Regression with Life Data

Table Of Contents

Regression with Life Data

Regression with Life Data Overview

Use Minitab's regression with life data commands to investigate the relationship between failure time and one or more predictors. For example, you might want to examine how a predictor affects the lifetime of a person, part, product, or organism. The goal is to come up with a model that predicts failure time. Based on these predictions you can estimate the reliability of the system.

- Accelerated Life Testing performs a simple regression with one or two predictors that is used to model failure times for highly reliable products. The first predictor is an accelerating variable; its levels exceed those normally found in the field. The second predictor can be either a second accelerating variable or a factor. The data obtained under the high stress conditions can then be used to extrapolate back to normal use conditions. In order to do this, you must have a good model of the relationship between failure time and the accelerating variable(s).
- Regression with Life Data performs a regression with one or more predictors. The model can include factors, covariates, interactions, and nested terms. This model will help you understand how different factors and covariates affect the lifetime of your part or product.

Both regression with life data commands differ from other regression commands in Minitab in that they use different distributions and accept censored data. You can choose to model your data on one of the following eight distributions: Weibull, smallest extreme value, exponential, normal, lognormal basee, logistic, and loglogistic.

Life data is often incomplete or censored in some way. Censored observations are those for which an exact failure time is unknown. Suppose you are testing how long a product lasts and you plan to end the study after a certain amount of time. Any products that have not failed **before** the study ends are right-censored, meaning that the part failed sometime after the present time. Similarly, you may only know that a product failed before a certain time, which is left-censored. Failure times that occur within a certain interval of time are interval-censored.

Minitab uses a modified Newton-Raphson algorithm to calculate maximum likelihood estimates of the model parameters.

Regression with Life Data

Stat > Reliability/Survival > Regression with Life Data

Use Regression with Life Data to see whether one or more predictors affect the failure time of a product. The goal is to come up with a model that predicts failure time. This model uses explanatory variables to explain changes in the response variable, for example why some products fail quickly and some survive for a long time. The model can include factors, covariates, interactions, and nested terms.

Regression with Life Data differs from Minitab's regression commands in that it accepts censored data and uses different distributions.

To do regression with life data, you must enter the following information:

- the **response variable** (failure times).
- **model terms**, which consist of any number of predictor variables and when appropriate, various interactions between predictors and nested terms. See How to specify the model terms. Some of these variables may be **factors**.

Dialog box items

Responses are uncens/right censored data: Choose if your data is uncensored or right censored.

Responses are uncens/ arbitrarily censored data: Choose if your data is uncensored or arbitrarily censored.

Variables/Start variables: Enter up to 10 columns (10 different samples) containing the start times.

End variables: If you have uncensored or arbitrarily censored data, enter up to 10 columns (10 different samples) of end times.

Freq. columns (optional): Enter a column for each variable containing the frequency data.

Model: Enter the model terms – see How to specify the model terms. If any of those predictors are factors, enter them again in **Factors**.

Factors (optional): Enter any variables in the model that are factors.

Assumed distribution: Choose one of eight common lifetime distributions: Weibull (default), smallest extreme value, exponential, normal, lognormal, logistic, and loglogistic.

Data - Regression with Life Data

Enter three types of columns in the worksheet:

- the **response variable** (failure times) – see Failure times.

- **censoring indicators** for the response variables, if needed
- **predictor variables**, which may be factors (categorical variables) or covariates (continuous variables). For factors, Minitab estimates the coefficients for k - 1 design variables (where k is the number of levels), to compare the effect of different levels on the response variable. For covariates, Minitab estimates the coefficient associated with the covariate to describe its effect on the response variable.

Unless you specify a predictor as a factor, the predictor is assumed to be a covariate. In the model, terms may be created from these predictor variables and treated as factors, covariates, interactions, or nested terms. The model can include up to 9 factors and 50 covariates. Factors may be crossed or nested. Covariates may be crossed with each other or with factors, or nested within factors. See How to specify the model terms.

You can enter up to ten samples per analysis.

Depending on the type of censoring you have, you will set up your worksheet in column or table form. You can also structure the worksheet as raw data, or as frequency data. For details, see Worksheet Structure for Regression with Life Data.

Factor columns can be numeric or text. Minitab by default designates the lowest numeric or text value as the **reference level**. To change the reference level, see Factor variables and reference levels.

Minitab automatically excludes all observations with missing values from all calculations.

How you run the analysis depend on whether your data are uncensored/right censored or uncensored/arbitrarily censored:

Uncensored/arbitrarily censored data

If you have any combination of exact failure times, right-, left- and interval-censored data, enter your data using a Start column and End column:

For this observation...	Enter in the Start column...	Enter in the End column...
Exact failure time	Failure time	Failure time
Right censored	time after which the failure occurred	* (missing value symbol)
Left censored	* (missing value symbol)	time before which the failure occurred
Interval censored	time at start of interval during which the failure occurred	time at the end of interval during which the failure occurred

This example uses a frequency column as well.

Start	End	Frequency
*	10000	20
10000	20000	10
20000	30000	10
30000	30000	2
30000	40000	20
40000	50000	40
50000	50000	7
50000	60000	50
60000	70000	120
70000	80000	230
80000	90000	310
90000	*	190

20 units are left censored at 10000 hours.

2 units are exact failures at 30000 hours.

50 units are interval censored between 50000 and 60000 hours.

190 units are right censored at 90000 hours.

Uncensored/right censored data

Enter two columns for each sample – one column of failure (or censoring) times and a corresponding column of censoring indicators.

The Time column contains failure times.

Time	Censor
53	F
60	F
53	F
40	F
51	F
99	C
35	F
53	F
.	.
.	.
.	.
etc.	etc.

The Censor column contains the corresponding censoring indicators: an F designates an actual failure time; a C designates a unit that was removed from the test, and was thus censored.

Censoring indicators can be numbers, text, or date/time values. If you do not specify which value indicates censoring in the Censor subdialog box, Minitab assumes the lower of the two values indicates censoring, and the higher of the two values indicates an exact failure.

The data column and associated censoring column must be the same length, although pairs of data/censor columns (each pair corresponds to a sample) can have different lengths.

Failure times

The response data you gather for the regression with life data commands are the individual failure times. For example, you might collect failure times for units running at a given temperature. You might also collect samples under different temperatures, or under varying conditions of any combination of accelerating variables.

Individual failure times are the same type of data used for Distribution Analysis.

Life data is often censored or incomplete in some way. Suppose you are monitoring air conditioner fans to find out the percentage of fans that fail within a three-year warranty period. The table below describes the types of observations you can have:

Type of observation	Description	Example
Exact failure time	You know **exactly** when the failure occurred.	The fan failed at exactly 500 days.
Right censored	You only know that the failure occurred **after** a particular time.	The fan had not yet failed at 500 days.
Left censored	You only know that the failure occurred **before** a particular time.	The fan failed sometime before 500 days.
Interval censored	You only know that the failure occurred **between** two particular times.	The fan failed sometime between 475 and 500 days.

How you set up your worksheet depends, in part, on the type of censoring you have:

- When your data consist of exact failures and right-censored observations, see Uncensored/right censored data.
- When your data have a varied censoring scheme, see Uncensored/arbitrarily censored data.

To perform regression with uncensored/right censored data

1 Choose **Stat > Reliability/Survival** > **Regression with Life Data**.

2 In **Variables/Start variables**, enter up to ten columns of failure times (ten different samples).

3 If you have frequency columns, enter them in **Freq. columns**.

4 In **Model**, enter the model terms – see How to specify the model terms. If any of those predictors are factors, enter them again in **Factors**.

Note If you have no censored data, you can skip steps 5 & 6.

5 Click **Censor**.

6 In **Use censoring columns**, enter the censoring columns. The first censoring column is paired with the first data column, the second censoring column is paired with the second data column, and so on.

By default, Minitab uses the lowest value in the censoring column to indicate a censored value. To use some other value, enter that value in **Censoring value**.

7 If you like, use any dialog box options, then click **OK**.

To perform regression with uncensored/arbitrarily censored data

1 Choose **Stat > Reliability/Survival > Regression with Life Data**.

2 Choose **Responses are uncens/arbitrarily censored data**.

3 In **Variables/Start variables**, enter up to ten columns of start times (ten different samples).

4 In **End variables**, enter up to ten columns of end times (ten different samples).

5 If you have frequency columns, enter them in **Freq. columns**.

6 In **Model**, enter the model terms – see How to specify the model terms. If any of those predictors are factors, enter them again in **Factors**.

7 If you like, use any dialog box options, then click **OK**.

Estimating the model parameters

Minitab uses a modified Newton-Raphson algorithm to estimate the model parameters. If you like, you can enter your own parameters. In this case, no estimation is done; all results – such as the percentiles – are based on these parameters.

When you let Minitab estimate the parameters from the data, you can optionally:

- enter starting values for the algorithm.
- change the maximum number of iterations for reaching convergence (the default is 20). Minitab obtains maximum likelihood estimates through an iterative process. If the maximum number of iterations is reached before convergence, the command terminates.
- estimate other model coefficients while holding the shape parameter (Weibull) or the scale parameter (other distributions) fixed at a specific value.

Why enter starting values for the algorithm? The maximum likelihood solution may not converge if the starting estimates are not in the neighborhood of the true solution, so you may want to specify what you think are good starting values for parameter estimates.

In all cases, enter a column with entries that correspond to the model terms in the order you entered them in the Model box. With complicated models, find out the order of entries for the starting estimates column by looking at the regression table in the output.

Factor variables and reference levels

You can enter numeric, text, or date/time factor levels. Minitab assigns one factor level to be the **reference level**, meaning that the estimated coefficients are interpreted relative to this level.

Regression with Life Data creates a set of design variables for each factor in the model. If there are k levels, there will be k - 1 design variables and the reference level will be coded as 0. Here are two examples of the default coding scheme:

Factor A with 4 levels (1 2 3 4)

reference level is 1		A1	A2	A3
	1	0	0	0
	2	1	0	0
	3	0	1	0
	4	0	0	1

Factor B with 3 levels (High Low Medium)

reference level is High		B1	B2
	High	0	0
	Low	1	0
	Medium	0	1

By default, Minitab designates the lowest numeric, date/time, or text value as the reference factor level. If you like, you can change this reference value in the Options subdialog box.

Multiple degrees of freedom test

When you have a term with more than one degree of freedom, you can request a multiple degrees of freedom test. This procedure tests whether or not the term is significant. In other words: Is at least one of the coefficients associated with this term significantly different than zero?

Regression with Life Data - Censor

Stat > Reliability/Survival > Regression with Life Data > Censor

Allows you to specify the censoring columns.

Dialog box items

Use censoring columns: If you have right censored data, enter the censoring columns. The first censoring column is paired with the first data column, the second censoring column is paired with the second data column, and so on.

Censoring value: By default, Minitab uses the lowest value in the censoring column to indicate a censored observation. To use some other value, enter that value in Censoring value. Text values must be contained in double quotes.

Regression with Life Data - Estimate

Stat > Reliability/Survival > Regression with Life Data > Estimate

Percentile and Probability Estimation

Enter new predictor values: Choose to specify values for predictors to estimate percentiles and/or survival probabilities. Most often, you would enter the design value. You can enter the values separated by spaces, or one or more columns. Values (or columns) should be listed in the order that the corresponding variables appear in the model.

Use predictor values in data (storage only): Choose to use the predictor values from the data to estimate percentiles and/or survival probabilities.

Estimate percentiles for percents: Enter the percents for which you want to estimate percentiles. By default, Minitab estimates the 50th percentile. If you want to look at the beginning, middle, and end of the product's lifetime for a given predictor value, enter 10 50 90 (the 10th, 50th, 90th percentiles). Minitab then estimates how long it takes for 10% of the units to fail, 50% of the units to fail, and 90% of the units to fail.

Store percentiles

Percentiles: Check to store the percentiles.

Standard error: Check to store the standard error of the percentiles.

Confidence limits: Check to store the confidence limits for the percentiles.

Estimate probabilities for times: Enter the times for which you want to estimate survival probabilities or cumulative failure probabilities.

Estimate survival probabilities: Choose to estimate survival probabilities.

Estimate cumulative failure probabilities: Choose to estimate cumulative failure probabilities.

Store probabilities

Probabilities: Check to store the survival probabilities or cumulative failure probabilities.

Confidence limits: Check to store the confidence limits for the survival probabilities or cumulative failure probabilities.

Confidence level: Enter the confidence level for **all** of the confidence intervals. The default is 95%.

Confidence intervals: Choose to use two-sided confidence intervals (the default) or just an upper or lower confidence interval.

To estimate percentiles and survival probabilities

You can estimate percentiles and survival probabilities for new predictor values, or the values in your data.

1 In the Regression with Life Data dialog box, click **Estimate**.

2 Do one of the following:

- To enter new predictor values: In **Enter new predictor values**, enter a set of predictor values (or columns containing sets of predictor values) for which you want to estimate percentiles or survival probabilities. The predictor values must be in the same order as the main effects in the model. For an illustration, see Example of regression with life data.
- To use the predictor values in the data, choose **Use predictor values in data (storage only)**. Because of the potentially large amount of output, Minitab stores the results in the worksheet rather then printing them in the Session window.

3 Do any of the following, then click **OK**:

- To estimate percentiles, enter the percents or a column of percents in **Estimate percentiles for percents**. By default, Minitab estimates the 50th percentile. If you want to look at the beginning, middle, and end of the product's lifetime for a given predictor value, enter 10 50 90 (the 10th, 50th, 90th percentiles). Minitab then estimates how long it takes for 10% of the units to fail, 50% of the units to fail, and 90% of the units to fail.
- To estimate survival probabilities, enter the times or a column of times in **Estimate probabilities for times**. For example, when you enter 70 (units in hours in this example), Minitab estimates the probability, for each predictor value, that the unit will survive past 70 hours.

Regression with Life Data - Graphs

Stat > Reliability/Survival > Regression with Life Data > Graphs

You can draw probability plots for the standardized and Cox-Snell residuals.

Dialog box items

Probability plot for standardized residuals: Check to display a probability plot for standardized residuals.

Exponential probability plot for Cox-Snell residuals: Check to display an exponential probability plot for Cox-Snell residuals.

Display confidence intervals on probability plots: Check to display confidence intervals on the probability plots.

Note To change the method for calculating probability plot points, see Tools > Options > Individual Graphs > Probability Plots.

Probability plots for regression with life data

The Regression with Life Data command draws probability plots for the standardized and Cox-Snell residuals. You can use these plots to assess whether a particular distribution fits your data. In general, the closer the points fall to the fitted line, the better the fit.

Minitab provides one goodness-of-fit measure: the Anderson-Darling statistic. The Anderson-Darling statistic is useful in comparing the fit of different distributions. It measures the distances from the plot points to the fitted line; therefore, a smaller Anderson-Darling statistic indicates that the distribution provides a better fit.

To draw a probability plot of the residuals

1 In the Accelerated Life Testing or Regression with Life Data dialog box, click **Graphs**.

2 Do any of the following, then click **OK**:

- To plot the standardized residuals, check **Probability plot for standardized residuals**
- To plot the Cox-Snell residuals, check **Exponential probability plot for Cox-Snell residuals**

Note To draw a probability plot with more options, store the residuals in the Storage subdialog box, then use the probability plot included with the Parametric Distribution Analysis commands.

Regression with Life Data - Options

Stat > Reliability/Survival > Regression with Life Data > Options

You can estimate the model parameters from the data or enter historical estimates – see Estimating the Model Parameters for more information. You can also change the reference level for a factor.

Dialog box items

Estimate model parameters: Choose to estimate the model parameters from the data.

Use starting estimates: If you have starting estimates, enter one column to be used for all of the response variables, or a separate column for each response variable. The column(s) should contain one value for each coefficient in the regression table, in the order that the coefficients appear in the regression table.

Maximum number of iterations: Enter a positive integer to specify the maximum number of iterations for the Newton-Raphson algorithm.

Set shape (Weibull) or scale (other distributions) at: To estimate other model coefficients while holding the shape or scale parameter fixed, enter one value to be used as the shape or scale parameter for all of the response variables, or a number of values equal to the number of response variables.

Use historical estimates: Choose to enter your own estimates for the model parameters. Enter one column to be used for all of the response variables, or a separate column for each response variable. The column(s) should contain one value for each coefficient in the regression table, in the order that the coefficients appear in the regression table.

Reference factor level (enter factor followed by level): For each factor you want to set the reference level for, enter a factor column followed by a value specifying the reference level. For text values, the value must be in double quotes. For date/time values, store the value as a constant and then enter the constant.

To control estimation of the parameters

1 In the Regression with Life Data dialog box, click **Options**.

2 Do one of the following:

- To estimate the model parameters from the data (the default), choose **Estimate model parameters**.
  - To enter starting estimates for the parameters: In **Use starting estimates**, enter one column to be used for all of the response variables, or a number of columns equal to the number of response variables.
  - To specify the **Maximum number of iterations**, enter a positive integer.
  - To estimate other model coefficients while holding the shape parameter (Weibull) or the scale parameter (other distributions) fixed: In **Set shape (Weibull) or scale (other distributions) at**, enter one value to be used for all of the response variables, or a number of values equal to the number of response variables.
- To enter your own estimates for the model parameters, choose **Use historical estimates** and enter one column to be used for all of the response variables, or a number of columns equal to the number of response variables.

3 Click **OK**.

To change the reference factor level

1 In the Regression with Life Data dialog box, click **Options**.

2 In **Reference factor level**, for each factor you want to set the reference level for, enter a factor column followed by a value specifying the reference level. For text values, the value must be in double quotes. For date/time values, store the value as a constant and then enter the constant.

3 Click **OK**.

Regression with Life Data - Results

Stat > Reliability/Survival > Regression with Life Data > Results

You can control the display of Session window output.

Dialog box items

Control the Display of Results

Display nothing: Choose to suppress all printed output, but do all requested storage and display graphs.

Response information, censoring information, regression table, log-likelihood, and goodness-of-fit: Choose to display the response and censoring information, the regression table, the log-likelihood, and goodness-of-fit measures.

In addition, table of percentiles and/or survival probabilities: Choose to display the output described above, plus a table of percentiles and/or survival probabilities (if requested in Regression with Life Data - Estimate).

In addition, list of factor level values, tests for terms with more than one degree of freedom: Choose to display all of the output described above, plus a list of the factor level values, and a multiple degrees of freedom test.

Show log-likelihood for each iteration of algorithm: Check to display the log-likelihood at each iteration of the parameter estimation process.

To perform multiple degrees of freedom tests

1 In the Regression with Life Data dialog box, click **Results**.

2 Choose **In addition, list of factor level values, tests for terms with more than 1 degree of freedom**, then click **OK**.

Regression with Life Data - Storage

Stat > Reliability/Survival > Regression with Life Data > Storage

You can store three types of residuals and information on the estimated equation.

Dialog box items

Residuals Check any of the residual types below to store them in the worksheet.

Ordinary

Standardized

Cox-Snell

Information on Estimated Equation

Estimated coefficients: Check to store the estimated coefficients.

Standard error of estimates: Check to store the standard error of the estimated coefficients.

Confidence limits for coefficients: Check to store the confidence limits for the coefficients.

Variance/covariance matrix: Check to store the variance/covariance matrix for the estimated coefficients.

Log-likelihood for last iteration: Check to store the log-likelihood for the last iteration.

Example of Regression with Life Data

Suppose you want to investigate the deterioration of an insulation used for electric motors. You want to know if you can predict failure times for the insulation based on the plant in which it was manufactured, and the temperature at which the motor runs. It is known that an Arrhenius relationship exists between temperature and $\log_e$ failure time.

You gather failure times at plant 1 and plant 2 for the insulation at four temperatures – 110, 130, 150, and 170° C. Because the motors generally run at between 80 and 100° C, you want to predict the insulation's behavior at those temperatures.

To see how well the model fits, you will draw a probability plot based on the standardized residuals.

1 Open the worksheet INSULATE.MTW.

2 Choose **Stat > Reliability/Survival > Regression with Life Data**.

3 In **Variables/Start variables**, enter *FailureT*.

4 In **Model**, enter *ArrTemp Plant*. In **Factors (optional)**, enter *Plant*.

5 Click **Censor**. In **Use censoring columns**, enter *Censor*, then click **OK**.

6 Click **Estimate**. In **Enter new predictor values**, enter *ArrNewT NewPlant*, then click **OK**.

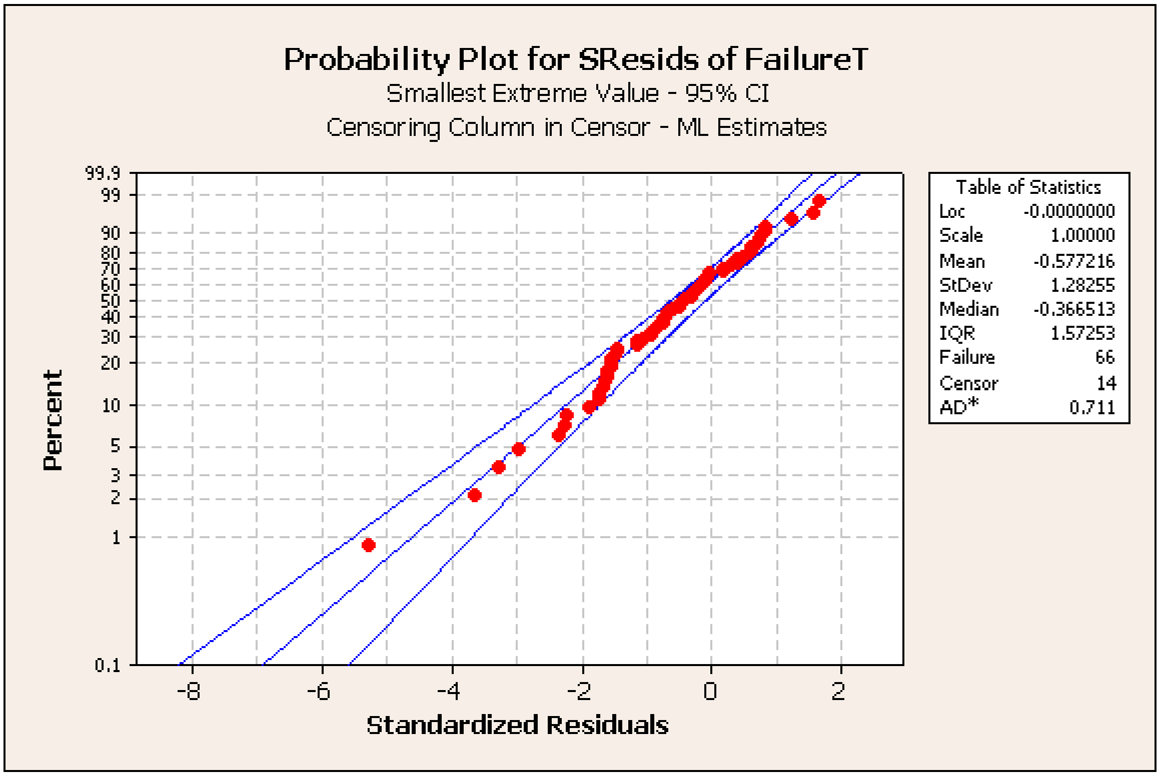
7 Click **Graphs**. Check **Probability plot for standardized residuals**, then click **OK** in each dialog box.

Session window output

Regression with Life Data: FailureT versus ArrTemp, Plant

```
Response Variable: FailureT

Censoring Information  Count
Uncensored value          66
Right censored value      14

Censoring value: Censor = C

Estimation Method: Maximum Likelihood

Distribution:   Weibull

Relationship with accelerating variable(s):   Linear


Regression Table

                       Standard                         95.0% Normal CI
Predictor        Coef     Error       Z      P       Lower        Upper
Intercept   -15.3411   0.950822  -16.13  0.000   -17.2047    -13.4775
ArrTemp     0.839255  0.0339710   24.71  0.000   0.772673    0.905837
Plant
 2         -0.180767  0.0845721   -2.14  0.033  -0.346525  -0.0150083
Shape        2.94309   0.270658                   2.45768     3.52439

Log-Likelihood = -562.525


Table of Percentiles

                                     Standard   95.0% Normal CI
Percent  ArrTemp  Plant  Percentile     Error    Lower    Upper
     50  32.8600      1      182094   32466.2   128390   258261
     50  32.8600      2      151981   25286.6   109690   210578
     50  31.0988      1     41530.4   5163.76  32548.4  52990.9
     50  31.0988      2     34662.5   3913.87  27781.0  43248.6
```

Graph window output:



Interpreting the results

From the Regression Table, you get the coefficients for the regression model. For the Weibull distribution, here is the equation that describes the relationship between temperature and failure time for the insulation for plant 1 and 2, respectively:

Loge (failure time) = –15.3411 + 0.83925 (ArrTemp) + 1/2.9431 ε_p

Loge (failure time) = –15.52187 + 0.83925 (ArrTemp) + 1/2.9431 ε_p

where ε_p = the pth percentile of the error distribution

ArrTemp = 11604.83/(Temp + 273.16)

The Table of Percentiles displays the 50th percentiles for the combinations of temperatures and plants that you entered. The 50th percentile is a good estimate of how long the insulation will last in the field:

- For motors running at 80° C, insulation from plant 1 lasts about 182093.6 hours or 20.77 years; insulation from plant 2 lasts about 151980.8 hours or 17.34 years.
- For motors running at 100° C, insulation from plant 1 lasts about 41530.38 hours or 4.74 years; insulation from plant 2 lasts about 34662.51 hours or 3.95 years.

As you can see from the low p-values, the plants are significantly different at the α = .05 level, and temperature is a significant predictor.

The probability plot for standardized residuals will help you determine whether the distribution, transformation, and equal shape (Weibull or exponential) or scale parameter (other distributions) assumption is appropriate. Here, the plot points fit the fitted line adequately; therefore you can assume the model is appropriate.

Default output

The default output consists of the regression table, which displays:

- the estimated **coefficients** for the regression model and their
  - standard errors.
  - Z-values and p-values. The Z-test tests that the coefficient is significantly different than 0; in other words, is it a significant predictor?
  - 95% confidence interval.
- the **Scale parameter**, a measure of the overall variability, and its
  - standard error.
  - 95% confidence interval.
- the **Shape parameter** (Weibull or exponential) or **Scale parameter** (other distributions), a measure of the overall variability, and its
  - standard error.
  - 95% confidence interval.
- the log-likelihood.