***Understanding Weibull Regression***

***Lecture 2: Power Law Models Viewed as Regression***

**Basic Power Law Model**

The basic power law model is fundamental to accelerated life testing. For the vast majority of products that require life testing, the time to failure at normal use conditions requires far more time than the testing constraints allow.

For example, carbon overwrapped pressure vessels (COPVs) are extremely important to NASA. They are vital components to all spacecraft. Typical expected mission life is ten – fifteen years. Of course, surviving mission life requires that none of the critical components fail. As a result, NASA requires an extremely high probability that each critical component can survive that period of time at the nominal use conditions.

The basic power law allows the engineers to test the critical component at much higher “stress” levels and then to extrapolate back to the use condition. There are serious consequences to this strategy, not the least of which is that the higher stress levels used in the test may create new failure mechanisms than the actual failure mechanism for the use condition. Engineers proceed under the assumption that the results of the power law model based test are “conservative.” By conservative the engineers mean that the true probability of surviving the expected mission life at the use conditions is actually greater than the test suggests. Thus, engineers believe that the power law model based probability of survival represents, in a meaningful sense, a lower bound on the probability of survival.

The classic power law model assumes that the stress applied to the test item is constant throughout the test, which this lecture assumes. Lecture 3 will discuss some alternatives.

Let be this constant stress applied to the test item, and let be the “reference” stress level. In many applications, is the stress associated with the use condition, which makes the extrapolation simpler. As we will see in this lecture’s example, NASA does not follow this convention. The key point is that the analyst uses to standardize the load. The stress ratio for the item tested, , is

which is dimensionless.

For convenience, let be the characteristic of interest. In our NASA example, is the time to failure for the test item whose constant stress ratio is . The power law model then is

The model is a power law in terms of the parameter . The resulting Weibull survivor function is

Many engineers call this relationship the classic engineering stress rupture model. By definition, a stress rupture test models the time to failure at a constant stress. The only parameter impacted by the choice of is . The maximum likelihood estimates for and are unaffected.

For many reliability engineers, Equation (1) defines their universe. An unfortunate consequence is that they then cannot see the relationship with standard regression models. In many cases, they have no concept that they are even remotely related, much less very intimately related.

Recall that the SEV parameter, . As a result, we can re-express the survivor function as

where is the residual term: the observed value, , minus its predicted value, . We next note that

Let . As a result,

is nothing more than the deterministic portion of the simple linear regression model where is the intercept, is the slope, and is the regressor. The parameters and define “location” given the specific value for . The parameter controls the “spread” given the location. Figure 1 illustrates the basic structure.

It is important to note that with this basic understanding we can create more complex models to explain the stress rupture data. This insight is important for understanding Lecture 3. A critical point, however, is that more complex models must respect the basic Weibull assumptions that the must be a free parameter and does not depend on the ’s and that the ’s must not depend on . This issue is extremely important for understanding the context of Lecture 3.



Figure 1: Illustration of the Basic Stress Rupture Model Parameters

**Estimation, Testing, and Model Selection**

The generally accepted method for estimating the model parameters is maximum likelihood. The maximum likelihood estimation of the three parameter classic stress rupture model is a straight-forward extension of the two-parameter case discussed in the first lecture. Expressing the likelihood function in terms of the scaled residuals greatly simplifies the derivatives.

Often, engineers assume that the classic stress rupture model is the true model. As a result, they never perform formal testing. Of necessity, and must be positive. The case where implies that the stress levels applied to the test items have no effect.

Many engineers have no idea that the classic stress rupture model is purely empirical in nature. There is no fundamental first principles basis for this model. Yet, many engineers take as an absolute and unquestionable fact that this model represents the “science.” Once again, this becomes an extremely important issue for Lecture 3.

We illustrate with the NASA example later in this lecture, that there always are alternative models that one could estimate. The key then is to understand the actual “mother” model within which the classic stress rupture model is nested. In the process, the analysts can discover whether there are terms that should be in the model but are not represented by the classic stress rupture model. It also sets the stage for evaluating other “interesting” models that some engineers declare represent the true science.

*Models Nested within a “Mother,” Formal Testing, and Model Selection*

The example that concludes this lecture involves three different stress ratios. We illustrate what we mean by the mother model within this context. In this example, we have both exact failures and right censored test items within each stress ratio. As a result, we can estimate separate two-parameter Weibull models for each stress ratio.

Consider a specific stress ratio. In our example, the three stress ratios are 0.80, 0.85, and 0.90. The experimental protocol is a completely randomized design where the test items are randomly allocated to the different stress ratios. The mother model assumes that each stress ratio has its own location parameter, . The mother model also assumes that each stress ratio has a different value for . The resulting overall model consists of three subset models whose parameters are:

 : and

 : and

 : and

The classic stress rupture model assumes:

* (a common )
*

The mother model require six distinct parameters; the classic stress rupture model requires only three.

This lecture uses Minitab to analyze the three models required for the mother as well as the single classic stress rupture model to generate the proper “raw material” (parameter estimates and the log-likelihood information) for the appropriate analyses.

We can test whether we need one or more of the extra three parameters required by the mother through a classic test based on the proper log-likelihood statistic. Let denote the likelihood function for the mother (full) model, and let denote the likelihood function for the classic stress rupture model. The experimental protocol allows us to calculate by summing together the individual log-likelihoods for three two-parameter Weibull models for each stress ratio. The first lecture illustrated how to perform each of these individual analyses. Asymptotically,

where in this case the degrees of freedom, . We conclude that at least one of the additional three parameters is important if the proper -value is sufficiently small. Otherwise, we conclude that the appropriate model is the classic stress rupture model by model parsimony.

The Akaike Information Criterion (), along with variants such as the Bayesian Information Criteria () provide additional insight about the relative predictive capability of a set of nested models for the observed data. These criteria use the log-likelihood as their basis and add penalties for the number of parameters required by the model. Proper analysis should include at least one of these criteria along with the classic test.

**Extended Residual Plots**

The clearest way to see the dependence of probability plots on the residuals for Weibull distributed data is through the smallest extreme value (SEV) representation. Let be the observed characteristic of interest with a Weibull scale parameter, , and a Weibull shape parameter, . The SEV representation of the Weibull distribution observes that follows a SEV distribution with a log-location parameter, , defined by

In this context, the raw residual is

and the scaled residual, , is

where the SEV scale parameter, .

An examination of the derivatives of the log-likelihood function required for the maximum likelihood estimation of the model clearly reveals that the estimation depends upon these residuals, not the in isolation from the model. The resulting Survivor function is

These concepts extend naturally to more complicated models. Consider the classic stress rupture model, which is an example of a power law model. The survivor function for this model is

where is a reference time, controls the relationship between the failure time and the stress ratio (), is the Weibull shape parameter, and . The expression involving reflects the smallest extreme value (SEV) reparameterization of the model. For this model,

Thus, the scaled residuals are

The scaled residuals are the proper predictions of the log probability for the specific observation. These scaled residuals are the formal basis for constructing the proper residual plots.

Meeker and Escobar (1998, pages 437-438) illustrate how to construct confidence intervals for specific values of the explanatory variable (in our case, ). Their approach requires the SEV parameterization of the classic Weibull stress rupture model. The SEV representation of the classic stress rupture model uses:

Let and be the maximum likelihood estimates of the classic Weibull stress rupture model. It follows that is a maximum likelihood estimate for which has a variance of

It also follows that is a maximum likelihood estimate of ; however, the proper variance of the maximum likelihood estimate of is not

As a result, we cannot use the Hessian from the classic Weibull stress rupture model for the appropriate variance of , the maximum likelihood estimate of . The proper variance requires the use of the SEV parameterization of the model. Consequently, it is not a simple procedure to use the Hessian from the classic Weibull stress rupture model to create proper confidence bands. The issue goes back to the use of the large sample property of maximum likelihood estimates following a normal distribution.

Fortunately, there is a very easy way to create the confidence bands for probability plots based on the scaled residuals. If the ’s follow a Weibull distribution, then the scaled residuals follow an SEV distribution which is a simple two-parameter model. Generally, this distribution has and . The key is that the scaled residual directly accounts for the impact of the more complicated model, mapping the result to a two-parameter space.

Meeker and Escobar (1998, page 189) discuss how to construct confidence intervals appropriate for probability plots. They use the large sample property that maximum likelihood estimates follow a normal distribution. Let be the inverse cdf function. For a Weibull distribution, as we saw in the first lecture,

Meeker and Escobar use the SEV parameterization of the Weibull distribution:

This relationship forms a straight line with a -intercept of and a slope of . This line represents the maximum likelihood prediction of the log times associated with the percentiles. The appropriate estimate of the standard error for is

where , , and are estimated from the inverse of the Hessian for the SEV parameterization of the scaled residuals.

Constructing the appropriate Weibull probability plot requires ordering the scaled residuals, calculating the resulting median ranks, and scaling these ranks appropriately. The scaled ranks are the empirical probabilities. The usual scaling used in the software is Benard’s approximation

where is the median rank for the ordered scaled residual and is the total number of observations. Let denote the “probability plot residual” defined by

This “residual” treats the log empirical probability as the observation, while the scaled residual reflects the appropriate predicted value. If the data follow a Weibull distribution, then a plot of the log of the empirical probabilities (-axis) and the scaled residuals (-axis) should form a straight-line. For moderate to large sample sizes, approximately 95% of the observed failure times should fall within the interval

which are the appropriate confidence limits displayed on the Weibull probability plots.

Typically, the software only generates the Weibull probability plot. However, one can generate analogs to all of the residual plots commonly used in OLS regression. The proper analogs standardize the probability plot residuals by

which asymptotically follows a standard normal distribution. This property allows the analyst to judge the “size” of the residual appropriately. The suggested cut-off for investigating possible outliers is an absolute value of greater than 3 (not 2). This suggested cut-off value accounts for the number of observed residuals flagged as potential outliers due to random chance depends on the total number of observations. Formal determination that an observed residual is a truly potential outlier actually requires cut-off values larger than 3, in some cases, much larger.

Typical OLS plots, in addition to the probability plot, include residuals versus predicted value, residuals versus time order, and the residuals versus all regressors, for example, the stress ratios. The example illustrates how one can create residual plots that are analogs of those commonly used for OLS regression for the Phase B data. It is critical to note that essentially, they are all different plots of the same residuals used in the standard probability plots.

**NASA Example**

This analysis compares the classic stress rupture model, model (Model A), to fitting the proper mother model for this situation (Model B). NASA defines the stress ratio in terms of the characteristic strength from the tensile strength test, . Let be the target constant load for the test item. The corresponding stress ratio, , is

The target stress ratios for the experiment are and

The parameters for Model A are: and . The parameters for Model B are:, , , , , and . This comparison thus allows us to see the impact from fitting the classic stress rupture model to the raw data.

Fitting the separate individual models to each stress ratio is the largest Weibull model and of necessity has the largest log-likelihood. The fundamental issue is whether the small improvement in the log-likelihood justifies the extra parameters. This comparison illustrates the value of the scaled log-likelihood statistic and in model selection. It also illustrates the need for good residual analysis when making the final decisions in model selection. Table 1 summarizes the results.

Table 1: The Classic Stress Rupture Model to the Full Model for the Hold Data

Model: Model A Model B

Number of Parameters: 3 6

Number of Observations: 708 708

True Log-Likelihood:

Log-Likelihood Statistic: 612.822 611.800

: 618.822 623.800

The -value associated with the based on the difference in the log-likelihood statistics is , which indicates that the none of the three extra parameters in the full model is significant. In such a situation, the model with the fewer parameters is deemed best based on model parsimony. In addition, the classic stress rupture model actually provides better insights into the fundamental nature of the relationship of stress ratio and failure time under a static load. The indicates a very distinct advantage to classic model through its correction for the number of parameters. The bottom-line from these comparisons is that classic stress rupture model is the clearly preferred model.

Figures 2-6 are the proper residual plots for the classic stress rupture model using the R code provided in a separate attachment. The confidence bands use three standard errors, not two. The residual plots reveal no issues with the data.



Blue: SR=.80

Red: SR=.85

Orange: SR=.90

Black:

Figure 2: Weibull Probability Plot Using R Code



Figure 3: Standardized Residual versus Predicted Residual Plot



Figure 4: Standardized Residual versus Stress Ratio



Figure 5: Standardized Residuals versus Serial Number (Time Plot in Order of Production)