15

Distribution Analysis

Distribution Analysis Overview

Use MINITAB's distribution analysis commands to understand the lifetime characteristics of a product, part, person, or organism. For instance, you might want to estimate how long a part is likely to last under different conditions, or how long a patient will survive after a certain type of surgery.

Your goal is to estimate the failure-time distribution of a product. You do this by estimating percentiles, survival probabilities, and distribution parameters and by drawing survival or hazard plots. You can use either parametric or nonparametric estimates. Parametric estimates are based on an assumed parametric distribution, while nonparametric estimates assume no parametric distribution.

Life data can be described using a variety of distributions. Once you have collected your data, you can use the commands in this chapter to select the best distribution to use for modeling your data, and then estimate the variety of functions that describe that distribution. These methods are called *parametric* because you assume the data follow a parametric distribution. If you cannot find a distribution that fits your data, MINITAB provides *nonparametric* estimates of the same functions.

Life data are often censored or incomplete in some way. Suppose you're testing how long a certain part lasts before wearing out and plan to cut off the study at a certain time. Any parts that did not fail before the study ended are censored, meaning their exact failure time is unknown. In this case, the failure is known only to be "on the right," or *after* the present time. This type of censoring is called right-censoring. Similarly, all you may know is that a part failed *before* a certain time (left-censoring), or *within* a certain interval of time (interval-censoring). When you know exactly when the part failed it is not censored, but is an exact failure.

Choosing a distribution analysis command

How do you know which distribution analysis command to use? You need to consider two things: 1) whether or not you can assume a parametric distribution for your data, and 2) the type of censoring you have.

- Use the **parametric distribution analysis** commands when you can assume your data follow a parametric distribution.
- Use the **nonparametric distribution analysis** commands when you cannot assume a parametric distribution.

Then, once you have decided which type of analysis to use, you need to choose whether you will use the right censoring or arbitrary censoring commands, which perform similar analyses.

- Use the **right-censoring** commands when you have exact failures and right-censored data.

- Use the **arbitrary-censoring** commands when your data include both exact failures and a varied censoring scheme, including right-censoring, left-censoring, and interval-censoring.

For details on creating worksheets for censored data, see *Distribution Analysis Data* on page 15-5.

Parametric distribution analysis commands

All parametric distribution analysis commands in this chapter can be used for both right censored and arbitrarily censored data. The parametric distribution analysis commands include Parametric Distribution Analysis, which performs the full analysis, and the specialty graphs, Distribution ID Plot and Distribution Overview Plot. The specialty graphs are often used before the full analysis to help choose a distribution or view summary information.

- **Distribution ID Plot—Right Censoring** and **Distribution ID Plot—Arbitrary Censoring** draw a layout of up to four probability plots, from your choice of eight common distributions: Weibull, extreme value, exponential, normal, lognormal base_e, lognormal base_{10}, logistic, and loglogistic. The layout helps you determine which, if any, of the parametric distributions best fits your data. See *Distribution ID Plot* on page 15-9.
- **Distribution Overview Plot—Right Censoring** and **Distribution Overview Plot—Arbitrary Censoring** draw a probability plot, probability density function, survival plot, and hazard plot on one page. The layout helps you assess the fit of the chosen distribution and view summary graphs of your data. See *Distribution Overview Plot* on page 15-19.
- **Parametric Distribution Analysis—Right Censoring** and **Parametric Distribution Analysis—Arbitrary Censoring** fit one of eight common parametric distributions to your data, then use that distribution to estimate percentiles and survival probabilities, and draw survival, hazard, and probability plots. See *Parametric Distribution Analysis* on page 15-27.

Nonparametric distribution analysis commands

The nonparametric distribution analysis commands include Nonparametric Distribution Analysis—Right Censoring and Nonparametric Distribution Analysis—Arbitrary Censoring, which perform the full analysis, and the specialty graph—Distribution Overview Plot—Right Censoring and Distribution Overview Plot—Arbitrary Censoring. Distribution Overview Plot is often used before the full analysis to view summary information.

- **Distribution Overview Plot** (uncensored/right censored data only) draws a Kaplan-Meier survival plot and hazard plot, or an Actuarial survival plot and hazard plot, on one page. See *Distribution Overview Plot* on page 15-19.

CONTENTS INDEX MEET MTB UGUIDE 1 UGUIDE 2 SC QREF HOW TO USE

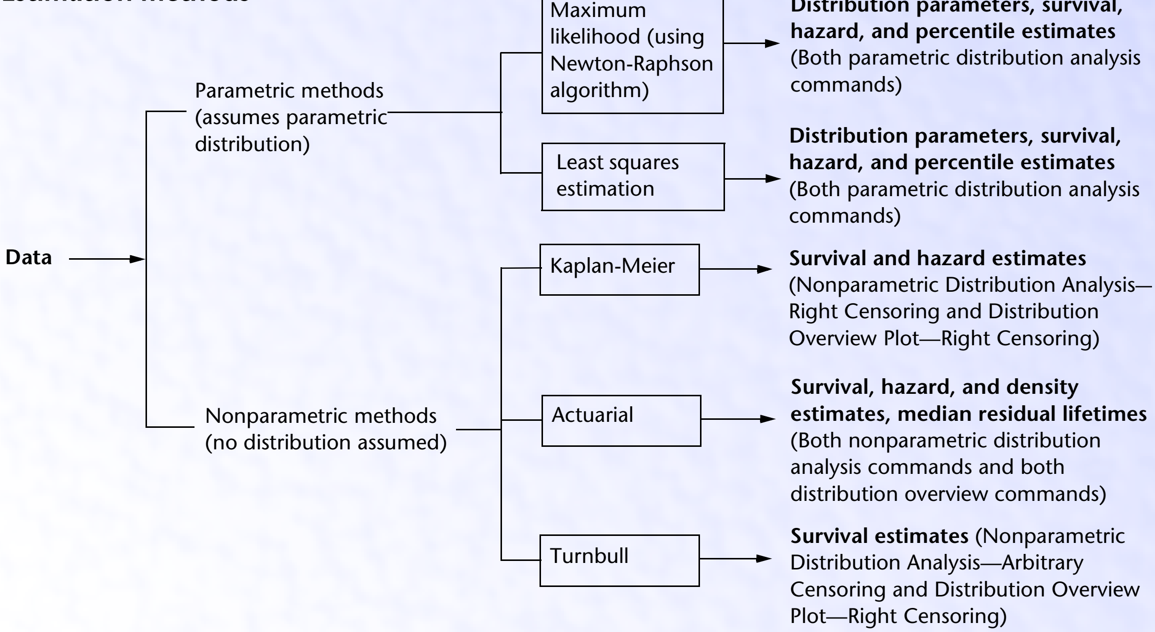
- **Distribution Overview Plot** (uncensored/arbitrarily censored data only) draws a Turnbull survival plot or an Actuarial survival plot and hazard plot. See *Distribution Overview Plot* on page 15-19.
- **Nonparametric Distribution Analysis—Right Censoring** and **Nonparametric Distribution Analysis—Arbitrary Censoring** give you nonparametric estimates of the survival probabilities, hazard estimates, and other estimates depending on the nonparametric technique chosen, and draw survival and hazard plots. When you have multiple samples, Nonparametric Distribution Analysis—Right Censoring also tests the equality of their survival curves. See *Nonparametric Distribution Analysis* on page 15-52.

Estimation methods

As described above, MINITAB provides both parametric and nonparametric methods to estimate functions. If a parametric distribution fits your data, then use the parametric estimates. If no parametric distribution adequately fits your data, then use the nonparametric estimates.

For the parametric estimates in this chapter, you can choose either the maximum likelihood method or least squares approach. Nonparametric methods differ, depending on the type of censoring. For the formulas used, see Help.

Estimation methods



Distribution Analysis Data

The data you gather for the commands in this chapter are individual failure times. For example, you might collect failure times for units running at a given temperature. You might also collect samples of failure times under different temperatures, or under different combinations of stress variables.

Life data are 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. This table 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 *Distribution analysis—right censored data* on page 15-5.
- when your data have exact failures and a varied censoring scheme, including right-censoring, left-censoring, and interval-censoring, see *Distribution analysis—arbitrarily censored data* on page 15-8.

Distribution analysis—right censored data

Right-censored data can be **singly** or **multiply censored**. Singly censored means that the censored items all ran for the same amount of time, and all of the exact failures occurred earlier than that censoring time. Multiply censored means that items were censored at different times, with failure times intermixed with those censoring times.

Multiply censored data are more common in the field, where units go into service at different times. Singly censored data are more common in controlled studies.

In these two examples, the Months column contains failure times, and the Censor column contains indicators that say whether that failure was censored (C) or an exact failure time (F):

These units had not failed and dropped out of the study before it finished. The data set is **multiply censored** because censoring times (C) intermix with failure times (F).

Months	Censor
50	F
53	F
60	C
65	C
70	F
70	F
50	F
53	F
⋮	⋮
etc.	etc.

Months	Censor
50	F
50	F
53	F
53	F
60	F
65	F
70	C
70	C
⋮	⋮
etc.	etc.

This data set is **singly censored**—specifically, it's **time censored** at 70 months, meaning any observation greater than or equal to 70 months is considered censored.

Singly censored data can be either:

- **time censored**, meaning that you run the study for a specified period of *time*. All units still running at the end time are time censored. This is known as Type I censoring on the right.
- **failure censored**, meaning that you run the study until you observe a specified *number of failures*. All units running from the last specified failure onward are failure censored. This is known as Type II censoring on the right.

Worksheet structure

Do one of the following, depending on the type of censoring you have:

Singly censored data

- to use a constant *failure time* to define censoring, enter a column of failure times for each sample. Later, when executing the command, you will specify the failure time at which to begin censoring.
- to use a specified *number of failures* to define censoring, enter a column of failure times for each sample. Later, when executing the command, you will specify the number of failures at which to begin censoring.

Singly or multiply censored data

- to use censoring columns to define censoring, enter two columns for each sample—one column of failure times and a corresponding column of censoring indicators. You must use this method for multiply censored data.

  Censoring indicators can be numbers or text. If you don't 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 and censor columns (from different samples) can have different lengths.

This data set uses censoring columns:

This column contains failure times for engine windings in a turbine assembly.

Months	Censor
50	F
60	F
53	F
40	F
51	F
99	C
35	F
55	F
⋮	⋮
etc.	etc.

This 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.

Using frequency columns

You can structure each column so that it contains individual observations (one row = one observation), as shown above, or unique observations with a corresponding column of frequencies (counts).

Here are the same data structured both ways:

Here we have four failures at 150 days.

Days	Censor
140	F
150	F
150	F
150	F
150	F
151	C
151	F
151	F
⋮	⋮
etc.	etc.

Days	Censor	Freq
140	F	1
150	F	4
151	C	1
151	F	35
153	F	42
161	C	1
170	F	39
199	F	1
⋮	⋮	⋮
etc.	etc.	etc.

Here we have four failures at 150 days.

Frequency columns are useful for data where you have large numbers of observations with common failure and censoring times. For example, warranty data usually includes large numbers of observations with common censoring times.

Stacked vs. unstacked data

In the discussion so far, we have shown illustrations of unstacked data: that is, data from different samples are in separate columns. You can optionally stack all of the data in one column, then set up a column of grouping indicators. The grouping indicators define each sample. Grouping indicators, like censoring indicators, can be numbers or text.

Here is the same data set structured both ways:

Unstacked data

Drug A	Drug B
20	2
30	3
43	6
51	14
57	24
82	26
85	27
89	31

Stacked data

Drug	Group
20	A
30	A
43	A
51	A
57	A
82	A
85	A
89	A
2	B
3	B
6	B
14	B
24	B
26	B
27	B
31	B

Note You cannot analyze more than one column of stacked data per analysis. So when you use grouping indicators, the data for each sample must be in one column.

Distribution analysis—arbitrarily censored data

Arbitrarily-censored data includes exact failure times and a varied censoring scheme, including right, left, and interval censored data. Enter your data in table form, 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 that the failure occurred after	the missing value symbol '*'
Left censored	the missing value symbol '*'	time that the failure occurred before
Interval censored	time at start of interval during which the failure occurred	time at end of interval during which the failure occurred

CONTENTS INDEX MEET MTB UGUIDE 1 UGUIDE 2 SC QREF HOW TO USE

This data set illustrates tabled data, as well as the use of a frequency column. Frequency columns are described in *Using frequency columns* on page 15-7.

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.

When you have more than one sample, you can use separate columns for each sample. Alternatively, you can stack all of the samples in one column, then set up a column of grouping indicators. Grouping indicators can be numbers or text. For an illustration, see *Stacked vs. unstacked data* on page 15-8.

Distribution ID Plot

Use Distribution ID Plot to plot up to four different probability plots (with distributions chosen from Weibull, extreme value, exponential, normal, lognormal $base_e$, lognormal $base_{10}$, logistic, and loglogistic) to help you determine which of these distributions best fits your data. Usually this is done by comparing how closely the plot points lie to the best-fit lines—in particular those points in the tails of the distribution.

MINITAB also provides two goodness-of-fit tests—Anderson-Darling for the maximum likelihood and least squares estimation methods and Pearson correlation coefficient for the least squares estimation method—to help you assess how the distribution fits your data. See *Goodness-of-fit statistics* on page 15-13.

The data you gather are the individual failure times, which may be censored. For example, you might collect failure times for units running at a given temperature. You might also collect samples of failure times under different temperatures, or under varying conditions of any combination of stress variables.

You can display up to ten samples on each plot. All of the samples display on a single plot, in different colors and symbols.

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

Data

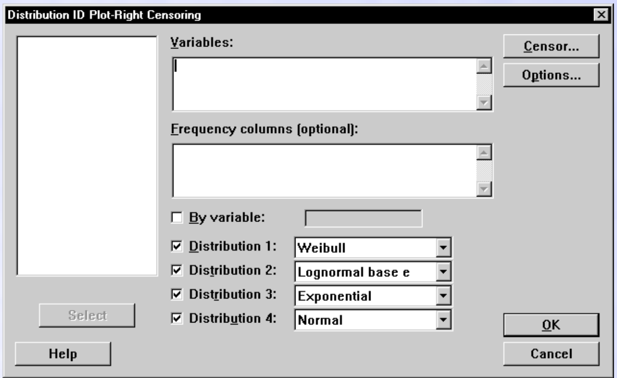
Distribution ID Plot accepts different kinds of data:

- Distribution ID Plot—Right Censoring accepts exact failure times and right censored data. For information on how to set up your worksheet see *Distribution analysis—right censored data* on page 15-5.
- Distribution ID Plot—Arbitrary Censoring accepts exact failure times and right-, left-, and interval-censored data. For information on how to set up your worksheet see *Distribution analysis—arbitrarily censored data* on page 15-8.

You can enter up to ten samples per analysis. For general information on life data and censoring, see *Distribution Analysis Data* on page 15-5.

▶ To make a distribution ID plot (uncensored/right censored data)

1 Choose **Stat ➤ Reliability/Survival ➤ Distribution ID Plot–Right Cens.**



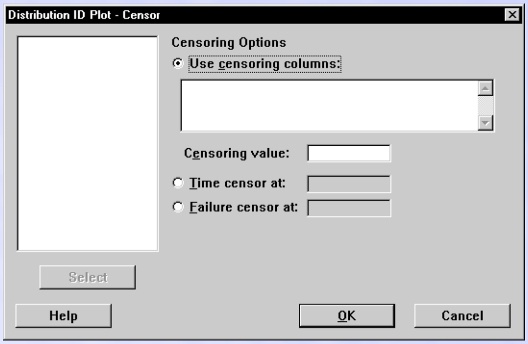
2 In **Variables**, enter the columns of failure times. You can enter up to ten columns (ten different samples).

3 If you have frequency columns, enter the columns in **Frequency columns**.

4 If all of the samples are stacked in one column, check **By variable**, and enter a column of grouping indicators in the box.

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

5 Click **Censor**.



6 Do one of the following, then click **OK**.

- For data with censoring columns: Choose **Use censoring columns**, then enter the censoring columns in the box. 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.

  If you like, enter the value you use to indicate censoring in **Censoring value**. If you do not enter a value, MINITAB uses the lowest value in the censoring column.

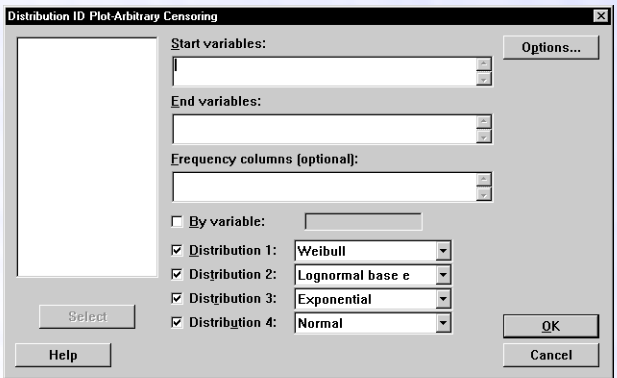
- For time censored data: Choose **Time censor at**, then enter a failure time at which to begin censoring. For example, entering *500* says that any observation from 500 time units onward is considered censored.

- For failure censored data: Choose **Failure censor at**, then enter a number of failures at which to begin censoring. For example, entering *150* says to censor all (ordered) observations from the 150th observed failure on, and leave all other observations uncensored.

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

▶ To make a distribution ID plot (arbitrarily censored data)

1 Choose **Stat ➤ Reliability/Survival ➤ Distribution ID Plot–Arbitrary Cens.**



2 In **Start variables**, enter the column of start times. You can enter up to ten columns (ten different samples).

3 In **End variables**, enter the column of end times. You can enter up to ten columns (ten different samples). The first start column is paired with the first end column, the second start column is paired with the second end column, and so on.

4 If you have frequency columns, enter the columns in **Frequency columns**.

5 If all of the samples are stacked in one column, check **By variable**, and enter a column of grouping indicators in the box.

6 If you like, use any of the options described below, then click **OK**.

Options

Distribution ID Plot dialog box

- choose to create up to four probability plots. The default is to create four plots.
- choose to fit up to four common lifetime distributions for the parametric analysis, including the Weibull, extreme value, exponential, normal, lognormal base$_e$, lognormal base$_{10}$, logistic, and loglogistic distributions. The four default distributions are Weibull, lognormal base$_e$, exponential, and normal.

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

Options subdialog box

- estimate parameters using the maximum likelihood (default) or least squares methods.
- estimate percentiles for additional percents. The default is 1, 5, 10, and 50.
- obtain the plot points for the probability plot using various nonparametric methods—see *Probability plots* on page 15-37.
  - With Distribution ID Plot—Right Censoring, you can choose the Default method, Modified Kaplan-Meier method, Herd-Johnson method, or Kaplan-Meier method. The Default method is the normal score for uncensored data; the modified Kaplan-Meier method for censored data.
  - With Distribution ID Plot—Arbitrary Censoring, you can choose the Turnbull or Actuarial method. The Turnbull method is the default.
- (Distribution ID Plot—Right Censoring only) handle ties by plotting all of the points (default), the maximum of the tied points, or the average (median) of the tied points.
- enter minimum and/or maximum values for the x-axis scale.
- replace the default graph title with your own title.

Output

The default output consists of:

- goodness-of-fit statistics for the chosen distributions—see *Goodness-of-fit statistics* on page 15-13
- table of percents and their percentiles, standard errors, and 95% confidence intervals
- table of MTTFs (mean time to failures) and their standard errors and 95% confidence intervals
- four probability plots for the Weibull, lognormal base$_e$, exponential, and normal distributions

Goodness-of-fit statistics

MINITAB provides two goodness-of-fit statistics—Anderson-Darling for the maximum likelihood and least squares estimation methods and Pearson correlation coefficient for the least squares estimation method—to help you compare the fit of competing distributions.

The Anderson-Darling statistic is a measure of how far the plot points fall from the fitted line in a probability plot. The statistic is a weighted squared distance from the plot points to the fitted line with larger weights in the tails of the distribution. Minitab uses an adjusted Anderson-Darling statistic, because the statistic changes when a different plot point method is used. A smaller Anderson-Darling statistic indicates that the distribution fits the data better.

For least squares estimation, Minitab calculates a Pearson correlation coefficient. If the distribution fits the data well, then the plot points on a probability plot will fall on a straight line. The correlation measures the strength of the linear relationship between the X and Y variables on a probability plot. The correlation will range between 0 and 1, and higher values indicate a better fitting distribution.

Use the Anderson-Darling statistic and Pearson correlation coefficient to compare the fit of different distributions.

▶ Example of a distribution ID plot for right-censored data

Suppose you work for a company that manufactures engine windings for turbine assemblies. Engine windings may decompose at an unacceptable rate at high temperatures. You want to know—at given high temperatures—the time at which 1% of the engine windings fail. You plan to get this information by using the Parametric Distribution Analysis—Right Censoring command, which requires you to specify the distribution for your data. Distribution ID Plot—Right Censoring can help you choose that distribution.

First you collect failure times for the engine windings at two temperatures. In the first sample, you test 50 windings at 80° C; in the second sample, you test 40 windings at 100° C. Some of the units drop out of the test for unrelated reasons. In the MINITAB worksheet, you use a column of censoring indicators to designate which times are actual failures (1) and which are censored units removed from the test before failure (0).

1 Open the worksheet RELIABLE.MTW.

2 Choose **Stat ➤ Reliability/Survival ➤ Distribution ID Plot—Right Cens.**

3 In **Variables**, enter *Temp80 Temp100*.

4 Click **Censor**. Choose **Use censoring columns** and enter *Cens80 Cens100* in the box. Click **OK** in each dialog box.

Session window output

Distribution ID Plot

```
Variable:  Temp80

Goodness of Fit

Distribution        Anderson-Darling
Weibull             67.64
Lognormal base e    67.22
Exponential         70.33
Normal              67.73
```


CONTENTS INDEX MEET MTB UGUIDE 1 UGUIDE 2 SC QREF HOW TO USE

Table of Percentiles

Distribution	Percent	Percentile	Standard	95.0% Normal	CI
			Error	Lower	Upper
Weibull	1	10.0765	2.78453	5.8626	17.3193
Lognormal base e	1	19.3281	2.83750	14.4953	25.7722
Exponential	1	0.8097	0.13312	0.5867	1.1176
Normal	1	-0.5493	8.37183	-16.9578	15.8592
Weibull	5	20.3592	3.79130	14.1335	29.3273
Lognormal base e	5	26.9212	3.02621	21.5978	33.5566
Exponential	5	4.1326	0.67939	2.9942	5.7037
Normal	5	18.2289	6.40367	5.6779	30.7798

------the rest of this table omitted for space-----

Table of MTTF

Distribution	Mean	Standard	95% Normal	CI
		Error	Lower	Upper
Weibull	64.9829	4.6102	56.5472	74.677
Lognormal base e	67.4153	5.5525	57.3656	79.225
Exponential	80.5676	13.2452	58.3746	111.198
Normal	63.5518	4.0694	55.5759	71.528

Variable: Temp100

Goodness of Fit

Distribution	Anderson-Darling
Weibull	16.60
Lognormal base e	16.50
Exponential	18.19
Normal	17.03

Table of Percentiles

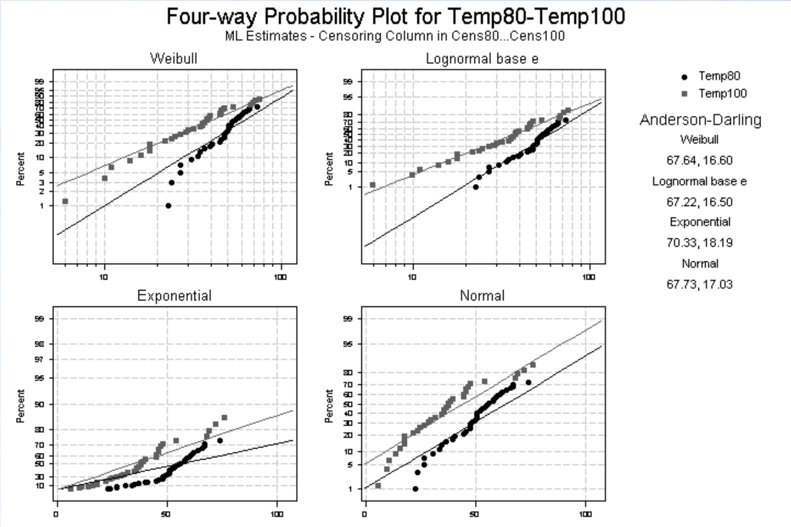
Distribution	Percent	Percentile	Standard	95.0% Normal	CI
			Error	Lower	Upper
Weibull	1	2.9819	1.26067	1.3020	6.8290
Lognormal base e	1	6.8776	1.61698	4.3383	10.9034
Exponential	1	0.5025	0.08618	0.3591	0.7033
Normal	1	-18.8392	8.80960	-36.1057	-1.5727
Weibull	5	8.1711	2.36772	4.6306	14.4189
Lognormal base e	5	11.3181	2.07658	7.8995	16.2162
Exponential	5	2.5647	0.43984	1.8325	3.5893
Normal	5	-0.2984	6.86755	-13.7585	13.1618

-----the rest of this table omitted for space-----

Table of MTTF

		Standard	95% Normal	CI
Distribution	Mean	Error	Lower	Upper
Weibull	45.9448	4.87525	37.3177	56.5663
Lognormal base e	49.1969	6.91761	37.3465	64.8076
Exponential	50.0000	8.57493	35.7265	69.9761
Normal	44.4516	4.37371	35.8793	53.0240

Graph window output



Interpreting the results

The points fall approximately on the straight line on the lognormal probability plot, so the lognormal base$_e$ distribution would be a good choice when running the parametric distribution analysis. You can also compare the Anderson-Darling goodness-of-fit values to determine which distribution best fits the data. A smaller Anderson-Darling statistic means that the distribution provides a better fit. Here, the Anderson-Darling values for the lognormal base$_e$ distribution are lower than the Anderson-Darling values for other distributions, thus supporting your conclusion that the lognormal base$_e$ distribution provides the best fit.

The table of percentiles and MTTFs allow you to see how your conclusions may change with different distributions.

▶ Example of a distribution ID plot for arbitrarily-censored data

Suppose you work for a company that manufactures tires. You are interested in finding out how many miles it takes for various proportions of the tires to “fail,” or wear down to 2/32 of an inch of tread. You are especially interested in knowing how many of the tires last past 45,000 miles. You plan to get this information by using the Parametric

Distribution Analysis—Arbitrary Censoring command, which requires you to specify the distribution for your data. Distribution ID Plot—Arbitrary Censoring can help you choose that distribution.

You inspect each good tire at regular intervals (every 10,000 miles) to see if the tire has failed, then enter the data into the MINITAB worksheet.

1 Open the worksheet TIREWEAR.MTW.

2 Choose **Stat ➤ Reliability/Survival ➤ Distribution ID Plot–Arbitrary Cens.**

3 In **Start variables**, enter *Start*. In **End variables**, enter *End*.

4 In **Frequency columns**, enter *Freq*.

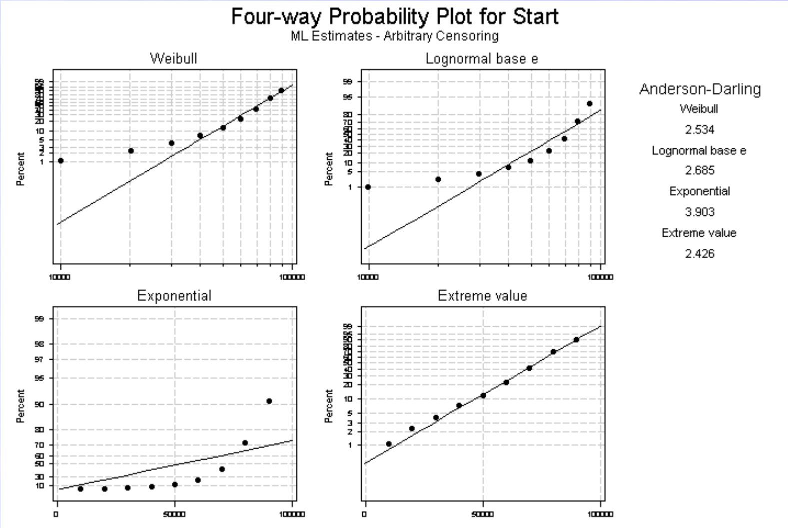
5 Under **Distribution 4**, choose **Extreme value**. Click **OK**.

Session window output

Distribution ID Plot

```
Variable
Start:  Start    End:  End
Frequency:  Freq

Goodness of Fit

Distribution       Anderson-Darling
Weibull            2.534
Lognormal base e   2.685
Exponential        3.903
Extreme value      2.426

Table of Percentiles

                                            Standard  95.0%  Normal  CI
Distribution       Percent  Percentile  Error     Lower          Upper
Weibull             1       27623.0      998.00   25734.6        29650.0
Lognormal base e    1       27580.2      781.26   26090.7        29154.8
Exponential         1         762.4       28.80     708.0          821.0
Extreme value       1       13264.5     2216.24    8920.8        17608.3

Weibull             5       39569.8      975.59   37703.1        41528.9
Lognormal base e    5       35793.9      795.52   34268.2        37387.6
Exponential         5        3891.0      146.96    3613.4         4190.0
Extreme value       5       36038.3     1522.71   33053.9        39022.8

-----the rest of this table omitted for space-----

Table of MTTF

                            Standard  95%  Normal  CI
Distribution       Mean     Error     Lower         Upper
Weibull            69545.4   629.34   68322.8       70789.9
Lognormal base e   72248.6  1066.42   70188.4       74369.3
Exponential        75858.8  2865.18   70446.0       81687.6
Extreme value      69473.3   646.64   68205.9       70740.7
```

Graph window output



Interpreting results

The points fall approximately on the straight line on the extreme value probability plot, so the extreme value distribution would be a good choice when running the parametric distribution analysis.

You can also compare the Anderson-Darling goodness-of-fit values to determine which distribution best fits the data. A smaller Anderson-Darling statistic means that the distribution provides a better fit. Here, the Anderson-darling values for the extreme value distribution are lower than the Anderson-Darling values for other distributions, thus supporting your conclusion that the extreme value distribution provides the best fit.

The table of percentiles and MTTFs allow you to see how your conclusions may change with different distributions.

Distribution Overview Plot

Use Distribution Overview Plot to generate a layout of plots that allow you to view your life data in different ways on one page. You can draw a *parametric* overview plot by selecting a distribution for your data, or a *nonparametric* overview plot.

The parametric display includes a probability plot (for a selected distribution), a survival (or reliability) plot, a probability density function, and a hazard plot. The nonparametric display depends on the type of data: if you have right-censored data MINITAB displays a Kaplan-Meier survival plot and a hazard plot or an Actuarial survival plot and hazard plot, and if you have arbitrarily-censored data, MINITAB displays a Turnbull survival plot or an Actuarial survival plot and hazard plot. These functions are all typical ways of describing the distribution of failure time data.

The data you gather are the individual failure times, some of which may be censored. For example, you might collect failure times for units running at a given temperature. You might also collect samples of failure times under different temperatures, or under various combination of stress variables.

You can enter up to ten samples per analysis. MINITAB estimates the functions independently for each sample. All of the samples display on a single plot, in different colors and symbols, which helps you compare their various functions.

To draw these plots with more information, see *Parametric Distribution Analysis* on page 15-27 or *Nonparametric Distribution Analysis* on page 15-52.

Data

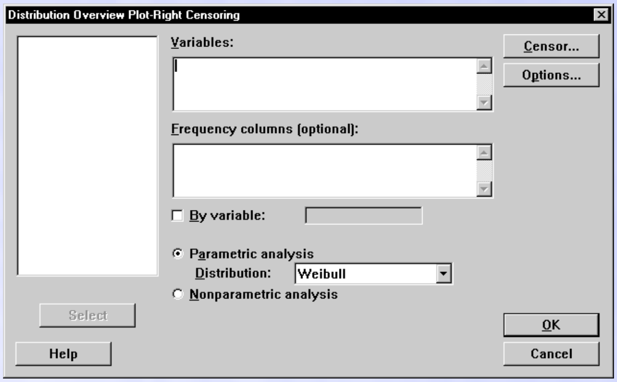
Distribution Overview Plot accepts different kinds of data:

- Distribution Overview Plot—Right Censoring accepts exact failure times and right censored data. For information on how to set up your worksheet, see *Distribution analysis—right censored data* on page 15-5.
- Distribution Overview Plot—Arbitrary Censoring accepts exact failure times and right-, left-, and interval-censored data. The data must be in tabled form. For information on how to set up your worksheet, see *Distribution analysis—arbitrarily censored data* on page 15-8.

You can enter up to ten samples per analysis. For general information on life data and censoring, see *Distribution Analysis Data* on page 15-5.

▶ To make a distribution overview plot (uncensored/right-censored data)

1 Choose **Stat ➤ Reliability/Survival ➤ Distribution Overview Plot—Right Censoring**.



2 In **Variables**, enter the columns of failure times. You can enter up to ten columns (ten different samples).

3 If you have frequency columns, enter the columns in **Frequency columns**.

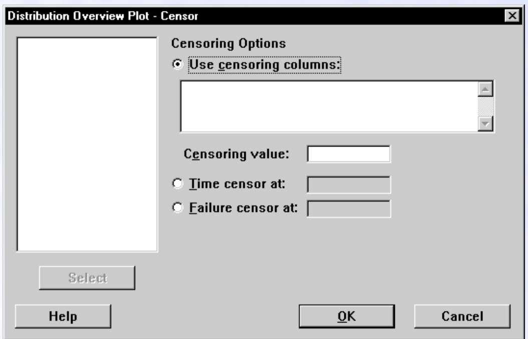
4 If all of the samples are stacked in one column, check **By variable**, and enter a column of grouping indicators in the box.

5 Choose to draw a parametric or nonparametric plot:

- *Parametric plot*—Choose **Parametric analysis**. From **Distribution**, choose to plot one of the eight available distributions.
- *Nonparametric plot*—Choose **Nonparametric analysis**.

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

6 Click **Censor**.



7 Do one of the following, then click **OK**.

- For data with censoring columns: Choose **Use censoring columns**, then enter the censoring columns in the box. 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.

  If you like, enter the value you use to indicate censoring in **Censoring value**. If you don't enter a value, by default MINITAB uses the lowest value in the censoring column.

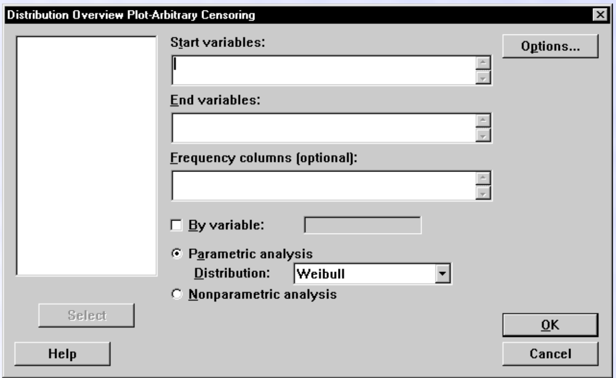
- For time censored data: Choose **Time censor at**, then enter a failure time at which to begin censoring. For example, entering *500* says that any observation from 500 time units onward is considered censored.

- For failure censored data: Choose **Failure censor at**, then enter a number of failures at which to begin censoring. For example, entering *150* says to censor all (ordered) observations from the 150th observed failure on, and to leave all other observations uncensored.

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

▶ To make a distribution overview plot (arbitrarily censored data)

1 Choose **Stat ➤ Reliability/Survival ➤ Distribution Overview Plot–Arbitrarily Censored**.



2 In **Start variables**, enter the columns of start times. You can enter up to ten columns (ten different samples).

3 In **End variables**, enter the columns of end times. You can enter up to ten columns (ten different samples).

4 If you have frequency columns, enter the columns in **Frequency columns**.

5 If all of the samples are stacked in one column, check **By variable**, and enter a column of grouping indicators in the box.

6 Choose to draw a parametric or nonparametric plot:

- *Parametric plot*—Choose **Parametric analysis**. From **Distribution**, choose to plot one of eight distributions.
- *Nonparametric plot*—Choose **Nonparametric analysis**.

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

Options

Distribution Overview Plot dialog box

- for the parametric display of plots, choose one of eight common lifetime distributions for the data—Weibull (default), extreme value, exponential, normal, lognormal base$_e$, lognormal base$_{10}$, logistic, or loglogistic.

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

- draw a nonparametric display of plots.

Options subdialog box (right censoring)

When you have chosen to conduct a parametric analysis

- estimate parameters using the maximum likelihood (default) or least squares methods
- obtain the plot points for the probability plot using various nonparametric methods—see *Probability plots* on page 15-37. You can choose the Default method, Modified Kaplan-Meier, Herd-Johnson, or Kaplan-Meier method. The Default method is the normal score for uncensored data; the modified Kaplan-Meier method is for censored data.
- handle ties by plotting all of the points (default), the maximum of the tied points, or the average (median) of the tied points.

When you have chosen to conduct a nonparametric analysis:

- estimate parameters using the Kaplan-Meier method (default) or Actuarial method.

For both parametric and nonparametric analyses:

- enter minimum and/or maximum values for the x-axis scale.
- replace the default graph title with your own title.

Options subdialog box (arbitrary censoring)

When you have chosen to conduct a parametric analysis:

- estimate parameters using the maximum likelihood (default) or least squares methods
- obtain the plot points for the probability plot using various nonparametric methods—see *Probability plots* on page 15-37. You can choose from the Turnbull method (default) or Actuarial method.

When you have chosen to conduct a nonparametric analysis:

- estimate parameters using the Turnbull method (default) or Actuarial method (default).

For both parametric and nonparametric analyses:

- enter minimum and/or maximum values for the x-axis scale.
- replace the default graph title with your own title.

Output

The distribution overview plot display differs depending on whether you select the parametric or nonparametric display.

When you select a parametric display, you get:

- goodness-of-fit statistics for the chosen distribution.
- a probability plot, which displays estimates of the cumulative distribution function F(y) vs. failure time—see *Probability plots* on page 15-37.
- a parametric survival (or reliability) plot, which displays the survival (or reliability) function 1–F(y) vs. failure time—see *Survival plots* on page 15-40.
- a probability density function, which displays the curve that describes the distribution of your data, or f(y).
- a parametric hazard plot, which displays the hazard function or instantaneous failure rate, f(y)/(1–F(y)) vs. failure time—see *Hazard plots* on page 15-41.

When you select a nonparametric display, you get:

- For right-censored data with Kaplan-Meier method
  - a Kaplan-Meier survival plot
  - a nonparametric hazard plot based on the empirical hazard function
- For right-censored data with Actuarial method
  - an Actuarial survival plot
  - a nonparametric hazard plot based on the empirical hazard function

- For arbitrarily-censored data with Turnbull method
  - a Turnbull survival plot
- For arbitrarily-censored data with Actuarial method
  - an Actuarial survival plot
  - a nonparametric hazard plot based on the empirical hazard function

The Kaplan-Meier survival estimates, Turnbull survival estimates, and empirical hazard function change values only at exact failure times, so the nonparametric survival and hazard curves are step functions. Parametric survival and hazard estimates are based on a fitted distribution and the curve will therefore be smooth.

▶ Example of a distribution overview plot with right-censored data

Suppose you work for a company that manufactures engine windings for turbine assemblies. Engine windings may decompose at an unacceptable rate at high temperatures. You want to know, at given high temperatures, at what time do 1% of the engine windings fail. You plan to get this information by using the Parametric Distribution Analysis—Right Censoring command, but you first want to have a quick look at your data from different perspectives.

First you collect data for times to failure for the engine windings at two temperatures. In the first sample, you test 50 windings at 80° C; in the second sample, you test 40 windings at 100° C. Some of the units drop out of the test due to failures from other causes. These units are considered to be right censored because their failures were not due to the cause of interest. In the MINITAB worksheet, you use a column of censoring indicators to designate which times are actual failures (1) and which are censored units removed from the test before failure (0).

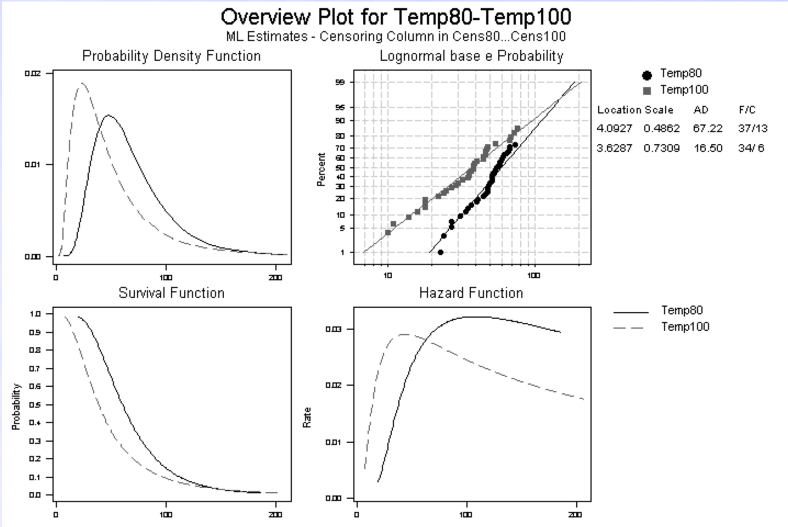
1. Open the worksheet RELIABLE.MTW.
2. Choose **Stat ➤ Reliability/Survival ➤ Distribution Overview Plot—Right Cens**.
3. In **Variables**, enter *Temp80 Temp100*.
4. From **Distribution**, choose **Lognormal base e**.
5. Click **Censor**. Choose **Use censoring columns** and enter *Cens80 Cens100* in the box. Click **OK** in each dialog box.

Session window output

Distribution Overview Plot

```
Distribution:    Lognormal base e
Variable  Anderson-Darling
Temp80    67.22
Temp100   16.50
```

Graph window output



Interpreting the results

These four plots describe the failure rate of engine windings at two different temperatures. With these plots, you can determine how much more likely it is that the engine windings will fail when running at 100° C as opposed to 80° C.

▶ Example of a distribution overview plot with arbitrarily-censored data

Suppose you work for a company that manufactures tires. You are interested in finding out how many miles it takes for various proportions of the tires to “fail,” or wear down to 2/32 of an inch of tread. You are especially interested in knowing how many of the tires last past 45,000 miles. You plan to get this information by using the Parametric Distribution Analysis—Arbitrary Censoring command, but first you want to have a quick look at your data from different perspectives.

You inspect each good tire at regular intervals (every 10,000 miles) to see if the tire has failed, then enter the data into the MINITAB worksheet.

1 Open the worksheet TIREWEAR.MTW.

2 Choose **Stat ➤ Reliability/Survival ➤ Distribution Overview Plot–Arbitrary Cens**.

3 In **Start variables**, enter *Start*. In **End variables**, enter *End*.

4 In **Frequency columns**, enter *Freq*.

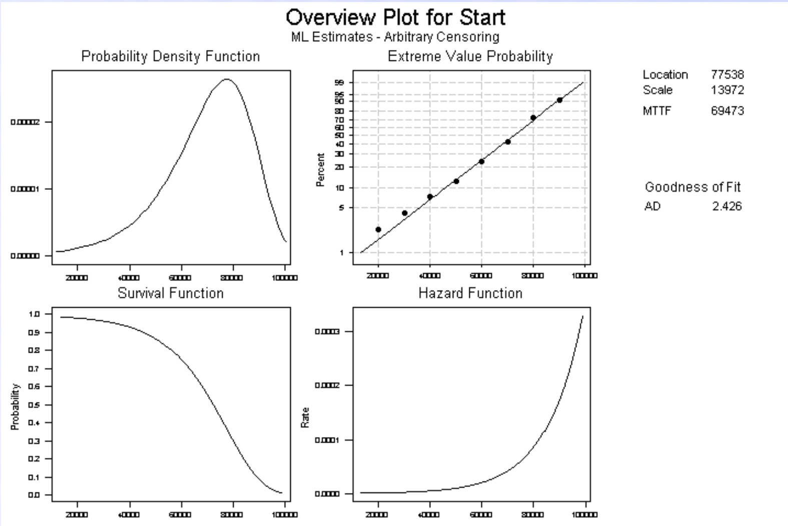
5 From **Distribution**, choose **Extreme value**. Click **OK**.

Session window output

Distribution Overview Plot

```
Variable
Start: Start    End: End
Frequency: Freq


Anderson-Darling
2.426
```

Graph window output



Interpreting the results

These four plots describe the failure rate for tires over time. With these plots, you can approximately determine how many tires last past 45,000 miles.

Parametric Distribution Analysis

Use the parametric distribution analysis commands to fit one of eight common distributions to your data, estimate percentiles and survival probabilities, evaluate the appropriateness of the distribution, and draw survival, hazard, and probability plots. The command you choose, Parametric Distribution Analysis—Right Censoring or Parametric Distribution Analysis—Arbitrary Censoring, depends on the type of data you have, as described in *Data* on page 15-27.

Use the probability plot to see if the distribution fits your data. To compare the fits of four different distributions, see *Distribution ID Plot* on page 15-9, which draws four probability plots on one page. If no parametric distribution fits your data, use *Nonparametric Distribution Analysis* on page 15-52.

The data you gather are the individual failure times, some of which may be censored. For example, you might collect failure times for units running at a given temperature. You might also collect failure times under different temperatures, or under various combinations of stress variables.

You can enter up to ten samples per analysis. MINITAB estimates the functions independently for each sample, unless you assume a common shape (Weibull) or scale (other distributions). All of the samples display on a single plot, in different colors and symbols, which helps you compare the various functions between samples.

To view your data in different ways on one page, see *Distribution Overview Plot* on page 15-19.

Data

The parametric distribution analysis commands accept different kinds of data:

- Parametric Distribution Analysis—Right Censoring accepts exact failure times and right-censored data. For information on how to set up your worksheet, see *Distribution analysis—right censored data* on page 15-5.
- Parametric Distribution Analysis—Arbitrary Censoring accepts exact failure times, right-, left-, and interval-censored data. The data must be in table form. For information on how to set up your worksheet, see *Distribution analysis—arbitrarily censored data* on page 15-8.

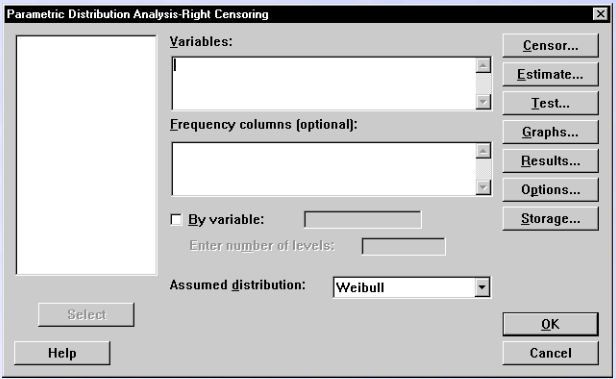
You can enter up to ten samples per analysis. For general information on life data and censoring, see *Distribution Analysis Data* on page 15-5.

Occasionally, you may have life data with no failures. Under certain conditions, MINITAB allows you to draw conclusions based on that data. See *Drawing conclusions when you have few or no failures* on page 15-33.

CONTENTS INDEX MEET MTB UGUIDE 1 UGUIDE 2 SC QREF HOW TO USE

▶ To do a parametric distribution analysis (uncensored/right censored data)

1 Choose **Stat ➤ Reliability/Survival ➤ Parametric Dist Analysis—Right Cens**.



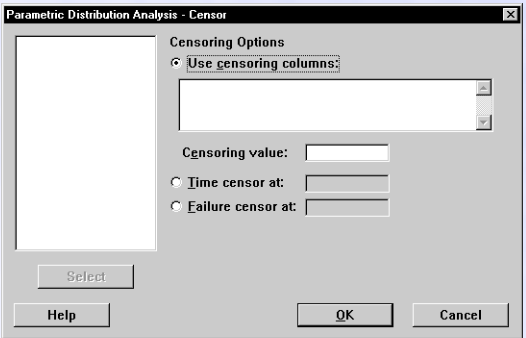
2 In **Variables**, enter the columns of failure times. You can enter up to ten columns (ten different samples).

3 If you have frequency columns, enter the columns in **Frequency columns**.

4 If all of the samples are stacked in one column, check **By variable**, and enter a column of grouping indicators in the box. In **Enter number of levels**, enter the number of levels the indicator column contains.

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

5 Click **Censor**.



6 Do one of the following, then click **OK**.

- For data with censoring columns: Choose **Use censoring columns**, then enter the censoring columns in the box. 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.

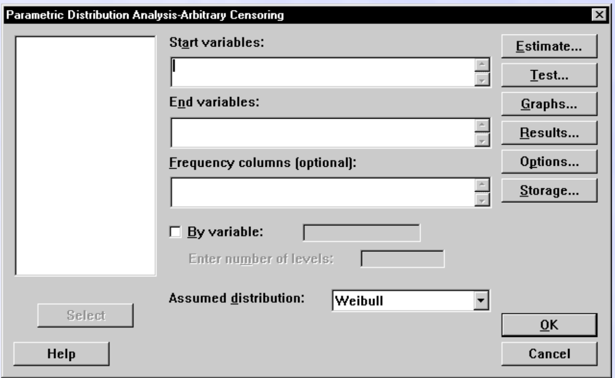
If you like, enter the value you use to indicate censoring in **Censoring value**. If you don't enter a value, by default MINITAB uses the lowest value in the censoring column.

- For time censored data: Choose **Time censor at**, then enter a failure time at which to begin censoring. For example, entering *500* says that any observation from 500 time units onward is considered censored.
- For failure censored data: Choose **Failure censor at**, then enter a number of failures at which to begin censoring. For example, entering *150* says to censor all (ordered) observations from the 150th observed failure on, and leave all other observations uncensored.

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

▶ To do a parametric distribution analysis (arbitrarily censored data)

1 Choose **Stat ➤ Reliability/Survival ➤ Parametric Dist Analysis–Arbitrary Cens**.



2 In **Start variables**, enter the columns of start times. You can enter up to ten columns (ten different samples).

3 In **End variables**, enter the columns of end times. You can enter up to ten columns (ten different samples).

4 If you have frequency columns, enter the columns in **Frequency columns**.

5 If all of the samples are stacked in one column, check **By variable**, and enter a column of grouping indicators in the box. In **Enter number of levels**, enter the number of levels the indicator column contains.

6 If you like, use any of the options described below, then click **OK**.

Options

Parametric Distribution Analysis dialog box

- fit one of eight common lifetime distributions for the parametric analysis, including Weibull (default), extreme value, exponential, normal, lognormal base_e, lognormal base_{10}, logistic, and loglogistic

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

Estimate subdialog box

- estimate parameters using the maximum likelihood (default) or least squares methods—see *Estimating the distribution parameters* on page 15-42.
- estimate parameters assuming a common shape (Weibull distribution) or scale (other distributions).
- estimate the scale parameter while holding the shape fixed (Weibull distribution), or estimate the location parameter while keeping the scale fixed (all other distributions)—see *Estimating the distribution parameters* on page 15-42.
- draw conclusions when you have few or no failures—*Drawing conclusions when you have few or no failures* on page 15-33.
- estimate percentiles for additional percents—see *Percentiles* on page 15-36.
- estimate survival probabilities for times (values) you specify—see *Survival probabilities* on page 15-39.
- specify a confidence level for all of the confidence intervals. The default is 95.0%.
- choose to calculate two-sided confidence intervals, or lower or upper bounds. The default is two-sided.

Test subdialog box

- test whether the distribution parameters (scale, shape, or location) are consistent with specified values—see *Comparing parameters* on page 15-34.
- test whether two or more samples come from the same population—see *Comparing parameters* on page 15-34.
- test whether the shape, scale, or location parameters from K distributions are the same—see *Comparing parameters* on page 15-34.

Graphs subdialog box

- obtain the plot points for the probability plot using various nonparametric methods—see *Probability plots* on page 15-37.

CONTENTS INDEX MEET MTB UGUIDE 1 UGUIDE 2 SC QREF HOW TO USE

With Parametric Distribution Analysis—Right Censoring, you can choose the Default method, Modified Kaplan-Meier method, Herd-Johnson method, or Kaplan-Meier method. The Default method is the normal score for uncensored data; the modified Kaplan-Meier method for censored data.

With Parametric Distribution Analysis—Arbitrary Censoring, choose the Turnbull or Actuarial method. Turnbull is the default method.

- (Parametric Distribution Analysis—Right Censoring only) handle tied failure times in the probability plot by plotting all of the points (default), the average (median) of the tied points, or the maximum of the tied points.
- draw a survival plot—see *Survival plots* on page 15-40.
- suppress confidence intervals on the probability and survival plots.
- draw a hazard plot—see *Hazard plots* on page 15-41.
- enter minimum and/or maximum values for the x-axis scale.
- enter a label for the x-axis.

Results subdialog box

- display the following Session window output:
  - no output
  - the basic output, which includes variable information, censoring information, estimated parameters, the log-likelihood, goodness-of-fit statistics, and tests of parameters
  - the above output, plus characteristics of the distribution, and tables of percentiles and survival probabilities
- show the log-likelihood for each iteration of the algorithm

Options subdialog box

- enter starting values for model parameters—see *Estimating the distribution parameters* on page 15-42.
- 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 distribution parameters* on page 15-42.
- use historical estimates for the parameters rather than estimate them from the data. In this case, no estimation is done; all results—such as the percentiles and survival probabilities—are based on these historical estimates. See *Estimating the distribution parameters* on page 15-42.

Storage subdialog box

- store characteristics of the fitted distribution:
  - percentiles and their percents, standard errors, and confidence limits
  - survival probabilities and their times and confidence limits
- store information on parameters:
  - estimates of the parameters and their standard errors and confidence limits
  - the variance/covariance matrix
  - the log-likelihood for the last iteration

Output

The default output for Parametric Distribution Analysis—Right Censoring and Parametric Distribution Analysis—Arbitrary Censoring consists of:

- the censoring information
- parameter estimates and their
  - standard errors
  - 95% confidence intervals
- log-likelihood and Anderson-Darling goodness-of-fit statistic—see *Goodness-of-fit statistics* on page 15-13
- characteristics of distribution and their
  - standard errors
  - 95% confidence intervals
- table of percentiles and their
  - standard errors
  - 95% confidence intervals
- probability plot

Fitting a distribution

You can fit one of eight common lifetime distributions to your data, 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 the smallest extreme value (Type 1).

CONTENTS INDEX MEET MTB UGUIDE 1 UGUIDE 2 SC QREF HOW TO USE

The Session window output includes two tables that describe the distribution. Here is some sample output from a default Weibull distribution:

```
Estimation Method:  Maximum Likelihood
Distribution:  Weibull

Parameter Estimates
                        Standard        95.0% Normal CI
Parameter    Estimate      Error       Lower       Upper
Shape          2.3175     0.3127      1.7790      3.0191
Scale          73.344      5.203      63.824      84.286

Log-Likelihood = -186.128

Goodness-of-Fit
Anderson-Darling = 67.6366

Characteristics of Distribution
                                        Standard        95.0% Normal CI
                            Estimate       Error       Lower       Upper
Mean(MTTF)                   64.9829      4.6102     56.5472     74.6771
Standard Deviation           29.7597      4.1463     22.6481     39.1043
Median                       62.6158      4.6251     54.1763     72.3700
First Quartile(Q1)           42.8439      4.3240     35.1546     52.2151
Third Quartile(Q3)           84.4457      6.2186     73.0962     97.5575
Interquartile Range(IQR)     41.6018      5.5878     31.9730     54.1305
```

- **Parameter Estimates** displays the maximum likelihood or the least squares estimates of the distribution parameters, their standard errors and approximate 95.0% confidence intervals, and the log-likelihood and Anderson-Darling goodness-of-fit statistic for the fitted distribution.
- **Characteristics of Distribution** displays common measures of the center and spread of the distribution with 95.0% lower and upper confidence intervals. The mean and standard deviation are not resistant to large lifetimes, while the median, Q1 (25th percentile), Q3 (75th percentile), and the IQR (interquartile range) are resistant.

The Newton-Raphson algorithm is used to calculate maximum likelihood estimates of the parameters. These parameters define the distribution. All resulting functions, such as the percentiles and survival probabilities, are calculated from that distribution. For computations, see [6].

Drawing conclusions when you have few or no failures

MINITAB allows you to use historical values for distribution parameters to improve the current analysis. Providing the shape (Weibull) or scale (other distributions) parameter makes the resulting analysis more precise, if your shape/scale is an appropriate choice. An added benefit of providing historical values, is that, when your data are from a Weibull or exponential distribution, you can do a Bayes analysis and draw conclusions when your data has few or no failures.

Sometimes you may collect life data and have no failures. MINITAB offers the ability to draw conclusions based on that data under certain conditions:

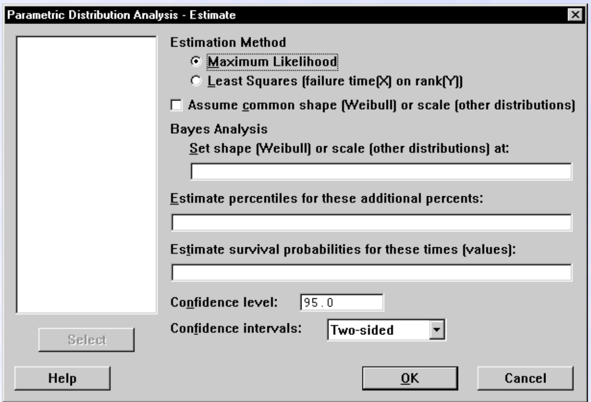
- The data come from a Weibull or exponential distribution.
- The data are right-censored.
- The maximum likelihood method will be used to estimate parameters.
- You provide a historical value for the shape parameter (Weibull or exponential).

MINITAB provides lower confidence bounds for the scale parameter (Weibull or exponential), percentiles, and survival probabilities. The lower confidence bound helps you to draw some conclusions; if the value of the lower confidence bound is better than the specifications, then you may be able to terminate the test.

For example, your reliability specifications require that the 5th percentile is at least 12 months. You run a Bayes analysis on data with no failures, and then examine the lower confidence bound to substantiate that the product is at least as good as specifications. If the lower confidence bound for the 5th percentile is 13.1 months, then you can conclude that your product meets specifications and terminate the test.

▶ To draw conclusions when you have no failures

1 In the main dialog box, click **Estimate**.



2 In **Set shape (Weibull) or scale (other distributions) at** enter the shape or scale value. Click **OK**.

Comparing parameters

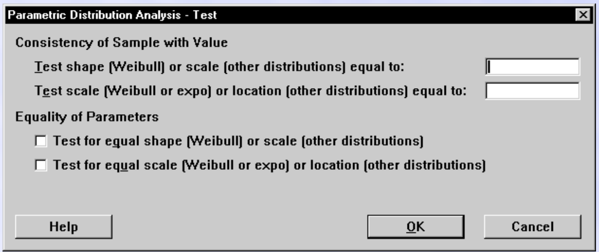
Are the distribution parameters for a sample equal to specified values; for example, does the scale equal 1.1? Does the sample come from the historical distribution? Do two or more samples come from the same population? Do two or more samples share the same shape, scale, or location parameters? To answer these questions you need to perform hypothesis tests on the distribution parameters.

MINITAB performs Wald Tests [7] and provides Bonferroni 95.0% confidence intervals for the following hypothesis tests:

- Test whether the distribution parameters (scale, shape, or location) are consistent with specified values
- Test whether the sample comes from the historical distribution
- Test whether two or more samples come from the same population
- Test whether two or more samples share the same shape, scale or location parameters

▶ To compare distribution parameters to a specified value

1 In the main dialog box, click **Test**.



2 In **Test shape (Weibull) or scale (other distributions) equal to** or **Test scale (Weibull or expo) or location (other distributions) equal to** enter the value to be tested. Click **OK**.

▶ To test whether a sample comes from a historical distribution

1 In the main dialog box, click **Test**.

2 In **Test shape (Weibull) or scale (other distributions) equal to** and **Test scale (Weibull or expo) or location (other distributions) equal to** enter the parameters of the historical distribution. Click **OK**.

▶ To determine whether two or more samples come from the same population

1 In the main dialog box, click **Test**.

2 Check **Test for equal shape (Weibull) or scale (other distributions)** and **Test for equal scale (Weibull or expo) or location (other distributions)**. Click **OK**.

▶ To compare the shape, scale, or location parameters from two or more samples

1 In the main dialog box, click **Test**.

2 Check **Test for equal shape (Weibull) or scale (other distributions)** or **Test for equal scale (Weibull or expo) or location (other distributions)**. Click **OK**.

CONTENTS INDEX MEET MTB UGUIDE 1 UGUIDE 2 SC QREF HOW TO USE

Percentiles

By what time do half of the engine windings fail? How long until 10% of the blenders stop working? You are looking for *percentiles*. The parametric distribution analysis commands automatically display a table of percentiles in the Session window. By default, MINITAB displays the percentiles 1–10, 20, 30, 40, 50, 60, 70, 80, and 90–99.

In this example, we entered failure times (in months) for engine windings.

Table of Percentiles

Percent	Percentile	Standard Error	95.0% Normal CI Lower	95.0% Normal CI Upper
1	10.0765	2.7845	5.8626	17.3193
2	13.6193	3.2316	8.5543	21.6834
3	16.2590	3.4890	10.6767	24.7601
4	18.4489	3.6635	12.5009	27.2270

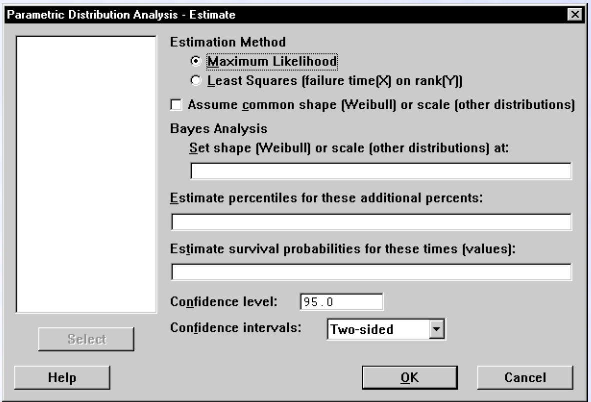
At about 10 months, 1% of the windings failed.

The values in the Percentile column are estimates of the times at which the corresponding percent of the units failed. The table also includes standard errors and approximate 95.0% confidence intervals for each percentile.

In the Estimate subdialog box, you can specify a different confidence level for all confidence intervals. You can also request percentiles to be added the default table.

▶ To request additional percentiles

1 In the main dialog box, click **Estimate**.



2 In **Estimate percentiles for these additional percents**, enter the additional percents for which you want to estimate percentiles. You can enter individual percents (0 < P < 100) or a column of percents. Click **OK**.

Probability plots

Use the probability plot to assess whether a particular distribution fits your data. The plot consists of:

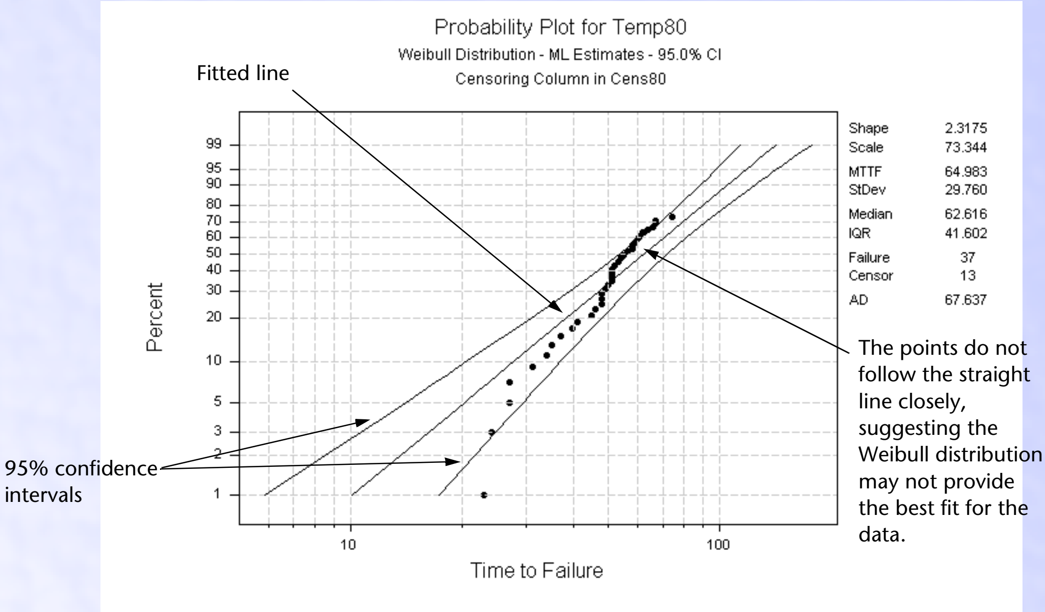
- *plot points*, which represent the proportion of failures up to a certain time. The plot points are calculated using a nonparametric method, which assumes no parametric distribution—for formulas, see Calculations in Help. The proportions are transformed and used as the y variable, while their corresponding times may be transformed and used as the x variable.
- the *fitted line*, which is a graphical representation of the percentiles. To make the fitted line, MINITAB first calculates the percentiles for the various percents, based on the chosen distribution. The associated probabilities are then transformed and used as the y variables. The percentiles may be transformed, depending on the distribution, and are used as the x variables. The transformed scales, chosen to linearize the fitted line, differ depending on the distribution used.
- a set of approximate *95.0% confidence intervals* for the fitted line.

Because the plot points do not depend on any distribution, they would be the same (before being transformed) for any probability plot made. The fitted line, however, differs depending on the parametric distribution chosen. So you can use the probability plot to assess whether a particular distribution fits your data. In general, the closer the points fall to the fitted line, the better the fit.

Tip To quickly compare the fit of up to four different distributions at once see *Distribution ID Plot* on page 15-9.

MINITAB provides two goodness of fit measures to help assess how the distribution fits your data: the Anderson-Darling statistic for both the maximum likelihood and the least squares methods and the Pearson correlation coefficient for the least squares method. A smaller Anderson-Darling statistic indicates that the distribution provides a better fit. A larger Pearson correlation coefficient indicates that the distribution provides a better fit. See *Goodness-of-fit statistics* on page 15-13.

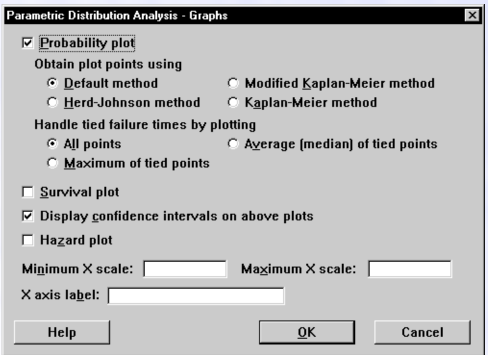
Here is a Weibull probability plot for failure times associated with running engine windings at a temperature of 80° C:

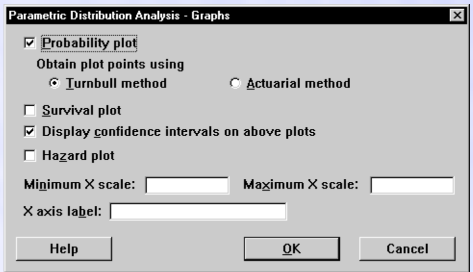


With the commands in this chapter, you can choose from various methods to estimate the plot points. You can also choose the method used to obtain the fitted line. The task below describes all the ways you can modify the probability plot.

▶ To modify the default probability plot

1 In the main dialog box, click **Graphs**.





2 Do any or all of the following:

- specify the method used to obtain the plot points—under **Obtain plot points using**, choose one of the following:
  – with Parametric Distribution Analysis—Right Censoring: **Default method**, **Modified Kaplan-Meier method**, **Herd-Johnson method**, or **Kaplan-Meier method**. The Default method is the normal score for uncensored data; the modified Kaplan-Meier method for censored data.
  – with Parametric Distribution Analysis—Arbitrary Censoring: **Turnbull method** or **Actuarial method**.
- Parametric Distribution Analysis—Right Censoring only: Choose what to plot when you have tied failure times—under **Handle tied failure times by plotting**, choose **All points** (default), **Maximum of the tied points**, or **Average (median) of tied points**.
- turn off the 95.0% confidence interval—uncheck **Display confidence intervals on above plots**.
- specify a minimum and/or maximum value for the x-axis scale.
- enter a label for the x axis.

3 Click **OK**.

4 If you want to change the confidence level for the 95.0% confidence interval to some other level, click **Estimate**. In **Confidence level**, enter a value. Click **OK**.

5 If you want to change the method used to obtain the fitted line, click Estimate. In **Estimation Method**, choose **Maximum Likelihood** (default) or **Least Squares**. Click **OK**.

Survival probabilities

What is the probability of an engine winding running past a given time? How likely is it that a cancer patient will live five years after receiving a certain drug? You are looking for *survival probabilities*, which are estimates of the proportion of units that survive past a given time.

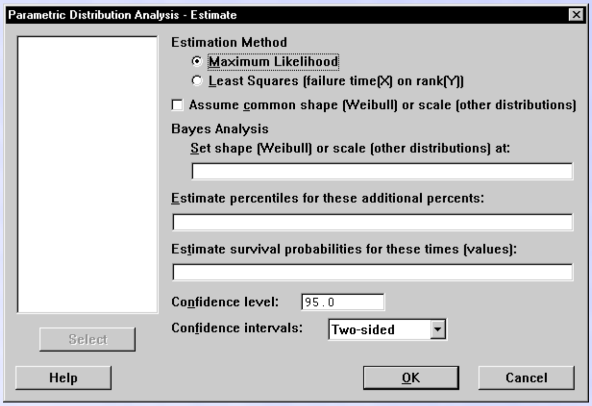
When you request survival probabilities in the Estimate subdialog box, the parametric distribution analysis commands display them in the Session window. Here, for example, we requested a survival probability for engine windings running at 70 months:

40.76% of the engine windings last past 70 months.

```
                 Table of Survival Probabilities
                                             95.0% Normal CI
             Time   Probability           Lower         Upper
          70.0000        0.4076          0.2894        0.5222
```

▶ To request parametric survival probabilities

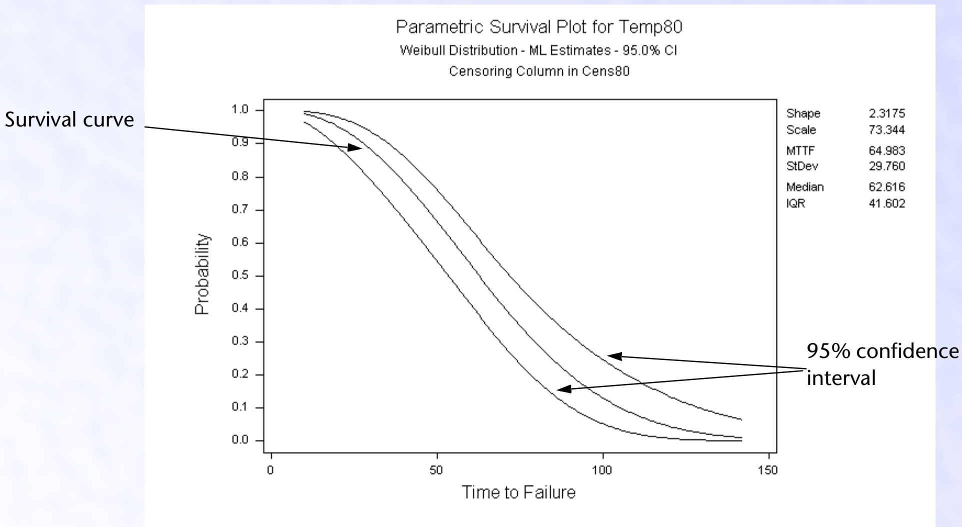
1 In the main dialog box, click **Estimate**.



2 In **Estimate survival probabilities for these times (values)**, enter one or more times or a column of times for which you want to calculate survival probabilities. Click **OK**.

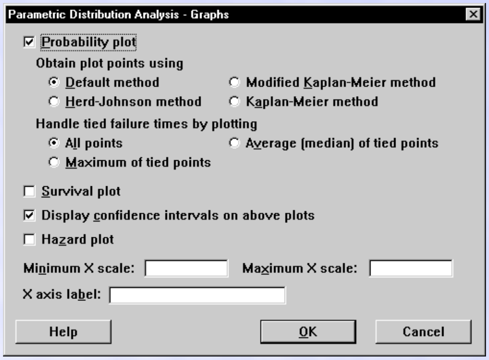
Survival plots

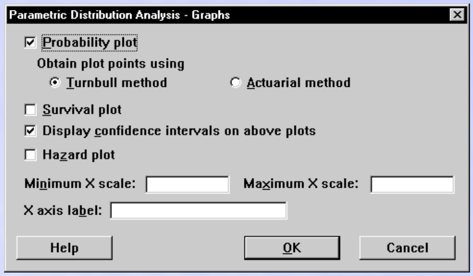
Survival (or reliability) plots display the survival probabilities versus time. Each plot point represents the proportion of units surviving at time t. The survival curve is surrounded by two outer lines—the approximate 95.0% confidence interval for the curve, which provide reasonable values for the “true” survival function.



▶ To draw a parametric survival plot

1 In the main dialog box, click **Graphs**.





2 Check **Survival plot**.

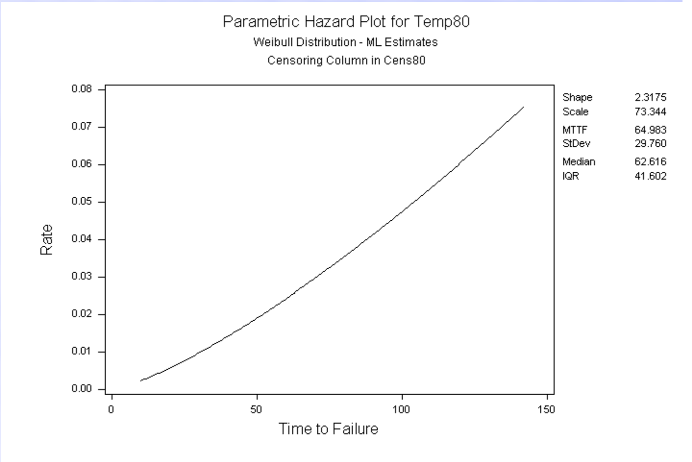
3 If you like, do any of the following:

- turn off the 95.0% confidence interval—uncheck **Display confidence intervals on above plots**. Click **OK**.
- change the confidence level for the 95.0% confidence interval. First, click **OK** in the Graphs subdialog box. Click **Estimate**. In **Confidence level for confidence intervals**, enter a value. Click **OK**.
- specify minimum and/or maximum values for the x-axis scale.
- enter a label for the x-axis.

Hazard plots

The hazard plot displays the instantaneous failure rate for each time t. Often, the hazard rate is high at the beginning of the plot, low in the middle of the plot, then high again at the end of the plot. Thus, the curve often resembles the shape of a bathtub. The early period with high failure rate is often called the infant mortality stage. The middle section of the curve, where the failure rate is low, is the normal life stage. The end of the curve, where failure rate increases again, is the wearout stage.

Note MINITAB's distributions will not resemble a bathtub curve. The failures at different parts of the bathtub curve are likely caused by different failure modes. MINITAB estimates the distribution of the failure time caused by one failure mode.



▶ To draw a parametric hazard plot

1 In the main dialog box, click **Graphs**.

2 Check **Hazard plot**.

3 If you like, do any of the following, then click **OK**.

- specify minimum and/or maximum values for the x-axis scale
- enter a label for the x-axis

Estimating the distribution parameters

MINITAB uses a the maximum likelihood estimations method (modified Newton-Raphson algorithm) or least squares (XY) method to estimate the parameters of the distribution. Or, if you like, you can use your own parameters. In this case, no estimation is done; all results—such as the percentiles—are based on the parameters you enter.

You can choose to estimate the parameters using either the maximum likelihood method or the least squares method—see *Maximum likelihood estimates versus least squares estimates* on page 15-44.

When you let MINITAB estimate the parameters from the data using the maximum likelihood method, 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.

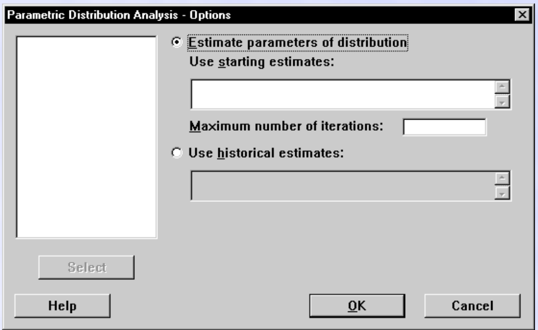
Why enter starting values for the algorithm? The maximum likelihood solution may not converge if the starting estimates are not in the neighborhood of the true solution, so you may want to specify what you think are good starting values for parameter estimates. In these cases, enter the distribution parameters. For the Weibull distribution, enter the shape and scale. For the exponential distribution, enter the scale. For all other distributions, enter the location and scale.

You can also choose to

- estimate the scale parameter while keeping the shape fixed (Weibull and exponential distributions)
- estimate the location parameter while keeping the scale fixed (other distributions)

▶ To control estimation of the parameters

1 In the main dialog box, click **Options**.



2 Do one of the following:

- To estimate the distribution parameters from the data (the default), choose **Estimate parameters of distribution**.

  If you like, do any of the following:

  – Enter starting estimates for the parameters: In **Use starting estimates,** enter one column of values to be used for all samples, or several columns of values that match the order in which the corresponding variables appear in the **Variables** box in the main dialog box.

  – Specify the maximum number of iterations: In **Maximum number of iterations**, enter a positive integer.

- To enter your own estimates for the distribution parameters, choose **Use historical estimates** and enter one column of values to be used for all samples, or several columns of values that match the order in which the corresponding variables appear in the **Variables** box in the main dialog box.

3 Click **OK**.

▶ **To choose the method for estimating parameters**

1 In the main dialog box, choose **Estimate**.

2 Under **Estimation Method**, choose **Maximum Likelihood** (the default) or **Least Squares**. Click **OK**.

▶ **To estimate one parameter while keeping the other parameter fixed**

You can estimate the scale parameter while keeping the shape parameter fixed (Weibull and exponential) or estimate the location parameter while keeping the scale fixed (other distributions).

1 In the main dialog box, click **Estimate**.

2 Do one of the following:

– Estimate the scale parameter while keeping the shape fixed (Weibull and exponential distributions): In **Set shape (Weibull) or scale (other distributions) at**, enter one value to be used for all samples, or a series of values that match the order in which the corresponding variables appear in the **Variables** box in the main dialog box.

– Estimate the location parameter while keeping the scale fixed (other distributions): In **Set shape (Weibull) or scale (other distributions) at**, enter one value to be used for all samples, or a series of values that match the order in which the corresponding variables appear in the **Variables** box in the main dialog box.

3 Click **OK**.

Maximum likelihood estimates versus least squares estimates

Maximum likelihood estimates are calculated by maximizing the likelihood function. The likelihood function describes, for each set of distribution parameters, the chance that the true distribution has the parameters based on the sample.

Least squares estimates are calculated by fitting a regression line to the points in a probability plot. The line is formed by regressing time (X) to failure (Y) or log (time to failure) on the transformed percent.

Here are the major advantages of each method:

Maximum likelihood (MLE)

- Distribution parameter estimates are more precise than least squares (XY).
- MLE allows you to perform an analysis when there are no failures. When there is only one failure and some right-censored observations, the maximum likelihood parameter estimates may exist for a Weibull distribution.
- The maximum likelihood estimation method has attractive mathematical qualities.

Least squares (LSXY)

- Better graphical display to the probability plot because the line is fitted to the points on a probability plot.
- For small or heavily censored sample, LSXY is more accurate than MLE. MLE tends to overestimate the shape parameter for a Weibull distribution and underestimate the scale parameter in other distributions. Therefore, MLE will tend to overestimate the low percentiles.

When possible, both methods should be tried; if the results are consistent, then there is more support for your conclusions. Otherwise, you may want to use the more conservative estimates or consider the advantages of both approaches and make a choice for your problem.

▶ Example of a parametric distribution analysis with exact failure/right-censored data

Suppose you work for a company that manufactures engine windings for turbine assemblies. Engine windings may decompose at an unacceptable rate at high temperatures. You decide to look at failure times for engine windings at two temperatures, 80 and 100°C. You want to find out the following information for each temperature:

- the times at which various percentages of the windings fail. You are particularly interested in the 0.1st percentile.
- the proportion of windings that survive past 70 months.

You also want to draw two plots: a probability plot to see if the lognormal$_e$ distribution provides a good fit for your data, and a survival plot.

In the first sample, you collect failure times (in months) for 50 windings at 80°C; in the second sample, you collect failure times for 40 windings at 100°C. Some of the windings drop out of the test for unrelated reasons. In the MINITAB worksheet, you use a column of censoring indicators to designate which times are actual failures (1) and which are censored units removed from the test before failure (0).

1. Open the worksheet RELIABLE.MTW.
2. Choose **Stat ➤ Reliability/Survival ➤ Parametric Dist Analysis–Right Cens**.
3. In **Variables**, enter *Temp80 Temp100*.
4. From **Assumed distribution**, choose **Lognormal base e**.
5. Click **Censor**. Choose **Use censoring columns** and enter *Cens80 Cens100* in the box. Click **OK**.
6. Click **Estimate**. In **Estimate percentiles for these additional percents**, enter *.1*.
7. In **Estimate survival probabilities for these times (values)**, enter *70*. Click **OK**.
8. Click **Graphs**. Check **Survival plot**. Click **OK** in each dialog box.

Session window output

Distribution Analysis: Temp80

Variable: Temp80

Censoring Information	Count
Uncensored value	37
Right censored value	13

Censoring value: Cens80 = 0

Estimation Method: Maximum Likelihood
Distribution: Lognormal base e

Parameter Estimates

Parameter	Estimate	Standard Error	95.0% Normal CI Lower	95.0% Normal CI Upper
Location	4.09267	0.07197	3.95161	4.23372
Scale	0.48622	0.06062	0.38080	0.62082

Log-Likelihood = -181.625

Goodness-of-Fit
Anderson-Darling = 67.2208

Characteristics of Distribution

	Estimate	Standard Error	95.0% Normal CI Lower	95.0% Normal CI Upper
Mean(MTTF)	67.4153	5.5525	57.3656	79.2255
Standard Deviation	34.8145	6.7983	23.7435	51.0476
Median	59.8995	4.3109	52.0192	68.9735
First Quartile(Q1)	43.1516	3.2953	37.1531	50.1186
Third Quartile(Q3)	83.1475	7.3769	69.8763	98.9392
Interquartile Range(IQR)	39.9959	6.3332	29.3245	54.5505

Table of Percentiles

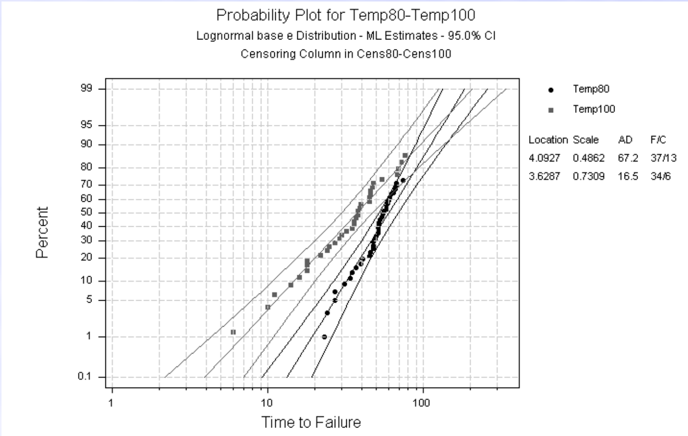
Percent	Percentile	Standard Error	95.0% Normal CI Lower	95.0% Normal CI Upper
0.1	13.3317	2.5156	9.2103	19.2975
1.0	19.3281	2.8375	14.4953	25.7722
2.0	22.0674	2.9256	17.0178	28.6154
3.0	24.0034	2.9726	18.8304	30.5975
4.0	25.5709	3.0036	20.3126	32.1906
5.0	26.9212	3.0262	21.5978	33.5566
6.0	28.1265	3.0440	22.7506	34.7727
7.0	29.2276	3.0588	23.8074	35.8819
8.0	30.2501	3.0717	24.7910	36.9113
9.0	31.2110	3.0833	25.7170	37.8788
10.0	32.1225	3.0941	26.5962	38.7970
20.0	39.7837	3.2100	33.9646	46.5999
30.0	46.4184	3.4101	40.1936	53.6073
40.0	52.9573	3.7567	46.0833	60.8568
50.0	59.8995	4.3109	52.0192	68.9735

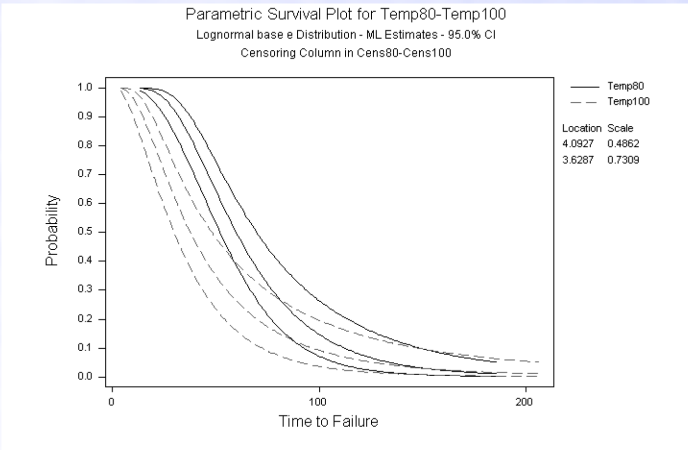
-----the rest of this table omitted for space-----

```
Table of Survival Probabilities
                                95.0% Normal CI
      Time  Probability       Lower       Upper
   70.0000       0.3743      0.2631      0.4971


Distribution Analysis: Temp100


Variable:  Temp100

Censoring Information                Count
Uncensored value                        34
Right censored value                     6
Censoring value:  Cens100 = 0

Estimation Method:  Maximum Likelihood
Distribution:  Lognormal base e

Parameter Estimates
                        Standard        95.0% Normal CI
Parameter    Estimate      Error       Lower       Upper
Location       3.6287     0.1178      3.3978      3.8595
Scale         0.73094    0.09198     0.57117     0.93540

Log-Likelihood = -160.688

Goodness-of-Fit
Anderson-Darling = 16.4987

Characteristics of Distribution
                                        Standard        95.0% Normal CI
                            Estimate       Error       Lower       Upper
Mean(MTTF)                   49.1969      6.9176     37.3465     64.8076
Standard Deviation           41.3431     11.0416     24.4947     69.7806
Median                       37.6636      4.4362     29.8995     47.4439
First Quartile(Q1)           23.0044      2.9505     17.8910     29.5791
Third Quartile(Q3)           61.6643      8.4984     47.0677     80.7876
Interquartile Range(IQR)     38.6600      7.2450     26.7759     55.8185

Table of Percentiles
                        Standard        95.0% Normal CI
 Percent  Percentile       Error       Lower       Upper
     0.1      3.9350      1.1729      2.1940      7.0577
     1.0      6.8776      1.6170      4.3383     10.9034
     2.0      8.3941      1.7942      5.5212     12.7619
     3.0      9.5253      1.9111      6.4283     14.1144
     4.0     10.4756      2.0015      7.2036     15.2338
     5.0     11.3181      2.0766      7.8995     16.2162
     6.0     12.0884      2.1419      8.5418     17.1076
     7.0     12.8069      2.2003      9.1453     17.9343
     8.0     13.4863      2.2538      9.7195     18.7129
     9.0     14.1354      2.3034     10.2707     19.4544
```

```
         10.0      14.7606       2.3502      10.8036      20.1667
         20.0      20.3589       2.7526      15.6197      26.5362
         30.0      25.6717       3.1662      20.1592      32.6916
         40.0      31.2967       3.6950      24.8316      39.4451
         50.0      37.6636       4.4362      29.8995      47.4439

-----the rest of this table omitted for space-----

Table of Survival Probabilities
                                 95.0% Normal CI
     Time  Probability        Lower        Upper
  70.0000       0.1982       0.1072       0.3248
```

Graph window output





Interpreting the results

To see the times at which various percentages of the windings fail, look at the Table of Percentiles. At 80° C, for example, it takes 19.3281 months for 1% of the windings fail.

You can find the .1st percentile, which you requested, within the Table of Percentiles. At 80° C, .1% of the windings fail by 13.3317 months; at 100° C, .1% of the windings fail by 3.9350 months. So the increase in temperature decreased the percentile by about 9.5 months.

What proportion of windings would you expect to still be running past 70 months? In the Table of Survival Probabilities you find your answer. At 80° C, 37.43% survive past 70 months; at 100° C, 19.82% survive.

▶ Example of parametric distribution analysis with arbitrarily censored data

Suppose you work for a company that manufactures tires. You are interested in finding out how many miles it takes for various proportions of the tires to “fail,” or wear down to 2/32 of an inch of tread. You are especially interested in knowing how many of the tires last past 45,000 miles.

You inspect each good tire at regular intervals (every 10,000 miles) to see if the tire has failed, then enter the data into the MINITAB worksheet.

1 Open the worksheet TIREWEAR.MTW.

2 Choose **Stat ➤ Reliability/Survival ➤ Parametric Dist Analysis–Arbitrary Cens**.

3 In **Start variables**, enter *Start*. In **End variables**, enter *End*.

4 In **Frequency columns**, enter *Freq*.

5 From **Assumed distribution**, choose **Extreme value**.

6 Click **Graphs**. Check **Survival plot**, then click **OK**.

7 Click **Estimate**. In **Estimate survival probabilities for these times (values)**, enter 45000. Click **OK** in each dialog box.

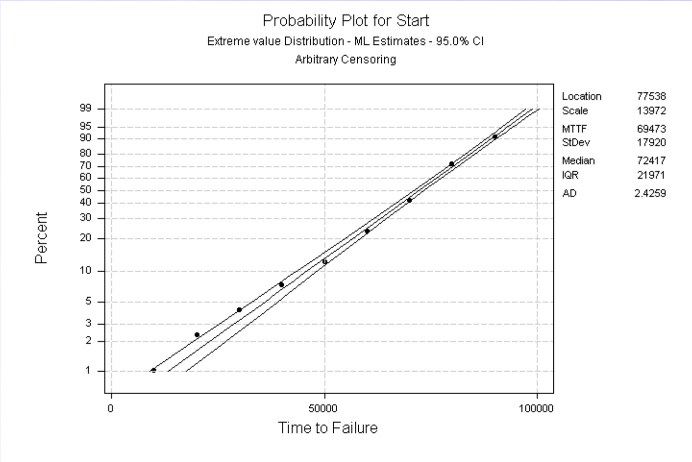
Session window output

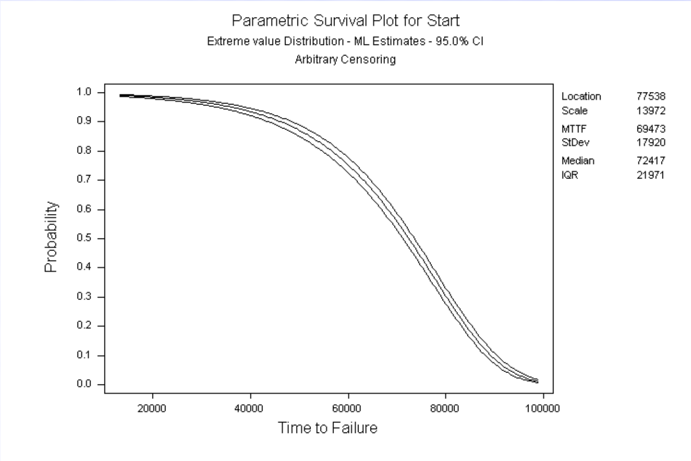
Distribution Analysis, Start = Start and End = End

```
Variable
Start:  Start    End:  End
Frequency:  Freq

Censoring Information                Count
Right censored value                    71
Interval censored value                694
Left censored value                      8

Estimation Method:  Maximum Likelihood
Distribution:  Extreme value

Parameter Estimates
                        Standard        95.0% Normal CI
Parameter    Estimate      Error       Lower        Upper
Location      77538.0      547.0     76465.8      78610.2
Scale         13972.0      445.0     13126.5      14872.1
```

```
Log-Likelihood = -1465.913

Goodness-of-Fit
Anderson-Darling = 2.4259

Characteristics of Distribution
                                        Standard       95.0% Normal CI
                              Estimate      Error       Lower       Upper
Mean(MTTF)                    69473.32   646.6352    68205.94    70740.70
Standard Deviation            17919.83   570.7594    16835.36    19074.15
Median                        72417.04   599.5413    71241.97    73592.12
First Quartile(Q1)            60130.23   849.0361    58466.15    61794.31
Third Quartile(Q3)            82101.72   538.9283    81045.44    83158.00
Interquartile Range(IQR)      21971.49   699.8078    20641.82    23386.80

Table of Percentiles
                          Standard       95.0% Normal CI
 Percent   Percentile        Error       Lower       Upper
       1     13264.55     2216.243    8920.791    17608.30
       2     23019.97     1916.275    19264.14    26775.80
       3     28756.49     1741.644    25342.93    32170.05
       4     32847.96     1618.183    29676.38    36019.54
       5     36038.31     1522.706    33053.87    39022.76
       6     38658.95     1444.905    35826.99    41490.91
       7     40886.63     1379.291    38183.26    43589.99
       8     42826.87     1322.593    40234.64    45419.11
       9     44547.76     1272.702    42053.31    47042.21
      10     46095.77     1228.182    43688.58    48502.97
      20     56580.77     939.3041    54739.76    58421.77
      30     63133.78     777.3208    61610.26    64657.30
      40     68152.58     670.9556    66837.54    69467.63
      50     72417.04     599.5413    71241.97    73592.12

-----the rest of this table omitted for space-----

Table of Survival Probabilities
                                95.0% Normal CI
      Time  Probability        Lower       Upper
  45000.00       0.9072       0.8903      0.9216
```

Graph window output





Interpreting the results

As shown in the Characteristics of Distribution table, the mean and median miles until the tires fail are 69,473.32 and 72,417.04 miles, respectively.

To see the times at which various percentages or proportions of the tires fail, look at the Table of Percentiles. For example, 5% of the tires fail by 36,038.31 miles and 50% fail by 72,417.04 miles.

In the Table of Survival Probabilities, you can see that 90.72% of the tires last past 45,000 miles.

Nonparametric Distribution Analysis

When no distribution fits your data, use the nonparametric distribution analysis commands to estimate survival probabilities, hazard estimates, and other functions, and draw survival and hazard plots.

When you have exact failure/right-censored data, you can request Kaplan-Meier or Actuarial estimates. When you have tabled data with a varied censoring scheme, you can request Turnbull or Actuarial estimates.

When you have exact failure/right-censored data and multiple samples, MINITAB tests the equality of survival curves.

The data you gather are the individual failure times, some of which may be censored. For example, you might collect failure times for units running at a given temperature. You might also collect failure times under different temperatures, or under various combination of stress variables.

You can enter up to ten samples per analysis. When you enter more than one sample, MINITAB estimates the functions independently. All of the samples display on a single plot, in different colors and symbols, which helps you compare the various functions between samples.

To make a quick Kaplan-Meier survival plot and empirical hazard plot, see *Distribution Overview Plot* on page 15-19.

If a distribution fits your data, use *Parametric Distribution Analysis* on page 15-27.

Data

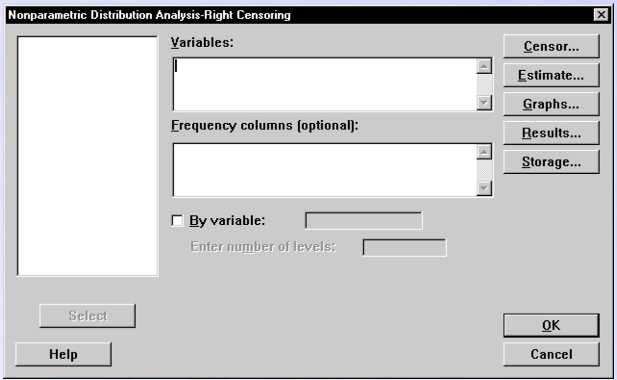
The nonparametric distribution analysis commands accept different kinds of data:

- Nonparametric Distribution Analysis—Right Censoring accepts exact failure times and right-censored data—for information on how to set up your worksheet, see *Distribution analysis—right censored data* on page 15-5.
- Nonparametric Distribution Analysis—Arbitrary Censoring accepts exact failure times, right-, left-, and interval-censored data. The data must be in table form. For information on how to set up your worksheet, see *Distribution analysis—arbitrarily censored data* on page 15-8.

You can enter up to ten samples per analysis. For general information on life data and censoring, see *Distribution Analysis Data* on page 15-5.

▶ To do a nonparametric distribution analysis (uncensored/right censored data)

1 Choose **Stat ➤ Reliability/Survival ➤ Nonparametric Distribution Analysis–Right Cens.**



2 In **Variables**, enter the columns of failure times. You can enter up to ten columns (ten different samples).

3 If you have frequency columns, enter the columns in **Frequency columns**.

4 If all of the samples are stacked in one column, enter a column of grouping indicators in **By variable**. In **Enter number of levels**, enter the number of levels the indicator column contains.

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

5 Click **Censor**.

6 Do one of the following, then click **OK**.

- For data with censoring columns: Choose **Use censoring columns**, then enter the censoring columns in the box. 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.

  If you like, enter the value you use to indicate a censored value in **Censoring value**. If you do not enter a censoring value, MINITAB uses the lowest value in the censoring column by default.

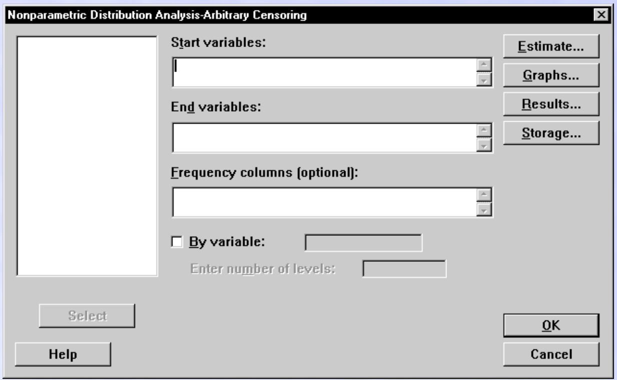
- For time censored data: Choose **Time censor at**, then enter a failure time at which to begin censoring. For example, entering *500* says that any observation from 500 time units onward is considered censored.

- For failure censored data: Choose **Failure censor at**, then enter a number of failures at which to begin censoring. For example, entering *150* says to censor all (ordered) observations starting with the 150th observed failures, and leave all other observations uncensored.

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

▶ To do a nonparametric distribution analysis (arbitrarily censored data)

1 Choose **Stat ➤ Reliability/Survival ➤ Nonparametric Dist Analysis–Arbitrary Cens**.



2 In **Start variables**, enter the columns of start times. You can enter up to ten columns (ten different samples).

3 In **End variables**, enter the columns of end times. You can enter up to ten columns (ten different samples).

4 When you have frequency columns, enter the columns in **Frequency columns**.

5 If all of the samples are stacked in one column, enter a column of grouping indicators in **By variable**. In **Enter number of levels**, enter the number of levels the indicator column contains.

6 If you like, use any of the options described below, then click **OK**.

Options

Estimate subdialog box

- estimate survival probabilities using the Kaplan-Meier or Actuarial method (Nonparametric Distribution Analysis—Right Censoring), or the Turnbull or Actuarial method (Nonparametric Distribution Analysis—Arbitrary Censoring)—see *Survival probabilities* on page 15-56.
- specify a confidence level for all confidence intervals. The default is 95.0%.
- choose to calculate two-sided confidence intervals, or lower or upper bounds. The default is two-sided.

Graphs subdialog box

- draw a survival plot, with or without confidence intervals—see *Nonparametric survival plots* on page 15-61.

- draw a hazard plot—see *Hazard plots* on page 15-62.
- specify minimum and/or maximum values for the x-axis scale.
- enter a label for the x axis.

Results subdialog box

- display the following Session window output:

  Nonparametric Distribution Analysis—Right Censoring
  - no output.
  - the basic output, which includes variable information, censoring information, characteristics of variable, and test statistics for comparing survival curves.
  - the basic output, plus the Kaplan-Meier survival probabilities or actuarial table.
  - the above output, plus hazard, density (actuarial method) estimates, and log-rank and Wilcoxon statistics. The log-rank and Wilcoxon statistics are used to compare survival curves when you have more than one sample—*Comparing survival curves (nonparametric distribution analysis—right censoring only)* on page 15-62.

  Nonparametric Distribution Analysis—Arbitrary Censoring
  - no output
  - the basic output, which includes variable information, censoring information, and characteristics of the variable (actuarial method)
  - the basic output, plus the Turnbull survival probabilities or actuarial table
  - the above output, plus hazard and density estimates (actuarial method)

Storage subdialog box

- store any of these nonparametric estimates:
  - survival probabilities and their times, standard errors, and confidence limits
  - hazard rates and their times

Output

The nonparametric distribution analysis output differs depending on whether your data are uncensored/right censored or arbitrarily censored.

When your data are uncensored/right censored you get

- the censoring information
- characteristics of the variable, which includes the mean, its standard error and 95% confidence intervals, median, interquartile range, Q1, and Q3
- Kaplan-Meier estimates of survival probabilities and their
  - standard error

  - 95% confidence intervals

When your data are arbitrarily censored you get

- the censoring information
- Turnbull estimates of the probability of failure and their standard errors
- Turnbull estimates of the survival probabilities and their standard errors and 95% confidence intervals

Survival probabilities

What is the probability of an engine winding running past a given time? How likely is it that a cancer patient will live five years after receiving a certain drug? You are looking for *survival probabilities*. Survival probabilities estimate the proportion of units surviving at time t.

You can choose various estimation methods, depending on the command.

- **Nonparametric Distribution Analysis—Right Censoring** automatically displays a table of Kaplan-Meier survival estimates. Alternatively, you can request Actuarial survival estimates. The two methods are very similar, but where the Kaplan-Meier method displays information for individual failure times, the Actuarial method displays information for *groupings* of failure times. The Actuarial method is generally used for large samples where you have natural groupings, such as human mortality data, which are commonly grouped into one-year intervals, or warranty data. The intervals may be equal or unequal in size.
- **Nonparametric Distribution Analysis—Arbitrary Censoring** automatically displays a table of Turnbull survival estimates. Alternatively, you can request Actuarial survival estimates.

To plot the survival probabilities versus time, see *Nonparametric survival plots* on page 15-61.

Kaplan-Meier survival estimates (Nonparametric Distribution Analysis—Right Censoring only)

With Nonparametric Distribution Analysis—Right Censoring, the default output includes the characteristics of the variable, and a table of Kaplan-Meier survival estimates. You can also request hazard estimates (empirical hazard function) in the Results subdialog box.

Here we entered failure times for engine windings:

```
Characteristics of Variable

Standard          95.0% Normal CI
Mean(MTTF)         Error        Lower        Upper
   55.7000        2.2069      51.3746      60.0254
Median =     55.0000
IQR =                  *  Q1 =     48.0000  Q3 =           *

Kaplan-Meier Estimates
Number     Number      Survival    Standard     95.0% Normal CI
      Time    at Risk     Failed  Probability       Error     Lower      Upper
   23.0000         50          1       0.9800      0.0198    0.9412     1.0000
   24.0000         49          1       0.9600      0.0277    0.9057     1.0000
   27.0000         48          2       0.9200      0.0384    0.8448     0.9952
   31.0000         46          1       0.9000      0.0424    0.8168     0.9832
   34.0000         45          1       0.8800      0.0460    0.7899     0.9701
   35.0000         44          1       0.8600      0.0491    0.7638     0.9562
   etc.
```

At 35 months, 86% of the units are still running.

- **Characteristics of Variable** displays common measures of the center and spread of the distribution. The mean is not resistant to large lifetimes, while the median, Q1 (25th percentile), Q3 (75th percentile) and the IQR (interquartile range) are resistant.
- **Kaplan-Meier Estimates** contains the *Survival Probability* column—estimates of the proportion of units still surviving at time t.

For each failure time t, MINITAB also displays the number of units at risk, the number failed, and the standard error and 95.0% confidence interval for the survival probabilities.

Additional output

You can request this additional output in the Results subdialog box:

```
Empirical Hazard Function

                  Hazard
     Time      Estimates
  23.0000        0.02000
  24.0000        0.02041
  27.0000        0.02128
  31.0000        0.02174
  34.0000        0.02222
  etc.
```

- *Hazard Estimates* are measures of the instantaneous failure rate for each time t.

Turnbull survival estimates (Nonparametric Distribution Analysis—Arbitrary Censoring only)

With Nonparametric Distribution Analysis—Arbitrary Censoring, the default output includes a table of Turnbull survival estimates.

Here we entered failure times (in miles) for tires.

Turnbull Estimates

Interval lower	Interval upper	Probability of Failure	Standard Error
*	10000.00	0.0103	0.0036
10000.00	20000.00	0.0129	0.0041
20000.00	30000.00	0.0181	0.0048
30000.00	40000.00	0.0323	0.0064
40000.00	50000.00	0.0479	0.0077
50000.00	60000.00	0.1125	0.0114
60000.00	70000.00	0.1876	0.0140
70000.00	80000.00	0.2988	0.0165
80000.00	90000.00	0.1876	0.0140
90000.00	*	0.0918	*

At 40,000 miles, 92.63% of the tires have survived.

Time	Survival Probability	Standard Error	95.0% Normal CI lower	95.0% Normal CI upper
10000.00	0.9897	0.0036	0.9825	0.9968
20000.00	0.9767	0.0054	0.9661	0.9873
30000.00	0.9586	0.0072	0.9446	0.9726
40000.00	0.9263	0.0094	0.9078	0.9447
50000.00	0.8784	0.0118	0.8554	0.9014
60000.00	0.7658	0.0152	0.7360	0.7957
70000.00	0.5783	0.0178	0.5435	0.6131
80000.00	0.2794	0.0161	0.2478	0.3111
90000.00	0.0918	0.0104	0.0715	0.1122

The *Probability of Failure* column contains estimates of the probability of failing during the interval.

The *Survival Probability* column contains estimates of the proportion of units still surviving at time t—in our case, the number of miles.

For each time t, the table also displays the standard errors for both the probability of failures and survival probabilities, and 95.0% approximate confidence intervals for the survival probabilities.

Actuarial survival estimates

Instead of the default Kaplan-Meier or Turnbull survival estimates, you can request Actuarial estimates in the Estimate subdialog box. Actuarial output includes median residual lifetimes, conditional probabilities of failure, and survival probabilities. When using the actuarial method, you can also request hazard estimates and density estimates in the Results subdialog box.

With Nonparametric Distribution Analysis—Right Censoring, you can request specific time intervals. In this example, we requested equally spaced time intervals from 0–110, in increments of 20:

Characteristics of Variable

Median	Standard Error	95.0% Normal CI lower	upper
56.1905	3.3672	49.5909	62.7900

Additional Time from Time T until 50% of Running Units Fail

Time T	Proportion of Running Units	Additional Time	Standard Error	95.0% Normal CI lower	upper
20.0000	1.0000	36.1905	3.3672	29.5909	42.7900
40.0000	0.8400	20.0000	3.0861	13.9514	26.0486

Actuarial Table

Interval lower	upper	Number Entering	Number Failed	Number Censored	Conditional Probability of Failure	Standard Error
0.000000	20.0000	50	0	0	0.0000	0.0000
20.0000	40.0000	50	8	0	0.1600	0.0518
40.0000	60.0000	42	21	0	0.5000	0.0772
60.0000	80.0000	21	8	4	0.4211	0.1133
80.0000	100.0000	9	0	6	0.0000	0.0000
100.0000	120.0000	3	0	3	0.0000	0.0000

Time	Survival Probability	Standard Error	95.0% Normal CI lower	upper
20.0000	1.0000	0.0000	1.0000	1.0000
40.0000	0.8400	0.0518	0.7384	0.9416
60.0000	0.4200	0.0698	0.2832	0.5568
80.0000	0.2432	0.0624	0.1208	0.3655
100.0000	0.2432	0.0624	0.1208	0.3655

Characteristics of Variable displays the median, its standard error, and 95% confidence interval.

Additional Time from Time T Until 50% of Running Units Fail

- *Additional Time* contains the median residual lifetimes, which estimate the additional time from Time t until half of the running units fail. For example, at 40 months, it will take an estimated additional 20 months until 42% (1/2 of 84%) of the running units fail.

Actuarial Table

- *Conditional Probability of Failure* displays conditional probabilities, which estimate the chance that a unit fails in the interval, given that it had not failed up to this point. For example, between 40 and 60 months, 0.5000 of the units failed, given the unit was running at 40 months.

- *Survival Probability* displays the survival probabilities, which estimate the probability that a unit is running at a given time. For example, 0.8400 of the units are running at 40 months.

For each estimate, MINITAB displays the associated standard errors and, for the survival probabilities, 95.0% approximate confidence intervals.

Additional output

You can request this additional output in the Results subdialog box:

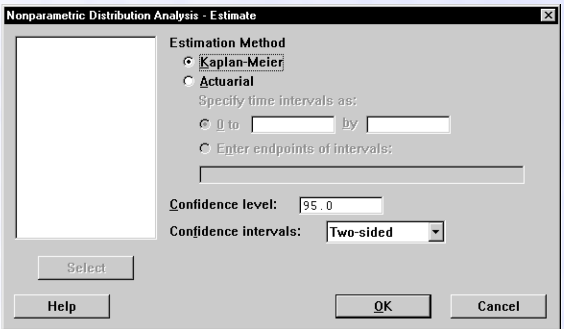
Time	Hazard Estimates	Standard Error	Density Estimates	Standard Error
10.0000	0.000000	*	0.000000	*
30.0000	0.008696	0.003063	0.008000	0.002592
50.0000	0.03333	0.006858	0.02100	0.003490
70.0000	0.02667	0.009087	0.008842	0.002796
90.0000	0.000000	*	0.000000	*
110.0000	0.000000	*	0.000000	*

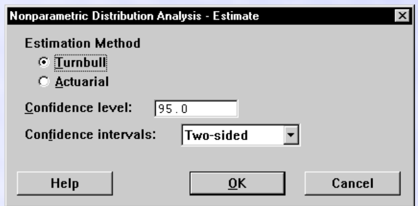
- *Hazard Estimates* estimate the hazard function at the midpoint of the interval. The hazard function is a measure of the instantaneous failure rate for each time t.
- *Density Estimates* estimate the density function at the midpoint of the interval. The density function describes the distribution of failure times.

For each estimate, MINITAB also displays the standard errors.

▶ To request actuarial estimates

1 In the main dialog box, click **Estimate**.





2 Under **Estimation Method**, check **Actuarial**.

3 With Nonparametric Distribution Analysis—Right Censoring, do one of the following:

- use equally spaced time intervals—choose **0 to_by_** and enter numbers in the boxes. For example, **0 to** 100 **by** 20 gives you these time intervals: 0–20, 20–40, and so on up to 80–100.
- use unequally spaced time intervals—choose **Enter endpoints of intervals**, and enter a series of numbers, or a column of numbers, in the box. For example, entering 0 4 6 8 10 20 30, gives you these time intervals: 0–4, 4–6, 6–8, 8–10, 10–20, and 20–30.

4 Click **OK**.

More To display hazard and density estimates in the Actuarial table, from the main dialog box, click **Results**. Do one of the following, then click **OK**:

- With Nonparametric Distribution Analysis—Right Censoring, choose **In addition, hazard, density (actuarial method) estimates, and log-rank and Wilcoxon statistics**.
- With Nonparametric Distribution Analysis—Arbitrary Censoring, choose **In addition, hazard and density estimates (actuarial method)**.

Nonparametric survival plots

Survival (or reliability) plots display the survival probabilities versus time. Each plot point represents the proportion of units surviving at time t. The survival curve is surrounded by two outer lines—the 95% confidence interval for the curve, which provide reasonable values for the "true" survival function.

You can choose from various estimation methods, depending on the command you use:

- With Nonparametric Distribution Analysis—Right Censoring, the survival plot uses Kaplan-Meier survival estimates by default, but you can choose to plot Actuarial estimates.
- With Nonparametric Distribution Analysis—Arbitrary Censoring, the survival plot uses Turnbull survival estimates by default, but you can choose to plot Actuarial estimates.

You can interpret the nonparametric survival curve in a similar manner as you would the parametric survival curve on page 15-40. The major difference is that the nonparametric survival curve is a step function while the parametric survival curve is a smoothed function.

To draw a nonparametric survival plot, check Survival plot in the Graphs subdialog box. By default, the survival plot uses Kaplan-Meier (Nonparametric Distribution Analysis—Right Censoring) or Turnbull (Nonparametric Distribution Analysis—Arbitrary Censoring) estimates of the survival function. If you want to plot Actuarial estimates, choose Actuarial method in the Estimate subdialog box. See *To request actuarial estimates* on page 15-60.

For computations, see Help.

Comparing survival curves (nonparametric distribution analysis—right censoring only)

When you enter more than one sample, Nonparametric Distribution Analysis—Right Censoring automatically compares their survival curves, and displays this table in the Session window:

```
Comparison of Survival Curves

Test Statistics
Method     Chi-Square    DF    P-Value
Log-Rank       7.7152     1     0.0055
Wilcoxon      13.1326     1     0.0003
```

This table contains measures that tell you if the survival curves for various samples are significantly different. A p-value $< \alpha$ indicates that the survival curves are significantly different.

To get more detailed log-rank and Wilcoxon statistics, choose **In addition, hazard, density (actuarial method) estimates and log-rank and Wilcoxon statistics** in the Results subdialog box.

Hazard plots

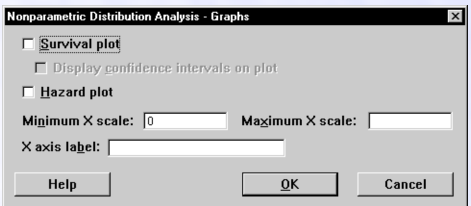
Nonparametric hazard estimates are calculated various ways:

- Nonparametric Distribution Analysis—Right Censoring automatically plots the empirical hazard function. You can optionally plot Actuarial estimates.
- Nonparametric Distribution Analysis—Arbitrary Censoring only plots Actuarial estimates. Since the Actuarial method is not the default estimation method, be sure to choose Actuarial method in the Estimate subdialog box when you want to draw a hazard plot.

For a general description, see *Hazard plots* on page 15-41. For computations, see Help.

▶ **To draw a hazard plot (nonparametric distribution analysis—right censoring command)**

1 In the Nonparametric Distribution Analysis—Right Censoring dialog box, click **Graphs**.



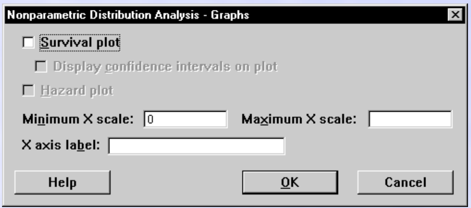
2 Check **Hazard plot**, then click **OK**.

More By default, Nonparametric Distribution Analysis—Right Censoring's hazard plot uses the empirical hazard function. If you want to plot Actuarial estimates, choose **Actuarial method** in the Estimate subdialog box. See *To request actuarial estimates* on page 15-60.

▶ To draw a hazard plot (nonparametric distribution analysis—arbitrary censoring command)

1 In the Nonparametric Distribution Analysis—Arbitrary Censoring dialog box, click **Estimate**. Choose **Actuarial**. Click **OK**.

2 Click **Graphs**.



3 Check **Hazard plot**, then click **OK**.

▶ Example of a nonparametric distribution analysis with exact failure/right censored data

Suppose you work for a company that manufactures engine windings for turbine assemblies. Engine windings may decompose at an unacceptable rate at high temperatures. You decide to look at failure times for engine windings at two temperatures, 80 and 100°C. You want to find out the following information for each temperature:

- the times at which half of the windings fail
- the proportion of windings that survive past various times

You also want to know whether or not the survival curves at the two temperatures are significantly different.

In the first sample, you collect times to failure for 50 windings at 80°C; in the second sample, you collect times to failure for 40 windings at 100°C. Some of the windings drop out of the test for unrelated reasons. In the MINITAB worksheet, you use a column of censoring indicators to designate which times are actual failures (1) and which are censored units removed from the test before failure (0).

1 Open the worksheet RELIABLE.MTW.

2 Choose **Stat ➤ Reliability/Survival ➤ Nonparametric Dist Analysis–Right Cens**.

3 In **Variables**, enter *Temp80 Temp100*.

4 Click **Censor**. Choose **Use censoring columns** and enter *Cens80 Cens100* in the box. Click **OK**.

5 Click **Graphs**. Check **Survival plot** and **Display confidence intervals on plot**. Click **OK** in each dialog box.

The output for the 100°C sample follows that of the 80°C sample. The comparison of survival curves shows up last.

Session window output

Distribution Analysis: Temp80

```
Variable:  Temp80

Censoring Information                 Count
Uncensored value                         37
Right censored value                     13
Censoring value:  Cens80 = 0

Nonparametric Estimates

Characteristics of Variable

              Standard        95.0% Normal CI
Mean(MTTF)       Error        Lower       Upper
   55.7000      2.2069      51.3746     60.0254

Median =     55.0000
IQR =              *  Q1 =     48.0000  Q3 =          *

Kaplan-Meier Estimates

              Number    Number    Survival    Standard     95.0% Normal CI
      Time   at Risk    Failed  Probability      Error     Lower     Upper
   23.0000        50         1       0.9800     0.0198    0.9412    1.0000
   24.0000        49         1       0.9600     0.0277    0.9057    1.0000
   27.0000        48         2       0.9200     0.0384    0.8448    0.9952
   31.0000        46         1       0.9000     0.0424    0.8168    0.9832
   34.0000        45         1       0.8800     0.0460    0.7899    0.9701
   35.0000        44         1       0.8600     0.0491    0.7638    0.9562
   37.0000        43         1       0.8400     0.0518    0.7384    0.9416
   40.0000        42         1       0.8200     0.0543    0.7135    0.9265
   41.0000        41         1       0.8000     0.0566    0.6891    0.9109
   45.0000        40         1       0.7800     0.0586    0.6652    0.8948

Distribution Analysis: Temp100

Variable:  Temp100

Censoring Information                 Count
Uncensored value                         34
Right censored value                      6
Censoring value:  Cens100 = 0
```

Nonparametric Estimates

Characteristics of Variable

	Standard	95.0% Normal CI	
Mean(MTTF)	Error	Lower	Upper
41.6563	3.4695	34.8561	48.4564

Median = 38.0000
IQR = 30.0000 Q1 = 24.0000 Q3 = 54.0000

Kaplan-Meier Estimates

Time	Number at Risk	Number Failed	Survival Probability	Standard Error	95.0% Normal CI Lower	Upper
6.0000	40	1	0.9750	0.0247	0.9266	1.0000
10.0000	39	1	0.9500	0.0345	0.8825	1.0000
11.0000	38	1	0.9250	0.0416	0.8434	1.0000
14.0000	37	1	0.9000	0.0474	0.8070	0.9930
16.0000	36	1	0.8750	0.0523	0.7725	0.9775
18.0000	35	3	0.8000	0.0632	0.6760	0.9240
22.0000	32	1	0.7750	0.0660	0.6456	0.9044
24.0000	31	1	0.7500	0.0685	0.6158	0.8842
25.0000	30	1	0.7250	0.0706	0.5866	0.8634
27.0000	29	1	0.7000	0.0725	0.5580	0.8420
29.0000	28	1	0.6750	0.0741	0.5299	0.8201
30.0000	27	1	0.6500	0.0754	0.5022	0.7978
32.0000	26	1	0.6250	0.0765	0.4750	0.7750
35.0000	25	1	0.6000	0.0775	0.4482	0.7518

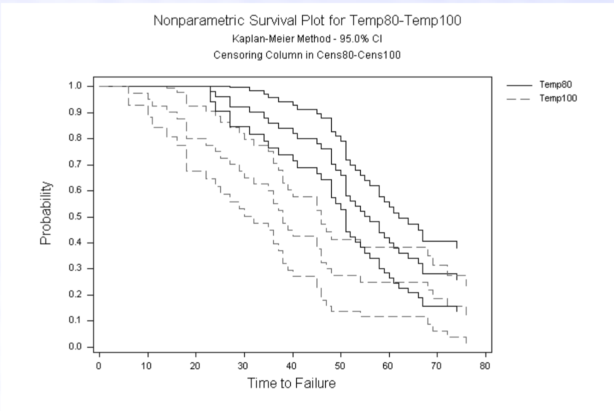
Distribution Analysis: Temp80, Temp100

Comparison of Survival Curves

Test Statistics

Method	Chi-Square	DF	P-Value
Log-Rank	7.7152	1	0.0055
Wilcoxon	13.1326	1	0.0003

Graph window output



Interpreting the results

The estimated median failure time for Temp80 is 55 months and 38 months for Temp100. So the increase in temperature decreased the median failure time by approximately 17 months.

The survival estimates are displayed in the Kaplan-Meier Estimates table. For example, at 80° C, 0.9000 of the windings survive past 31 months, while at 100° C, 0.9000 of the windings survive past 14 months.

Are the survival curves for Temp80 and Temp100 significantly different? In the Test Statistics table, a p-value $< \alpha$ indicates that the survival curves are significantly different. In this case, the small p-values (0.0055 and 0.003) suggest that a change of 20° C plays a significant role in the breakdown of engine windings.

▶ Example of a nonparametric distribution analysis with arbitrarily censored data

Suppose you work for a company that manufactures tires. You are interested in finding out how likely it is that a tire will "fail," or wear down to 2/32 of an inch of tread, within given mileage intervals. You are especially interested in knowing how many of the tires last past 45,000 miles.

You inspect each good tire at regular intervals (every 10,000 miles) to see if the tire fails, then enter the data into the MINITAB worksheet.

1 Open the worksheet TIREWEAR.MTW.

2 Choose **Stat ➤ Reliability/Survival ➤ Nonparametric Dist Analysis–Arbitrary Cens**.

3 In **Start variables**, enter *Start*. In **End variables**, enter *End*.

4 In **Frequency columns**, enter *Freq,* then click **OK**.

Session window output

Distribution Analysis, Start = Start and End = End

```
Variable
Start:  Start    End:  End
Frequency:  Freq

Censoring Information                Count
Right censored value                    71
Interval censored value                694
Left censored value                      8

Turnbull Estimates

         Interval         Probability    Standard
     Lower        Upper    of Failure       Error
         *     10000.00        0.0103      0.0036
  10000.00     20000.00        0.0129      0.0041
  20000.00     30000.00        0.0181      0.0048
  30000.00     40000.00        0.0323      0.0064
  40000.00     50000.00        0.0479      0.0077
  50000.00     60000.00        0.1125      0.0114
  60000.00     70000.00        0.1876      0.0140
  70000.00     80000.00        0.2988      0.0165
  80000.00     90000.00        0.1876      0.0140
  90000.00            *        0.0918           *

                Survival     Standard      95.0% Normal CI
      Time   Probability        Error     Lower      Upper
  10000.00        0.9897       0.0036    0.9825     0.9968
  20000.00        0.9767       0.0054    0.9661     0.9873
  30000.00        0.9586       0.0072    0.9446     0.9726
  40000.00        0.9263       0.0094    0.9078     0.9447
  50000.00        0.8784       0.0118    0.8554     0.9014
  60000.00        0.7658       0.0152    0.7360     0.7957
  70000.00        0.5783       0.0178    0.5435     0.6131
  80000.00        0.2794       0.0161    0.2478     0.3111
  90000.00        0.0918       0.0104    0.0715     0.1122
```

Interpreting the results

The Turnbull Estimates table displays the probabilities of failure. For example, between 60,000 and 70,000 miles, 18.76% of the tires fail.

You can see in the column of survival probabilities that 92.63% of the tires last past 40,000 miles.

References

[1] R.B. D'Agostino and M.A. Stephens (1986). *Goodness-of-Fit Techniques*, Marcel Dekker.

[2] J.D. Kalbfleisch and R.L. Prentice (1980). *The Statistical Analysis of Failure Time Data*, John Wiley & Sons.

[3] D. Kececioglu (1991). *Reliability Engineering Handbook*, Vols 1 and 2, Prentice Hall.

[4] J.F. Lawless (1982). *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, Inc.

[5] W.Q. Meeker and L.A. Escobar (1998). *Statistical Methods for Reliability Data*, John Wiley & Sons, Inc.

[6] W. Murray, Ed. (1972). *Numerical Methods for Unconstrained Optimization*, Academic Press.

[7] W. Nelson (1982). *Applied Life Data Analysis*, John Wiley & Sons.

[8] R. Peto (1973). "Experimental Survival Curves for Interval-censored Data," *Applied Statistics* 22, pp. 86-91.

[9] B.W. Turnbull (1976). "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society* 38, pp. 290-295.

[10] B.W. Turnbull (1974). "Nonparametric Estimation of a Survivorship Function with Doubly Censored Data," *Journal of the American Statistical Association* 69, 345, pp. 169-173.