Brazilian Journal of Probability and Statistics (2001), 15, pp. 169–200.

SOFTWARE FOR RELIABILITY DATA ANALYSIS AND TEST PLANNING

William Q. Meeker[1] and **Luis A. Escobar**[2]

[1] *Department of Statistics, Iowa State University, Ames, IA 50011, USA. Email: wqmeeker@iastate.edu*

[2] *Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA 70803, USA. Email: luis@lsu.edu*

Summary
Increasingly, statisticians and reliability engineers in industry are being asked to analyze reliability data. Because of the complicated nature of the data and models that are often encountered in reliability studies, statistical methods and corresponding software needed for appropriate analyses are not developed as well as methods and software needed for the analysis of standard experimental designs and observational studies. This paper outlines the needs of practitioners and researchers in this area and describes a software tool that is being developed to satisfy the most important needs facing reliability analysts.

Key Words: Censored data; degradation data; experimental design; life data analysis; maximum likelihood; recurrence data; statistical computing.

1 Introduction

1.1 Quality and reliability

Over the past twenty years, manufacturing industries have been faced with the need to improve quality, productivity, and reliability. Much of this need has been driven by the expanding global marketplace and the resulting increased competition. Manufacturers of high quality and high reliability products have a competitive advantage. Traditional methods of process monitoring and experimental design for product and process improvement have, for some companies, proven to be highly successful. Evidence for this includes the phenomenal success of companies such as General Electric in their implementation of Six-Sigma (e.g., see Hahn, Doganaksoy, and Hoerl 2000).

Because reliability can be defined as "quality over time," improving quality has also had the effect of improving product reliability. It has been recognized, however, that achieving and improving high reliability requires tools that lie beyond the standard tools used in quality improvement. The purpose of this paper is to describe some of these tools and to outline and illustrate a (partially complete) software implementation of them.

1.2 Reliability data analysis

Reliability data arise from a number of different sources, including laboratory life tests, field tracking studies, and warranty data bases. Reliability data occur in a number of different forms, often complicated by features like censoring, truncation, and multiple failure modes. Physically based or physically motivated models are generally required when the amount of information in a reliability data set is limited (e.g., because the number of observed failures is small). Also, almost all reliability analysis problems involve various forms of extrapolation, requiring a strong model basis. This results in the need to handle a wide range of standard and nonstandard models (e.g., nonnormal distributions and nonlinear relationships). As such, standard statistical methods and software are often inadequate for the proper analysis of reliability data. Relatedly, standard methods for planning experiments often have to be extended or otherwise modified to account for the complications that arise in reliability data (e.g., censoring and truncation). Also, reliability inferences generally extend beyond the standard moments and regression coefficients upon which other kinds of statistical studies focus. Instead, reliability analysts need estimates of quantities like failure probabilities, distribution quantiles, hazard rates, and so on.

This paper describes the general needs that arise across a wide range of applications in reliability data analysis. It also describes a software tool, SPLIDA, that is being designed and developed to provide a tool for statisticians and reliability engineers who need to properly analyze reliability data. SPLIDA is a collection of S-PLUS functions and a graphical user interface (GUI) for reliability data analysis and test planning. The S-PLUS functions were originally developed to provide a means to apply the methodology presented in Meeker and Escobar (1998) and to allow users of that book to do their own analyses. The most up-to-date version of SPLIDA is always available at

www.public.iastate.edu/~stat533/splida.html.

1.3 General needs in reliability data analysis software

The goal of the SPLIDA development project is to provide a software system to support the work of statisticians and engineers involved in reliability work. SPLIDA is also an extendible tool to support computing needed for applications and research in the area of reliability data analysis and test planning. In general, SPLIDA provides:

- Basic capabilities to fit appropriate models (physical or empirical) to the commonly occurring types of reliability data, as described in Section 2.
- Tools for planning various kinds of reliability studies that are used in applications.
- Procedures that make effective use of the modern computing tools of simulation, graphical display, and visualization.
- An intuitive GUI that will make it easy for statisticians and reliability engineers to efficiently, properly, and effectively use the tools in the system without the need to constantly refer to program documentation.
- A collection of low-level functions that can be used as building blocks for developing new methods.

1.4 Reliability data analysis and test planning literature and available software

There is a large amount of literature and numerous books on the subject of reliability data analysis. Pioneering books include Mann, Schafer, and Singpurwalla (1974), Nelson (1982), Lawless (1982), and Cox and Oakes (1984). Some more recent additions include Crowder, Kimber, Smith and Sweeting (1991), Tobias and Trindade (1995), and Meeker and Escobar (1998). Books dealing with more specialized topics include Nelson (1990) on the important subject of accelerated testing, Rigdon and Basu (2000) on repairable system methods, and Nelson (2002) on statistical methods for recurrence data (methods for recurrence data are important for repairable systems analysis).

The first major software system for reliability data analysis was designed by Wayne Nelson. This package, called STATPAC, was described in Nelson and Hendrickson (1972) and Strauss (1980). STATPAC was far ahead of its time and contained a combination of capabilities for graphical analysis and fitting of general statistical models with censored data that are available today in only the most advanced statistical packages. Subsequently, SAS incorporated a limited number of the models and methods described in Lawless (1982), Nelson (1982), and Nelson (1990) into their general purpose system. More recently, JMP, MINITAB, and S-PLUS have also incorporated a limited number of the most widely used of these methods and models.

In addition, some special-purpose software packages have been developed to do analyses for the most common types of reliability data analysis (fitting single distributions and common accelerated life test models). These include Weibull++ and WinSmith. Existing packages, however, provide little or nothing in the way of capabilities to extend the system to perform operations that were not envisioned by the developers. Because of its S-PLUS base, SPLIDA is extendible.

1.5 Overview

The remainder of this paper is organized as follows. Section 2 describes the different types of reliability data and illustrates the basic estimation methods for each of these types. Section 3 provides an overview of the key ideas behind the SPLIDA project. Section 4 describes and illustrates some of the technical capabilities of SPLIDA for data analysis. Section 5 explains the philosophy for experimental design and reliability test planning used in SPLIDA. Section 6 makes some concluding remarks and describes future work planned for SPLIDA.

2 Types of reliability data and basic analysis tools

The purpose of this section is to provide background that we need to describe reliability data analysis software. The section presents and illustrates examples of the wide range of reliability data that can arise in different applications and describes corresponding basic analysis tools.

2.1 Failure-time data and basic analysis methods

The vast majority of the literature on statistical analysis of reliability data is concerned with failure-time data. The assumption behind the commonly used methods is that failure times, perhaps conditional on observed values of explanatory variables, can be modeled as independent random variables from a continuous distribution. Such data arise frequently in laboratory life tests, accelerated life tests, and in the analysis of field and warranty data. Time is a generic term that may measure hours of service, number of cycles, number of miles, and so on, depending on the application and what can be recorded.

In life testing, it is common to start a sample of units on test and then to terminate the test before all units have failed. Early termination is used to save time or to provide testing resources for other studies. If a test is terminated after a particular amount of time, the resulting data are known as “Type I” censored data and if the test is terminated after a particular number of failures, the data are known as "Type II" censored data. If a test is terminated before all units fail and if the number of survivors is not recorded, the data are said to be "right truncated." Of course, there are many other possible variations in the way that censoring or truncation can occur.

With product field and warranty data, time is generally measured from the time of sale or the beginning of service of a unit. End of life also requires a specific definition. In some applications, the definition is obvious (e.g., when a light bulb burns out), but in other applications, the definition is more arbitrary (e.g., gloss loss of a coating). See, for example, page 25 of Nelson (1990) for discussion of these issues. Censoring patterns in field data are usually more complicated because of entry into service at different points of time (staggered entry) and differing use rates. Generally, some (usually most) of the units that are still in service have not failed at the time of the analysis.

Systems or subsystems for which there is interest in quantifying reliability may be either repairable or not. In either case, the most useful and important information is on the particular failures mode or components that cause a system to fail. Without information at the component/cause level, the data provide little or no information for improving reliability.

Figure 1 is an "event plot" showing field failure data for a bearing cage used in a jet engine. The data are from Abernethy, Breneman, Medlin, and Reinman (1983). Time is measured in hours of service. There were only a few failures (indicated by *), but large numbers of failure-free units with differing amounts of time in service (counts larger than one are shown in the right margin of the figure). The variation in the amounts of time in service among the units was primarily due to differences in dates that the jet engines started service, but also due to different use rates for different engines.

Figure 2 is a lognormal probability plot of the bearing cage data. Probability plots are useful for assessing the adequacy of distributional assumptions and presenting the results of maximum likelihood (ML) analyses. As shown in Section 6.2 of Meeker and Escobar (1998), a lognormal cdf plots as a straight line on lognormal paper. The slope and intercept of the line are related to the lognormal scale and shape parameters, respectively. The straight line in Figure 2 is a ML estimate of the bearing cage lognormal cdf. When the points in the plot (corresponding to a nonparametric estimated of fraction failing versus time) fall approximately along a straight line on a lognormal plot (except in the tails of a distribution where deviations are expected), the data fit the lognormal distribution.

It is possible to construct probability plots for any given distribution (although in some cases the construction of a probability plot may depend on one or more unknown parameters that must be estimated from the data). Chapters 3 and 6 of Meeker and Escobar (1998), for example, describe nonparametric estimation and probability plotting for censored data. They also extend the ideas to truncated data in Chapter 11.

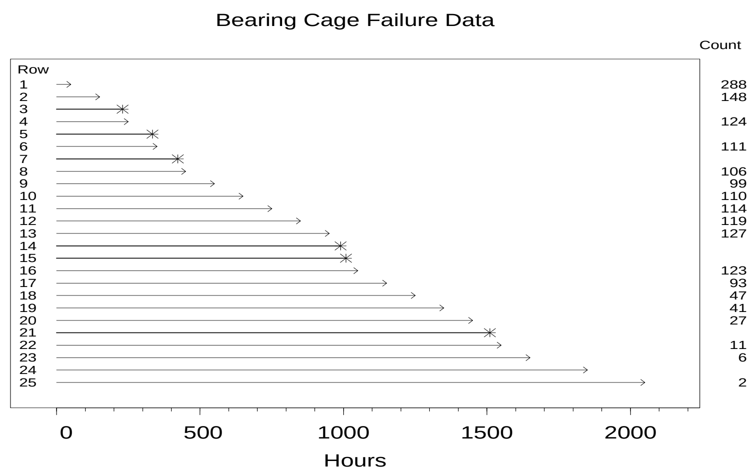

Figure 1
Bearing cage failure-time data.

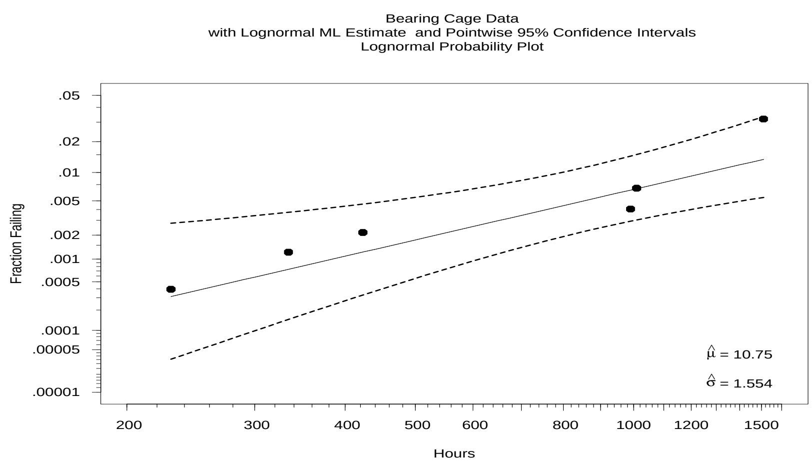

Figure 2
Bearing cage failure-time data lognormal probability plot and ML cdf estimate.

2.2 Degradation data and basic analysis methods

With modern high-reliability components and products that are designed to last for many years, failures might not be expected to be observed in a life test of reasonable length. In such cases, it may be possible to measure degradation as a function of time to get timely useful information about progression toward failure. Such data are called “repeated measures degradation data.” In other cases, only one degradation measurement can be taken on each unit. Such data are known as “destructive degradation data.” Degradation data are found most commonly in laboratory tests.

Consider a life test on 2000 units of an electronic component. If, after 4000 hours of testing there are only 2 failures, the life test provides very little information about reliability, especially for times beyond 4000 hours. If, however, it were possible to look inside of the 1,998 surviving units to see how far they had progressed toward failing, there would be much more information available to assess the reliability of the component.

Figure 3 shows repeated measures degradation data taken over time on a sample of lasers. The lasers were designed with a feedback system to maintain constant light output. Over time, the amount of current required to maintain constant light output will increase. For these lasers, failure was defined as the point in time when the required current has increased by 10 percent.

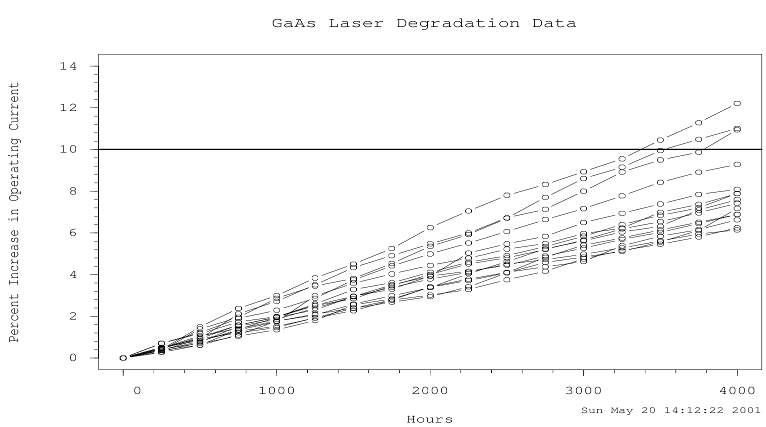

Figure 3
GaAs laser data showing the amount of current needed to maintain constant light output.

The time to first crossing of the failure-definition boundary induces a failure time probability distribution. There are several ways to analyze degradation data and estimate this distribution. One approach is to model the paths with a parametric function and use random parameters to describe unit-to-unit variability. For the lasers in Figure 3, an appropriate model would have a common intercept but a distribution for the slopes. The failure-time distribution could be deduced from the empirical distribution of the slopes. For example, if the slopes of a zero-intercept linear degradation process follow a lognormal distribution, then the life times will also follow a lognormal distribution. This general approach is described in Chapters 13 and 21 of Meeker and Escobar (1998).

Another simpler approximate analysis method extrapolates the paths of units that have not failed to obtain “pseudo failure times.” Then standard life data techniques can be used to analyze the pseudo failure times. This method gives reasonable results if the sample paths are well-behaved with only a small amount of measurement error and stochastic variability. Chapter 13 of Meeker and Escobar (1998) and Meeker, Doganaksoy, and Hahn (2001) illustrate the simple method by analyzing the laser data in Figure 3, under varying assumptions.

2.3 Recurrence data and basic analysis methods

In many applications, a collection of units is monitored over a period of time, and events of interest are recorded for the individual units (e.g., failure, need for adjustment and other maintenance actions, and so on). Such monitoring processes result in “recurrence data.” Recurrence data arise from populations of repairable systems such as fleets of trucks, automobiles, and jet engines. Data on events for such fleets are collected for purposes of monitoring costs, developing strategies for improving system availability, providing information for pricing maintenance contracts, and so on.

Figure 4 is an event plot indicating the times of maintenance events for a fleet of earth-moving machines. The data came from Chapter 16 of Meeker and Escobar (1998). At each recurrence, information is usually recorded on variables like the cause and/or cost of the event. For the earth moving machine, the number of labor hours was recorded for each event. In an automobile warranty data base, the “labor code” indicating the type of repair, and cost are recorded for each repair to an automobile.

For some purposes, the analysis of recurrence data uses models and methods that differ from those used in life data or degradation analysis. Recurrence data are also known as point process data and there are a number of important general references describing point process models and data analysis. These include Snyder (1975), Cox and Lewis (1966), and Thompson (1988). Ascher and Feingold (1984) and Rigdon and Basu (2000) are more specialized references, focusing on parametric models for repairable system reliability. Lawless and Nadeau (1995) and Nelson (1995) present useful nonparametric methods for estimating the mean cumulative function (MCF) in the presence of censoring (the derivative of the MCF can be interpreted as the intensity rate of events). These methods are described and illustrated in Chapter 16 of Meeker and Escobar (1998). Nelson (2002) presents analysis methods and a variety of applications that involve recurrence data.

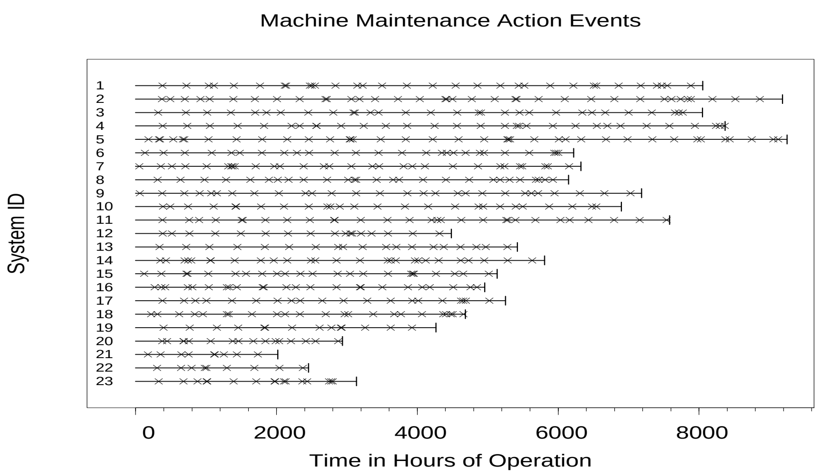

Figure 4

Earth-moving machine maintenance actions event plot.

Figure 5 is the plot of the empirical MCF for the earth-moving machines. The curve in this plot was estimated from the recurrence data using methods in Nelson (2002). The MCF in this application can be interpreted as an estimate of the number of labor-hours (a surrogate for cost), as a function of time, needed to maintain one of these machines. The dashed lines are confidence intervals that reflect the statistical uncertainty of the empirical MCF when used to estimate the the MCF of earth-moving machine maintenance process.

Parametric methods can also be useful in the analysis of recurrence data, particularly when one needs to make predictions beyond the range of the data (e.g., for the earth moving machines it would be necessary to use a model to describe the MCF beyond 9000 hours of operation). Nonhomogeneous Poisson Process (NHPP) models are often useful for this purpose and are described, for example, in Cox and Isham (1980), Ascher and Feingold (1984), Thompson (1988), and Rigdon and Basu (2000).

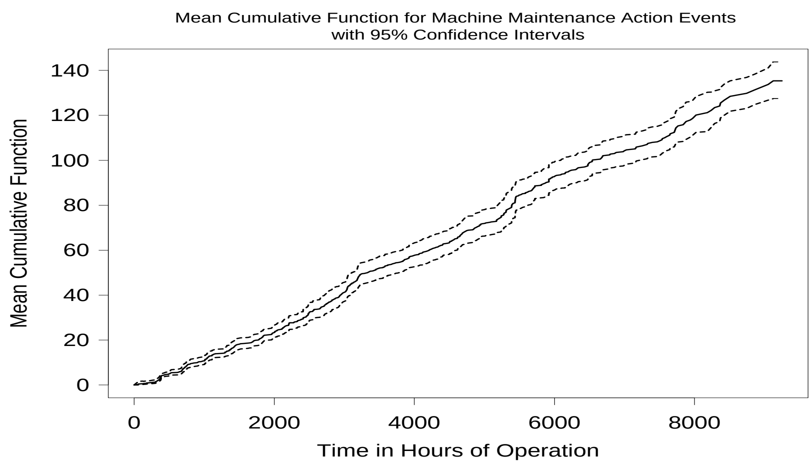

Figure 5
Earth-moving machine maintenance actions event plot.

2.4 Accelerated test and other data with explanatory variables

Accelerated testing is a widely used technique for obtaining reliability information quickly. The basic idea is to test units at higher-than-usual levels of variables like use rate, temperature, voltage, humidity, or some combination of such variables. Then through the use of a model, results are extrapolated to make inferences at levels of these variables that are close to use conditions. Models used for this purpose generally come from theoretical knowledge about the physical/chemical failure mechanism or large amounts of previous experience with the particular failure mode(s) being studied.

Figure 6 is a scatter plot of failure times for the results of an accelerated life test on an electronic device (where time to failure was the response). All units were tested simultaneously in four different temperature chambers. The test was terminated after 5000 hours, causing some observations to be censored at each level of temperature, as shown on plot. The data were first presented in Hooper and Amster (1990).

There are a number of statistical tools available for analyzing accelerated life data. Figure 7 is a multiple lognormal probability plot of the Device-A data with the ML estimate of the Arrhenius-lognormal model superimposed on the plot.

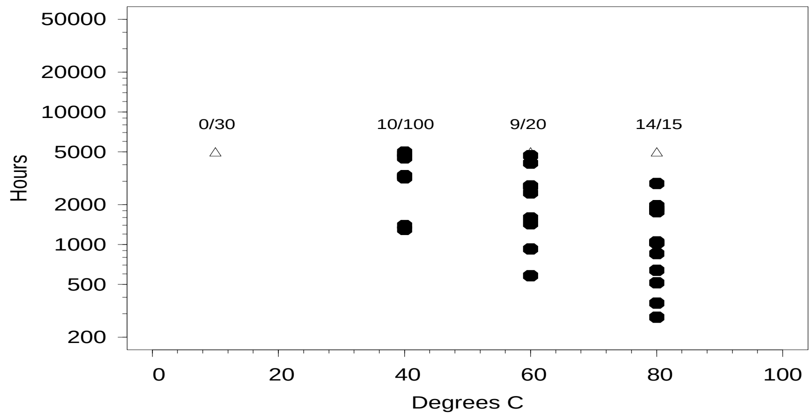

Figure 6
Scatter-plot of lifetime versus temperature from a temperature-accelerated life test for Device-A. Censored observations are indicated by Δ*. The number of failures/tested units were 0/30, 10/100, 9/20, 14/15 at 10, 40, 60, and* 80°C*, respectively.*

Such multiple probability plots are useful summary/diagnostic tools that provide an assessment of the distributional fit (the points representing the nonparametric estimate fall along straight lines), the constant variance assumption (departures of the points from parallel lines are not large and do not deviate systematically), and the fit to the Arrhenius model (the model lines agree with the nonparametric estimates). Figure 8, showing life as a function of temperature, provides another view of the same data, along with the fitted model.

As described in Section 2.2, degradation data can provide much more information than life test data. Degradation tests also can be accelerated. Indeed, the laser test results shown in Figure 3 came from an accelerated test that used increased temperature and humidity. Another example of a repeated measures degradation data is given in Figure 9 where the increase in resistance of carbon-film resistors is accelerated by using higher-than-usual temperatures.

The same two analysis methods described in Section 2.2 can be used with accelerated repeated measures degradation data. Chapter 21 of Meeker and Escobar (1998) describes these methods in detail and provides additional examples.

In some situations measuring degradation of a unit is a destructive process (or the measurement changes the degradation process itself). In such

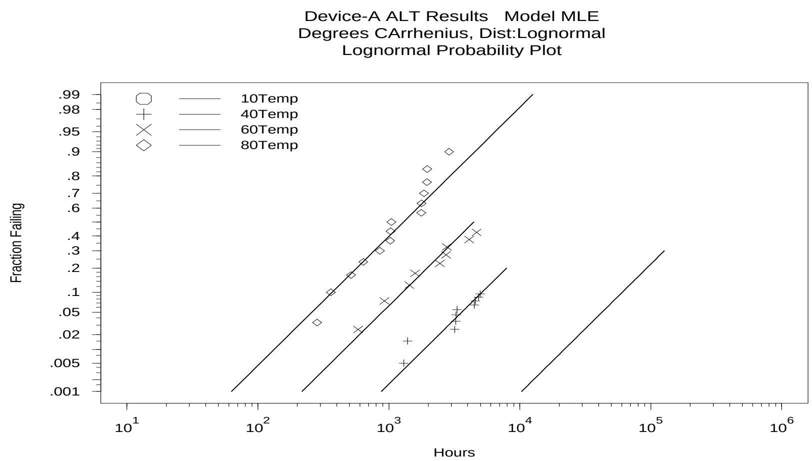

Figure 7

Multiple lognormal probability plot for the Device-A data with the Arrhenius acceleration model ML estimates at 10, 40, 60, and 80°C.

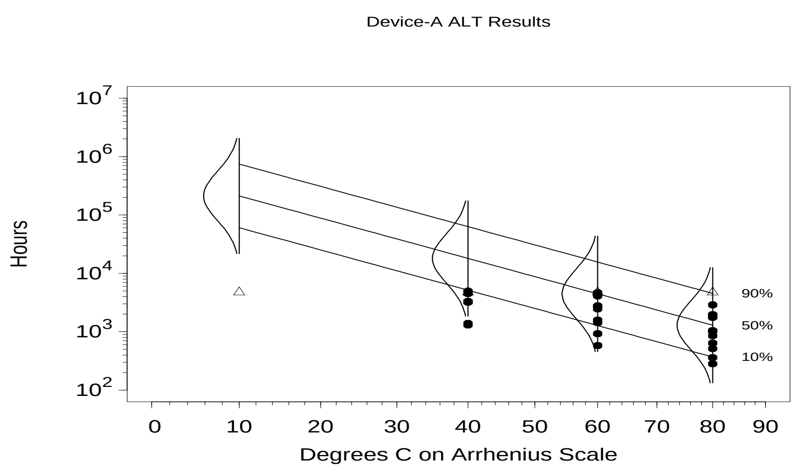

Figure 8

Device-A data Arrhenius-lognormal acceleration model.

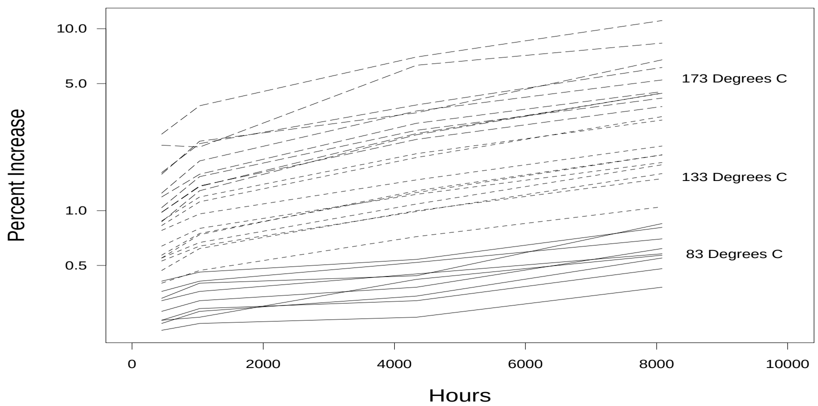

Figure 9
Percent increase in resistance over time for a sample of carbon-film resistors.

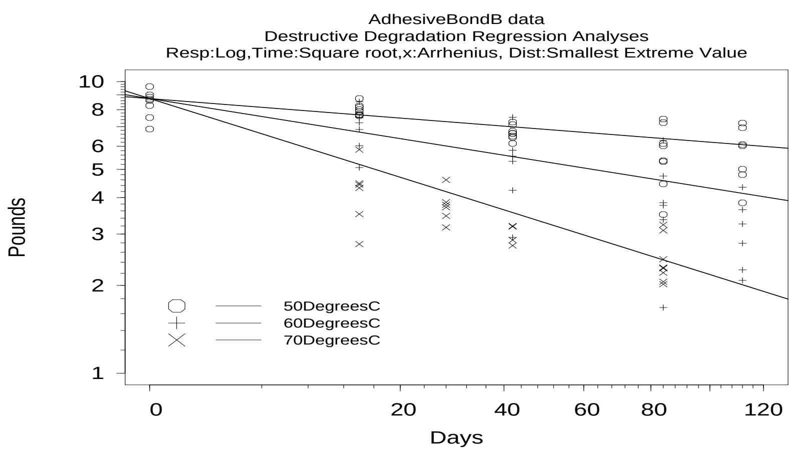

Figure 10
Destructive degradation data on the strength of an adhesive bond.

situations, only one measurement can be taken on each unit, and the resulting data are called "destructive degradation data." Special models and methods are needed for analyzing such data. These have been described in Chapter 11 of Nelson (1990) and extended in Escobar, Meeker, and Kugler (2002a). Figure 10 shows destructive degradation data for an adhesive bond. Specimens were aged at different temperatures and for different amounts of time before being destructively tested for strength.

3 SPLIDA overview and concepts

SPLIDA was designed as a tool for doing reliability data analysis for all of the different kinds of data described in Section 2, as well as test planning. This section describes some of the common features and concepts behind the SPLIDA project to provide reliability data analysis capabilities in an easy-to use software package. Table 1 shows the first-level SPLIDA menu for version 5.8 (August 2001), giving an overview of SPLIDA's different capabilities.

Table 1
Top level of the SPLIDA *menu*

Make/summary/view/modify data object ⇒
—LIFE DATA SINGLE DISTRIBUTION—
Plan single a distribution study ⇒
Single distribution life data analyses ⇒
Single distribution Bayes analysis ⇒
Multiple failure mode life data analysis ⇒
—LIFE DATA COMPARISON AND REGRESSION—
Comparison of distributions life data analysis ⇒
Plan an accelerated life test (ALT) ⇒
Simple regression (ALT) data analysis ⇒
Multiple regression (ALT) life data analysis ⇒
—RECURRENCE DATA—
Recurrence (point process) data analysis ⇒
—REPEATED MEASURES DEGRADATION DATA—
Degradation (repeated measures) data analysis ⇒
—ACCELERATED DESTRUCTIVE DEGRADATION TESTS (ADDT)—
Plan an ADDT ⇒
ADDT data analysis ⇒
—SPLIDA SPECIAL TOOLS AND MODELS—
Special models ⇒
SPLIDA tools ⇒
Preferences (change SPLIDA default options)

3.1 User interface

The S-PLUS system, upon which SPLIDA has been built, was originally a command-driven system. Starting with Version 4, however, S-PLUS has had a graphical user interface (GUI). Although many S-PLUS experts still prefer the command-line approach for providing inputs to S-PLUS and doing data analysis (because it allows flexibility and easy extendibility), there are a number of important advantages for also having a GUI. The GUI is especially convenient when analyses involve complicated specification (e.g., requiring the use of multiple options and typing long descriptive strings) or when users only need to use a software system infrequently. The S-PLUS GUI is also extendible.

Developments in SPLIDA generally start by designing and writing a command-line interface for a desired procedure. Then a dialog box is designed and an appropriate menu item is given a position in the two-level SPLIDA menu. Invisible to the user, the dialog, when exercised, will generate a command-line call to the corresponding underlying S-PLUS/SPLIDA function. It is possible for GUI users to capture the function calls (using the S-PLUS history feature) so that analysis scripts can be saved and run in the future, possibly with minor changes to the inputs. All of the SPLIDA functionality is available from the command line, but the commands are documented only through a collection of files (called echapters) containing commands to do all of the examples in Meeker and Escobar (1998). A large fraction of the command-level methodology is also available through the GUI. There is a user's manual for the SPLIDA GUI (Meeker and Escobar 2002b), distributed in electronic form with the SPLIDA.

In line with modern statistical analysis methods, most of the output from SPLIDA is in the form of graphical display of data, fitted models, and various diagnostics. Tabular output is, however, sometimes also provided, and there are generally options to request additional detailed numerical output for important information displayed in SPLIDA graphics.

Other important features of a good GUI that we have tried to incorporate into SPLIDA include:

- Menus should be organized according to the way that an analyst is expected to use a program. In SPLIDA, the first-level entries (Table 1) are organized by kind of data/analysis to be done (roughly corresponding to parts or chapters of Meeker and Escobar 1998) and second level entries are ordered according to the sequence in which the corresponding methods would be expected to be used for a given analysis. Table 2 shows, for example, the second-level menu for a single distribution Bayesian analysis.
- Flexibility requires that users be given a wide range of options in conducting analyses. To make the system easy to use, however, the amount of input required to be specified by the user should be minimized. This can be accomplished by using carefully chosen defaults,

when ever possible, and by providing an easy-to-use mechanism for changing defaults when needed.

- When there is no natural default for an input, intelligent, dynamic option lists should be provided, whenever possible. In some cases, options are limited by system design (e.g., available distributions to fit). In other cases, appropriate option lists for inputs are deduced from information available to SPLIDA (e.g., by remembering recent objects that were used, examining data that has been selected for analysis, etc.). For example, when the user needs to choose a value of an explanatory variable at which to estimate a failure probability (or some other quantity), a list can be prepared providing options in the range of the values in the data set. The list serves as a reminder to the user of the range and scaling of the data, but the user is free to choose any value (even if not in the list) as input (unless the explanatory variable is categorical, in which case choices should be limited to the categories in the variable).
- Dialog boxes should be designed carefully to lead the user through the use of the dialog. Required specifications and commonly-used options should be on the first page. Less frequently used options should be placed on back pages. Dialog components can be enabled dynamically to lead the user through a dialog. Parts of the dialog can be disabled to keep from distracting the user by irrelevant dialog elements (e.g., if a table of fraction failing $F(t)$ is not being requested, the dialog elements allowing choice of points of evaluation should be disabled).
- It is important to catch input errors or inconsistencies as soon as possible and to provide appropriate messages and diagnostics so that the user can effect a correction and proceed.

Table 2
Second-level menu for a single distribution Bayesian analysis

Specify prior information
Make a posterior distribution
Summarize a posterior distribution

3.2 S-PLUS **and object oriented systems**

S-PLUS is an "object-oriented" computing environment. In simple terms, object-oriented means that system developers define "objects" of different types along with "methods." The methods operate on these objects. Objects serve as inputs to methods and methods can create new objects as outputs. Objects generally contain self-describing information that reduces the need for the user to specify information to the system. Once an

S-PLUS object has been created, it remains on the S-PLUS "data base" until it is explicitly deleted (or overwritten).

Within a system of methods and object definitions, the user is generally relieved from the task of having to specify, in detail, which particular method to use. For example, the print method (or command) knows how to effectively format and print widely different kinds of objects. Methods are designed to be intelligent enough to use information in the attributes of the input object to decide how to take appropriate actions.

SPLIDA uses these same concepts to simplify operation for its users. Also, SPLIDA chooses default names for objects that are created (relieving the user from having to do this). The names are self-describing (and thus sometimes rather long, but this is not a problem when option lists containing object names are created automatically).

3.3 Data objects

This section describes SPLIDA data objects. Section 3.4 describes other kinds of SPLIDA objects.

Data are usually imported into S-PLUS from a text file or an Excel spreadsheet. The resulting simple rectangular S-PLUS object is known as a data set (in the S-PLUS GUI) or a data frame (in the S-PLUS language). As is traditional in statistical software, a data set has columns corresponding to variables (numeric or character) and rows corresponding to cases. Although such data sets are the natural input for many S-PLUS functions, in SPLIDA, the object definition is richer, containing much more information about the data. In particular, a SPLIDA data object contains definitions (in the form of object attributes) that specify which column is the response and (optionally) which columns correspond to censor definition, case weights (if there are a thousand units censored at the same time, only one row is needed in the data matrix), truncation information, explanatory variables, data title, response units, comments on the data, etc. These specifications are made once and for all when an object is created. Then the same data object can be used over and over again as input to different methods used in the analysis of the data, greatly simplifying the subsequent tasks for the user.

There are a number of different kinds of data objects defined in SPLIDA. These include:

- Life data objects of different types including
  - Single distribution life data (no explanatory variables).
  - Single distribution life data with information on the cause of failure (multiple failure mode data).
  - Life data with a categorical explanatory variable (usually for the purpose of comparing categories such a manufacturer, processing method, vendor, and so on).

  - Life data with general explanatory variables that may be either continuous or categorical (used for accelerated testing and covariate adjustment of field data). For regression modeling, SPLIDA takes advantage of the powerful model-specification methods in S-PLUS.

- Recurrence (point process) data.

- Repeated measures degradation data.

- Single measure (destructive) degradation data.

In analysis dialog boxes, SPLIDA will, by default, display only the data objects that are appropriate for the analysis method. When using an analysis method, the user generally has only to choose the data object from a list of appropriate objects and perhaps the distribution or model to be fitted to the data. Other specifications are optional with defaults generated automatically through the information contained in the data object (e.g., ranges of the variables) or by SPLIDA default options (e.g., the confidence level default is 95%). For example, the default title on each graph is constructed by pasting together the title in the data object and a description of the method chosen for analysis. Also, it is easy to override default options. This system structure allows the user to rapidly and accurately perform multiple analyses for a given data set.

3.4 Results objects and other SPLIDA objects

In addition to the different data objects described above, SPLIDA uses a number of other kinds of objects. Some of these include

- Results objects that are the output of a model-fitting method and serve as the input to various diagnostic or model summarization methods.

- Prior distribution objects that describe the prior information available for the parameters of given model (i.e., parameters of prior distributions).

- Posterior distribution objects that contain simulated samples from the posterior distribution for a given model/data/prior combination.

- Simulation output objects that serve as input to simulation summarization methods. Such summarization methods allow users to view the results of the simulation in different ways (simulation results should be viewed as data and the simulation summarization methods are special data analysis tools).

- Planning value objects that describe a model and parameters for a particular testing situation and that serve as input to test planning methods.

- Test plan objects that serve as input (along with the plan values) to test plan evaluation or simulation methods.

As with data objects, the appropriate object types appear in the option lists of various method dialogs. For example, the dialog to evaluate an accelerated life test plan allows the user to choose among available plan value objects from one list and available test plans from another list. It is easy to evaluate various test plans by changing the test plan choice, without having to re-specify the plan values or other options.

4 Technical capabilities in SPLIDA

Section 1.3 outlined general needs for reliability data analysis software and how these have been addressed in the design of SPLIDA. This section describes some of the particular capabilities that are available in SPLIDA.

4.1 Graphics and visualization

Today it is widely accepted that graphical methods provide essential tools for data analysis. In addition to presenting data and the results of model fitting, graphical presentation of simulation results of statistical phenomena provides a powerful tool for visualizing and obtaining insights into statistical variability and sensitivity to model specification. Such graphics are used both with bootstrap samples for inferential problems and for simulating proposed test plans (as described in Section 5.2). Graphical methods are also useful for studying the behavior of likelihood functions.

4.2 Likelihood analysis methods

Likelihood methods provide powerful, versatile tools for data analysis and inference. Virtually all of the procedures in SPLIDA are based, in one way or another, on likelihood. There is well developed theory for maximum likelihood (ML) estimation. The technical justification for ML methods is based on large-sample theory. In particular, under mild regularity conditions, ML estimators have desirable statistical properties, such as minimum variance. ML methods, however, generally perform very well even in small samples and provide a basis for constructing confidence intervals. The concepts behind likelihood inference methods are intuitive and easy to present to engineers. The basic idea behind likelihood inference is that regions in the parameter space with relatively high likelihood are more plausible than those with relatively small likelihood.

Figures 12 and 13 show the likelihood for the bearing cage data. The contours in Figure 12 are labeled to correspond to joint confidence regions for the parameters. The wire-frame plot in Figure 13 may be easier for some to visualize. For the trained eye, however, contour plots are more informative.

4.3 Quantifying statistical uncertainty

Presentation of the results of a statistical study often requires quantification of uncertainty, especially in reliability and safety applications. It is useful to divide uncertainty into "statistical uncertainty" and other uncertainties which we will collectively call "model uncertainty."

Statistical uncertainty arises because of limited data. Confidence intervals are the commonly accepted method to quantify *statistical* uncertainty. The confidence level (e.g., 95%) associated with a confidence interval is a *property of the procedure* that was used to construct the interval. Confidence interval procedures may not have exactly the nominal confidence level. Such procedures are called "approximate." Examples of confidence intervals are shown in Figures 2, 4 and 7.

For complicated data and models (e.g., including censoring and truncation), "exact" methods of constructing confidence intervals are not available. Approximate methods are used instead. In recent years it has been recognized that the normal (or Wald) approximations can be seriously deficient when the amount of information in one's data is limited, even if the sample size is large. With failure-time data, the adequacy of large sample approximations is best described in terms of the (expected) number of failures. For example, it may be necessary to have between 50 to 100 failures in a censored sample before a normal approximation interval provides an adequate approximation (e.g., within 1% of the nominal). Alternative methods based on likelihood or bootstrap simulation are recommended because they perform much better. Meeker and Escobar (1995) illustrate the difference between likelihood and normal-approximation intervals. Further discussion of these issues, references for further reading, and a detailed evaluation of alternative methods is given in Jeng and Meeker (2000).

Modern statistical software should allow the user to use one of these better methods for constructing confidence intervals. In SPLIDA, normal approximation intervals are provided by default. At the present time, methods for likelihood and bootstrap confidence intervals are available by special request, but only for single distribution analysis. In some future release, there will be a transparent option to allow alternative methods for users willing to pay the increased cost for computation time and data storage.

4.4 Quantifying model uncertainty

It is important to recognize that statistical confidence intervals, as described in Section 4.3, reflect *only* statistical uncertainty. This fact is critically important, especially when making inferences outside of the range of one's data. Sensitivity analysis is the most commonly used method for assessing model uncertainty. The approach for sensitivity analysis in SPLIDA is to make it easy for users to compare alternative models. If the results of the models give approximately the same answers, then there need be little concern for model deviations within the class of models used in the sensitivity analysis. When differences are important, and there is no basis outside of the data that can be used to narrow the selection of the model, the deviations seen in the sensitivity analysis must be acknowledged by describing the uncertainty in the results of the analysis.

Figure 11 shows the output of a SPLIDA procedure that allows comparison between two different distributions. This particular example compares fitting lognormal and Weibull distributions to the bearing cage data described in Section 2.1. The plot is given on lognormal probability paper and, correspondingly, the fitted lognormal distribution plots as a straight line. The curved line is the estimate of the Weibull cdf. The plot shows that, within the range of the data (roughly between 200 and 1600 hours of service), there is very little difference between the two distributions. Beyond 2000 hours, however, the Weibull gives much more pessimistic estimates of product reliability, relative to the lognormal distribution. Indeed, the Weibull estimate of the fraction failing approaching the upper confidence intervals for the lognormal distribution as time approaches 10,000 hours (underlining the statement above that confidence intervals reflect *only* statistical uncertainty).

Meeker and Escobar (2002a) provide a detailed example to illustrate the use of SPLIDA's sensitivity analysis tools with respect to an underlying regression model used for extrapolation in the analysis of accelerated life test data. They present an example estimating the fatigue life distribution of a mechanical spring.

4.5 Bayesian analysis methods

There are many reliability applications for which it is essential that engineering information be incorporated into the analysis. Often this is done by assuming that a particular parameter (e.g., the Weibull shape parameter) is known on the basis of previous experience or physical theory (e.g., Nelson 1985). Incorporating such knowledge into an analysis can have a profound effect on the needed experimental resources to answer a question or to the resulting precision, for a given amount of data.

Usually, the engineering information is not known precisely. In such cases Bayesian methods can be used to combine the engineering information with information in the data. Instead of trying to describe the tech-

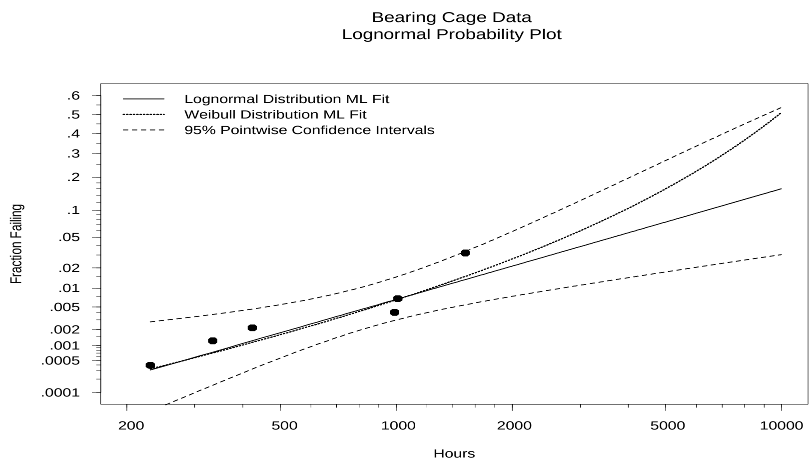

Figure 11
Bearing cage failure-time data Weibull-lognormal comparison.

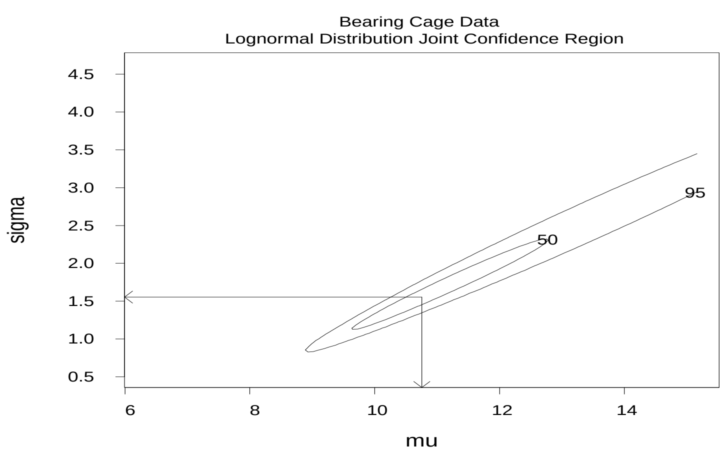

Figure 12
Bearing cage failure-time data lognormal joint confidence region.

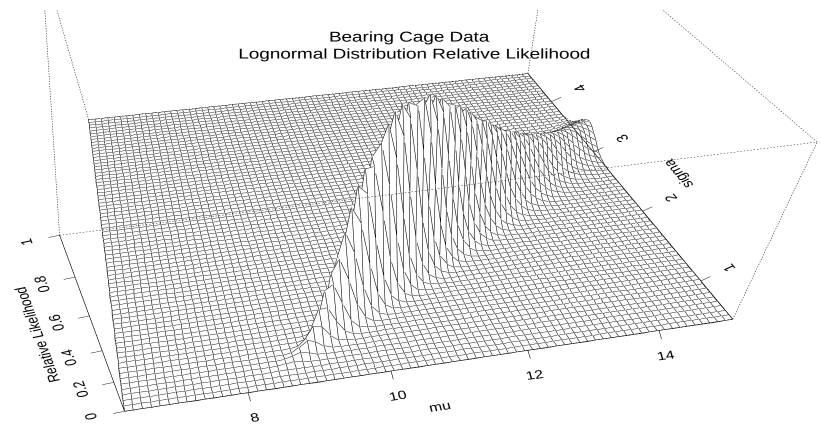

Figure 13
Bearing cage failure-time data lognormal relative likelihood.

nical details of the use of Bayes methodology in reliability data analysis, we provide an example and visualization of how prior information can be combined with data to create a posterior distribution reflecting uncertainty in a reliability characteristic. For simplicity and to provide a convenient visualization of the procedure, SPLIDA uses a Monte Carlo method given by Smith and Gelfand (1992) to implement Bayes theorem.

The points plotted in Figure 14 represent a sample from the joint prior distribution for the bearing cage 0.01 quantile and the Weibull shape parameter β. The prior distribution was rather diffuse for the 0.01 quantile (log uniform from 700 to 3000 hours), but concentrated for the Weibull shape parameter (lognormal with 99% of the probability between 1.7 and 2.2), reflective of strong information such as might be available from previous experience with the same material, geometry, and operating conditions. The contours are the relative likelihood $R(t_{.10}, \beta) = L(t_{.10}, \beta)/L(\widehat{t}_{.10}, \widehat{\beta})$ from the bearing cage data, reflecting the information in the data.

As shown in Chapter 14 of Meeker and Escobar(1998), a sample from the posterior distribution can be obtained by filtering a prior distribution sample with a probability of keeping a point being equal to the relative likelihood at that point. Figure 15 shows such a posterior sample (there were actually more than 6000 points in the posterior sample, but only a subset are plotted).

The density of the points from the posterior reflect the combination of the data and the prior distribution. We can see that from the data alone,

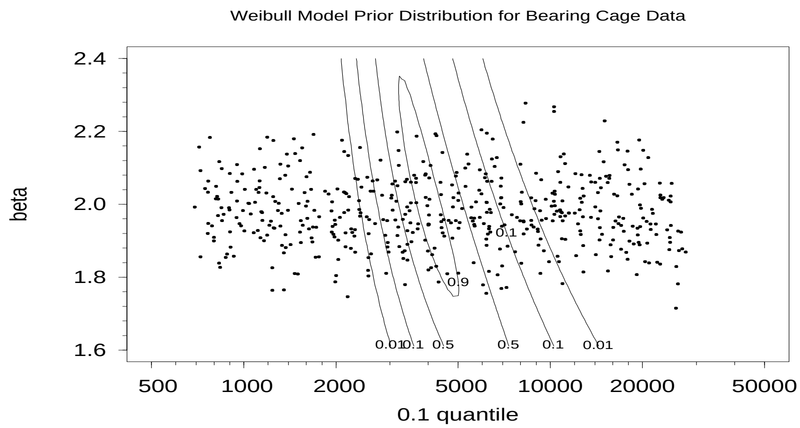

Figure 14
Bearing cage prior distribution sample and relative likelihood contours.

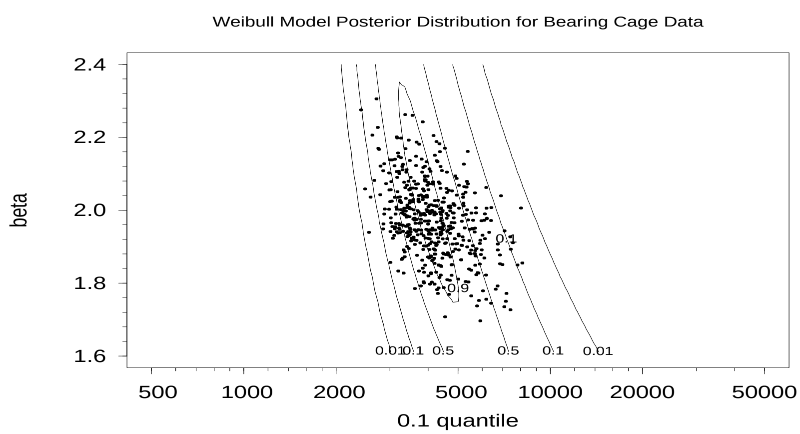

Figure 15
Bearing cage posterior distribution sample and relative likelihood contours.

there was little information about the value of the Weibull shape parameter. After filtering, however, the location of the Weibull distribution is known with much more precision. SPLIDA provides a number of useful methods for summarizing such posterior samples.

4.6 Numerical methods

It is essential that high-quality numerical algorithms be used in developing methods for reliability data analysis. Needed numerical methods include matrix algebra, computation of probability and other special functions, numerical integration, numerical differentiation, and numerical maximization. Derivatives should be programmed explicitly whenever practicable. Numerical differentiation is inherently an unstable operation and needs to be done with great care.

S-PLUS generally uses excellent numerical algorithms and we take advantage of these algorithms whenever possible. At the Fortran level, we use a number of excellent algorithms that were obtained from Netlib at www.netlib.org.

Iterations to maximize the likelihood can fail for a number of different reasons, including poor starting values, poor parameterization, or a lack of identifiability (in which case a unique maximum does not exist and there are ridges or other indications of constant likelihood in places where the likelihood is relatively high). Finding a stable parameterization (described, for example in Ross 1990) is vitally important to having a robust ML algorithm. Good starting values can generally be obtained from crude methods of estimation like the method of moments. More sophisticated methods would use a few iterations of an EM algorithm to correct for censoring or truncation. If parameters are not identifiable, there is little that numerical methods can do to solve the problem. Even in such cases, however, a good algorithm should be able to detect the identifiability problem and provide appropriate diagnostics.

5 Planning reliability studies and sample size choice

Although many of the key ideas of classical experimental design are important and useful in planning reliability studies, the special needs arising from issues like censoring, extrapolative inference, and prediction require special methods and tools. The general principles that we use to guide users in planning all types of reliability studies are as follows.

- Have well defined goals for what is to be estimated and at least a rough idea about the degree of precision needed for estimation.

- Some information is required about the underlying model and the values of the parameters of the model. Generally, in nonlinear estimation, the best experimental plan depends importantly on this information. Such information has to be obtained from some combination of previous experience with similar products, physical/chemical knowledge relating the the failure mechanism, or engineering judgment.

- Before planning a reliability study, analysts should take steps to anticipate the kind of results that are expected in the study (e.g., by using Monte Carlo simulation) and to understand well the methods of analysis that are likely to be used in analysis.

The philosophy that we follow in SPLIDA, with respect to experimental design and test planning, is to use technical methods to anticipate and investigate the possible range of results that one would expect to obtain for a given test plan. In this way, the experimenter can assess the amount of information (or precision) that might be obtained from a given experiment and then make informed decisions about the advantages and disadvantages of alternative test plans.

Once information about the test setting (i.e., the model and planning values for the model parameters) has been specified/obtained, it is possible to evaluate the properties of proposed test plans by using one or both of the following two methods.

1. Direct evaluation of properties or approximate properties of proposed test plans (e.g., standard errors of ML estimates of quantities to be estimated or related precision metrics). In all but the simplest situations, these evaluations require the use of large sample approximations. When planning life tests, the approximations tend to be adequate as long as the probability of zero failures is not too small.

2. Evaluation of test plans by means of simulation. For a given test plan and set of planning values, the "test" is simulated a large number of times. The results of the simulation, then properly plotted, provide a visualization of sampling error and an assessment of potential estimation precision.

5.1 Test planning information

Information about the model and its parameters is needed for test planning. It can be shown theoretically that certain planning decisions are invariant to particular model properties. For example the optimum proportional allocation of test units to levels of an accelerating variable will, in certain accelerated testing problems, be invariant to the scale parameter of the underlying location-scale distribution. The needed sample size to achieve a specified degree of precision will, however, depend on this parameter.

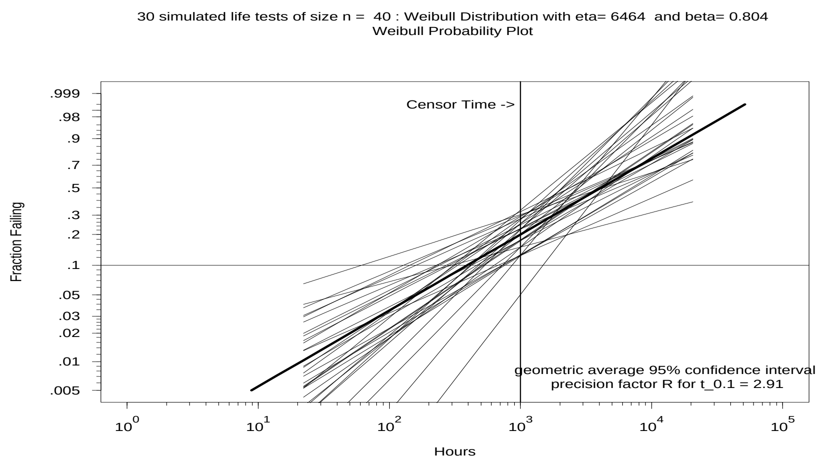

Figure 16
Simulated life test $n = 40$.

Planning information is usually obtained from some combination of previous experience, engineering judgment, physical theory, or preliminary tests. Due to the uncertainty in these inputs, it is important to assess the effect on proposed test plans of perturbations to these inputs in the same manner as the sensitivity analysis describes in Section 4.4.

For a simple life test to estimate a failure time distribution, the planning information can be displayed graphically by drawing a line on a probability plot similar to the one in Figure 2 (but without the data).

5.2 Simulation of a proposed test plan

Figures 16 and 17 present the results of a simulation of a life test experiment to estimate the failure time distribution of an insulator (from Examples 10.1 and 10.7 of Meeker and Escobar 1998), using samples of size 40 and 160, respectively. The planning information specified a Weibull distribution with about 20% of the units expected to fail after 1000 hours and a Weibull shape parameter in the neighborhood of 0.8037. The corresponding Weibull distribution is given by the longer dark solid lines in these figures.

The life test plans under comparison both have censoring times of 1000 hours. The expected number failing for the two plans are 8 and 24, respectively. The thinner, shorter lines in the figures represent ML estimates from 30 simulated life tests. These lines allow the test planner to visualize the amount of sampling error that will be associated with a given test plan. The horizontal line at 0.1, for example, allows an assessment of

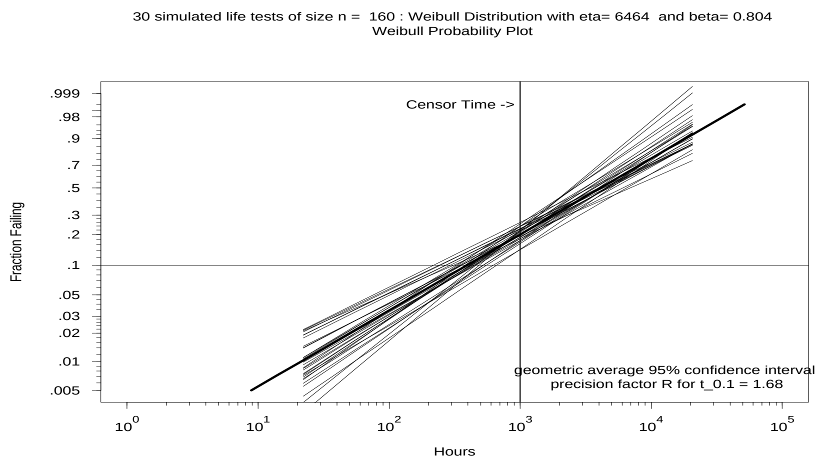

Figure 17
Simulated life test $n = 160$.

the sampling distribution of the ML estimate of the 0.10 quantile of the failure-time distribution. Numerical summaries of the simulation results are also useful and some of these are printed in the plot itself. Others are available in the form of tabular output.

5.3 Large-sample approximations and sample-size tools

For most reliability data analysis problems it is possible to derive expressions and develop algorithms to compute large-sample approximations for standard errors of ML estimators (e.g., Escobar and Meeker 1994, 1998). Such algorithms can be used to develop tools to allow an assessment of the relationship between sample size and precision. Figure 18 provides such a tool for the insulation evaluation life test, showing the sample size needed to have a specified precision factor target.

A precision factor of $R = 2$ implies, for example, that the upper endpoint of a confidence interval will be two times the ML estimate (and that the lower endpoint will be half of the ML estimate).

5.4 Other test planning methods

SPLIDA also has methods for a number of other planing problems that arise in reliability testing, including

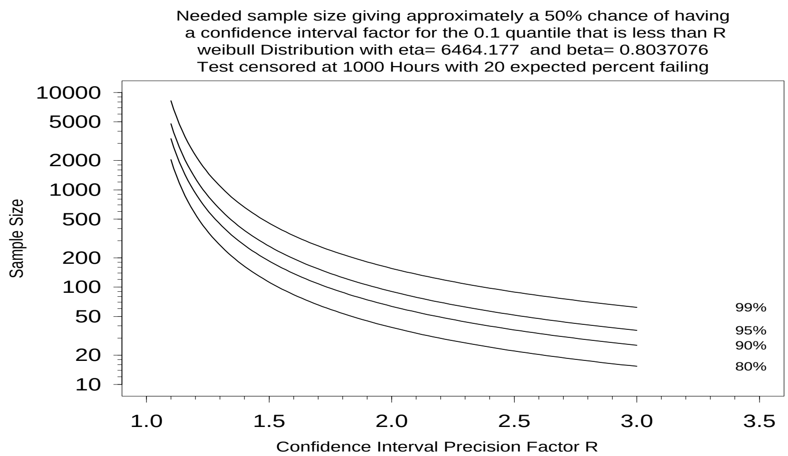

Figure 18
Needed sample size for a life test.

- Probability of successful demonstration (McKane, Escobar and Meeker 2001).
- Probability of correct selection (Pascual, Escobar and Meeker 2002).
- Planning accelerated life tests (Chapter 20 of Meeker and Escobar 1998).
- Planning accelerated destructive degradation tests (Escobar, Meeker, and Kugler 2002b).

These methods use simulation and/or large-sample approximation in a manner that is similar to the single distribution methods described in Sections 5.2 and 5.3.

Some of the underlying functions in SPLIDA (e.g., computation of the Fisher information matrix for censored observations and procedures for doing simulation with censored data) provide building blocks allowing advanced users to develop their own analysis and test planning methods for special situations.

6 Concluding remarks and future work

This paper has outlined the important computing needs that we have encountered in our consulting, teaching, and research in the area of reliability data analysis. We have also described and illustrated some of the capabilities of the SPLIDA computing system that we have developed to meet these needs. Much remains to be done, even on the landscape that we have described. For example more accurate approximate confidence intervals and a more general implementation of Bayesian methods would find immediate applications. Only the simple method of repeated measures degradation analysis has been programmed at the GUI level. Methods for accelerated testing with multiple failure modes would also be useful.

SPLIDA, like S-PLUS, is written largely in the S-PLUS language itself. Correspondingly, the SPLIDA system, especially at the command level, can be extended to go beyond the existing capabilities. Some of the underlying functions in SPLIDA (e.g., computation of the Fisher information matrix for censored observations and procedures for doing simulation with censored data) provide building blocks allowing advanced users to develop their own analysis and test planning methods for special situations.

Most of the commonly used reliability models are based on log-location-scale distributions and linear relationships between distribution parameters and explanatory variables. Particularly because so much of reliability statistics depends on the use of physical models, it is important that analysts be able to deal with new models as they are developed. Greater flexibility in model choice is an important goal that deserves more effort.

Acknowledgments

We would like to thank Victor Chan and Katherine Meeker for helpful comments on a previous version of this paper.

References

Abernethy, R. B., Breneman, J. E., Medlin, C. H., and Reinman, G. L. (1983). *Weibull Analysis Handbook*. Air Force Wright Aeronautical Laboratories Technical Report AFWAL-TR-83-2079. Available from the National Technical Information Service, Washington, DC.

Ascher, H., and Feingold, H. (1984). *Repairable Systems Reliability*. New York: Marcel Dekker.

Cox, D. R., and Isham, V. (1980). *Point Processes*. London: Chapman & Hall.

Cox, D. R., and Lewis, P. A. W. (1966). *The Statistical Analysis of a Series of Events*. London: Chapman & Hall.

Cox, D. R., and Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman & Hall.

Crowder, M. J., Kimber, A. C., Smith, R. L., and Sweeting, T. J. (1991). *Statistical Analysis of Reliability Data*. New York: Chapman & Hall.

Escobar, L. A., Meeker, W. Q., and Kugler, D. (2002a). Models and methods for accelerated destructive degradation tests. Preprint, Iowa State University.

Escobar, L. A., Meeker, W. Q., and Kugler, D. (2002b). Planning accelerated destructive degradation tests. Preprint, Iowa State University.

Hahn, G. J., Doganaksoy, N., and Hoerl, R. (2000). The evolution of six-sigma. *Quality Engineering*, **12**, 317–326.

Hooper, J. H., and Amster, S. J. (1990). Analysis and presentation of reliability data. In *Handbook of Statistical Methods for Engineers and Scientists*, Harrison M. Wadsworth, Editor. New York: McGraw Hill.

Jeng, S. L., and Meeker W.Q. (2000). Comparisons of Weibull distribution approximate confidence intervals procedures for type I Censored data. *Technometrics* , **42**, 135–148.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: John Wiley & Sons.

Lawless, J. F., and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics*, **37**, 158–168.

Mann, N. R., Schafer, R. E., and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data*. New York: John Wiley & Sons.

McKane, S.W., Meeker, W.Q., and Escobar, L.A. (2001). Sample size and number of failure requirements for demonstration tests with location-scale distributions and type II censoring. Preprint, Iowa State University.

Meeker, W. Q., and Escobar, L. A. (1995). Teaching about approximate confidence regions based on maximum likelihood estimation. *The American Statistician*, **49**, 48–53.

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

Meeker, W. Q., and Escobar, L. A. (2002a). Use of sensitivity analysis to assess the effect of model uncertainty in analyzing accelerated life test data. In *Case Studies in Reliability and Maintenance*, Blischke, W. R. and Murthy, D. N. P.. New York: John Wiley & Sons.

Meeker, W. Q., and Escobar, L. A. (2002b). SPLIDA User's Manual. Available at www.public.iastate.edu/~stat533/splida.html.

Meeker, W. Q., Doganaksoy, N., and Hahn, G. J. (2001). Using Degradation data for product life analysis. *Quality Progress* **32**, 60–65.

Nelson, W. (1982). *Applied Life Data Analysis.* New York: John Wiley & Sons.

Nelson, W. (1985). Weibull analysis of reliability data with few or no failures. *Journal of Quality Technology*, **17**, 140–146.

Nelson, W. (1988). Graphical analysis of system repair data. *Journal of Quality Technology*, **20**, 24–35.

Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses.* New York: John Wiley & Sons.

Nelson, W. (1995). Confidence limits for recurrence data—applied to cost or number of product repairs. *Technometrics*, **37**, 147–157.

Nelson, W. (2002). *Analysis of Recurrent-Events Data on Product Repairs, Disease Recurrences, and Other Applications.* To appear in the ASA-SIAM Series on Statistics and Applied Probability.

Nelson, W., and Hendrickson, R. (1972). 1972 User manual for STATPAC-a general purpose program for data analysis and for fitting statistical models to data. General Electric CR&D Technical Report 72GEN009.

Pascual, F., Escobar, L.A., and Meeker, W.Q. (2002). Probability of correct selection in life test comparisons. Preprint, Iowa State University.

Rigdon, S., and Basu, A. (2000). *Statistical Methods for Reliability of Repairable Systems.* New York: John Wiley & Sons.

Ross, G. J. S. (1990). *Nonlinear Estimation.* New York: Springer-Verlag.

Smith, A. F. M., and Gelfand, A. E. (1992). Bayesian statistics without tears: a sampling-resampling perspective. *The American Statistician*, **46**, 84–88.

Snyder, D. L. (1975). *Random Point Processes.* New York: John Wiley & Sons.

Strauss, S. (1980). STATPAC: A general purpose program for data analysis and for fitting statistical models to data. *The American Statistician*, **34**, 59–60.

Thompson, W. A. (1988). *Point Process Models with Applications to Safety and Reliability.* London: Chapman & Hall.

Tobias, P. A., and Trindade, D. C. (1995). *Applied Reliability.* 2nd ed.. New York: Van Nostrand Reinhold.