



# 16

## Regression with Life Data

- [Regression with Life Data Overview, 16-2](#)
- [Worksheet Structure for Regression with Life Data, 16-3](#)
- [Accelerated Life Testing, 16-6](#)
- [Regression with Life Data, 16-19](#)





# 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 predictor that is used to model failure times for highly reliable products. The predictor is an accelerating variable; its levels exceed those normally found in the field. 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.
- **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, extreme value, exponential, normal, lognormal base<sub>e</sub>, lognormal base<sub>10</sub>, 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.





# Worksheet Structure for Regression with Life Data

The basic worksheet structure for regression with life data is three columns, although you may have more than three. The three columns are:

- the *response variable* (failure times)—see *Failure times* on page 16-4
- *censoring indicators* (for the failure times, if needed)—see *Uncensored/right censored data* on page 16-4
- *predictor variables*
  - For Accelerated Life Testing, enter one predictor column containing various levels of an accelerating variable. For example, an accelerating variable may be stresses or catalysts whose levels exceed normal operating conditions.
  - For Regression with Life Data, enter one or more predictor columns. These predictor variables may be treated as factors or covariates in the model. For more information, see *How to specify the model terms* on page 16-24.

Structure each column so that it contains individual observations (one row = one observation), or unique observations with a corresponding column of frequencies. Frequency columns are useful when you have large numbers of data with common failure or censoring times, and identical predictor values. Here is the same worksheet structured both ways:

**Raw Data:** one row for each

**Frequency Data:** one row for each combination of response, censoring indicator,

C1	C2	C3	C4	
Response	Censor	Factor	Covar	
29	F	1	12	—1
31	F	1	12	}—19
31	F	1	12	
⋮	⋮	⋮	⋮	
37	F	1	12	—1
37	C	2	12	—1
41	F	2	12	}—19
⋮	⋮	⋮	⋮	
⋮	⋮	⋮	⋮	

C1	C2	C3	C4	C5
Response	Censor	Covar	Factor	Count
29	F	12	1	1
31	F	12	1	19
37	F	12	1	1
37	C	12	2	1
41	F	12	2	19

Text categories (factor levels) are processed in alphabetical order by default. If you wish, you can define your own order—see *Ordering Text Categories* in the *Manipulating Data* chapter of *MINITAB User's Guide 1* for details.

The way you set up the worksheet depends on the type of censoring you have, as described in *Failure times* below.



## Failure times

The response data you gather for the commands in this chapter 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* on page 15-1.

Life data is often censored or incomplete in some way. Suppose you are monitoring air conditioner fans to find out the percentage of fans which 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 <i>exactly</i> when the failure occurred.	The fan failed at exactly 500 days.
Right censored	You only know that the failure occurred <i>after</i> a particular time.	The fan failed sometime after 500 days.
Left censored	You only know that the failure occurred <i>before</i> a particular time.	The fan failed sometime before 500 days.
Interval censored	You only know that the failure occurred <i>between</i> 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* on page 16-4.
- When your data have a varied censoring scheme, see *Uncensored/arbitrarily censored data* on page 16-5.

### Uncensored/right censored data

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

This column contains failure times.

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

This columns contains the corresponding censoring indicators:  
an F designates an exact 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.

### Uncensored/arbitrarily censored data

When 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	the missing value symbol '*'
Left censored	the missing value symbol '*'	time before which the failure occurred
Interval censored	time at start of interval during which the failure occurred	time at end of interval during which the failure occurred

This example uses a frequency column as well.

Start	End	Frequency	
*	10000	20	← 20 units are left censored at 10000 hours.
10000	20000	10	
20000	30000	10	
30000	30000	2	← Two units are exact failures at 30000 hours.
30000	40000	20	
40000	50000	40	
50000	50000	7	
50000	60000	50	← 50 units are interval censored between 50000 and 60000 hours.
60000	70000	120	
70000	80000	230	
80000	90000	310	
90000	*	190	← 190 units are right censored at 90000 hours.





# Accelerated Life Testing

Use Accelerated Life Testing to investigate the relationship between failure time and one predictor. The most common application of accelerated life testing is for studies in which you impose a series of variable levels far exceeding normal field conditions to accelerate the failure process. The variable is thus called the *accelerating variable*. Accelerated tests are performed to save time and money, since, under normal field conditions, it can take a very long time for a unit to fail. Accelerated life testing requires knowledge of the relationship between the accelerating variable and failure time.

Here are the steps:

- 1 Impose levels of the accelerating variable on the units.
- 2 Record the failure (or censoring) times.
- 3 Run the Accelerated Life Testing analysis, asking MINITAB to extrapolate to the *design value*, or common field condition. This way, you can find out how the units behave under normal field conditions.

You can request an Arrhenius, inverse temperature,  $\log_e$ ,  $\log_{10}$  transformation, or no transformation for the accelerating variable. By default, MINITAB assumes the relationship is linear (no transformation).

The simplest output includes a regression table, relation plot, and probability plot for each level of the accelerating variable based on the fitted model. The relation plot displays the relationship between the accelerating variable and failure time by plotting percentiles for each level of the accelerating variable. By default, lines are drawn at the 10th, 50th, and 90th percentiles. The 50th percentile is a good estimate for the time a part will last when exposed to various levels of the accelerating variable. The probability plot is created for each level of the accelerating variable based on the fitted model (line) and based on a nonparametric model (points).

## Data

See *Worksheet Structure for Regression with Life Data* on page 16-3.

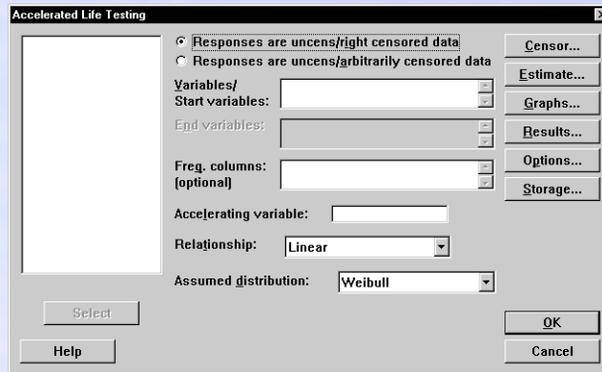
MINITAB automatically excludes all observations with missing values from all calculations.

How you run the analysis depends on whether your data is uncensored/right censored or uncensored/arbitrarily censored.





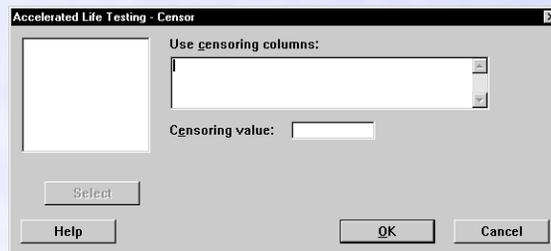
- ▶ To perform accelerated life testing with uncensored/right censored data
  - 1 Choose **Stat** ▶ **Reliability/Survival** ▶ **Accelerated Life Testing**.



- 2 In **Variables/Start variables**, enter the columns of failure times. You can enter up to ten columns (ten different samples).
- 3 If you have frequency columns, enter them in **Freq. columns**.
- 4 In **Accelerating variable**, enter the column of predictors.

**Note** | If you have no censored values, 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.

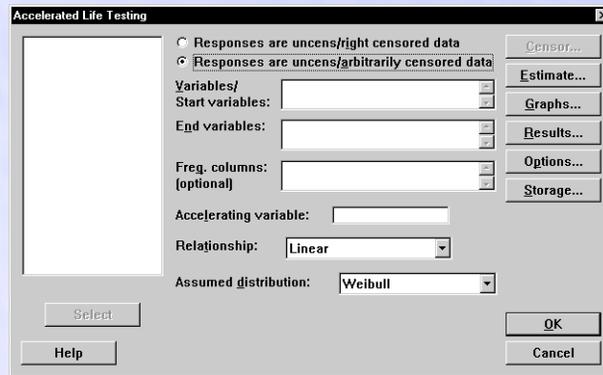
By default, MINITAB uses the lowest value in the censoring column to indicate a censored observation. To use some other value, enter that value in **Censoring value**.

**Note** | If your censoring indicators are date/time values, store the value as a constant and then enter the constant.

- 7 If you like, use any of the options listed below, then click **OK**.



- ▶ To perform accelerated life testing with uncensored/arbitrarily censored data
  - 1 Choose **Stat** ▶ **Reliability/Survival** ▶ **Accelerated Life Testing**.
  - 2 Choose **Responses are uncens/arbitrarily censored data**.



- 3 In **Variables/Start variables**, enter the columns of start times. You can enter up to ten columns (ten different samples).
- 4 In **End variables**, enter the columns of end times. You can enter up to ten columns (ten different samples).
- 5 If you have frequency columns, enter them in **Freq. columns**.
- 6 In **Accelerating variable**, enter the column of predictor values.
- 7 If you like, use any of the options described below, then click **OK**.

## Options

### Accelerated Life Testing dialog box

- transform the accelerating variable one of four common ways: Arrhenius, inverse temperature,  $\log_e$ , or  $\log_{10}$ . By default, MINITAB uses no transformation (linear). See *Transforming the accelerating variable* on page 16-12.
- choose one of eight common lifetime distributions for the error distribution, including the Weibull (default), extreme value, exponential, normal, lognormal base $_e$ , lognormal base $_{10}$ , logistic, and loglogistic distributions.

**More** | MINITAB's extreme value distribution is the smallest extreme value (Type 1).

### Estimate subdialog box

- enter predictor values (levels of accelerating variable) for which to estimate percentiles and/or survival probabilities. Most often, you would enter the design



value. You can also use the predictor values (levels of accelerating variable) from the data.

- estimate percentiles for the percents you specify—see *Percentiles and survival probabilities* on page 16-16. By default, MINITAB estimates the 50th percentile.
- store the percentiles, their standard errors, and confidence intervals.
- estimate survival probabilities for the times you specify—see *Percentiles and survival probabilities* on page 16-16. For example, if you enter 10 hours, MINITAB estimates (for each predictor value) the proportion of units that survive at least 10 hours.
- store the survival probabilities and confidence intervals.
- specify a confidence level for all of the confidence intervals. The default is 95.0%.

### Graphs subdialog box

- enter a design value to include on the plots based on the fitted model (relation plot and probability plot for each accelerating level).
- draw a relation plot to display the relationship between an accelerating variable and failure time—see *Relation plot* on page 16-11. You can:
  - plot percentiles for the percents you specify. By default, MINITAB plots the 10th, 50th, and 90th percentiles.
  - display confidence intervals for all of the percentiles or the middle percentiles only. You can also suppress their display.
  - display points for failure times (exact failure time or midpoint of interval for interval censored observation) on the plot.
- draw a probability plot for each level of the accelerating variable based on the fitted model—see *Probability plots* on page 16-14. You can:
  - display confidence intervals for the design value or for all levels of the accelerating variable. You can also suppress their display.
- draw a probability plot for each level of the accelerating variable based on the individual fits—see *Probability plots* on page 16-14.
- draw a probability plot for the standardized residuals and an exponential probability plot for the Cox-Snell residuals—see *Probability plots* on page 16-14.
- include confidence intervals on the diagnostic plots.

### Results subdialog box

- display the following Session window output:
  - no output
  - the basic output, which includes the response information, censoring information, regression table, the log-likelihood, and goodness-of-fit measures





- the basic output, plus the table of percentiles and/or survival probabilities (the default)
- show the log-likelihood for each iteration of the algorithm

### Options subdialog box

- enter starting values for model parameters for the Newton-Raphson algorithm—see *Estimating the model parameters* on page 16-28.
- 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—see *Estimating the model parameters* on page 16-28.
- estimate other model coefficients while holding the shape parameter (Weibull) or the scale parameter (other distributions) fixed at a specific value—see *Estimating the model parameters* on page 16-28.
- use historical estimates for the parameters rather than estimate them from the data—see *Estimating the model parameters* on page 16-28. In this case, no estimation is done; all results—such as the percentiles—are based on these parameters.

### Storage subdialog box

- store the ordinary, standardized, and Cox-Snell residuals.
- store information on the estimated equation, including the estimated coefficients, their standard errors and confidence intervals, the variance/covariance matrix, and the log-likelihood for the last iteration.

## Output

The default output consists of the regression table, a relation plot, a probability plot for each level of the accelerating variable based on the fitted model, and Anderson-Darling goodness-of-fit statistics for the probability plot.

### Regression table

The regression table 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 zero; in other words, is it a significant predictor?
  - 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.
- Anderson-Darling goodness-of-fit statistics for each level of the accelerating variable based on the fitted model.

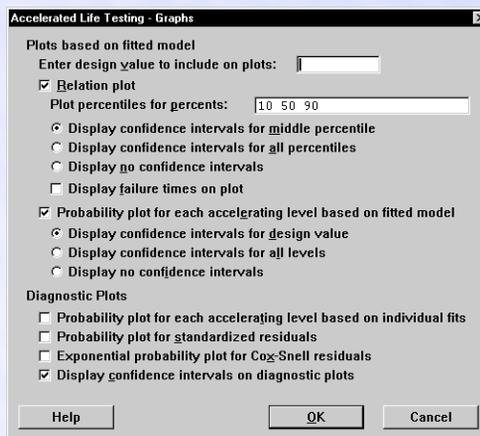
### Relation plot

The relation plot displays failure time versus an accelerating variable. By default, lines are drawn at the 10th, 50th, and 90th percentiles. The 50th percentile is a good estimate for the time a part will last for the given conditions. For an illustration, see *Example of accelerated life testing* on page 16-17.

You can optionally specify up to ten percentiles to plot and display the failure times (exact failure time or midpoint of interval for interval censored observation) on the plot. You can enter a design value to include on the plot.

### ► To modify the relation plot

- 1 In the Accelerated Life Testing dialog box, click **Graphs**.



- 2 Do any of the following:

- To include a design value on the plot, specify a value in **Enter design value to include on plots**.
- To plot percentiles for the percents you specify, enter the percents or a column of percents in **Plot percentiles for percents**. For example, to plot the 30th percentile (how long it takes 30% of the units to fail), enter 30. By default, MINITAB plots the 10th, 50th, and 90th percentiles.
- Choose one:
  - **Display confidence intervals for middle percentile**



- **Display confidence intervals for all percentiles**
  - **Display no confidence intervals**
  - To include failure times (exact failure time or midpoint of interval for interval censored observation) on the plot, check **Display failure times on plot**.
- 3 Click **OK**.
  - 4 If you like, change the confidence level for the intervals (default = 95%): Click **Estimate**. In **Confidence level**, enter a value, then click **OK**.

### Probability plot for each accelerating level based on fitted model

The probability plot displays the percents for each level of the accelerating variable based on the fitted model (line) and a nonparametric model (points). By default, the probability plot includes the shape and scale parameters (Weibull and exponential distributions) or the location and scale parameters (other distributions). These parameters are based on the fitted model. This plot assumes that the observations for each accelerating variable level share a common shape (Weibull and exponential distributions) or scale (other distributions).

The probability plot also includes the Anderson-Darling statistic, which is a goodness-of-fit measure. A count of failures and right-censored data appears when your data are exact failures/right-censored.

You can choose to display confidence intervals for the design value, for all levels of the accelerating variable, or no confidence intervals. For more information on creating and interpreting probability plots see *Probability plots* on page 16-14.

## Transforming the accelerating variable

If you assume a linear relationship then no transformation is needed. Any change in failure time or log failure time is directly proportional to the change in the accelerating variable.

A log relationship is used to model the life of products running under constant stress. The log relationship is most often used in combination with a log-based failure time distribution. When it is used in combination with a log-based failure time distribution, an **inverse power** relationship results. Common applications of the log transformations include electrical insulations, metal fatigue, and ball bearings.

Based on the Arrhenius Rate Law, the rate of a simple chemical reaction depends on the temperature. This relationship is often used to describe failures due to degradation caused by a chemical reaction. Common applications of the Arrhenius transformation include electrical insulations, semiconductor devices, solid state devices, and plastics.

$$\text{Arrhenius transformation} = \frac{11604.83}{C^{\circ} + 273.16}$$



The inverse temperature transformation is a simple relationship that assumes that failure time is inversely proportional to Kelvin temperature. The inverse and Arrhenius transformations have similar results, but the coefficients have different interpretations.

$$\text{Inverse temperature transformation} = \frac{1}{C^\circ + 273.16}$$

## Interpreting the regression equation

The regression model estimates the percentiles of the failure time distribution:

$$Y_p = \beta_0 + \beta_1 X + \sigma \varepsilon_p$$

where:

$Y_p$  = pth percentile of the failure time distribution, either failure time or log (failure time)

$\beta_0$  = y-intercept (constant)

$\beta_1$  = regression coefficient

$X$  = predictor values (may be transformed)

$\sigma$  = scale parameter

$\varepsilon_p$  = pth percentile of the error distribution

Depending on the distribution,  $Y_p$  = failure time or log (failure time):

- For the Weibull, exponential, lognormal base<sub>e</sub>, lognormal base<sub>10</sub> and loglogistic distributions,  $Y_p$  = log (failure time)
- For the normal, extreme value, and logistic distributions,  $Y_p$  = failure time

When  $Y_p$  = log (failure time), MINITAB takes the antilog to display the percentiles on the original scale.

You can find the values for the y-intercept, the regression coefficient(s), and the shape or scale parameter in the regression table. When you enter predictor values in the Estimate subdialog box, the percentiles are displayed in the table of percentiles.

**Note** | You will often have more than one regression coefficient and predictor (X) with *Regression with Life Data* on page 16-19.

The value of the error distribution  $\varepsilon_p$  also depends on the distribution chosen.

- For the normal distribution, the error distribution is the standard normal distribution—normal (0,1). For the lognormal base<sub>10</sub> and lognormal base<sub>e</sub> distributions, MINITAB takes the log base<sub>10</sub> or log base<sub>e</sub> of the data, respectively, and uses a normal distribution.
- For the logistic distribution, the error distribution is the standard logistic distribution—logistic (0, 1). For the loglogistic distribution, MINITAB takes the log of the data and uses a logistic distribution.



- For the extreme value distribution, the error distribution is the standard extreme value distribution—extreme value (0, 1). For the Weibull distribution and the exponential distribution (a type of Weibull distribution), MINITAB takes the log of the data and uses the extreme value distribution.

## Probability plots

The Accelerated Life Testing command draws several probability plots to help you assess the fit of the chosen distribution. You can draw 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.

You can also choose to draw probability plots for each level of the accelerating variable based on individual fits or on the fitted model. You can use these plots to assess whether the distribution, transformation, and assumption of equal shape (Weibull or exponential) or scale (other distributions) are appropriate. The probability plot based on the fitted model includes fitted lines that are based on the chosen distribution and transformation. If the points do not fit the lines adequately, then consider a different transformation or distribution.

The probability plot based on the individual fits includes fitted lines that are calculated by individually fitting the distribution to each level of the accelerating variable. If the distributions have equal shape (Weibull or exponential) or scale (other distributions) parameters, then the fitted lines should be approximately parallel. The points should fit the line adequately if the chosen distribution is appropriate.

MINITAB provides one goodness-of-fit measure: the Anderson-Darling statistic. A smaller Anderson-Darling statistic indicates that the distribution provides a better fit. You can use the Anderson-darling statistic to compare the fit of competing models.

For a discussion of probability plots, see *Probability plots* on page 15-37.





► To draw a probability plot for each level of the accelerating variable

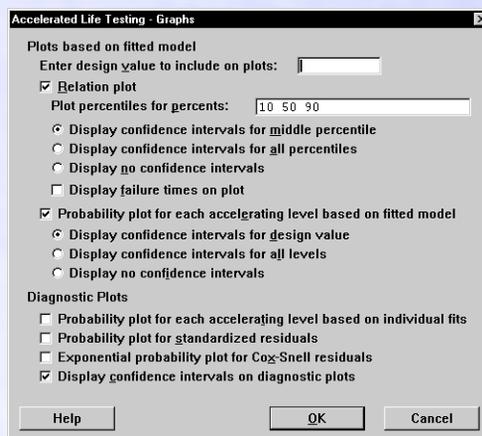
- 1 In the Accelerated Life Testing dialog box, click **Graphs**.
- 2 Do any of the following, then click **OK**.
  - To plot based on the fitted model, check **Probability plot for each accelerating level based on fitted model**. Choose one of the following:
    - Display confidence intervals for design value
    - Display confidence intervals for all levels
    - Display no confidence intervals

To include a design value on the fitted model plot, enter a value in **Enter a design value to include on plots**.

- To plot based on the individual fits, check **Probability plot for each accelerating level based on individual fits**.

► To draw a probability plot of the residuals

- 1 In the Accelerated Life Testing 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**

**Tip**

To draw a probability plot with more options, store the residuals in the Storage subdialog box, then use the probability plot included with *Parametric Distribution Analysis* on page 15-27.



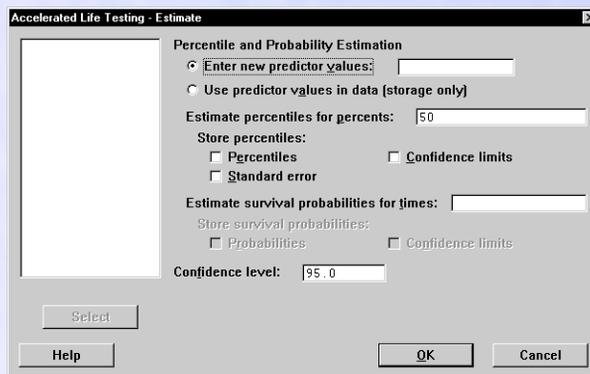
## Percentiles and survival probabilities

When doing accelerated life testing, you subject units to levels of an accelerating variable far exceeding normal field conditions to accelerate the failure process. But most likely, the information you ultimately want is, How do the units behave under normal field conditions?

In the Estimate subdialog box, you can ask MINITAB to extrapolate information gained from the accelerated situation to the *design value*, or common field condition.

### ► To estimate percentiles and survival probabilities

- 1 In the **Accelerated Life Testing** dialog box, click **Estimate**.



- 2 In **Enter new predictor values**, enter one new value or column of new values. Often you will enter the design value, or common running condition, for the units.
- 3 Do any of the following, then click **OK**:
  - To estimate percentiles, enter the 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 in **Estimate survival probabilities for times**. For example, when you enter 70 (units in hours), MINITAB estimates the probability, for each predictor value, that the unit will survive past 70 hours.

#### More

Sometimes you may want to estimate percentiles or survival probabilities for the accelerating variable levels used in the study:

In the Estimate subdialog box, choose **Use predictor values in data (storage only)**. Because of the potentially large amount of output, MINITAB stores the results in the worksheet rather than printing them in the Session window.



► **Example of accelerated life testing**

Suppose you want to investigate the deterioration of an insulation used for electric motors. The motors normally run between 80 and 100° C. To save time and money, you decide to use accelerated life testing.

First you gather failure times for the insulation at abnormally high temperatures—110, 130, 150, and 170° C—to speed up the deterioration. With failure time information at these temperatures, you can then extrapolate to 80 and 100° C. It is known that an Arrhenius relationship exists between temperature and failure time.

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** ► **Accelerated Life Testing**.
- 3 In **Variables/Start variables**, enter *FailureT*. In **Accelerating variable**, enter *Temp*.
- 4 From **Relationship**, choose **Arrhenius**.
- 5 Click **Censor**. In **Use censoring columns**, enter *Censor*, then click **OK**.
- 6 Click **Graphs**. In **Enter design value to include on plot**, enter *80*. Click **OK**.
- 7 Click **Estimate**. In **Enter new predictor values**, enter *Design*, then click **OK** in each dialog box.

*Session window output*

**Regression with Life Data: FailureT versus Temp**

Response Variable: FailureT

Censoring Information	Count
Uncensored value	66
Right censored value	14
Censoring value: Censor = C	

Estimation Method: Maximum Likelihood  
 Distribution: Weibull  
 Transformation on accelerating variable: Arrhenius

Regression Table

Predictor	Coef	Standard Error	Z	P	95.0% Normal CI	
					Lower	Upper
Intercept	-15.1874	0.9862	-15.40	0.000	-17.1203	-13.2546
Temp	0.83072	0.03504	23.71	0.000	0.76204	0.89940
Shape	2.8246	0.2570			2.3633	3.3760

Log-Likelihood = -564.693



Anderson-Darling (adjusted) Goodness-of-Fit

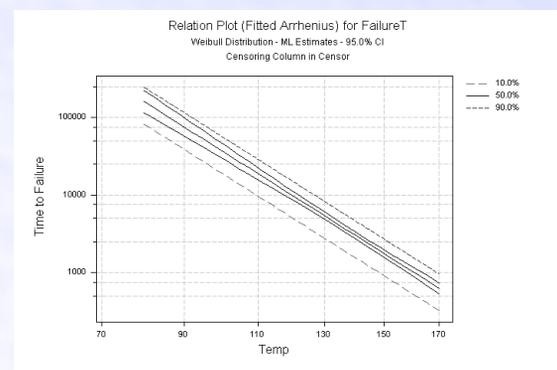
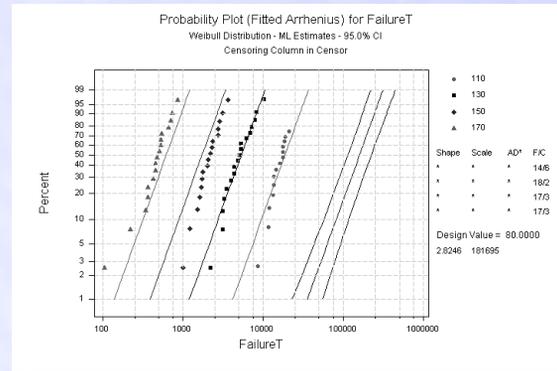
At each accelerating level

Level	Fitted Model
110	*
130	*
150	*
170	*

Table of Percentiles

Percent	Temp	Percentile	Standard Error	95.0% Normal CI	
				Lower	Upper
50	80.0000	159584.5	27446.85	113918.2	223557.0
50	100.0000	36948.57	4216.511	29543.36	46209.94

Graph window output





### Interpreting the results

From the Regression Table, you get the coefficients for the regression model. For a Weibull distribution, this model describes the relationship between temperature and failure time for the insulation:

$$\text{Log}_e(\text{failure time}) = -15.1874 + 0.83072 (\text{ArrTemp}) + 2.8246 \varepsilon_p$$

where  $\varepsilon_p$  = the  $p$ th percentile of the standard extreme value distribution

$$\text{ArrTemp} = \left( \frac{11604.83}{\text{Temp} + 273.16} \right)$$

The Table of Percentiles displays the 50th percentiles for the temperatures that you entered. The 50th percentile is a good estimate of how long the insulation will last in the field. At 80° C, the insulation lasts about 159,584.5 hours, or 18.20 years; at 100° C, the insulation lasts about 36,948.57 hours, or 4.21 years.

With the relation plot, you can look at the distribution of failure times for each temperature—in this case, the 10th, 50th, and 90th percentiles.

The probability plot based on the fitted model can help you determine whether the distribution, transformation, and assumption of equal shape (Weibull) at each level of the accelerating variable are appropriate. In this case, the points fit the lines adequately, thereby verifying that the assumptions of the model are appropriate for the accelerating variable levels.

## 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 which 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. See *How to specify the model terms* on page 16-24. Some of these terms may be *factors*.





## Data

Enter three types of columns in the worksheet:

- the *response variable* (failure times)—see *Failure times* on page 16-4.
- *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 estimate 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* on page 16-24.

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* on page 16-3.

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* on page 16-25.

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

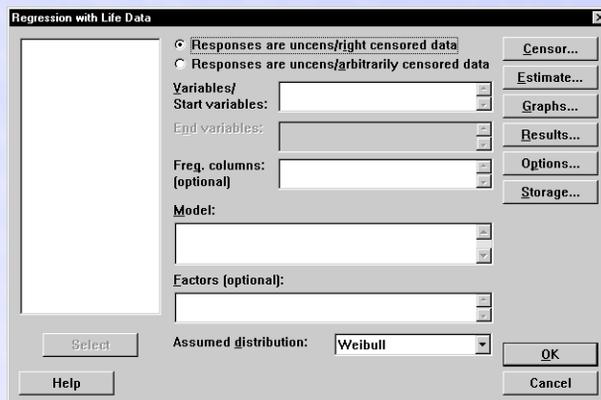
How you run the analysis depend on whether your data are uncensored/right censored or uncensored/arbitrarily censored.





► To perform regression with uncensored/right censored data

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



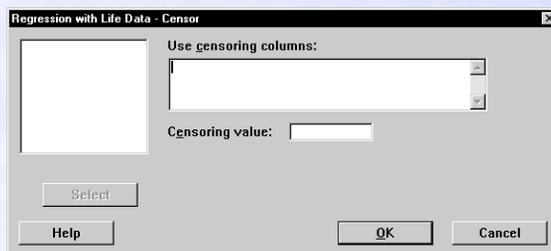
2 In **Variables/Start variables**, enter up to ten columns of failure times (10 different samples).

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* on page 16-24. If any of those predictors are factors, enter them again in **Factors**.

**Note** | If you have no censored values, 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.

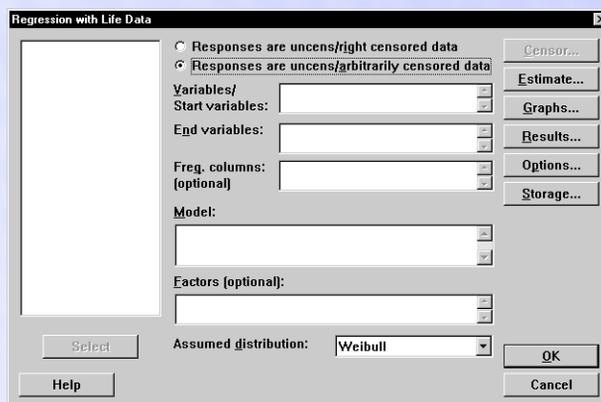
By default, MINITAB uses the lowest value in the censoring column to indicate a censored value. To use some other value, enter that value in **Censoring value**.

**Note** | If your censoring indicators are date/time values, store the values as constants and then enter them as constants.

7 If you like, use any of the options listed below, 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* on page 16-24. If any of those predictors are factors, enter them again in **Factors**.
- 7 If you like, use any of the options described below, then click **OK**.

## Options

### Regression with Life Data dialog box

- choose one of eight common lifetime distributions for the error distribution, including the Weibull (default), extreme value, exponential, normal, lognormal base<sub>e</sub>, lognormal base<sub>10</sub>, logistic, and loglogistic distributions.

**More** | MINITAB's extreme value distribution is the smallest extreme value (Type 1).

### Estimate subdialog box

- enter new predictor values for which to estimate percentiles and/or survival probabilities—see *Percentiles and survival probabilities* on page 16-27. You can also use the predictor values from the data.
- estimate percentiles for the percents you specify—see *Percentiles and survival probabilities* on page 16-27. By default, MINITAB estimates the 50th percentile.



- store the percentiles, their standard errors, and confidence intervals.
- estimate survival probabilities for the times you specify—see *Percentiles and survival probabilities* on page 16-27. For example, if you enter ten hours, MINITAB estimates (for each predictor value) the proportion of units that survive past ten hours.
- store the survival probabilities and confidence intervals.
- specify a confidence level for all of the confidence intervals. The default is 95.0%.

### Graphs subdialog box

- draw a probability plot for the standardized residuals or an exponential probability plot for the Cox-Snell residuals—*Probability plots* on page 16-26.
- choose not to include confidence intervals on the plots.

### Results subdialog box

- display the following Session window output:
  - no output.
  - the basic output, which includes the response information, censoring information, regression table, the log-likelihood, and goodness-of-fit measures.
  - the basic output, plus the table of percentiles and/or survival probabilities (the default).
  - the previous levels of output, plus the list of factor level values, and the tests for terms with more than one degree of freedom—see *Multiple degrees of freedom test* on page 16-28.
- show the log-likelihood for each iteration of the algorithm.

### Options subdialog box

- enter starting values for model parameters for the Newton-Raphson algorithm—see *Estimating the model parameters* on page 16-28.
- 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—see *Estimating the model parameters* on page 16-28.
- estimate other model coefficients while holding the shape parameter (Weibull) or the scale parameter (other distributions) fixed at a specific value—see *Estimating the model parameters* on page 16-28.
- use historical estimates for the parameters rather than estimate them from the data—see *Estimating the model parameters* on page 16-28. In this case, no estimation is done; all results—such as the percentiles—are based on these parameters.
- change the reference levels for the factors—see *Factor variables and reference levels* on page 16-25.





### Storage subdialog box

- store the ordinary, standardized, and Cox-Snell residuals.
- store the information on the estimated equation, including the coefficients, their standard errors and confidence intervals, the variance/covariance matrix, and the log-likelihood for the last iteration.

### 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 zero; in other words, is it a significant predictor?
  - 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.

### How to specify the model terms

You can fit models with:

- up to 9 factors and up to 50 covariates
- crossed or nested factors—see *Crossed vs. nested factors* on page 3-19
- covariates that are crossed with each other or with factors, or nested within factors

Here are some examples. A is a factor and X is a covariate.

#### Model terms

A X A*X	fits a full model with a covariate crossed with a factor
A   X	an alternative way to specify the previous model
A X X*X	fits a model with a covariate crossed with itself making a squared term
A X(A)	fits a model with a covariate nested within a factor

This model is a generalization of the model used in MINITAB's general linear model (GLM) procedure. Any model fit by GLM can also be fit by the life data procedures. For a general discussion of specifying models, see *Specifying the model terms* on page 3-20 and *Specifying reduced models* on page 3-21. In the regression with life data commands,





MINITAB assumes any variable in the model is a covariate unless the variable is specified as a factor. In contrast, GLM assumes any variable in the model is a factor unless the variable is specified as a covariate.

### Model restrictions

Life data models in MINITAB have the same restrictions as general linear models:

- The model must be *full rank*, meaning there must be enough data to estimate all the terms in your model. Suppose you have a two-factor crossed model with one empty cell. You can then fit the model with terms A B, but not A B A\*B. Do not worry about figuring out whether or not your model is of full rank. MINITAB will tell you if it is not. In most cases, eliminating some of the high order interactions in your model (assuming, of course, they are not important) will solve your problem.
- The model must be hierarchical. In a hierarchical model, if an interaction term is included, all lower order interactions and main effects that comprise the interaction term must appear in the model.

## 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)				Factor B with 3 levels (High Low Medium)		
reference level →	A1	A2	A3	reference level →	B1	B2
1	0	0	0	High	0	0
2	1	0	0	Low	1	0
3	0	1	0	Medium	0	1
4	0	0	1			

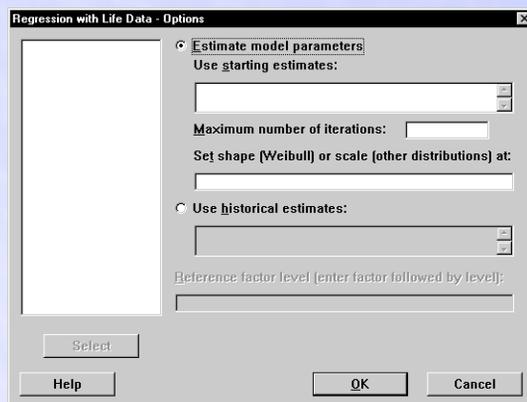
By default, MINITAB designates the lowest numeric, date/time, or text value as the reference factor level. If you like, you can change this reference value in the Options subdialog box.





► **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. Click **OK**.

## Probability plots

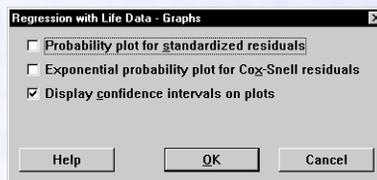
The Regression with Life Data command draws probability plots for the standardized and Cox-Snell residuals. You can use these plots to assess whether a particular distribution fits your data. In general, the closer the points fall to the fitted line, the better the fit.

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.

For a discussion of probability plots, see *Probability plots* on page 15-37.

► **To draw a probability plot of the residuals**

- 1 In the 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**

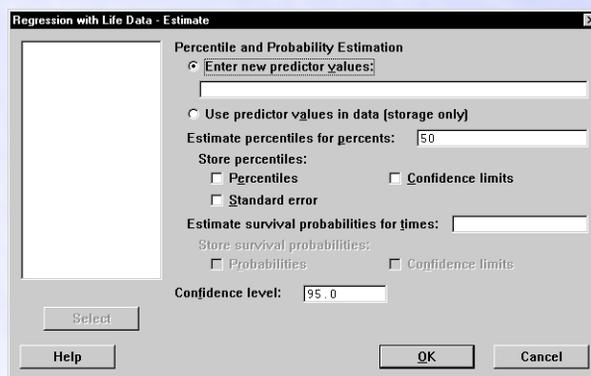
**Tip** | To draw a probability plot with more options, store the residuals in the Storage subdialog box, then use the probability plot included with *Parametric Distribution Analysis* on page 15-27.

## Percentiles and survival probabilities

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

### ► To estimate percentiles and survival probabilities

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* on page 16-30.
- 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 than 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.

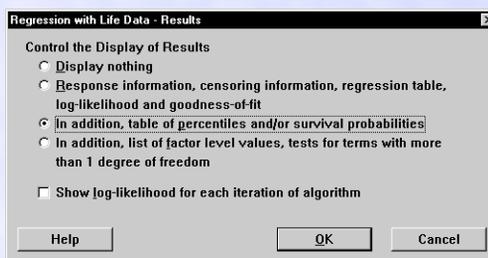
- To estimate survival probabilities, enter the times or a column of times in **Estimate survival probabilities for times**. For example, when you enter 70 (units in hours in this example), MINITAB estimates the probability, for each predictor value, that the unit will survive past 70 hours.

## 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?

### ► 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**.

## 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.



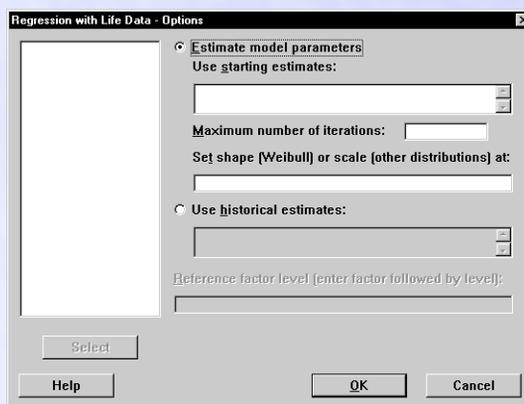
- estimate other model coefficients while holding the shape parameter (Weibull) or the scale parameter (other distributions) fixed at a specific value.

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 which 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.

► **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 (Weibull) or the scale (other distributions) parameter fixed: In **Set shape (Weibull) or scale parameter (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**.



### ► 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 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

Regression Table

Predictor	Coef	Standard Error	Z	P	95.0% Normal CI	
					Lower	Upper
Intercept	-15.3411	0.9508	-16.13	0.000	-17.2047	-13.4775
ArrTemp	0.83925	0.03397	24.71	0.000	0.77267	0.90584
Plant						
2	-0.18077	0.08457	-2.14	0.033	-0.34652	-0.01501
Shape	2.9431	0.2707			2.4577	3.5244

Log-Likelihood = -562.525

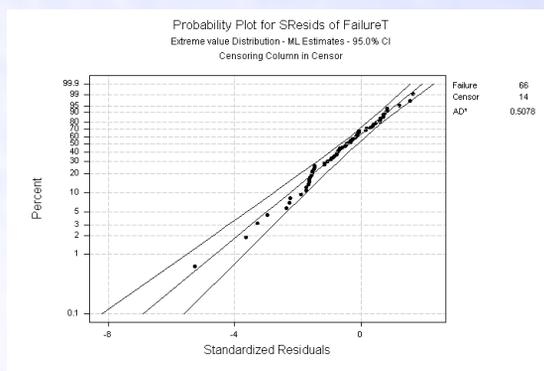
Anderson-Darling (adjusted) Goodness-of-Fit

Standardized Residuals = 0.5078

Table of Percentiles

Percent	Predictor		Standard Error	95.0% Normal CI	
	Row Number	Percentile		Lower	Upper
50	1	182093.6	32466.16	128389.8	258260.9
50	2	151980.8	25286.65	109689.6	210577.6
50	3	41530.38	5163.756	32548.44	52990.94
50	4	34662.51	3913.866	27781.00	43248.61

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:

$$\text{Log}_e(\text{failure time}) = -15.3411 + 0.83925 (\text{ArrTemp}) + 2.9431 \varepsilon_p$$

$$\text{Log}_e(\text{failure time}) = -15.52187 + 0.83925 (\text{ArrTemp}) + 2.9431 \varepsilon_p$$

where  $\varepsilon_p$  = the pth percentile of the error distribution

$$\text{ArrTemp} = \left( \frac{11604.83}{\text{Temp} + 273.16} \right)$$

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  $\alpha = .05$  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.

## References

- [1] J.D. Kalbfleisch and R.L. Prentice (1980). *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, Inc.
- [2] J.F. Lawless (1982). *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, Inc.
- [3] W.Q. Meeker and L.A. Escobar (1998). *Statistical Methods for Reliability Data*, John Wiley & Sons, Inc.
- [4] W. Murray (Ed.) (1972). *Numerical Methods for Unconstrained Optimization*, Academic Press.
- [5] W. Nelson (1990). *Accelerated Testing*, John Wiley & Sons, Inc.