

An Overview of Designing Experiments for Reliability Data

G. Geoffrey Vining, Laura J. Freeman, and Jennifer L.K. Kensler

Abstract The reliability of products and processes will become increasingly important in the near future. One definition of reliability is “quality over time.” Customers increasing will make their purchasing decisions on how long they can expect their products and processes to deliver high quality results. As a result, there will be increasing demands for manufacturers to design appropriate experiments to improve reliability. This paper begins with a review of the current practice for planning reliability experiments. It then reviews some recent work that takes into proper account the experimental protocol. A basic issue is that most reliability engineers have little training in planning experiments while most experimental design experts have little background in reliability data.

Keywords Censored data • Skewed distributions

1 Introduction

Consumers are demanding more reliable products and services. A popular definition of reliability is “quality over time.” Consumers expect that products will continue to meet or exceed their expectations for at least the advertised lifetime, if not longer. One reason for the rise of the Japanese automotive industry within North America since the 1980s is the far better reliability of their cars and trucks.

Just as companies needed to apply experimental design concepts to improve quality, so too will they need to apply these same concepts to improve reliability. Current practice almost exclusively restricts the use of experimental protocols in

G.G. Vining (✉)
Department of Statistics, Virginia Tech, Blacksburg, VA 24060, USA
e-mail: vining@vt.edu

L.J. Freeman
Institute for Defense Analyses, Alexandria, VA, USA
e-mail: lfreeman@ida.org

J.L.K. Kensler
Shell Global Solutions (USA) Inc., Houston, TX 77210, USA
e-mail: Jennifer.Kensler@shell.com

reliability testing to completely randomized accelerated life tests. The future will see more broad scale use of basic experimental designs, analyses, protocols, and concepts.

A major problem facing this transition to more use of experimental design in reliability is the nature of reliability data. Classical experimental designs and analyses assume that the data at least roughly follow normal distributions. Reliability data tend to follow highly skewed distributions, better modeled by such distributions as the Weibull. Another complication is that typical reliability experiments censor large amounts of the data, which stands in stark contrast to classical experimental design and analysis. The issue then becomes that the people who routinely work with reliability data apply very different tools and methods than people who routinely do classical experimental design and analysis. The only proper way to apply classical experimental design approaches is to understand at a fundamental level the nature of reliability data. Unfortunately, very few people understand both fields well enough to bridge the gap.

This paper first presents an introduction to reliability data, with a special focus on the Weibull distribution and censoring. It then gives an overview of designing experiments for lifetime data. The next section introduces a motivating example analyzed in Meeker and Escobar (1998), who analyze the results as if they came from a completely randomized design. However, the data actually reflect sub-sampling. We then introduce a naive two-stage analysis that takes into account the actual experimental protocol. The next section discusses a more statistically rigorous approach to the data analysis. We then extend these basic results to the situation where we have sub-sampling within random blocks. The final section offers some conclusions and some future research directions.

2 Introduction to Reliability Data

Typically, reliability data focus on lifetimes. In some cases, these data are cycles until failure, which is a surrogate for time. Engineering examples include extremely complex systems, such as aircraft engines, as well as relatively simple parts such as metal braces. Often, engineers must build reliability models on the relatively simple components in order to develop a reasonable model for the complex system.

The most common distributions used by reliability engineers to model reliability data are the lognormal, the exponential, and the Weibull. Of these three, the Weibull tends to dominate, especially since the exponential is a special case. The lognormal distribution transforms highly skewed data to the normal distribution. The exponential distribution has a constant hazard function, which is associated with true random failure behavior, i.e. there is no specific failure mechanism associated with the failure. The biggest value of the Weibull distribution is its ability to model the times to failure for specific failure mechanisms.

Some textbooks discuss the gamma distribution for reliability applications. The gamma distribution is extremely important in queueing theory for modeling inter-arrival times. Most reliability engineers reject the basic concept of modeling times to failures as an inter-arrival problem. The primary reason is that failures have fundamental causes as opposed to truly random events. The physics based interpretation of the Weibull distribution is that it models the time to failure when the failure is due to the “weakest link,” which is a common failure mechanism. This paper purely focuses on the Weibull distribution for modeling reliability data because of its overwhelming popularity among reliability engineers.

Most reliability data involve censoring, which does complicate the analysis. The basic types of censoring are:

- Right, where the test stops before all specimens fail
 - Type I, where the test stops at a pre-specified time,
 - Type II, where the test stops after a pre-specified number of failures,
- Left, where the first failure time recorded is after failures have begun to occur,
- Interval, where the analyst only knows that the failure occurred between two times.

Censoring reflects the reality that failures typically are rare events, even under accelerated conditions. By far, the most common approaches for estimation and inference for reliability data are maximum likelihood and log-likelihood.

The likelihood for Type I and Type II censored data is:

$$L(\beta, \gamma) = \mathcal{C} \prod_{i=1}^N [f(t_i)]^{\delta_i} [1 - F(t_i)]^{1-\delta_i},$$

where $\delta_i = 1$ if the data point is observed and $\delta_i = 0$ if the data point is right censored. Additionally, $f(t_i)$ is the probability distribution function (PDF) for the assumed distribution, $F(t_i)$ is the cumulative distribution function (CDF), and \mathcal{C} is a constant which varies based on the censoring type. However, \mathcal{C} does not impact the maximum likelihood estimators (MLEs). Therefore we use $\mathcal{C} = 1$ for simplicity. The log-likelihood for the right censored data case is then given by:

$$\ell_i(\beta, \gamma) = \sum_{i=1}^N \delta_i \log(f(t_i)) + \sum_{i=1}^N [1 - \delta_i] \log(1 - F(t_i)).$$

The PDF for the Weibull distribution is

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

where $\beta > 0$ is the shape parameter, $\eta > 0$ is the scale parameter, and $t > 0$ is the time to failure. The Weibull distribution is popular because the shape parameter allows it to model several different mechanisms of failure. The CDF is

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

The hazard function represents the instantaneous probability of failure, which is quite important for reliability engineers. The Weibull hazard function is

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

We note that the hazard function is a constant when $\beta = 1$ (the exponential distribution). As a result, reliability engineers view the exponential distribution as modeling purely random failure, which often is of limited interest. For $\beta < 1$, the hazard function is monotonically decreasing, which corresponds to infant mortality. For $\beta > 1$, the hazard function is monotonically increasing, which corresponds to wear-out. As a result, the Weibull shape parameter, β , has a specific relationship to the specific failure mechanism.

The Weibull distribution is a special case of the smallest extreme value distribution, which is a member of the log-location-scale family of distributions. Let $\mu = \log(\eta)$, and let $z_i = \beta [\log(t_i) - \mu]$. We note that

$$\begin{aligned} \log[f(t_i)] &= \log\left[\frac{\beta}{t_i}\right] + z_i - e^{z_i} \quad \text{and} \\ \log[1 - F(t_i)] &= -e^{z_i}. \end{aligned}$$

As a result, the log-likelihood for right censored Weibull data reduces to:

$$\ell(\beta, \gamma) = \sum_{i=1}^N \left(\delta_i \left[\log\left(\frac{\beta}{t_i}\right) + z_i \right] - e^{z_i} \right).$$

3 Current Approaches to Planning Experiments with Reliability Data

Reliability engineers conduct life tests to develop models for the product/process lifetimes at the use conditions. In some cases, the use conditions produce sufficient failures to estimate the model well. In most cases, however, the engineers must use a stress factor (in some cases, more than one stress factor) to increase the

probability of failures. Such experiments are called accelerated life tests. Common stress factors include temperature, voltage, humidity, etc. The engineer uses the estimated model to project back to the use conditions. Inherently, accelerated life tests involve extrapolation. As a result, the experimenter must exercise caution in choosing the appropriate levels for these stress factors. If the level is too extreme, the failure mechanism can change, which then nullifies the ability to extrapolate back to the expected behavior at the use conditions. In such a case, the cause of the failure may never occur at the use conditions, which is a problem.

The current literature for designed experiments with reliability data almost exclusively uses a completely randomized design, even when the actual protocol is something different. The focus of the current literature is on planning optimal designs. The basic issues are how many levels to use for the stress factors, how to allocate the available units to these levels, and how long to run the test.

Typically, accelerated life tests use a single stress factor with at least three levels. The linear models theory underlying the analysis suggests that the optimal design should use only two levels. The rationale for at least three is practical. Often the level for the stress factor closest to the use conditions does not produce enough failures to estimate the model well. Using more than two levels helps to mitigate that risk. Similarly, using more than two levels can help to mitigate the risk of inducing a new failure mechanism.

The typical analysis of life and accelerated life tests uses the reparameterization of the Weibull distribution to the smallest extreme value distribution. The basic idea is to use a linear model for the log-location parameter, μ . As a result, the basic model is

$$\mu_i = \mathbf{x}'_i \boldsymbol{\theta},$$

where μ_i is the log-location parameter for the i th experimental run, \mathbf{x}_i is the specific value for the i th setting for the experimental factors (taking the model into account), and $\boldsymbol{\theta}$ is the corresponding vector of regression coefficients. Typically, the model does include the y -intercept term. Engineers then use maximum likelihood to estimate the model parameters and log-likelihood to perform inference.

4 Motivating Example

Zelen (1959) discusses a factorial experiment to determine the effect of voltage and temperature on the lifespan of a glass capacitor. Zelen describes the experiment as “ n components are simultaneously placed on test.” Table 1 summarizes the experimental results. Zelen uses eight capacitors per test stand and Type II censoring after the fourth failure. Each test stand receives a different combination

Table 1 Life test results of capacitors, adapted from Zelen (1959)

Temperature (C)	Applied voltage			
	200	250	300	350
170	439	572	315	258
	904	690	315	258
	1,092	904	439	347
	1,105	1,090	628	588
180	439	572	315	258
	904	690	315	258
	1,092	904	439	347
	1,105	1,090	628	588

of temperature and voltage. It is quite clear that the actual experimental protocol involves sub-sampling. The actual experimental units are the test stands since the treatment combinations are applied to the stand, not to the individual capacitors. Each capacitor in a test stand receives the exact same combination of the two factors. Two capacitors within a stand cannot have different temperatures or voltages. As a result, the capacitors within a stand are correlated with each other.

Meeker and Escobar (1998) use these data to illustrate how to analyze a reliability experiment using regression. They treat each capacitor as independent, thus ignoring the fact that capacitors within cells are correlated. Meeker and Escobar analyze the experiment as if there are 64 experimental units when in fact there are only 8. As a result, they treat the data as if they came from a completely randomized design in the capacitors replicated a total of eight times. The problem is that the actual protocol is an unreplicated completely randomized design in the test stands.

5 Naive Two-Stage Analysis of Reliability Data with Sub-Sampling

Freeman and Vining (2010) propose a naive two-step approach to this problem that assumes:

- Lifetimes within a test stand follow the same Weibull distribution.
- The failure mechanism remains the same across the test stands.
- The impact of treatments is through the scale parameter.
- Test stands are independent and, given the scale parameter, the observations within a test stand are independent.
- The experimental variability between scale parameters is log-normal.

5.1 First Stage of the Naive Analysis

Let t_{ij} be the observed lifetime for the j th item within the i th test stand. The failure times follow a Weibull distribution within a test stand, therefore:

$$f(t_{ij}) = \frac{\beta}{\eta_i} \left(\frac{t_{ij}}{\eta_i}\right)^{\beta-1} e^{-\left(\frac{t_{ij}}{\eta_i}\right)^\beta},$$

where $\beta > 0$ is the constant shape parameter and η_i is the scale parameter for i th test stand. The likelihood for an individual test stand with right censoring present is:

$$\mathcal{L}(\beta, \mu_i) = \mathcal{C} \prod_{j=1}^n [f(t_{ij})]^{\delta_{ij}} [1 - F(t_{ij})]^{1-\delta_{ij}},$$

where $\delta_{ij} = 1$ if the item fails and $\delta_{ij} = 0$ if the item is censored. Again, \mathcal{C} is a constant dependent on the type of censoring but can be taken as $\mathcal{C} = 1$ when calculating maximum likelihood estimates. The joint log-likelihood for data with right censoring then becomes:

$$\ell(\beta, \mu_1, \dots, \mu_m) = \sum_{i=1}^m \sum_{j=1}^n \left(\delta_{ij} \log\left(\frac{\beta}{t_{ij}}\right) + \delta_{ij} z_{ij} - e^{z_{ij}} \right).$$

One then can find the MLEs for β and the η_i s by maximizing the joint likelihood function. Many standard statistical software packages such as Minitab and SAS-JMP do this analysis.

Meeker and Escobar (1998) show that the Weibull distribution meets the regularity conditions to derive the asymptotic variance–covariance matrix from the maximum likelihood estimates. The resulting estimated variance matrix for the maximum likelihood estimates is:

$$\hat{\Sigma}_{\hat{\theta}} = \begin{bmatrix} \widehat{\text{Var}}(\hat{\beta}) & \widehat{\text{Cov}}(\hat{\beta}, \hat{\mu}_1) & \dots & \widehat{\text{Cov}}(\hat{\beta}, \hat{\mu}_m) \\ \widehat{\text{Cov}}(\hat{\beta}, \hat{\mu}_1) & \widehat{\text{Var}}(\hat{\mu}_1) & & \vdots \\ \vdots & & \ddots & \widehat{\text{Cov}}(\hat{\mu}_{m-1}, \hat{\mu}_m) \\ \widehat{\text{Cov}}(\hat{\beta}, \hat{\mu}_m) & \dots & \widehat{\text{Cov}}(\hat{\mu}_m, \hat{\mu}_{m-1}) & \widehat{\text{Var}}(\hat{\mu}_m) \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} -\frac{\partial^2 \ell(\beta, \mu_1, \dots, \mu_m)}{\partial \beta^2} & -\frac{\partial^2 \ell(\beta, \mu_1, \dots, \mu_m)}{\partial \beta \partial \mu_1} & \dots & -\frac{\partial^2 \ell(\beta, \mu_1, \dots, \mu_m)}{\partial \beta \partial \mu_m} \\ -\frac{\partial^2 \ell(\beta, \mu_1, \dots, \mu_m)}{\partial \beta \partial \mu_1} & -\frac{\partial^2 \ell(\beta, \mu_1, \dots, \mu_m)}{\partial \mu_1^2} & & \vdots \\ \vdots & & \ddots & -\frac{\partial^2 \ell(\beta, \mu_1, \dots, \mu_m)}{\partial \mu_{m-1} \partial \mu_m} \\ -\frac{\partial^2 \ell(\beta, \mu_1, \dots, \mu_m)}{\partial \beta \partial \mu_m} & \dots & -\frac{\partial^2 \ell(\beta, \mu_1, \dots, \mu_m)}{\partial \mu_m \partial \mu_{m-1}} & -\frac{\partial^2 \ell(\beta, \mu_1, \dots, \mu_m)}{\partial \mu_m^2} \end{bmatrix}^{-1}.$$

From the log-likelihood one can establish:

$$\begin{aligned}
 -\frac{\partial^2 \ell(\beta, \mu_1, \dots, \mu_m)}{\partial \beta^2} &= \sum_{i=1}^m \sum_{j=1}^n \left[\frac{\delta_{ij}}{\beta^2} + \left(\frac{z_{ij}}{\beta} \right)^2 \exp(z_{ij}) \right] \\
 -\frac{\partial^2 \ell(\beta, \mu_1, \dots, \mu_m)}{\partial \beta \partial \mu_i} &= -\sum_{j=1}^n \left(2 \frac{z_{ij}}{\beta} \exp(z_{ij}) \right) \\
 -\frac{\partial^2 \ell(\beta, \mu_1, \dots, \mu_m)}{\partial \mu_i^2} &= \sum_{j=1}^n (\beta^2 \exp(z_{ij})).
 \end{aligned}$$

Additionally, the second derivatives between all pairs of μ_i and μ_j are zero. This variance matrix will be used in the second stage of the model.

5.2 The Second Stage: The Model Between Experimental Units

This step uses the estimates for the shape parameter and the scale parameters and their corresponding variances for each experimental unit to model the estimates of the scale parameters as a linear function of the factors. The most appropriate way to estimate the model takes into account the variances on the scale parameter estimates through weighted least squares. In this case, the second stage model is:

$$\hat{\boldsymbol{\mu}} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon},$$

where \mathbf{X} is the matrix containing the treatment levels of the factors, $\boldsymbol{\theta}$ is the vector of the corresponding regression coefficients, and $\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \mathbf{V})$. The variance matrix, \mathbf{V} , accounts for the scale parameter variance estimates. Since the covariances are essentially 0, we can simplify the analysis by assuming that \mathbf{V} is diagonal with the non-zero elements simply being $\langle \widehat{\text{Var}}(\hat{\mu}_i) \rangle$. The resulting parameter estimates are:

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \hat{\boldsymbol{\mu}}.$$

The big advantage to this approach is that one can correctly model the experimental error in current statistical packages that have the ability to fit lifetime distributions and linear models.

The key to this analysis is a proper understanding of how one can deal with sub-sampling under normal theory. Once again, the observations within the experimental unit are correlated. However, one can take into proper consideration the correlation by replacing the individual observations within the experimental unit by their average, as long as each experimental unit has the same number of observations. The

Table 2 Stage one analysis results

Voltage	Temperature	$\hat{\eta}$	$\hat{\mu}_i = \log(\hat{\eta}_i)$	$\widehat{\text{Var}}(\hat{\mu}_i)$
200	170	1,262.35	7.141	0.1387
200	180	1,292.78	7.165	0.1390
250	170	1,207.58	7.096	0.1386
250	180	532.85	6.278	0.1387
300	170	683.61	6.527	0.1385
300	180	431.04	6.066	0.1388
350	170	633.86	6.452	0.1384
350	180	510.10	6.235	0.1386

Table 3 Analysis from minitab, stage two analysis result

Predictor	Coefficient	Standard error	T	p
Constant	14.613	3.249	4.50	0.056
Voltage	-0.005638	0.001644	-3.43	0.019
Temperature	-0.03682	0.01838	-2.00	0.102

proposed two-stage analysis extends this basic idea to the sub-sampling situation with Weibull data. The first step uses the Weibull distribution to estimate the common shape parameter and the different scale parameters, one for each test stand. The second step models the log transform of the different scale parameters using weighted least squares where the experimental error terms are given by the asymptotic experimental error derived in the first step for the log-scale parameters.

Table 2 presents the results from Minitab estimating the eight different scale parameter assuming constant β . The estimate of the shape parameter is $\hat{\beta}_{\text{New}} = 3.62$, which is a dramatically different estimate from the shape parameter estimate in the traditional reliability analysis, which is $\hat{\beta}_{\text{Trad.}} = 2.75$. This difference in the shape parameter estimate is the first practical implication of taking the experimental design into account.

The second step of our proposed new analysis models the resulting maximum likelihood estimates for the μ_i 's using a weighted regression linear model where the weights are determined by the asymptotic variances from the first step of this model. The second stage of this analysis can be done in any standard statistical package. Note that the variance estimates on the different μ_i are essentially equal in Table 2. This is a nice result because in the second stage of the model using a weighted regression is essentially equivalent to standard least squares regression further simplifying this two-stage analysis method. This is the case because we have assumed a constant shape parameter, β and the shape parameter is the driving parameter in the Fisher Information matrix calculations for the variances on the scale parameters. The results from running the analysis in Minitab are displayed in Table 3. Table 4 gives the analysis from Meeker and Escobar (1998), which assumes that all the capacitors are independent.

Several practical differences emerge comparing the results of the new analysis back to the traditional analysis. First, the Meeker and Escobar analysis overstates the true experimental degrees of freedom by treating each observation as an independent

Table 4 Analysis from Meeker and Escobar

Predictor	Coefficient	Standard error	Z	p
Intercept	13.4070	2.29584	5.84	0.000
Voltage	-0.0059	0.0010398	-5.68	0.000
Temperature	-0.0289	0.0129	-2.24	0.0250
Weibull shape	2.74869	0.418739		

data point. As a result, their standard errors of the coefficients are all smaller. The increase in standard error results in the temperature not being a significant factor at $\alpha = 0.05$ level for the new analysis. Additionally, the estimates of the shape parameter are dramatically different between the two analysis methods. The coefficient estimates for the linear relationship between the log scale parameter and temperature and pressure are also slightly different.

6 Joint Likelihood Approach

Kensler (2012) performed a simulation study comparing the two stage approach to the Meeker and Escobar approach. The type I error rate for the two stage was very close to the nominal. On the other hand, the Meeker and Escobar approach produced Type I error rates much higher than the nominal. In many cases, so high, in fact, as to invalidate the analysis.

A major problem with the naive two-stage analysis is that it cannot generate a joint likelihood for β and the coefficients for temperature and voltage. As a result, one cannot generate confidence or prediction intervals for such predicted values as percentiles, which often are of prime importance to a reliability analysis.

Freeman and Vining (2013) propose a joint likelihood approach that requires a variance component to take into proper account the test stand-to-test stand variability. If we have $i = 1, \dots, m$ independent experimental units and $j = 1, \dots, n_i$ sub-samples or observational units per experimental unit, one can specify the nonlinear mixed model for the Weibull distribution with sub-sampling as:

$$\begin{aligned}
 t_{ij}|u_i &\sim \text{Indep. Weib}(\beta, \eta_i) \\
 F_1(t_{ij}|\beta, \eta_i, u_i) &= 1 - \exp\left[-\left(\frac{t_{ij}}{\eta_i}\right)^\beta\right] \\
 \log(\eta_i) &= \mu_i = \mathbf{x}_i^T \boldsymbol{\theta} + u_i \\
 f_2(u_i) &\sim iidN(0, \sigma_u^2),
 \end{aligned}$$

where \mathbf{x}_i is the $p \times 1$ vector of fixed factor levels, $\boldsymbol{\theta}$ is the vector of fixed effect coefficients, and u_i are $i = 1, \dots, m$ independent random effects. Since the random

effects are independent, we can write the likelihood as:

$$\mathcal{L}(\beta, \theta | \text{Data}) = \prod_{i=1}^m \int_{-\infty}^{\infty} \left[\prod_{j=1}^{n_i} [f_1(t_{ij}|u_i)]^{\delta_{ij}} [1 - F_1(t_{ij}|u_i)]^{1-\delta_{ij}} f_2(u_i) \right] du_i,$$

where $f_1(t_{ij}|u_i)$ is the Weibull PDF for the data within an experimental unit and $f_2(u_i)$ is the normal PDF for the random effect.

Random effects models, especially nonlinear models, pose computational issues since it is necessary to integrate over the random effect u_i to maximize the likelihood. Gauss-Hermite (G-H) quadrature is an especially effective technique when the random effect follows a normal distribution. Quadrature involves weighting the sum of a function values at specific points over the domain of integration. G-H quadrature uses the roots of the Hermite polynomials to provide these specific points. G-H quadrature requires the random effect to have the form e^{-x^2} . As a result, a change of variables is necessary to apply G-H quadrature to our likelihood function. Let $u_i = \sqrt{2}\sigma_u v_i$, then the likelihood before the change of variables is:

$$\mathcal{L}(\beta, \theta | \text{Data}) = \prod_{i=1}^m \int_{-\infty}^{\infty} \left[\prod_{j=1}^{n_i} g(t_{ij}|u_i) \frac{e^{-\frac{u_i^2}{2\sigma_u^2}}}{\sqrt{2\pi\sigma_u^2}} \right] du_i,$$

where $g(t_{ij}|u_i) = [f_1(t_{ij}|u_i)]^{\delta_{ij}} [1 - F_1(t_{ij}|u_i)]^{1-\delta_{ij}}$ for right censored data. Executing the change in variables results in the following likelihood:

$$\mathcal{L}(\beta, \theta | \text{Data}) = \prod_{i=1}^m \int_{-\infty}^{\infty} \left[\prod_{j=1}^{n_i} g(t_{ij}|\sqrt{2}\sigma_u v_i) \frac{e^{-v_i^2}}{\sqrt{\pi}} \right] dv_i$$

G-H quadrature results in the following approximation of the likelihood:

$$\mathcal{L}(\beta, \theta | \text{Data}) \approx \prod_{i=1}^m \frac{1}{\sqrt{\pi}} \left\{ \sum_{k=1}^{n_k} \left[\prod_{j=1}^{n_i} g(t_{ij}|\sqrt{2}\sigma_u q_k) w_k \right] \right\},$$

where n_k is the number of quadrature points, q_k are the evaluation points found from the roots of the Hermite polynomials, and w_k are the corresponding weights to the evaluation points:

$$w_k = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [H_{n-1}(q_k)]^2}.$$

A common recommendation for the number of quadrature points to minimize bias is 20 points. Pinheiro and Bates (1995) show that G-H quadrature with 100 points

is as good as any other solution they investigated to the numerical optimization problem. In this research, we use 20 quadrature points in all of our analyses, unless otherwise stated, as an appropriate compromise on computation time, especially in the simulation studies. The log-likelihood is:

$$\ell(\beta, \theta | \text{Data}) \approx \sum_{i=1}^m \log \left(\frac{1}{\sqrt{\pi}} \sum_{k=1}^{n_k} \left[\prod_{j=1}^{n_i} g(t_{ij} | \sqrt{2}\sigma_u q_{k_i}) w_k \right] \right).$$

The approximate log-likelihood is maximized through standard maximization techniques.

A major advantage of using G-H quadrature is that it results in a closed-form approximate log-likelihood which allows for one to derive a Hessian matrix and the corresponding asymptotic covariance matrix. Maximum likelihood theory states that under certain regularity conditions that $\sqrt{(n)}(\hat{\theta} - \theta)$ converges in distribution to a multivariate normal. Let $\theta^{*T} = [\beta, \theta]$, then $\hat{\theta}^* \sim$ Asymptotically $MVN(\theta^*, I(\theta^*)^{-1})$, where

$$I(\theta^*) = \begin{bmatrix} -\frac{\partial^2 \ell(\beta, \theta_1, \dots, \theta_p)}{\partial \beta^2} & -\frac{\partial^2 \ell(\beta, \theta_1, \dots, \theta_p)}{\partial \beta \partial \theta_1} & \dots & -\frac{\partial^2 \ell(\beta, \theta_1, \dots, \theta_p)}{\partial \beta \partial \theta_p} \\ -\frac{\partial^2 \ell(\beta, \theta_1, \dots, \theta_p)}{\partial \beta \partial \theta_1} & -\frac{\partial^2 \ell(\beta, \theta_1, \dots, \theta_p)}{\partial \theta_1^2} & & \vdots \\ \vdots & & \ddots & -\frac{\partial^2 \ell(\beta, \theta_1, \dots, \theta_p)}{\partial \theta_{p-1} \partial \theta_p} \\ -\frac{\partial^2 \ell(\beta, \theta_1, \dots, \theta_p)}{\partial \beta \partial \theta_p} & \dots & -\frac{\partial^2 \ell(\beta, \theta_1, \dots, \theta_p)}{\partial \theta_p \partial \theta_{p-1}} & -\frac{\partial^2 \ell(\beta, \theta_1, \dots, \theta_p)}{\partial \theta_p^2} \end{bmatrix}.$$

The estimated covariance for the parameter estimates can be found by substituting the MLEs for the parameters they estimate into the information matrix, $I(\theta^*)$. Meeker and Escobar (1998, p. 622) note that the regularity conditions hold for the Weibull distribution. See Freeman (2010) for the derivation of the information matrix for the random effects Weibull model. Alternatively, the standard errors for the model parameters could be calculated through a bootstrapping procedure.

Table 5 summarizes the analysis of the Zelen data using the joint likelihood approach. The standard errors for the intercept, voltage, and temperature are similar to those from the naive two-stage analysis. However, the estimate of the shape parameter from the joint analysis is very similar to that from the Meeker and Escobar analysis rather than the two-stage analysis, which is much higher. This result suggests that the two-stage method may be susceptible to bias in the estimation of the shape parameter.

Freeman and Vining (2013) perform a simulation study to investigate the properties of the joint likelihood analysis. Their basic conclusions are

- The joint likelihood approach results in Weibull shape parameter estimates that are robust to model misspecification and random effect variation.
- Weibull scale parameter estimates are consistently good for both the joint likelihood analysis and the Meeker and Escobar— independent analysis.

Table 5 Joint likelihood analysis of the Zelen data

Parameter	Estimate	Standard error	T	p -value
Weibull shape	2.7753	0.6622	4.19	0.0041
Intercept	13.5257	3.0636	4.42	0.0031
Voltage	-0.00589	0.001154	-5.10	0.0014
Temperature	-0.02964	0.01808	-1.64	0.1451
Log(σ_μ)	-3.0184	9.0655	-0.33	0.7489

- The joint likelihood approach poorly estimates the true value of σ_μ , primarily due to the small number of degrees of freedom available in realistic reliability experiments to estimate this error term.
- The two-stage analysis provides a ready solution for practitioners with sub-sampling data.
- The joint likelihood approach provides a comprehensive solution for analyzing data from life test designs that are not completely randomized.
- Both analyses provide a motivation for thinking about design in a reliability context more comprehensively.

7 Extensions to Random Blocks with Sub-sampling

Kensler et al. (2014) extend the two-stage approach to the situation where we have test stands within random blocks. Like Freeman and Vining (2010), Kensler et al. estimate the log scale parameter for each test stand assuming a constant shape parameter across all the test stands in the first stage. They then perform a standard random block analysis using the estimated log scale parameters as the response. The model for the second stage analysis is

$$\hat{\mu} = X\boldsymbol{\gamma} + Z\boldsymbol{\rho} + \boldsymbol{\omega},$$

where $\boldsymbol{\mu}$ is the vector of estimated log-scale parameters from the first stage, X is the model matrix for the treatment effects, $\boldsymbol{\gamma}$ is the vector of model coefficients, Z is the incidence matrix for the blocks, $\boldsymbol{\rho}$ is the vector of random block effects, and $\boldsymbol{\omega}$ is the vector of random test stand errors. The second stage analysis assumes that the ρ 's are independent and normally distributed with a mean of 0 and a variance of σ_ρ^2 , that the ω 's are independent and normally distributed with a mean of 0 and a variance of σ_ω^2 , and that the ρ 's and the ω 's are independent.

They adapted the battery data from Montgomery (2005, p. 165) as the basis for their illustrative example. The experimental objective was to determine the effect of operating temperature on battery life. The batteries came from three batches, assumed to be random. Each test stand had eight batteries. The researchers used Type II censoring after the fourth failure.

Table 6 The battery life data (in hours)

Batch	Temperature		
	15	70	125
1	74	34	20
	130	40	58
	155	75	80
	180	80	82
2	126	106	25
	150	115	45
	159	122	58
	188	136	70
3	110	120	25
	138	139	82
	160	150	96
	168	174	104

Table 7 The first stage analysis of the battery life data

Batch	Temperature	η_{ij}	$\mu_{ij} = \log(\eta_{ij})$	$\widehat{\text{Var}}(\mu_{ij})$
1	15	197.79	5.287	0.0158
1	70	87.98	4.477	0.0158
1	125	89.85	4.498	0.0158
2	15	208.88	5.342	0.0160
2	70	153.56	5.034	0.0160
2	125	76.28	4.334	0.0158
3	15	189.22	5.243	0.0160
3	70	193.85	5.267	0.0160
3	125	116.05	4.754	0.0159

Table 8 The second stage analysis of the battery life data

Source	df	MS	F	p-value
Temperature	1	0.8704	16.96	0.009
Block	2	0.0839	1.63	0.284
Residual	5	0.05132		

Table 6 summarizes the battery data. Table 7 summarizes the estimates of the scale parameters, the log scale parameters, and the variances of the log scale parameters. The estimate of the shape parameter, β is 4.03, which indicates a wear-out failure mode. Once again, the estimated variances for the estimated log scale parameters are essentially constant like the example in Freeman and Vining (2010). As a result, Kensler et al. use standard ordinary least squares in stage two of their analysis. Table 8 summarizes the second stage regression analysis from Minitab. The estimates of the two variance components are $\sigma_{\omega}^2 = 0.05132$ and $\sigma_{\omega}^2 = 0.01086$. The second stage analysis does show that the temperature does have an important effect on the battery life.

Kensler et al. (2014) performed a simulation study to examine the performance of the two-stage method. They found that the two-stage method did an excellent job preserving the Type I error rate. They also found that power for the two-stage method was close to the nominal for larger β and number of failures; however, the actual power was less than the nominal for small β 's and only a few failures. Just as in the Freeman and Vining (2010) paper, the estimates for β from the first stage tend to be biased.

8 Conclusions and Future Research

The basic conclusions reached so far by this research are:

- The two-stage approach provides a straightforward basis for analyzing reliability experiments with sub-sampling that practitioners can apply with current standard statistical software.
- The two-stage approach does a good job preserving the nominal Type I error rates and a much better job than the traditional analyses.
- The two-stage approach does produce biased estimates of β .
- The two-stage approach does not allow the analyst to compute confidence or prediction intervals around predicted values.
- The joint likelihood approach has much less bias in its estimates of β .
- The joint likelihood approach does allow analysts to generate appropriate confidence and prediction intervals.

Future research includes

- submitting the paper that uses the joint likelihood approach for the random blocks case.
- extending the two-stage and joint likelihood approaches to split-plot experiments.

References

- Freeman, L. J. (2010). *Statistical methods for reliability data from designed experiments*. Unpublished Ph.D. dissertation, Virginia Tech, Department of Statistics.
- Freeman, L. J., & Vining, G. G. (2010). Reliability data analysis for life test experiments with subsampling. *Journal of Quality Technology*, 42(3), 233–241.
- Freeman, L. J., & Vining, G. G. (2013). Reliability data analysis for life test designed experiments with subsampling. *Quality and Reliability Engineering International*, 29, 509–519.
- Kensler, J. L. K. (2012). *Analyzing reliability experiments with random blocks and subsampling*. Unpublished Ph.D. dissertation, Virginia Tech, Department of Statistics.
- Kensler, J. L. K., Freeman, L. J., & Vining, G. G. (2014). A practitioner's guide to analyzing reliability experiments with random blocks and subsampling. *Quality Engineering*, 26(3), 359–369.

- Meeker, W. Q., & Escobar, L. A. (1998). *Statistical methods for reliability data*. Hoboken, NJ: Wiley.
- Montgomery, D. C. (2005). *Design and analysis of experiments*. Hoboken, NJ: Wiley.
- Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4(1), 12–35.
- Zelen, M. (1959). Factorial experiments in life testing. *Technometrics*, 1(3), 269–288.