for each bootstrap sample is computed both parametrically and nonparametrically, specified by the first column of Table 4 as "nonparametric", "Weibull", etc.

To avoid confusion of terminology in step 1.), we stress again that all the resampling schemes in this paper are assumed to be nonparametric. The term "nonparametric" (Table 4) refers to the "totally nonparametric bootstrap method" (compare Martinez and Martinez 2002 and their notation which we use). Not only is the resampling scheme nonparametric in the "totally nonparametric method," but the population parameter θ (here the first percentile) is calculated nonparametrically as $\hat{\theta}$; the same nonparametric computation of estimate of θ repeats to each bootstrap sample, producing the empirical bootstrap distribution of $\hat{\theta}^*$, where $\hat{\theta}^{*b}$ is the *b*th bootstrap estimate. All the other confidence intervals in Table 4, which are not labeled under the "nonparametric model assumption", are obtained in the parametric way: a ML estimated model is used to generalize the sample data and statistical inference is drawn from the model parameters. For other different general details on asymptotic normality of percentiles, see Serfling (1980).

Under each model assumption, there are different confidence interval methods, noted by "interval method" as the last column of Table 4, to construct a confidence interval for the desired statistic, namely the first percentile. "Normal-approximation" refers to the pointwise normal-approximation confidence intervals under the nonparametric model assumption (Section 3.4.2, Meeker and Escobar, 1998), or to a log-percentile normal-approximation confidence interval under respective parametric model assumptions (Section 7.3.3, Meeker and Escobar, 1998). When using bootstrap method, one can select either "bootstrap-t" or "bootstrap-percentile" method to compute the confidence intervals from the simulated sampling distribution of bootstrap step 3.). If appropriately used, the bootstrap-t confidence intervals can be expected to usually be more accurate than the normal-approximation ones. The mathematical descriptions of these confidence intervals can be found, for example, in Section 3.6, 7.3.3, and 9.3, respectively, of Meeker and Escobar (1998), or compare Edwards, Guess, Young (2004). Splida has provided GUI modules to compute all but the nonparametric bootstrap confidence intervals for the first percentile. A MATLAB code was written as part of this thesis to compute the bootstrap-t and bootstrap-percentile confidence intervals under the nonparametric model assumption.

There is no significant difference in the nonparametric confidence intervals of first percentile between the complete and median censored data, or bootstrap and non-bootstrap method. The nonparametric method only makes use of the data points local to the first percentile. These nonparametric confidence intervals are much wider, however, than the ones obtained under parametric model assumptions. Although these nonparametric intervals can serve as fairly broad, robust comparisons for intervals obtained by other methods, they do not allow for practical precision of more importance in the real world.

Because the parametric model is built to best generalize a whole bulk of data and extract information in terms of a few parameters, the computation of the normal-approximation confidence interval under a parametric model may come quick and be conditionally useful only at the cost of a local approximation, especially at the extremes. Such an approach may be correct when the model fit is good globally over the data range; however, when the globally good fit disagrees with the local data, the estimates become very unreliable. In the case of Type 1 product, the complete data set includes outliers and multiple failure modes. The normal-approximation confidence intervals from the Weibull and S.E.V. ML fits tend to severely underestimate the lower tail, compared to the generally more accurate bootstrap estimates (Meeker and Escobar, 1998). The gap between the bootstrap and normal-approximation confidence intervals ranges from a few to more than twenty pounds per square inch.

Not surprisingly, due to the speculations about overall physical breakdowns made in Chapter III, the normal ML fit may seem to produce close confidence intervals between the bootstrap and non-bootstrap results. This may occur by consistently ignoring the smallest extreme values and fitting the majority. The consequence, therefore, is that the ML fit tends towards overestimating the lowest percentile.

The bootstrap estimates are, to a certain extent, resistant to the influence of outliers, but not unconditionally. Even though empirically better than the approximate method, the bootstrap confidence intervals computed from the complete data might be as misleading in the complete data case. During step 3 of the bootstrap procedures, a histogram of the statistics from bootstrap samples can be drawn out as a simulation of the true sample distribution of the statistic. Such bootstrap histograms can warn us of potentially false structure in the complete data or reassure us in the censoring case of their likely usefulness. Figure 7 from the complete data shows much more variations in the first percentile nonparametric estimates of bootstrap samples, compared to the other percentiles, which corresponds to Figure 2 normal plot and causes the estimates of lowest percentiles to be difficult as discussed previously. Figure 8, also generated from the complete data, further shows a strong sign of ambiguity lying in the estimation of first percentile from Weibull ML fit of bootstrap samples. There are apparently two peaks in the histogram-simulated distribution of bootstrap first percentile estimates, caused potentially by different failure modes, or even possibly two different-shaped Weibull's over different failure range, that are previously speculated in this paper. Outliers in the data could be another reason that affected the bootstrapping histograms. The bootstrap estimates reaffirm that a simple complete data ML fit is insufficient to capture the failure

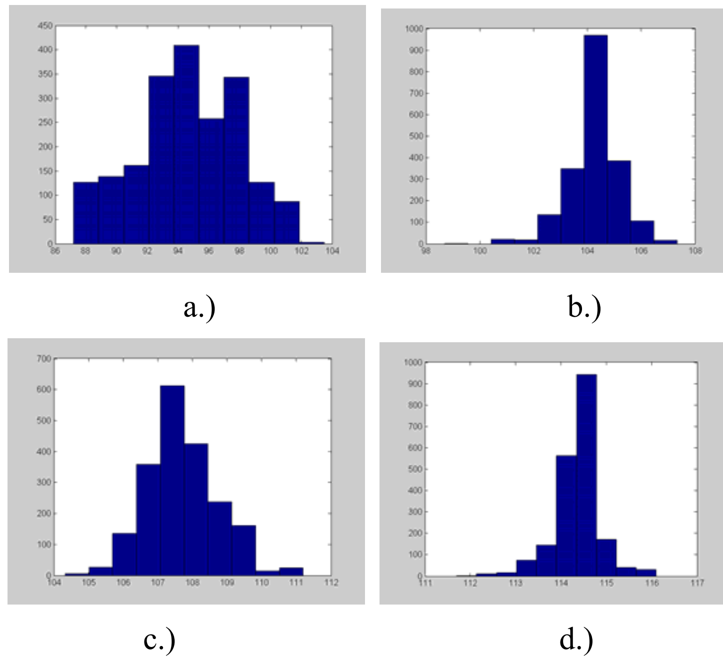


Figure 7. The Histograms of a.) 1st, b.) 5th, c.) 10th, and d.) 25th Percentile Nonparametric Estimates from Bootstrap Samples of the Complete Data for Type 1 Product

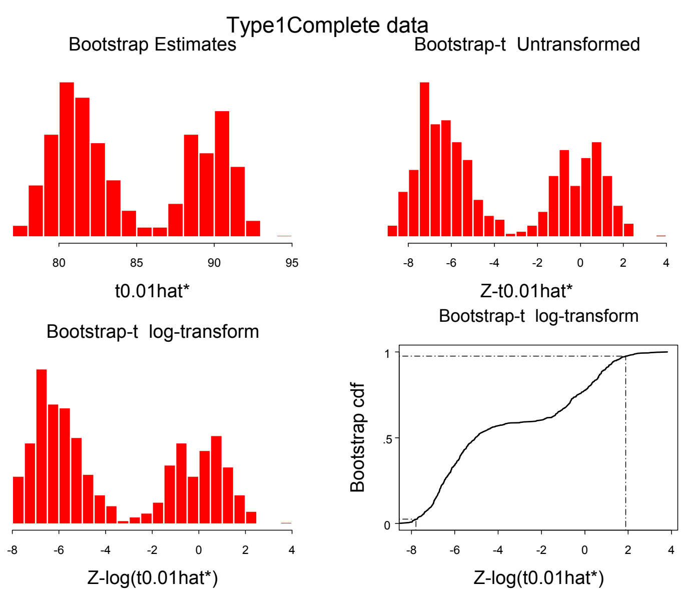


Figure 8. The Histograms of First Percentile Weibull ML Estimates from Bootstrap Samples of the Complete Data for Type 1 Product

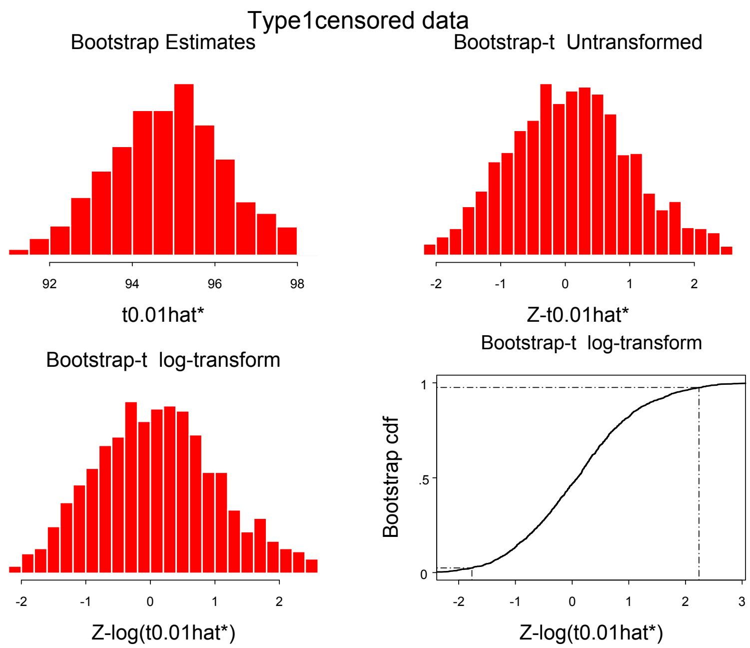


Figure 9. The Histogram of First Percentile Weibull ML Estimates from Bootstraps Samples of the Median Censored Data for Type 1 Product

mode of Type 1 product and produce reliable estimates of the lowest percentiles.

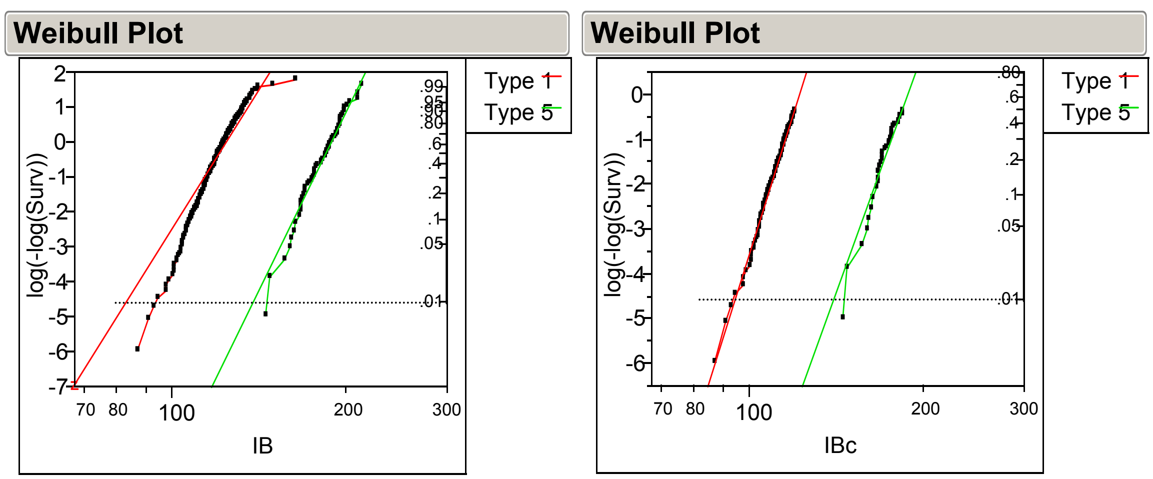
The bootstrap method supports the methodology of the median censoring technique, i.e., the data is reengineered by different weights so that a simple model can fit the observed data very well. Moreover, the desired information of the lower percentiles is protected from the influence of overall failure complexity as well as upper outliers in the complete data.

As a comparison, the histograms of bootstrap estimate on the median censored data in Figure 9 show no such bimodal patterns. Also, note carefully the scale is different in Figure 9 for the normal to not be as spread out as the other previous Figures. If we look at the computed confidence intervals from the median censored data in Table 4, all three types of estimation methods, normal-approximation, bootstrap-t, and bootstrap-percentile, produce very close results under the a simple model assumption. We slightly favor the Weibull model because the S.E.V. has the tendency of underestimating the data, and because the Weibull model is further supported by the information model selection criterion. On different occasions the choice between the Weibull and S.E.V. fit may depend on whether a more accurate or conservative estimate is preferred.

CHAPTER VI. PARAMETRIC MODELS FOR TYPE 1 AND TYPE 5 PRODUCTS

A sound statistical analysis always starts with graphical explorations of the data. We now take advantage of the powerful graphics provided by the JMP® Survival/Reliability platform and present comparisons of both Type 1 and Type 5 products side by side in the next several figures:

A few important observations can be made from the graphical analysis presented in Figure 10. First, the forced censoring technique provides a closer fit to the focus portion of data for Type 1. Second, there is departure on the lower tail of Type 5 which



a.) Weibull fit to uncensored data of both product types

b.) Weibull fit to the median censored data of both product types

Figure 10. Comparisons of Type 1 and Type 5 on the Weibull Probability Plots in JMP®

the median censoring does not improve upon as much.

Note, the manufacturing process for these MDF products is not a batch process and is continuous flow i.e., there is a gradual transition from Type 1 to Type 5. The variations observed in the upper percentiles of Type 1 and lower percentiles of Type 5 may likely be the result of this gradual production transition phase. We have successfully applied the median censoring technique to Type 1 to reduce the upper-tail influence in the case of modeling lower percentiles. In Type 5 an undesired outcome is that the lower-tail variance will affect our estimation on the small percentiles.

A practical strategy is to have a relatively conservative estimate of the percentile. Figure 11 illustrates the 95% simultaneous confidence interval for both types of products, generated in JMP® 6.0 beta test version. In both Figures 10 and 11, the median censoring technique helpfully improves the fit of the lower tail of Type 1 product to the Weibull model. For Type 5 product, even though Figure 10 does not show much difference between the fits of uncensored data and censored data of Type 5 product, it can be seen in Figure 11 that the Weibull fit to median-censored Type 5 product data renders a relatively wider confidence interval that realistically accommodates the rather large variations on the lower tail that is inherent there with the smaller sample. In fact, had the Type 5 data not been censored, one crucial lower percentile data point would be beyond the Weibull 95% confidence bands of the uncensored Type 5 product data (not shown in Figure 11). Based on the above reasoning, we prefer the more conservative results from the censored

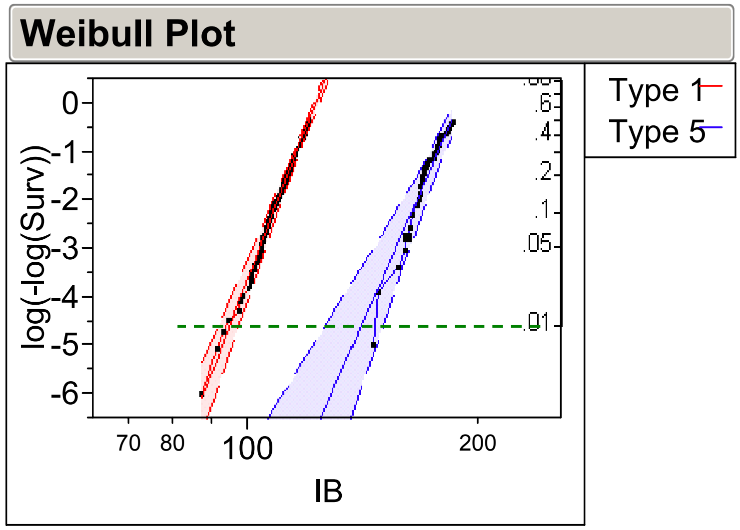


Figure 11. 95% Simultaneous Confidence Intervals of Median-censored Type 1 and 5 Products on the Weibull Probability Plot

Type 5 data.

Recall that the censoring technique does not truncate the data; rather the data portion of interest is given more weight for modeling. In the case of Type 5, the result is that more leeway is given to the lower percentile estimate given the relatively large local variations. (Aside: The plot option of fitted confidence interval is a newly-added feature of JMP® Survival/Reliability Platform in the beta 6.0 version we are reviewing.)

Table 5 shows the estimates of Weibull model fit to the both product types before median censoring. Note that the 95% confidence intervals for the shape parameter β of each product type do not even overlap. However, a refit of the Weibull model to the median censored Types 1 and 5 failure data produced similar range of confidence interval estimates of the shape parameter β (Table 6).

Table 5. Weibull Parameter Estimates Based on Uncensored Types 1 and 5 Product Data

Product type	Parameter	Estimate	Lower 95%	Upper 95%	N Tests
Type 1	α	124.76	123.61	125.90	396
Type 1	β	11.38	10.66	12.10	396
Type 5	α	190.03	186.83	193.18	74
Type 5	β	14.60	12.18	17.22	74

Table 6. Weibull Parameter Estimates Based on Median Censored Types 1 and 5 Product Data

Product type	Parameter	Estimate	Lower 95%	Upper 95%	N Tests
Type 1	α	122.71	121.70	123.87	198
Type 1	β	17.79	15.59	20.18	198
Type 5	α	189.51	185.73	194.88	37
Type 5	β	15.55	11.38	20.57	37

The results from individual model fits are not sufficient to conclude that the two product types have the same shape parameters, or the same type of failure modes. In the next chapter, we conduct a rigorous statistical hypothesis test to determine whether the two products had a common shape parameter (similar to the strategy for analyzing accelerated life test data). We will consider density for each product type as the accelerated variable.

CHAPTER VII. COMMON SHAPE WEIBULL MODEL FOR TYPE 1 AND TYPE 5 PRODUCTS

We fit both type 1 and 5 product data to a common shape model, as shown in Figure 12, considering the similarity between the values of shape parameters from separate Weibull model fits. For comparisons, individual model fits of both product types are shown in Figure 13. The increased overall sample size leads to confidence interval bandwidth which is indeed narrower (compare Figures 12 and 13). It is apparent the common shape Weibull model provides a very good coverage of data for both product types.

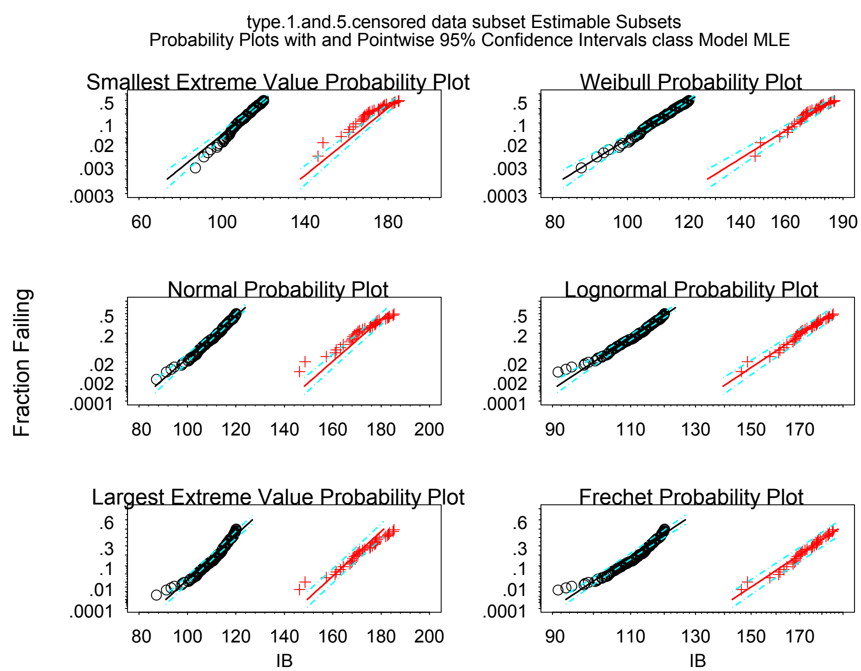


Figure 12. Fitting Both Type 1 and 5 Product Data to Common Shape Location-Scale Models

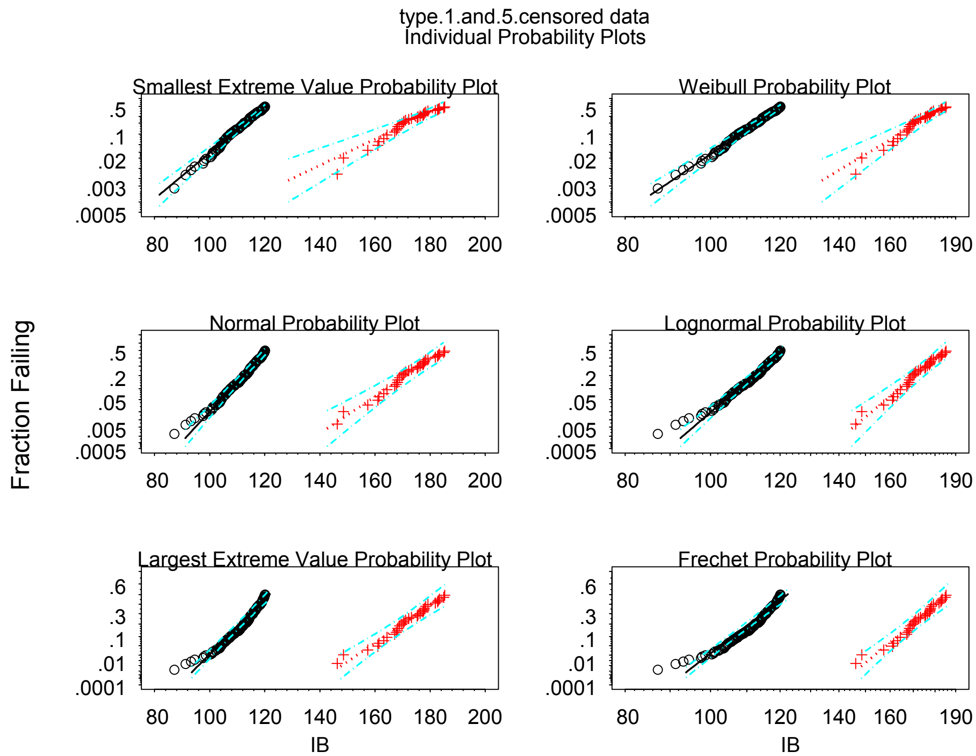


Figure 13. Fitting Both Type 1 and 5 Product Data to Individual Location-Scale Models

The graphical exploration (Figure 12) and parameter β confidence interval estimates (Table 6) both suggest a common shape, or similarity of failure modes between product types. Recall the Weibull model and linearized Weibull CDF as seen on the Weibull probability plot. The Weibull CDF can be often written as

$$P(T \le t;\alpha;\beta) = 1 - \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right], t > 0.$$

$\beta > 0$ is the shape parameter and $\alpha > 0$ is the scale parameter as well as 0.632 quantile (Weibull 1939, 1951). Meeker and Escobar (1998) have pointed out that the practical value of the Weibull distribution is to describe failure distributions with many different commonly occurring shapes. To better display or compare parametric models such as Weibull, we linearize a model CDF on the probability plot. In the Weibull case, one can derive the *p* quantile from the above CDF function: $t_p = \alpha\left[-\log\left(1-p\right)\right]^{1/\beta}$. This leads to

$$\log\left(t_p\right) = \log\alpha + \log\left[-\log\left(1-p\right)\right]\frac{1}{\beta}$$

If we use special scales to t_p and *p* on the probability, which is to take $\log\left(t_p\right)$ and $\log\left[-\log\left(1-p\right)\right]$ on the x and y axis, there is a linear relationship between $\log\left[-\log\left(1-p\right)\right]$ and $\log\left(t_p\right)$ provided a perfect Weibull distribution where the shape parameter β is the slope of the straight line. This justification underlies all the Weibull probability plots shown so far. Hence, if two models appear to have similar slopes on the Weibull probability plots, we may hypothesize that the two models have the same shape parameter, which is also an indicator of failure mode. In our case study, the Weibull probability plots for Types 1 and 5 failure modes have similar slopes (Figures 14).

We assume a constant-shape parameter assumption that is an overall constrained Weibull model {α_1, α_2, and common shape β} for the replacement of two individual unconstrained Weibull models {W_1: α_1, β_1} and {W_2: α_2, β_2}. The total likelihood for the unconstrained models is always larger than the likelihood of the constrained model. A likelihood ratio test is conducted to determine if the total likelihood for the unconstrained

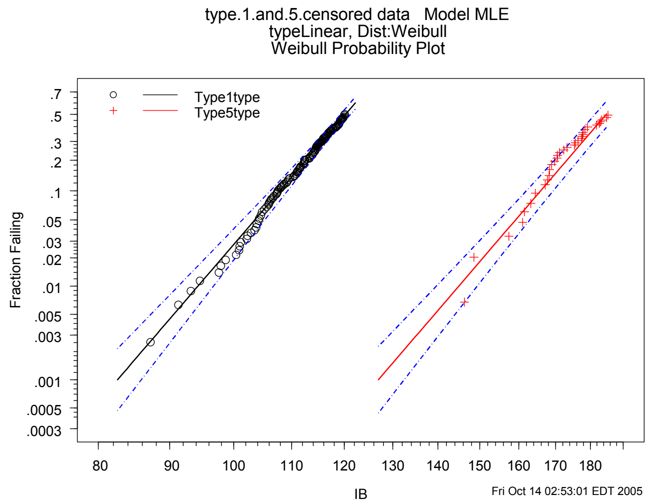


Figure 14. Median Censored Type 1 and 5 Product Data Fitted by Common Shape Weibull Models Plotted on Weibull Probability Plot

models is large enough to indicate lack of fit for the constrained model. The null and alternate hypotheses for the likelihood ratio test are:

H_0: The shape parameters are the same.

H_1: The shape parameters are different; the unconstrained models are better.

The test statistic $Q = -2\left(L_{constrained} - L_{unconstrained}\right) = -2\left[L_{constrained} - \left(L_{W_1} + L_{W_2}\right)\right]$, L denoting the log likelihood of each model, follows a χ_1^2 distribution, in which the one degree of freedom comes from the difference between the number of parameters in constrained and unconstrained models.

Table 7. Demonstration of Likelihood Ratio Test Based on JMP® "Fit Parametric Survival" Output

Model	Log likelihood	ChiSquare	d.f.	Prob>ChiSq
W_1 (Type 1)	-66.3138			
W_2 (Type 5)	-7.4132			
Unconstrained (W_1+ W_2)	-73.7271		4	
Constrained (common shape β)	-73.3324		3	
Test Statistic Q		0.7894	1	0.3744

An individual survival model is built using JMP® "Fit Parametric Survival" platform from its "Survival and Reliability" submenu. The accelerated Weibull model, we include the accelerating variable (density) as the regressor or "model effect" with the model specified. Table 7 illustrates the log likelihood values from three models: $L_{W_1}, L_{W_2}, L_{constrained}$ and the chi-square test results.

The estimated value of Q = -2 × [-73.3324 - (-66.3138 - 7.4132)] = 0.7894 is less than the critical $\chi^2_{0.95,1}$ = 3.84 (p-value = 0.3744), indicating no evidence of inadequacy of the constrained model. Based on this test, there is insufficient evidence to reject the null hypothesis that shape parameters for Types 1 and 5 were the same. (Aside: Another way to check the model adequacy is to use Akaike's Information Criterion (AIC): $-2L+2k$, k being the number of parameters in the model Akaike (1973).) The conclusion was the same using the AIC test, i.e., the common shape model is adequate for modeling both Type 1 and Type 5 products.

Constructing data tables, fitting separate models, and extracting log likelihood results from different reports for statistical testing can be very tedious and may be subject to human error even when an easy-to-use interactive interface such as JMP® is used. We develop a JMP® script to automate the data preparation and model computing process to complement graphical exploration and model building (see Figure 15 which is a JMP® output of a customized report of likelihood ratio test for common shape model using our customized JSL). See Young and Guess (2002) for more on process automation and storage of data in a relational database. Also, see English (1999) on designing a high information quality model for less information scrap and rework.

We are more interested in this investigation of the practical implications suggested by the common shape model associated with the Weibull distribution. For this

Common Shape Model

Model	-LogLikelihood	ChiSquare	DF	Prob
Type 1	-66.314	.	.	.
Type 5	-7.4133	.	.	.
Unconstrained Model	-73.727	.	4	.
Common Shape Model	-73.332	.	3	.
Likelihood Ratio Test Q	0.3947	0.78939	1	0.37428

Figure 15. Customized JMP® Report of Likelihood Ratio Test for Common Shape Weibull Model

data, Type 5 was a product of high value to the producer and consumer but is not sampled at the same level of intensity as Type 1, another important product. To understand the confidence in the estimates for Type 5 key parameters for the common shape Weibull model we investigate several methods to ensure product reliability. Given that dissimilar sample sizes of Types 1 and 5, we use the abundant information from Type 1 to assist in the model building and prediction of Type 5. A comparison of the various percentile estimates for each product is presented in Table 8.

Table 8. 95% Confidence Intervals of First Percentiles Computed Under Various Model Assumptions With and Without Median Censoring

a.) Type 1 product

Model Assumption	With median censoring		W/O median censoring		Interval Method
	95%_Lower	95%_Upper	95%_Lower	95%_Upper	
Weibull	91.834	97.392	88.085	97.164	Bootstrap-t
Weibull	91.206	98.424	81.276	85.312	JMP® Individual Model
Weibull	90.886	97.656	82.359	86.061	JMP® Common Shape Model

b.) Type 5 product

Model Assumption	With median censoring		W/O median censoring		Interval Method
	95%_Lower	95%_Upper	95%_Lower	95%_Upper	
Weibull	139.6	154.46	130.38	148.54	Bootstrap-t
Weibull	127.31	155.60	131.71	146.00	JMP® Individual Model
Weibull	139.36	150.79	123.60	131.07	JMP® Common Shape Model

Table 8 shows consistent confidence interval estimates from the bootstrapped and JMP® common shape models, for both product types after being median censored. One exception is the estimate for product Type 5 from the JMP® individual model. Recall the relatively large variations on the lower tail of Type 5 in Figures 2 and 3 due to production transition phase. The JMP® common shape model performs as well as the bootstrap method, even though the methodologies are different. However, because of the relatively large variation right at the percentile point of interest, more evidence and cross-validation results are needed to enhance our confidence in recommending one of these estimates.

Finally, Figure 16 strongly suggests the adequacy of common shape Weibull model assumptions as the residuals show a linear pattern on the Weibull probability plot.

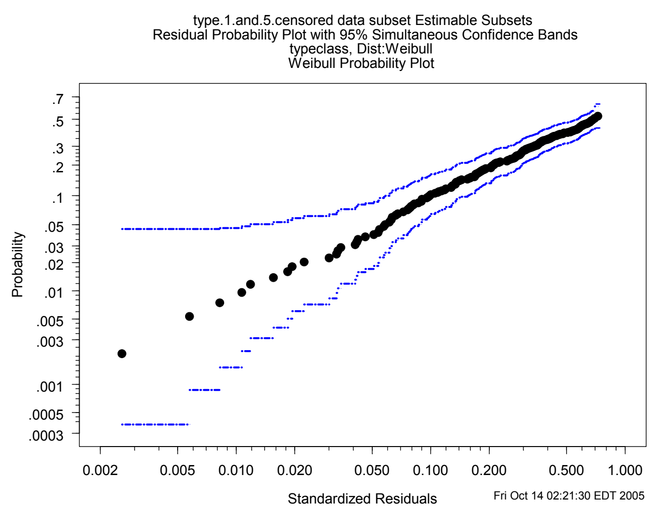


Figure 16. Residual Plot of the Common Shape Weibull Model

CHAPTER VIII. BAYESIAN METHODS FOR PERCENTILE ESTIMATION

There is evidence as illustrated in the last chapter that Type 5 product shares the same shape parameter for the Weibull distribution model, with Type 1 product if we performed median censoring. This is probably the result of a common failure mode. Therefore, even though the sample size of Type 5 product is relatively small, we are able to make reliable predictions of lower percentiles given the abundance of information for Type 1 product.

Yet, it may not always be the case with other product types. There may be a situation when we do not have data on another comparable product such as Type 1 to Type 5 product. Furthermore, there may be only limited amount of data available, such as Type 5 product data; however, we may feel that we know something about the likely range of values of the shape parameter. Bayesian methods come to mind in this situation as a promising approach.

The basic idea of Bayesian methods comes from Beyes' Theorem (Papoulis 1984):

$$P(A_i \mid A) = \frac{P(A_i \cap A)}{P(A)} = \frac{P(A_i)P(A \mid A_i)}{\sum_{j=1}^{N} P(A_j)P(A \mid A_j)}\text{, S}=\bigcup_{i=1}^{N} A_i \text{ and } A_i \cap A_j = \varnothing \text{ for } i \neq j$$

The philosophy is that there exists a set of mutually exclusive and exhaustive states (A_i);

one and only one of these states actually happens at a time. The uncertainty revolves exactly which one of the A_i's the outcome (A) would result from. If one obtains some additional information on the occurrence of A, the new information will more than likely improve assessing the probability of one of the states (A_i). We call the probabilities of these states, $P(A_i)$, priors, and updated probability $P(A_i|A)$ posterior.

The Bayesian probabilities can be described in probability density functions similarly, as following:

$$f(\theta \mid DATA) = \frac{L(DATA \mid \theta) f(\theta)}{\int L(DATA \mid \theta) f(\theta) d\theta} = \frac{R(\theta) f(\theta)}{\int R(\theta) f(\theta) d\theta}$$

$f(\theta)$ is the prior subjective probability of parameter(s) θ; $L(\theta)$ is the likelihood for the available data and specified model; $R(\theta) = L(\theta) / L(\hat{\theta})$ is the relative likelihood. $f(\theta \mid DATA)$ is the posterior probability density of θ given the update of newly available data. Meeker and Escobar (1998) have described simulation-based numerical methods to evaluate the posterior probability.

The prior information can be expert opinion or a noninformative (diffuse) prior distribution. Meeker and Escobar (1998) suggest eliciting the prior information for a straightforward parameter, such as the first percentile and the shape parameter of a Weibull model, preferably with physical or practical meaning for which the prior can be asserted independently. Also, because it is difficult to construct a meaningful joint prior distribution, marginal distributions for individual parameters are sufficient and one should avoid potential dependences among parameters. For example, the shape parameter β and location parameter η in the Weibull model would not be a good choice of prior pair because the two parameters are often dependent. Instead, a quantile on either tail of the distribution and the shape parameter would be approximately independent; and it would be meaningful to survey the field experts on the likely value of these parameters. Splida has a built-in module of single distribution Bayesian analysis.

There are a total of 74 samples of Type 5 product collected. The median censoring technique has proved useful on a relatively large dataset, e.g., Type 1 product. This censoring technique is also helpful when building the Weibull model with Type 5 product. Half of the Type 5 product data is censored so that the common shape Weibull is marginally more robust. But there is not much difference about the Weibull parameters between uncensored and censored data. In the Bayesian analysis, we can entertain both and compare the results later.

First, we specify the prior distribution of shape parameter β as lognormal between 5 and 25, reasonably based on Table 6. Also, the first percentile $t_{0.01}$ is estimated to fall between 100 and 160, according to elicited information. It is then decided to describe the uncertainty in $\log(t_{0.01})$ as a conservative and relatively wide range: UNIF[log(100), log(160)], in which we do not give any particular preference for the point of interest. Such a noninformative prior distribution has worked well on the Weibull 0.01 quartile (Meeker and Escobar 1998).

Splida uses the inverse cdf method to simulate the prior distribution. Figure 17 shows the simulated points from the joint prior distribution for $t_{0.01}$ and β. Figure 18 shows the simulated prior, transformed from the points in Figure 17, plus the histograms of sample shape and location parameters denoted as β and η. Splida uses an algorithm to retain a random sample of prior points and computes relative likelihood $R(\theta_i)$ on these selected sample points.

Figure 19 shows the same prior points given in Figure 17 with the relative likelihood contour superimposed. These contours filter out the prior points with very low relative likelihood (with probability equal to the relative likelihood at that point). The remaining prior points within the contours, shown in Figure 20, are computed to provide the Monte Carlo approximation to the posterior function of ($t_{0.01}$, β).

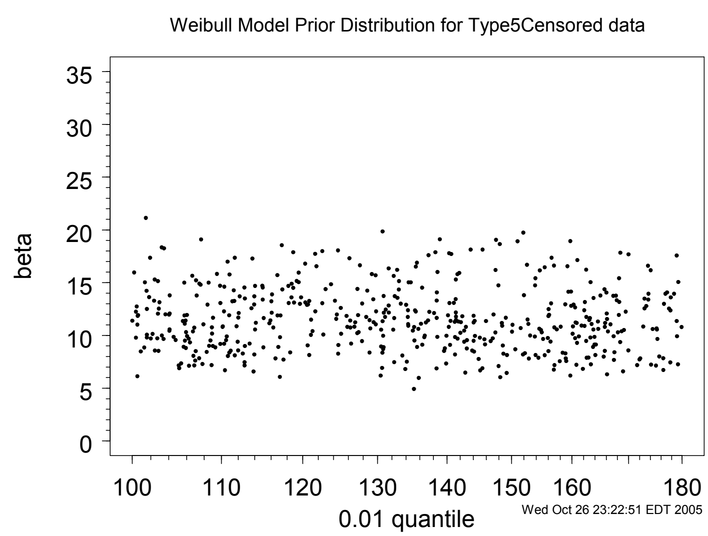


Figure 17. Simulated Points from the Joint Prior for $t_{0.01}$ and β

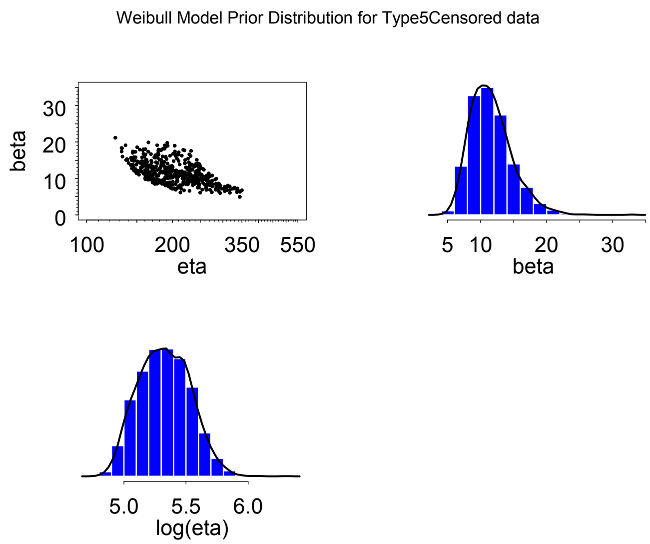


Figure 18. Simulated Points from the Joint and Corresponding Marginal Prior Distributions for η and β

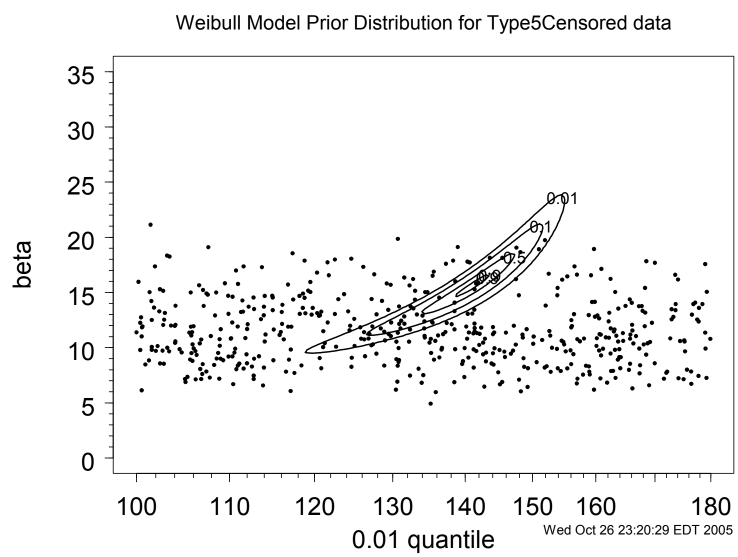


Figure 19. Simulated Points from the Join Prior Distribution with Weibull Relative Likelihood Contour Superimposed

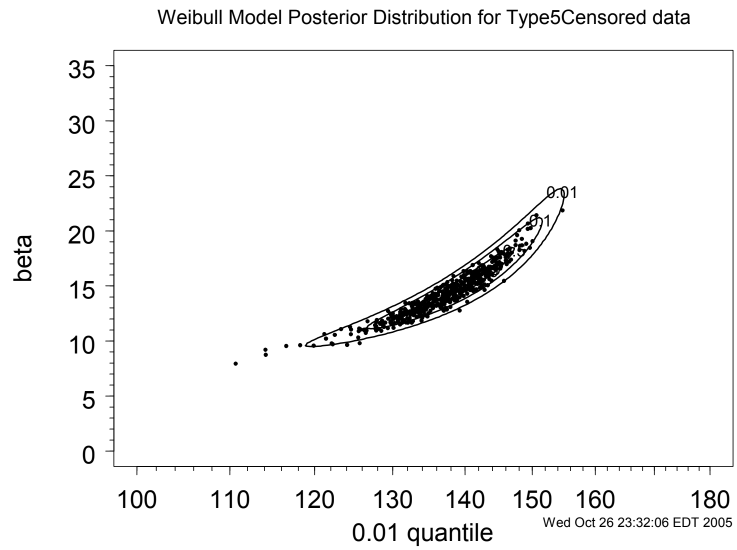


Figure 20. Simulated Points from the Joint Posterior for $t_{0.01}$, β

Figure 21 and 22 show the marginal posterior distributions of parameters of our interests, β and $t_{0.01}$, respectively. Note that the vertical dashed lines on both figures indicate the Bayesian 95% prediction intervals. In this manner, we obtain Bayesian credibility intervals for the shape parameter and first percentile of both censored and uncensored data, shown in Table 9. Traditional confidence and bootstrapped intervals for the same parameters are also provided in the table. However, the fundamental difference between these approaches is that prediction interval speaks about the uncertainty of the parameter, while the confidence interval is interpreted as the resampling coverage probabilities assuming fixed parameters. While these two methods should produce close results in large samples, it appears that the Bayesian prediction interval bounds tend to be

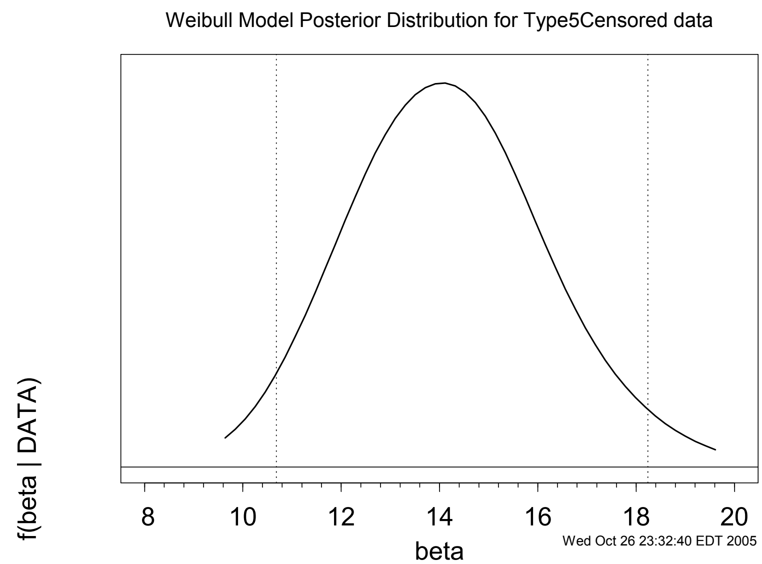


Figure 21. Marginal Posterior Distribution for Shape Parameter β

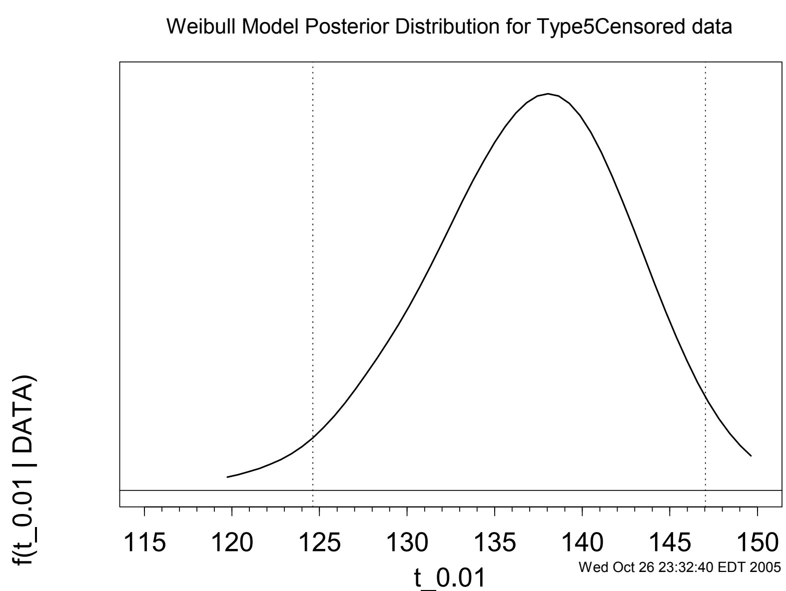


Figure 22. Marginal Posterior Distribution for the First Percentile $t_{0.01}$

Table 9. 95% Prediction and Confidence Intervals for β and $t_{0.01}$ of Type 5 Product

Parameter	With median censoring		W/O median censoring		Interval Method
	95%_Lower	95%_Upper	95%_Lower	95%_Upper	
β	10.68	18.24	11.8	16.47	Bayesian NormBeta
β	11.65	21.01	12.28	17.34	Individual Model
β	11.45	20.69	12.18	17.22	Relative Likelihood
β	11.97	19.35	11.91	16.80	Bootstrap-t
β	11.08	20.18	12.06	17.06	Bayesian UniformBeta
β	11.1	19.37	12.05	16.83	Bayesian NormQuantile
β	11	20	12	17	Bayesian widerQprior

Parameter	With median censoring		W/O median censoring		Interval Method
	95%_Lower	95%_Upper	95%_Lower	95%_Upper	
$t_{0.01}$	124.6	147	127.6	144.7	Bayesian NormBeta
$t_{0.01}$	130.4	152.8	130.37	147.5	Individual Model
$t_{0.01}$	128.1	150.8	129.3	146.5	Relative Likelihood
$t_{0.01}$	130.44	148.92	128.52	145.29	Bootstrap-t
$t_{0.01}$	126.5	150.3	128.9	146.2	Bayesian UniformBeta
$t_{0.01}$	127.1	149	128.9	145.8	Bayesian NormQuantile
$t_{0.01}$	126.6	150	128.6	146.1	Bayesian widerQprior

slightly smaller than the confidence interval bounds for Type 5 product with a relatively small sample size.

Table 9 also includes the 95% Bayesian prediction intervals computed from other priors. "UniformBeta" indicates a loguniorm prior distribution for the shape parameter, while the prior for first percentile and the range of both priors remain the same. Similarly, "NormQuantile" refers to changing the shape of first percentile prior distribution to lognormal, while keeping everything else constant. "WiderQprior" only widens the range of loguniform $t_{0.01}$ to [50, 250]. It appears that the diffuseness of prior does not affect the posterior prediction much, if any. The shape of prior distribution seems to have more impact on the prediction; change of prior distribution from lognormal to loguniform provides larger predictions.

In earlier chapters, small percentile confidence intervals are computed using nonparametric, normal-approximation maximum likelihood (ML) including both individual and common shape model, and bootstrapping methods. More detailed references on these types of confidence intervals can be found in Meeker and Escobar (1998) with bootstrap intervals discussed by Davison and Hinkley (1997), Chernick (1999), Efron and Tibshirani (1993), and Efron (2003). As shown in Tables 4 and 9, the results computed from different methods (even philosophies) are quite consistent and agreeable.

Note that Polansky (1999, 2000) warned of using bootstrap estimates when the percentiles are very small such as 1% or 5% and when the sample size is also small, e.g., less than 100. In our case study, there are an adequate number (396; 198 after median censoring) of samples for Type 1 product, and we observe consistent Bootstrap estimates; however, the sample size (74) for Type 5 product should raise our concerns for the bootstrap method. As a result of small sample size, the median censoring technique does not seem an appropriate and necessary preparation procedure for bootstrapping (Table 9). Instead, we should compare results from all methods from simple parametric model fit to relative likelihood estimates to Bayesian methods, and to bootstrapping without applying median censoring in advance and look for consistency in the estimates.

It is reassuring to have different methods of confidence interval estimation agreeing so closely. Besides the sample size issue, though we generally trust the bootstrap-t estimates more because of its resampling mechanism, the ML fit normal estimates are very close to them given the improvement in data quality by the median censoring technique. From a practitioner's point of view, even if a bootstrap-t macro or a computationally-intensive environment is unavailable, the conventional ML fit approach can still be acceptable as long as the median censoring technique has been applied. Such a conclusion also helps the tasks that demand online feedback or timely solutions.

As computing power has grown exponentially over the past two decades, all the simulation-based and Bayesian methods are more feasible and accessible for personal computing. Statistical knowledge and the power of quantifying future uncertainties are greatly enhanced by this PC accessibility. As shown in the results of low percentile estimates, we gain confidence in believing the predictions, as long as the underlying physics or chemistry mechanism remains stable under operating conditions. And as is learned from the Bayesian philosophy, we are forever going to incorporate new information to our knowledge and make our decisions of best action.

CHAPTER IX. SUMMARY

Be it the observed complexity within the complete data set of the Type 1 MDF product, or the limited observations of Type 5 MDF product, real world data will often present some “non-textbook” difficulties, therefore require careful evaluation and unconventional solutions. The nonparametric methods are easy to implement but may not apply the full benefits of available information, and may be difficult to interpret for the practitioner. Simply fitting a parametric model to primitive data may be problematic given the inadequate weighting of the most crucial information, e.g., lower percentiles. The resulting estimates for questionable assumptions of normality may lead to product failures at the plant and product failures in the field. Product failures detected in the plant lead to rework and higher costs of manufacturing, product failures in the field lead to claims and loss of customer value. Poor product information and knowledge result in poor product reliability, and as a result, poor product quality.

Rather than building a complicated model to match every portion of the observed data, or being misled to unnecessarily collect expensive test data, we introduce a new technique of median censoring which places more weight on the lower tail of the data for critical estimates of the smallest percentile. It is shown both graphically and quantitatively that with high quality data, a simple as well as empirical failure model like Weibull fits the lower tail exceptionally well and produces consistently reliable estimates of the small percentiles. Evidence presents that the forced censoring technique can enhance analyses of non-normal or highly complex data.

Probability plots and ML fits are very supportive of the median censoring technique. What is also crucial is the confirmation provided by bootstrapping. We have shown that not only is the median censoring technique supported, but enhanced by the bootstrap method. The bootstrap simulated sampling distribution reveals different failure modes existing in the complete data set, and that the median censoring technique resolves the bimodality difficulty in the ML fit. The high degree of agreement between the normal-approximation C.I. and the bootstrapped C.I. is strong evidence that the median censoring technique is superior. The exception is when the sample size is relatively small, e.g., less than 100, one must use the bootstrapping method with extreme caution. Other methods like Bayesian approach, restrained models (common shape, etc.) should be explored to leverage various sources of information into predictive modeling.

Graphic exploration and interactive discovery helps identify patterns in the data that may be hidden by descriptive statistics alone. We have further investigated an accelerated Weibull model to help increase the accuracies of extremely small percentile estimates which may be important methods for understanding product reliability and be helpful for improved product quality and lower manufacturing costs, especially when the samples are costly. The easy to use JMP® platform facilitated the implementation of a sound statistical strategy in the context of process improvement in reliability engineering that can be readily adopted by a large number of industrial users.

Finally, we caution practitioners that as straightforward as the practice seems to be by fitting a commonly known or accepted model to the raw lifetime data, it is dangerous and costly to draw any immediate or convenient inference merely from that type of preliminary analysis, which may mislead to over-engineering a product or over-sampling. We suggest that the data structure be examined via various probability plots first. If these plots suggest deviations from the ML fit or possible outliers or curvatures, it is advised to apply the forced median censoring technique to put more weights on the part of data of best interest. Then, refit a parametric model for better estimates of small percentiles.

It is important that different methods, bootstrapping in particular, be used to validate the model and improve the estimates. Under limited situations, the model fitting methods without bootstrapping may perform just fine and render quick and satisfactory results because of the critically improved data quality by the median censoring technique. Overall, our approach to analyzing complex real-world lifetime data is empirically successful, parsimonious, and suitable for real-time manufacturing settings. It does not depend on the underlying distribution being Weibull, lognormal, or otherwise. This approach is also applicable to lifetime or strength failure data for small sample sizes that are common during mill startups and new product development. The methods of this thesis could also be useful in “time to submission for rebates” or “times to return” a product common in marketing analyses.

LIST OF REFERENCES

LIST OF REFERENCES

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In B.N. Petrov and F. Csaki (eds.), *Second International Symposium on Information Theory*, Acaderniai Kiado, Budapest, 267-281.

Akaike, H. (1974). “Factor Analysis and AIC,” *Pschychometrika*, 52, 317–332.

Akaike, H. (1987). “A new Look at the Statistical Identification Model,” *IEEE Transactions on Automatic Control*, 19, 716–723.

Bozdogan, H. (2004). *Lecture notes for Statistics 564*, the University of Tennessee, Knoxville, TN.

Chen, W., León, R., Young, T., and Guess, F., (2005). Applying a Forced Censoring Technique with Accelerated Modeling for Better Estimation of Extremely Small Percentiles of Strengths. *International Journal of Reliability and Application*. In Review.

Chernick, M. R. (1999). *Bootstrap Methods: A Practitioner's Guide*. Wiley, New York, NY.

Condra, L.W. (1993). *Reliability Improvement with Design of Experiments*. Marcel Dekker, Inc. New York.

Crosby, P. (1979). *Quality Is Free*. Penguin Group, New York.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge, U.K. and New York, NY.

Deming, W. E. (1986). *Out of Crisis*. Massachusetts Institute of Technology Center for Advanced Engineering Study, Cambridge.

Edwards, D. J., Guess, F. M., and Young T. M., (2004). Bootstrap Confidence Intervals for Percentiles of Reliability of Modern Engineered Wood, *IIE Transactions*. In Review.

Efron, B. (2003). Second Thoughts on the Bootstrap. *Statistical Science*, **18** (2), 135-140.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York, NY.

English, L.P. (1999). *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*. Wiley, New York, NY.

Guess, F. M., Edwards D. J., Pickrell T. M. and Young T. M. (2003). Exploring Graphically and Statistically the Reliability of Medium Density Fiberboard. *International Journal of Reliability and Application*, **4** (4), 97-109.

Guess, F., Hollander, M., and Proschan, F. (1986). Testing Exponentiality Versus a Trend Change in Mean Residual Life. *Annals of Statistics*, **14** (4), 1388-1398.

Guess F. M., Leon R.V., Chen W., and Young T. M. (2004). Forcing a closer fit in the lower tails of a distribution for better estimating extremely small percentiles of strengths. *International Journal of Reliability and Application*, **5** (4), 129-145.

Guess, F., and Proschan, F. (1988). Mean Residual Life: Theory and Applications. *Handbook of Statistics: Quality Control and Reliability*, **7**, 215-224.

Guess, F., Walker, E., and Gallant, D. (1992). Burn-in to Improve which Measure of Reliability? *Microelectronics and Reliability,* **32**, 759-762.

Hoerl, R. W. and Snee, R. D. (2002). *Statistical thinking: improving business performance*. Duxbury Press, Pacific Grove, CA.

Huang, K-T., Lee, Y. L. and Wang, R. Y. (1999). *Quality Information and Knowledge*. Prentice Hall, New York, NY.

Imai, M. (1997). *Gemba Kaizen: A Commonsense, Low Cost Approach to Management*. McGraw-Hill, New York, NY.

Imai, M. (1989). *Kaizen: The Key to Japan's Competitive Success*. Random House, New York, NY.

Ishikawa, K., trans. David Lu. (1988). *What is Total Quality Control?---The Japanese Way*. Prentice-Hall, Englewood Cliffs.

Juran, J. M. (1988). *Juan on Planning For Quality*. Macmillan Free Press, New York, NY.

Kim, K. O. and Kuo, W. (2003). Percentile Life and Reliability as Performance Measures in Optimal System Design. *IIE Transactions*, **35**, 1133-1142.

Kuo, W., Chien, W. T. K., and Kim, T. (1998). *Reliability, Yield, and Stress Burn-In*. Kluwer Academic Publishers, Norwell, MA.

Kuo, W., Prasad, V. R., Tillman, F. A., and Hwang, C. L. (2000). *Optimal Reliability Design: Fundamentals and Applications*. Cambridge University Press, Cambridge, UK.

Martinez, W. L. and Martinez, A. R. (2002). *Computational Statistics Handbook with Matlab*. Chapman and Hall/CRC, Boca Raton, LA.

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data.* Wiley, New York, NY.

Mendenhall, W. and Sincich, T. (1995). *Statistics for Engineering and the Sciences*, 4th ed., Prentice Hall, Inc. Upper Saddle River, NJ.

Papoulis, A. (1984). Bayes' Theorem in Statistics" and "Bayes' Theorem in Statistics (Reexamined). §3-5 and 4-4 in *Probability, Random Variables, and Stochastic Processes, 2nd ed.* New York: McGraw-Hill, pp. 38-39, 78-81, and 112-114.

Polansky, A. M. (1999). Upper Bounds on the True Coverage of Bootstrap Percentile Type Confidence Intervals. *The American Statistician*, 53(4), 362-369.

Polansky A. M. (2000). Stabilizing Bootstrap-t Confidence Intervals for Small Samples. *The Canadian Journal of Statistics.* 28(3), 501-526.

Redman, T. C. (2001). *Data Quality: The Field Guide.* Butterworth-Heinemann Digital Press, Boston, MA.

SAS Institute Inc. (2004). *JMP® 5.1 Manual*, SAS Publications. Cary, NC.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York, NY.

Shewhart, W. (1986). *Statistical Methods from the Viewpoint of Quality Control*. Dover Publications, New York, NY.

Snee, R. D. and Hoerl, R. W. (2003). *Leading Six Sigma: a step-by-step guide based on experience with GE and other Six Sigma companies*. FT Prentice Hall, Upper Saddle River, NJ.

Stanard, C. L. and Osborn, B. E. (2002). Six Sigma Quality Beyond the Normal. *GE Global Research Technical Report*, 2002GRC119.

Taguchi, G. (1986). *Introduction to Quality Engineering, Designing Quality into Products and Processes*. Unipub, White Pains, NY.

Walker, E. and Guess, F. (2003). Comparing Reliabilities of the Strength of Two Container Designs: A Case Study. *Journal of Data Science*, **1**, 185-197.

Weibull, W. (1939). A statistical theory of the strength of materials. *Ing. Vetenskap Akad. Handl.*,**151** (1), 1-45.

Weibull, W. (1951). A statistical distribution function of wide applicability, *Journal of Applied Mechanics*, **18** (1), 293-297.

Young, T.M. and Guess, F.M. (2002). Mining Information in Automated Relational Databases for Improving Reliability in Forest Products Manufacturing. *International Journal of Reliability and Applications*, **3** (4), 155-164.

APPENDIX

Appendix

JMP® script of the forced censoring technique:

```
dt = Open();
/* Dialog to choose the censoring quantile */
  sdlg = Dialog(
                    "Enter censor quantile",
                    censorQt = EditNumber(0.50),
                    "Enter product type",
                    pType=EditText("Type 1"),
                    Button("OK"),
                    Button("Cancel")
  );
  If (sdlg["Button"]==-1,
          Throw("!Cancelled")
  );
  show(sdlg["censorQt"]);
  show(sdlg["pType"]);

  censorValue = Col Quantile(:IB, sdlg["censorQt"]);

/* Create new columns for censored data, cersor label, and product type info. */
  dt << New Column ("IBc",
                              Numeric,
                              Continuous
  );
  dt << New Column ("NewCensor",
                              Numeric,
                              Nominal
  );
  dt << New Column ("Type",
                              Char,
                              Nominal
  );

/* Forced censoring from the specified quantile for each row */
  For Each Row (
```

```
	If(
		:IB <= censorValue, :NewCensor=0; :IBc=:IB,
		:IB > censorValue, :NewCensor=1; :IBc=censorValue
	);
	:Type=sdlg["pType"]
 );

/* Create new data table; not overwrite the initial data file */
 dtnew = dt << Subset(
			Output Table(sdlg["pType"]||" censored"),
			Columns(:IB, :Censor, :IBc, :NewCensor, :Type)
 );

 close(dt, no save);
 close(dtnew);
```

VITA

Weiwei Chen was born on October 31st, 1979 in Huangshan, Anhui, People's Republic of China. He moved to Shanghai later and received a Bachelor's degree in Electrical Engineering from Northeastern University (Shenyang, Liaoning) in July 2001. Weiwei completed a Master of Science degree in the Department of Electrical & Computer Engineering at the University of Tennessee, Knoxville, in December 2003. Since then, he has been pursuing a second Master's degree in Statistics at UT. Weiwei has served as a Graduate Teaching Associate, Graduate Research Assistant, UT Thornton Athletics Student Life Center Tutor, and proud member of the UT Cultural Attractions Committee. He is currently a student member of American Statistics Association (ASA) and International Association of Information and Data Quality (IAIDQ).