Chapter 566

Parametric Survival (Weibull) Regression

Introduction

This module fits the regression relationship between a positive-valued dependent variable (often time to failure) and one or more independent variables. The distribution of the residuals (errors) is assumed to follow the exponential, extreme value, logistic, log-logistic, lognormal, lognormal10, normal, or Weibull distribution. The data may include failed, left censored, right censored, and interval observations. This type of data often arises in the area of *accelerated life testing*.

When testing highly reliable components at normal stress levels, it may be difficult to obtain a reasonable amount of failure data in a short period of time. For this reason, tests are conducted at higher than expected stress levels. The models that predict failure rates at normal stress levels from test data on items that fail at high stress levels are called *acceleration models*.

The basic assumption of acceleration models is that failures happen faster at higher stress levels. That is, the failure mechanism is the same, but the time scale has been changed (shortened).

Technical Details

The linear regression equation is

$$Y = B_0 + B_1 X_1 + B_2 X_2 + \cdots + Se$$

Here, S represents the value of a constant standard deviation, Y is a transformation of time (either *ln(t), log(t),* or just *t*), the *X's* are one or more independent variables, the *B*'s are the regression coefficients, and *e* is the residual (error) that is assumed to follow a particular probability distribution. The problem reduces to estimating the *B*'s and *S*. The density functions of the eight distributions that are fit by this module were given in the Distribution Fitting section and will not be repeated here.

So that you can get the general idea, we will give detailed results for the lognormal distribution. The results for other distributions follow a similar pattern.

The lognormal probability density function may be written as

$$f(t \mid M, S) = \frac{1}{tS\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(t) - M}{S}\right)^2}$$

If we replace the location parameter, *M*, with the regression model, the density now becomes

$$f(t \mid B_0 \cdots B_p, S) = \frac{1}{tS\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{\ln(t) - \sum_{i=0}^{p} B_i X_i}{S} \right)^2 \right\}$$

Maximum likelihood estimation consists of finding the values of the distribution parameters that maximize the log-likelihood of the data values. Loosely speaking, these are the values of the parameters which maximize the probability that the current set of data values occur.

The general form of the log-likelihood function is given by

$$L(\underline{P}) = \sum_F \ln\left(f(\underline{P}, t_k)\right) + \sum_R \ln\left(S(\underline{P}, t_k)\right) + \sum_L \ln\left(F(\underline{P}, t_k)\right) + \sum_I \ln\left(f(\underline{P}, t_{uk}) - f(\underline{P}, t_{lk})\right)$$

where *F* represents the set of failed items, *R* represents the set of right censored items, *L* represents the set of left censored items, and *I* represents the set of interval censored items. In the case of interval censored observations, t_{lk} represents the first time of the interval and t_{uk} represents the last time of the interval. Also, $\underline{P}$ represents the parameters, including *S* and the *B's*.

We employ the Newton-Raphson algorithm with numerical differentiation to obtain the maximum likelihood estimates. These estimates have been shown to have optimality characteristics in large samples (number of failures greater than 20). They have been shown to be competitive estimates even for sample sizes less than 20.

Data Structure

Survival data are somewhat more difficult to enter because of the presence of various types of censoring.

Time Variable(s)

One (or two in the case of interval data) variable is needed to contain the time values.

Censor Variable

Another variable is needed to indicate the type of censoring.

Failed or Complete

A failed observation is one in which the time until the terminal event was measured exactly; for example, the machine stopped working or the mouse died of the disease being studied.

Right Censored

A right censored observation provides a lower bound for the actual failure time. All that is known is that the failure occurred (or will occur) at some point after the given time value. Right censored observations occur when a study is terminated before all items have failed. They also occur when an item fails due to an event other than the one of interest.

Left Censored

A left censored observation provides an upper bound for the actual failure time. All we know is that the failure occurred at some point before the time value. Left censoring occurs when the items are not checked for failure until some time after the study has begun. When a failed item is found, we do not know exactly when it failed, only that it was at some point before the left censor time.

Interval Censored or Readout

An interval censored observation is one in which we know that the failure occurred between two time values, but we do not know exactly when. This type of data is often called *readout* data. It occurs in situations where items are checked periodically for failures.

Independent Variable(s)

One or more independent variables must be supplied also.

Sample Data

The following data, found in Nelson (1990), are quoted in many books and articles on accelerated testing. These data come from a temperature-accelerated life test of a Class-B insulation for electric motors. Ten motorettes were tested at each of four temperatures. When the testing was stopped, the following failure times were recorded. These data are stored in the Motors dataset.

Motors dataset

Hours	Censor	Count	Temperature
	1	1	130
8064	0	10	150
1764	1	1	170
2772	1	1	170
3444	1	1	170
3542	1	1	170
3780	1	1	170
4860	1	1	170
5196	1	1	170
5448	0	3	170
408	1	2	190
1344	1	2	190
1440	1	1	190
1680	0	5	190
408	1	2	220
504	1	3	220
528	0	5	220

Procedure Options

This section describes the options available in this procedure.

Variables Tab

This panel specifies the probability distribution that is fit and the variables used in the analysis.

Time Variables

Y: Time Variable

This variable contains the time values for both failed and censored observations. When interval (readout) data are used, this variable specifies the ending time of the interval.

Negative time values are treated as missing values. Zero time values are not allowed. They may be automatically replaced by the value in the Zero field.

These time values represent elapsed times. If your dataset is made up of dates (such as the failure date), you must subtract the starting date from the failure date so that you can analyze the elapsed time.

Start Time Variable

This variable contains the starting time for interval (readout) data. Hence its value is only used when the row's censor value indicates an interval data type.

Negative time values are treated as missing values. Rows with zero starting time values are reset to left censored type data.

Zero Time Replacement Value

Under normal conditions, a respondent beginning the study is "alive" and cannot "die" until after some small period of time has elapsed. Hence, a time value of zero is not defined and is ignored (treated as a missing value). If a zero time value does occur, it is replaced by this positive amount. If you do not want zero time values replaced, enter a "0.0" here.

This option is used when a "zero" on the database does not actually mean zero time. Instead, it means that the response occurred before the first reading was made and so the actual survival time is only known to be less.

Censor Variable

Censor Variable

The values in this optional variable indicate the type of censoring active for each row. Four possible data types may be entered: failed (complete), right censored, left censored, or interval. The values used to indicate each data type are specified in the four boxes to the right. These values may be text or numeric.

Failed

When this value is entered in the Censor Variable, the corresponding time value is treated as a failed observation. The value may be a number or a letter. We suggest the letter "F" when you are in doubt as to what to use.

A failed observation is one in which the time until the event of interest was measured exactly; for example, the machine stopped working or the mouse died of the disease being studied. The exact failure time is known.

Right

When this value is entered in the Censor Variable, the corresponding time value is treated as a right censored data value. The value may be a number or a letter. We suggest the letter "R" when you are in doubt as to what to use.

A right censored observation provides a lower bound for the actual failure time. All that is known is that the failure time occurred (or will occur) at some point after the given time value. Right censored observations occur when a study is terminated before all items have failed. They also occur when an item fails due to an event other than the one of interest.

Left

When this value is entered in the Censor Variable, the corresponding time value is treated as a left censored data value. The value may be a number or a letter. We suggest the letter "L" when you are in doubt as to what to use.

A left censored observation provides an upper bound for the actual failure time. All we know is that the failure time occurred at some point before the time value. Left censoring occurs when the items are not checked until some time after the study has begun. When a failed item is found, we do not know exactly when it failed, only that it was at some point before the left censor time.

Interval

When this value is entered in the Censor Variable, the corresponding time value is treated as an interval censored data value. The value may be a number or a letter. We suggest the letter "I" when you are in doubt as to what to use. When interval censoring is specified, the program uses both the Time Variable and the Start Time Variable.

An interval censored observation is one in which we know that the failure occurred between the two time values, but we do not know exactly when. This type of data is often called *readout* data. It occurs in situations where items are checked periodically for failures.

Note that when interval data are obtained, the first observation is usually left censored and the last observation is usually right censored.

Independent Variables

X's: Independent Variables

Specify additional independent variables. You can leave this option blank or you can leave the Stress Variable blank, but you cannot leave both blank.

These variables may be thought of as additional variables for which statistical adjustment is desired. They can contain both discrete and continuous variables. If discrete variables are specified, it is up to you to specify the appropriate number of dummy variables. For example, suppose you have three suppliers. Since this has three possible values, two indicator variables will be needed to specify the appropriate information.

Stress Variable

Stress Variable

This variable contains the values of the independent variable that will be transformed according to either the Arrhenius or Power transformation. This variable is optional, although when it is not specified, several of the reports and graphs will not be displayed.

You can leave this option blank or you can leave the Independent Variables option blank, but you cannot leave both blank.

Stress A

Specify the value of A in the Arrhenius transformation $X=A/(Stress+B)$. If A is zero, the logarithm of stress is used and the model is X=log(stress).

Set A to 1000 for Arrhenius model. Set A to 0 for Power model.

Stress B

Specify B in the transformation $X=A/(Stress+B)$. Ignore it by setting B to zero.

For the Arrhenius model, B is set to 273.16 when stress is measured in degrees Celsius. Usually, A and B are set to convert temperature values to degrees Kelvin.

For the Power acceleration model, this value is ignored.

Frequency Variable

Frequency Variable

This variable gives the count, or frequency, of the time displayed on that row. When omitted, each row receives a frequency of one. Frequency values should be positive integers. This is usually used to indicate the number of right censored values at the end of a study or the number of failures occurring within an interval. It may also be used to indicate ties for failure data.

Probability Distribution

Distribution

This option specifies which probability distribution is fit. All results are for the specified probability distribution. Usually, you will use one of the distributions that is based on the logarithm of time such as the lognormal, Weibull, exponential, or log-logistic.

Alpha Level

Alpha Level

This is the value of alpha used in the calculation of confidence limits. For example, if you specify 0.04 here, then 96% confidence limits will be calculated.

Estimation Tab

The following options control the searching algorithms used during parameter estimation.

Estimation Options

Maximum Iterations

Many of the parameter estimation algorithms are iterative. This option assigns a maximum to the number of iterations used in any one algorithm. We suggest a value of at least 100. This should be large enough to let the algorithm converge, but small enough to avoid a large delay if convergence cannot be obtained. If the number of iterations reaches this amount, you should re-run your analysis with a larger value.

Minimum Relative Change

This value is used to control the iterative algorithms used in maximum likelihood estimation. When the relative change in all of the parameters is less than this amount, the iterative procedure is terminated.

Parameter Adjustment

Newton's method calculates a change for each parameter value at each step. Instead of taking the whole parameter change, this option lets you take only a fraction of the indicated change. For datasets that diverge, taking only partial steps may allow the algorithm to converge. In essence, the algorithm tends to over correct the parameter values. This factor allows you to dampen this over correction. We suggest a value of about 0.2. This may increase the number of iterations (and you will have to increase the Maximum Iterations accordingly), but it provides a greater likelihood that the algorithm will converge.

Starting Sigma

Specify a starting value for *S*, the standard deviation of the residuals (errors*)*. Select '0 - Data' to calculate an appropriate value from the data. If convergence fails, try a different value.

Derivatives

This value specifies the machine precision value used in calculating numerical derivatives. Slight adjustments to this value can change the accuracy of the numerical derivatives (which impacts the variance/covariance matrix estimation).

Remember from calculus that the derivative is the slope calculated at a point along the function. It is the limit found by calculating the slope between two points on the function curve that are very close together. Numerical differentiation mimics this limit by calculating the slope between two function points that are very close together and then computing the slope. This value controls how close together these two function points are.

Numerical analysis suggests that this distance should be proportional to the machine precision of the computer. We have found that our algorithm achieves four-place accuracy in the variance-covariance matrix no matter what

value is selected here (within reason). However, increasing or decreasing this value by two orders of magnitude may achieve six or seven place accuracy in the variance-covariance matrix. We have found no way to find the optimal value except trial and error.

Note that the parameter estimates do not seem to be influenced a great deal, only their standard errors.

Reports Tab

The following options control which reports are displayed and the format of those reports.

Select Reports

Data Summary Report ... Residual Report

Each of these options specifies whether the indicated report is calculated and displayed.

Percent Failing and Failure Time Percentile Reports

These options indicate whether to display the corresponding report and which rows are to be shown. Usually, you will add rows with missing time values at the end of the database to be reported on. The percent failing will then be estimated for those values.

You can choose to omit these reports entirely, show only those rows with missing time values, or show the results for all rows.

Report Options

Precision

Specify the precision of numbers in the report. A single-precision number will show seven-place accuracy, while a double-precision number will show thirteen-place accuracy. Note that the reports are formatted for single precision. If you select double precision, some numbers may run into others. Also, note that all calculations are performed in double precision regardless of which option you select here. Single precision is for reporting purposes only.

Variable Names

This option lets you select whether to display only variable names, variable labels, or both.

Percent Failing Report Calculation Times

This option specifies a list of times at which the percent failing is reported on the Percent Failing Section report. Individual values are separated by commas. You can specify a sequence by specifying the minimum and maximum separated by a colon and putting the increment inside parentheses. For example: 5:25(5) means 5,10,15,20,25. Avoid 0 and negative numbers. Use '(10)' alone to specify ten values between zero and the maximum value found in the data.

Note that each time is used for all selected observations.

Failure Time Report and Stress-Time Plot Percentiles

This option specifies a list of percentiles (range 1 to 99) at which the failure time is reported, one percentile per line on the Failure Time Percentile report. It also specifies which percentiles are shown on the Stress plot. The values should be separated by commas.

You can specify sequences with a colon, putting the increment inside parentheses after the maximum in the sequence. For example: 5:25(5) means 5,10,15,20,25 and 1:5(2),10:20(2) means 1,3,5,10,12,14,16,18,20. Note that this option also controls which percentiles are displayed on the Stress - Time plot.

Report Options – Decimal Places

Time ... Stress Decimals

This option specifies the number of decimal places shown on reported time, probability, and stress values.

Plots Tab

These options control the attributes of the plots.

Select Plots

Stress - Time Plot ... X - Residual Plots

Each of these options specifies whether the indicated plot is displayed. Click the plot format button to change the plot settings.

Number Stresses or Number Predicted

This options sets resolution of the plot along the horizontal axis. A value near 50 is usually adequate.

Example 1 – Lognormal Regression

This section presents an example of how to fit a lognormal regression. The data used were shown above and are found in the Motors dataset.

You may follow along here by making the appropriate entries or load the completed template **Example 1** by clicking on Open Example Template from the File menu of the Parametric Survival (Weibull) Regression window.

1   Open the Motors dataset.
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Motors.NCSS**.
- Click **Open**.

2   Open the Parametric Survival (Weibull) Regression window.
- Using the Analysis menu or the Procedure Navigator, find and select the **Parametric Survival (Weibull) Regression** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3   Specify the variables.
- On the Parametric Survival (Weibull) Regression window, select the **Variables tab**.
- Double-click in the **Y: Time Variable** box. This will bring up the variable selection window.
- Select **Hours** from the list of variables and then click **Ok**.
- Double-click in the **Censor Variable** box. This will bring up the variable selection window.
- Select **Censor** from the list of variables and then click **Ok**.
- Double-click in the **Stress Variable** box. This will bring up the variable selection window.
- Select **Temp** from the list of variables and then click **Ok**. Note that the default values of Stress A and Stress B are appropriate for this problem.
- Double-click in the **Frequency Variable** box. This will bring up the variable selection window.
- Select **Count** from the list of variables and then click **Ok**.
- Set the **Distribution** to **Log10normal**.

4   Specify the reports.
- On the Parametric Survival (Weibull) Regression window, select the **Reports tab**.
- Enter **10000:100000(10000)** in the **Percent Failing Report Calculation Times** box.
- On the **Plots** tab, **Stress – Time Plot**, set the **Y Axis Scale** to **Log**, with the **Format** set to **Powers**.

5   Run the procedure.
- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

Data Summary Section

Data Summary Section

Type of Observation	Rows	Count	Hours Minimum	Hours Maximum
Missing or Prediction	1			
Failed	12	17	408.0	5196.0
Right Censored	4	23	528.0	8064.0
Left Censored	0	0		
Interval Censored	0	0		
Total (Nonmissing)	16	40	408.0	8064.0

Means for Rows with Failures

Variable	Mean
Hours	1919.412
Temp	190.5882

This report displays a summary of the data that were analyzed. Scan this report to determine if there are any obvious data-entry errors by double-checking the counts and the minimum and maximum.

The means given for each variable are for the noncensored rows.

Parameter Estimation Section

Maximum Likelihood Parameter Estimation Section

Parameter Name	Parameter Estimate	Standard Error	Z Value	Prob Level	Lower 95.0% C.L.	Upper 95.0% C.L.
Intercept	-6.018404	0.9471051	-6.3545	0.000000	-7.874695	-4.162112
Temp	4.31048	0.4367811	9.8687	0.000000	3.454405	5.166555
Sigma	0.2591772	0.0472949	5.4800	0.000000	0.1812461	0.3706166

R-Squared	0.531405
Log Likelihood	-148.5373
Iterations	55

This report displays parameter estimates along with standard errors, significance tests, and confidence limits. Note that the significance levels and confidence limits all use large sample formulas. How large is a large sample? We suggest that you only use these results when the number of failed items is greater than twenty.

Parameter Estimates

These are the maximum likelihood estimates (MLE) of the parameters. They are the estimates that maximize the likelihood function. Details are found in Nelson (1990) pages 287 - 295.

Standard Error

The standard errors are the square roots of the diagonal elements of the estimated Variance Covariance matrix.

Z Value

The z value is equal to the parameter estimate divided by the estimated standard error. This ratio, for large samples, follows the normal distribution. It is used to test the hypothesis that the parameter value is zero. This value corresponds to the t value that is used in multiple regression.

Prob Level

This is the two-tailed p-value for testing the significance of the corresponding parameter. You would deem independent variables with small p-values (less than 0.05) important in the regression equation.

Upper and Lower 100(1-Alpha)% Confidence Limits

These are the lower and upper confidence limits for the corresponding parameters. They are large sample limits. They should be ignored when the number of failed items is less than thirty. For the regression coefficients B, the formulas are

$$CL_i = \hat{B}_i \pm z_{1-\alpha/2}\hat{\sigma}_{\hat{B}_i} \quad i = 0,\cdots,p$$

where $\hat{B}_i$ is the estimated regression coefficient, $\hat{\sigma}_{\hat{B}_i}$ is its standard error, and z is found from tables of the standard normal distribution.

For the estimate of sigma, the formula is

$$CL = \hat{S}\exp\left\{\frac{\pm z_{1-\alpha/2}\hat{\sigma}_{\hat{S}}}{\hat{S}}\right\}$$

R-Squared

R-Squared reflects the percent of variation in log(time) explained by the independent variables in the model. A value near zero indicates a complete lack of fit, while a value near one indicates a perfect fit.

Note that this R-Squared value is computed for the failed observations only. Censored observations are ignored.

Log Likelihood

This is the value of the log likelihood function. This is the value being maximized. It is often used as a goodness-of-fit statistic. You can compare the log likelihood value from the fits of your data to several distributions and select as the best fitting the one with the largest value.

Iterations

This is the number of iterations that were required to solve the likelihood equations. If this is greater than the maximum you specified, you will receive a warning message. You should then increase the Maximum Iterations and rerun the analysis.

Variance Covariance Matrix

Inverse of Fisher Information Matrix

	Intercept	Temp	Sigma
Intercept	0.897008	-0.4129681	-0.008282286
Temp	-0.4129681	0.1907777	0.004405599
Sigma	-0.008282286	0.004405599	0.002236808

This table gives an estimate of the asymptotic variance covariance matrix which is the inverse of the Fisher information matrix. The elements of the Fisher information matrix are calculated using numerical differentiation.

Percent Failing Section

Percent Failing Section

Row	Temp	Hours	Percent Failing	Lower 95.0% C.L.	Upper 95.0% C.L.
1	130.0	10000.0	0.4689	0.0159	12.2408
1	130.0	20000.0	7.5434	0.9592	40.7346
1	130.0	30000.0	22.4510	4.5202	63.9044
1	130.0	40000.0	39.1662	9.9834	78.8917
1	130.0	50000.0	53.9401	15.9731	87.8264
1	130.0	60000.0	65.7053	21.6733	92.9902
1	130.0	70000.0	74.6251	26.7623	95.9463
1	130.0	80000.0	81.2324	31.1796	97.6388
1	130.0	90000.0	86.0786	34.9767	98.6126
1	130.0	100000.0	89.6239	38.2423	99.1768

This report displays the estimated percent failing at the time values that were specified in the Report Times box of the Reports Tab for each observation with a missing time value. In our example, the first row of the Motors database is missing. The value of Temp (the stress variable) equal to 130 degrees. Reliability is one minus probability of failure. Thus, the reliability at 80,000 hours at a temperature of 130 degrees is 100-81.2324 which is 18.7676%. The confidence limits for reliability may also be converted from the percent failing confidence limits by subtracting from 100.

Percent Failing

The percent failing at a particular temperature is calculated as

$$100 \times \hat{F}(t \mid X_i) = 100 \times \hat{F} = 100 \times F\left(\frac{\log(t) - \sum_{i=0}^{p} x_{ki}\hat{B}_i}{\hat{S}}\right)$$

where $F(z)$ is the cumulative distribution of $f(z)$, the probability density function. That is,

$$F(z) = \int_0^z f(t \mid B_0, B_1, S, X)dt$$

Confidence Limits for Percent Failing

The confidence limits for this estimate are computed using the following formulas from Nelson (1990) page 296. Note that these estimates are large sample estimates based on the assumption that the distribution of F is asymptotically normal. We recommend that the number of failures be at least thirty when using these estimates.

$$\hat{F}_{lower}(t \mid X_i) = \frac{\hat{F}}{\hat{F} + (1 - \hat{F})\exp\left\{\frac{z_{1-\alpha/2}\sigma_{\hat{F}}}{\hat{F}(1 - \hat{F})}\right\}}$$

$$\hat{F}_{upper}(t \mid X_i) = \frac{\hat{F}}{\hat{F} + (1 - \hat{F})\exp\left\{\frac{-z_{1-\alpha/2}\sigma_{\hat{F}}}{\hat{F}(1 - \hat{F})}\right\}}$$

where

$$\sigma_{\hat{F}} = \sqrt{\sum_{i=0}^{p+1}\sum_{j=0}^{p+1} h_i h_j vc_{ij}}$$

$$h_i = \frac{-x_i g\left(\dfrac{y(t) - \sum x_i \hat{B}_i}{\hat{S}}\right)}{\hat{S}} \quad i = 0, \cdots, p$$

$$h_{p+1} = -\frac{y(t) - \sum x_i \hat{B}_i}{\hat{S}^2}$$

and vc_{ij} is the corresponding element from the variance covariance matrix. The function $y(t)$ is $ln(t)$ for the Weibull, log-logistic, exponential, and lognormal distributions, $log(t)$ for the lognormal10 distribution, and simply t for the normal, extreme value, and logistic distributions. The value of $g(x)$ depends on the distribution. For the Weibull, exponential, and extreme value distributions

$$g(z) = e^{z - e^z}$$

For the normal, lognormal, and lognormal10 distributions

$$g(z) = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}$$

For the logistic and log-logistic distributions

$$g(z) = \frac{e^z}{\left(1 + e^z\right)^2}$$

Failure Time Percentile Section

Failure Time Percentile Section

Row	Temp	Percentile	Estimated Hours	Lower 95.0% C.L.	Upper 95.0% C.L.
1	130.0	10.0000	21937.2	11779.1	40855.8
1	130.0	50.0000	47133.6	24099.1	92185.1
1	130.0	90.0000	101269.8	44862.3	228601.2

This report displays failure time percentiles and confidence intervals for those percentiles specified in the Report Percentiles box of the Report tab. For example, the median failure time is 47,135.1 hours. The 95% confidence limits for the median time are 24,106.6 to 92,162.2 hours.

The confidence limits rely on the asymptotic normality of the distribution of the percentiles. The sample size should be greater than thirty failed items before you use these confidence limits. The formulas for these limits are given in Nelson (1990) page 295.

Percentile

This is the percentile being found. For example, the value of 50 here refers to the median failure time.

Estimated Hours

The estimated time value (dependent variable) at which *100P* of the items are expected to fail. The percentile is found by solving the equation

$$P = F\left(\frac{y(t) - \sum_{i=0}^{p} x_{ki}\hat{B}_i}{\hat{S}} \right)$$

for *y(t)*. The function *y(t)* is *ln(t)* for the Weibull, log-logistic, exponential, and lognormal distributions, *log(t)* for the lognormal10 distribution, and simply *t* for the normal, extreme value, and logistic distributions. *F(t)* is the cumulative distribution function.

Confidence Limits for a Percentile

The confidence limits are computed as follows. First compute

$$u_p = F^{-1}(P)$$

Next compute

$$\sigma_{\hat{t}_p} = \sqrt{\sum_{i=0}^{p+1}\sum_{j=0}^{p+1} x_i x_j vc_{ij}}$$

where vc_{ij} is the corresponding element of the variance covariance matrix and

$$x_0 = 1$$
$$x_1 = X_1$$
$$.$$
$$.$$
$$.$$
$$x_p = X_p$$
$$x_{p+1} = u_P$$

Finally, for the lognormal, Weibull, exponential, and log-logistic distributions, compute

$$\hat{t}_{lower,p} = e^{\left(\sum_{i=0}^{p} x_i B_i + u_p \hat{S} - z_{1-\alpha/2}\sigma_{\hat{t}_p} \right)}$$

$$\hat{t}_{upper,p} = e^{\left(\sum_{i=0}^{p} x_i B_i + u_p \hat{S} + z_{1-\alpha/2}\sigma_{\hat{t}_p} \right)}$$

For the lognormal10 distribution, compute

$$\hat{t}_{lower,p} = 10^{\left(\sum_{i=0}^{p} x_i B_i + u_p \hat{S} - z_{1-\alpha/2}\sigma_{\hat{t}_p}\right)}$$

$$\hat{t}_{upper,p} = 10^{\left(\sum_{i=0}^{p} x_i B_i + u_p \hat{S} + z_{1-\alpha/2}\sigma_{\hat{t}_p}\right)}$$

For the normal, extreme value, and logistic distributions, compute

$$\hat{t}_{lower,p} = \sum_{i=0}^{p} x_i B_i + u_p \hat{S} - z_{1-\alpha/2}\sigma_{\hat{t}_p}$$

$$\hat{t}_{upper,p} = \sum_{i=0}^{p} x_i B_i + u_p \hat{S} + z_{1-\alpha/2}\sigma_{\hat{t}_p}$$

Residual Section

Residual Section

Row	(T) Hours	Log10(T)	Predicted Log10(T)	Raw Residual	Standardized Residual	Cox-Snell Residual
1			4.673331			
2R	8064.0	3.90655	4.168003	-0.2614523	-1.008778	0.1702434
3	1764.0	3.246499	3.708286	-0.4617874	-1.781744	3.811264E-02
4	2772.0	3.442793	3.708286	-0.2654927	-1.024368	0.1658548
5	3444.0	3.537063	3.708286	-0.1712228	-0.6606401	0.293595
6	3542.0	3.549248	3.708286	-0.1590374	-0.6136243	0.3143435
7	3780.0	3.577492	3.708286	-0.1307942	-0.5046515	0.3665836
8	4860.0	3.686636	3.708286	-2.164969E-02	-8.353239E-02	0.6286976
9	5196.0	3.715669	3.708286	7.383185E-03	2.848702E-02	0.7161357
10R	5448.0	3.736237	3.708286	2.795114E-02	0.1078457	0.7829427
11	408.0	2.61066	3.288272	-0.6776116	-2.614473	4.47828E-03
12	1344.0	3.128399	3.288272	-0.1598725	-0.6168463	0.3128878
13	1440.0	3.158362	3.288272	-0.1299092	-0.5012372	0.3683169
14R	1680.0	3.225309	3.288272	-6.296246E-02	-0.2429321	0.5175633
15	408.0	2.61066	2.722126	-0.1114662	-0.4300773	0.4058197
16	504.0	2.70243	2.722126	-1.969583E-02	-7.599371E-02	0.6343351
17R	528.0	2.722634	2.722126	5.075522E-04	1.958322E-03	0.6947111

This report displays the predicted value and residual for each row. If the analysis is being run on logarithms of time, all values are in logarithms. The report provides predicted values for all rows with values for the independent variables. Hence, you can add rows of data with missing time values to the bottom of your database and obtain the predicted values for them from this report. The report also allows you to obtain predicted values for censored observations.

You should ignore the residuals for censored observations, since the residual is calculated as if the time value was a failure.

Row

This is the number of the observation being reported on. Censored observations have a letter appended to the row number.

(T) Hours

This is the original value of the dependent variable.

Log10(T)

This is the transformed value of the dependent variable.

Predicted Log10(T)

This is the predicted transformed value of the dependent variable (usually time). Note that y depends on the distribution being fit. For the Weibull, exponential, lognormal, and log-logistic distributions, the y is $ln(t)$. For the lognormal10 distribution, y is $log(t)$. For the extreme value, normal, and logistic distributions, y is t. The formula for y is

$$\hat{y} = \sum_{i=0}^{p} x_i B_i$$

Raw Residual

This is the residual in the y scale. The formula is

$$r_k = y_k - \sum_{i=0}^{p} x_i B_i$$

Note that the residuals of censored observations are not directly interpretable, since there is no obvious value of y. The row is displayed so that you can see the predicted value for this censored observation.

Standardized Residual

This is the residual standardized by dividing by the standard deviation. The formula is

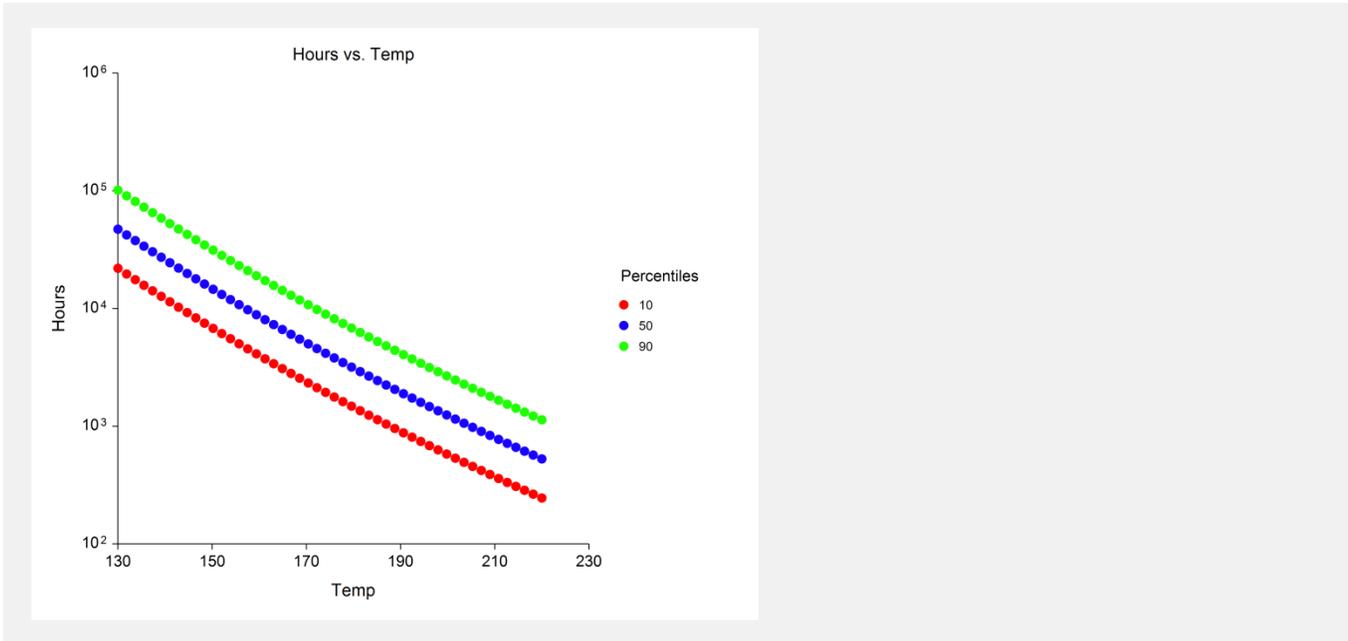$$r_k' = \frac{y_k - \sum_{i=0}^{p} x_i B_i}{\hat{S}}$$

Cox-Snell Residual

The Cox-Snell residual is defined as

$$r_k'' = -\log\left\{1 - F\left(\frac{y_k - \sum_{i=0}^{p} x_i B_i}{\hat{S}}\right)\right\}$$

Here again, the residual does not have a direct interpretation for censored values.

Stress Plot

This plot displays the time on the vertical axis and the stress variable on the horizontal axis. The plotted lines represent the percentiles specified on the Reports tab window. This allows you to quickly view the percentiles for a wide range of stress values.

X-Y Plots
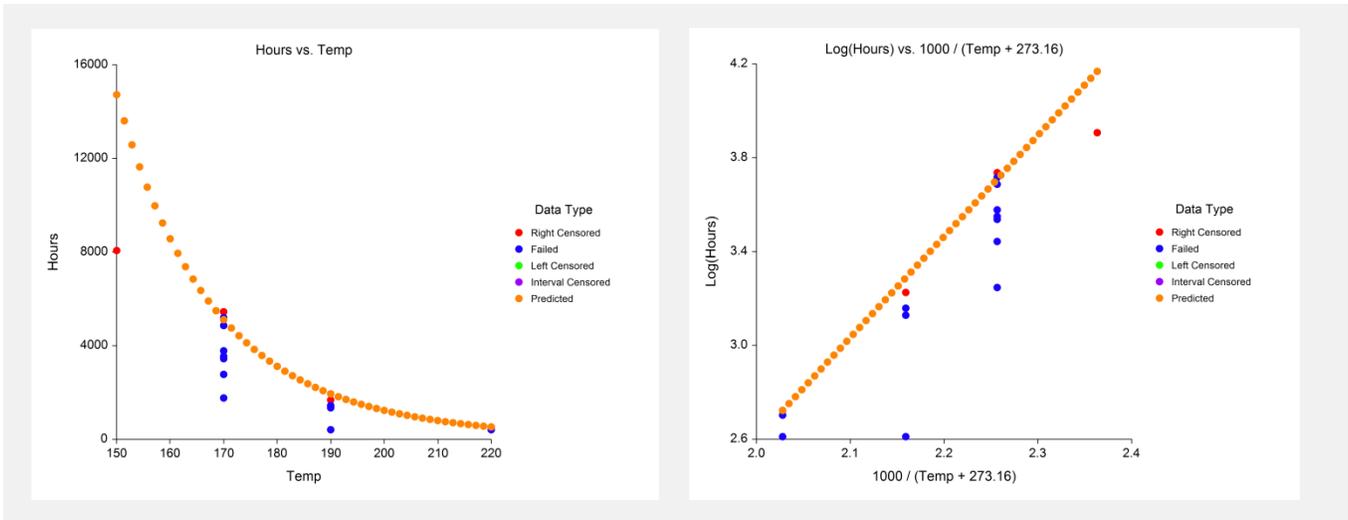
These plots show the data values from which the analysis was run. The plot on the left shows time versus stress in the original scale. The plot on the right shows time versus stress in the transformed metric. The prediction equation is also show on the chart. This lets you decide whether predictions are accurate. It also lets you study the goodness of fit.

Residual Plots
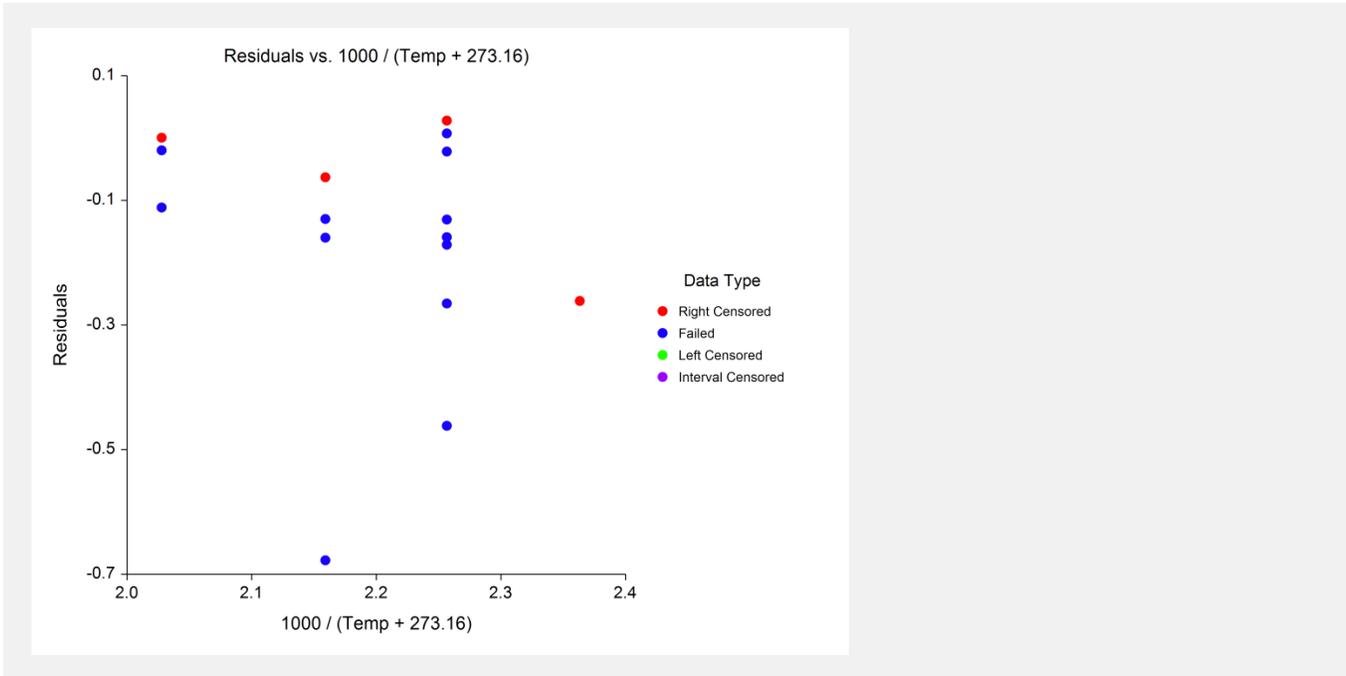
This plot shows the residuals in the transformed scale. You would study this chart just as you would any other residual versus independent variable plot from a multiple regression analysis. You are especially interested in finding outliers, as they can distort your results.

Discussion of Example

This example will look at an analysis of the electric motor data that was presented above and for which all of the above sample reports were generated. As mentioned earlier, a temperature-accelerated life test of a Class-B insulation was conducted using ten motors tested at each of four temperatures. When the testing was stopped, the failure times were recorded. These data are stored in the MOTORS.S0 database.

The purpose of this study was to determine the reliability of these motors at the normal operating temperature of 130°C by testing the reliability at higher temperatures. Note that at 150°C, none of the motors failed during the duration of the test.

The first step in the analysis is to determine if the fit is adequate. We look at the plots, the value of R-Squared, and the estimated value of sigma to determine this. The plots do not show any alarming points, although the residual plots show what might be a mild outlier in the 190°C batch.

Once the adequacy of the fit has been substantiated, we look at the Failure Time Percentile Section. This report provides the 10th, 50th, and 90th percentiles. The estimated failure times for these percentiles are 21,937 hours, 47,134 hours, and 101,270 hours. That is, we would expect about 10% of the motors to fail by 22,000 hours and 90% of the machines to have failed by 100,000 hours. No further calculations are necessary.