SAS/QC® 14.2 User's Guide

The RELIABILITY Procedure

This document is an individual chapter from *SAS/QC® 14.2 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2016. *SAS/QC® 14.2 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/QC® 14.2 User's Guide

Copyright © 2016, SAS Institute Inc., Cary, NC, USA

All Rights Reserved. Produced in the United States of America.

For a hard-copy book: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

For a web download or e-book: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication, or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a), and DFAR 227.7202-4, and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

November 2016

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

Chapter 18
The RELIABILITY Procedure

Contents

Overview: RELIABILITY Procedure **1174**
Getting Started: RELIABILITY Procedure **1176**
- Analysis of Right-Censored Data from a Single Population 1176
- Weibull Analysis Comparing Groups of Data 1180
- Analysis of Accelerated Life Test Data 1184
- Weibull Analysis of Interval Data with Common Inspection Schedule 1189
- Lognormal Analysis with Arbitrary Censoring 1194
- Regression Modeling 1199
- Regression Model with Nonconstant Scale 1205
- Regression Model with Two Independent Variables 1208
- Weibull Probability Plot for Two Combined Failure Modes 1211
- Analysis of Recurrence Data on Repairs 1215
- Comparison of Two Samples of Repair Data 1220
- Analysis of Interval Age Recurrence Data 1227
- Analysis of Binomial Data 1230
- Three-Parameter Weibull 1233
- Parametric Model for Recurrent Events Data 1236
- Parametric Model for Interval Recurrent Events Data 1238

Syntax: RELIABILITY Procedure **1241**
- Primary Statements 1241
- Secondary Statements 1241
- Graphical Enhancement Statements 1243
- PROC RELIABILITY Statement 1243
- ANALYZE Statement 1243
- BY Statement 1249
- CLASS Statement 1249
- DISTRIBUTION Statement 1250
- EFFECTPLOT Statement 1251
- ESTIMATE Statement 1252
- FMODE Statement 1253
- FREQ Statement 1254
- INSET Statement 1254
- LOGSCALE Statement 1258
- LSMEANS Statement 1258
- LSMESTIMATE Statement 1260
- MAKE Statement 1261

MCFPLOT Statement 1261
MODEL Statement 1272
NENTER Statement 1280
NLOPTIONS Statement 1280
PROBPLOT Statement 1281
RELATIONPLOT Statement 1295
SLICE Statement 1308
STORE Statement 1309
TEST Statement 1309
UNITID Statement 1309
Details: RELIABILITY Procedure **1310**
Abbreviations and Notation 1310
Types of Lifetime Data 1310
Probability Distributions 1310
Probability Plotting 1313
Nonparametric Confidence Intervals for Cumulative Failure Probabilities 1322
Parameter Estimation and Confidence Intervals 1324
Regression Model Statistics Computed for Each Observation for Lifetime Data 1340
Regression Model Statistics Computed for Each Observation for Recurrent Events Data 1345
Recurrence Data from Repairable Systems 1346
ODS Table Names 1358
ODS Graphics 1360
References **1361**

Overview: RELIABILITY Procedure

The RELIABILITY procedure provides tools for reliability and survival data analysis and for recurrent events data analysis. You can use this procedure to

- construct probability plots and fitted life distributions with left-censored, right-censored, and interval-censored lifetime data
- fit regression models, including accelerated life test models, to combinations of left-censored, right-censored, and interval-censored data
- analyze recurrence data from repairable systems

These tools benefit reliability engineers and industrial statisticians working with product life data and system repair data. They also aid workers in other fields, such as medical research, pharmaceuticals, social sciences, and business, where survival and recurrence data are analyzed.

Most practical problems in reliability data analysis involve right-censored, left-censored, or interval-censored data. The RELIABILITY procedure provides probability plots of uncensored, right-censored, interval-censored, and arbitrarily censored data.

Features of the RELIABILITY procedure include

- probability plotting and parameter estimation for the common life distributions: Weibull, three-parameter Weibull, exponential, extreme value, normal, lognormal, logistic, and log-logistic. The data can be complete, right censored, or interval censored.
- maximum likelihood estimates of distribution parameters, percentiles, and reliability functions
- both asymptotic normal and likelihood ratio confidence intervals for distribution parameters and percentiles. Asymptotic normal confidence intervals for the reliability function are also available.
- estimation of distribution parameters by least squares fitting to the probability plot
- Weibayes analysis, where there are no failures and where the data analyst specifies a value for the Weibull shape parameter
- estimates of the resulting distribution when specified failure modes are eliminated
- plots of the data and the fitted relation for life versus stress in the analysis of accelerated life test data
- fitting of regression models to life data, where the life distribution location parameter is a linear function of covariates. The fitting yields maximum likelihood estimates of parameters of a regression model with a Weibull, exponential, extreme value, normal, lognormal, logistic and log-logistic, or generalized gamma distribution. The data can be complete, right censored, left censored, or interval censored. For example, accelerated life test data can be modeled with such a regression model.
- nonparametric estimates and plots of the mean cumulative function for cost or number of recurrences and associated confidence intervals from data with exact or interval recurrence ages
- maximum likelihood estimation of the parameters of parametric models for recurrent events data
- horizontal plots of failure times for recurrent events data

Some of the features provided in the RELIABILITY procedure are available in other SAS procedures.

- You can construct probability plots of life data with the CAPABILITY procedure; however, the CAPABILITY procedure is intended for process capability analysis rather than reliability analysis, and the data must be complete (that is, uncensored).
- The LIFEREG procedure fits regression models with life distributions such as the Weibull, lognormal, and log-logistic to left-, right-, and interval-censored data. The RELIABILITY procedure fits the same distributions and regression models as the LIFEREG procedure and, in addition, provides a graphical display of life data in probability plots.

Lawless (2003), Meeker and Escobar (1998), Nelson (1982, 1990), Abernethy (2006), and Tobias and Trindade (1995) provide many examples taken from diverse fields and describe the analyses provided by the RELIABILITY procedure.

The features of the procedure that deal with the nonparametric analysis of recurrent events data from repairable systems are based on the work of Doganaksoy and Nelson (1998), Nelson (1988, 1995, 2002), and Nelson and Doganaksoy (1989), who provide examples of repair data analysis. Meeker and Escobar (1998), Rigdon and Basu (2000), Cook and Lawless (2007), Abernethy (2006), Tobias and Trindade (1995), Crowder et al. (1991), and US Army (2000) provide details of parametric models for recurrent events data.

Getting Started: RELIABILITY Procedure

This section introduces the RELIABILITY procedure with examples that illustrate some of the analyses that it performs.

Analysis of Right-Censored Data from a Single Population

The Weibull distribution is used in a wide variety of reliability analysis applications. This example illustrates the use of the Weibull distribution to model product life data from a single population. The following statements create a SAS data set containing observed and right-censored lifetimes of 70 diesel engine fans (Nelson 1982, p. 318):

```
data fan;
   input Lifetime censor @@;
   Lifetime = Lifetime/1000;
   label lifetime='Fan Life (1000s of Hours)';
   datalines;
 450 0    460 1   1150 0   1150 0   1560 1
1600 0   1660 1   1850 1   1850 1   1850 1
1850 1   1850 1   2030 1   2030 1   2030 1
2070 0   2070 0   2080 0   2200 1   3000 1
3000 1   3000 1   3000 1   3100 0   3200 1
3450 0   3750 1   3750 1   4150 1   4150 1
4150 1   4150 1   4300 1   4300 1   4300 1
4300 1   4600 0   4850 1   4850 1   4850 1
4850 1   5000 1   5000 1   5000 1   6100 1
6100 0   6100 1   6100 1   6300 1   6450 1
6450 1   6700 1   7450 1   7800 1   7800 1
8100 1   8100 1   8200 1   8500 1   8500 1
8500 1   8750 1   8750 0   8750 1   9400 1
9900 1  10100 1  10100 1  10100 1  11500 1
;
```

Some of the fans had not failed at the time the data were collected, and the unfailed units have right-censored lifetimes. The variable Lifetime represents either a failure time or a censoring time in thousands of hours. The variable Censor is equal to 0 if the value of Lifetime is a failure time, and it is equal to 1 if the value is a censoring time.

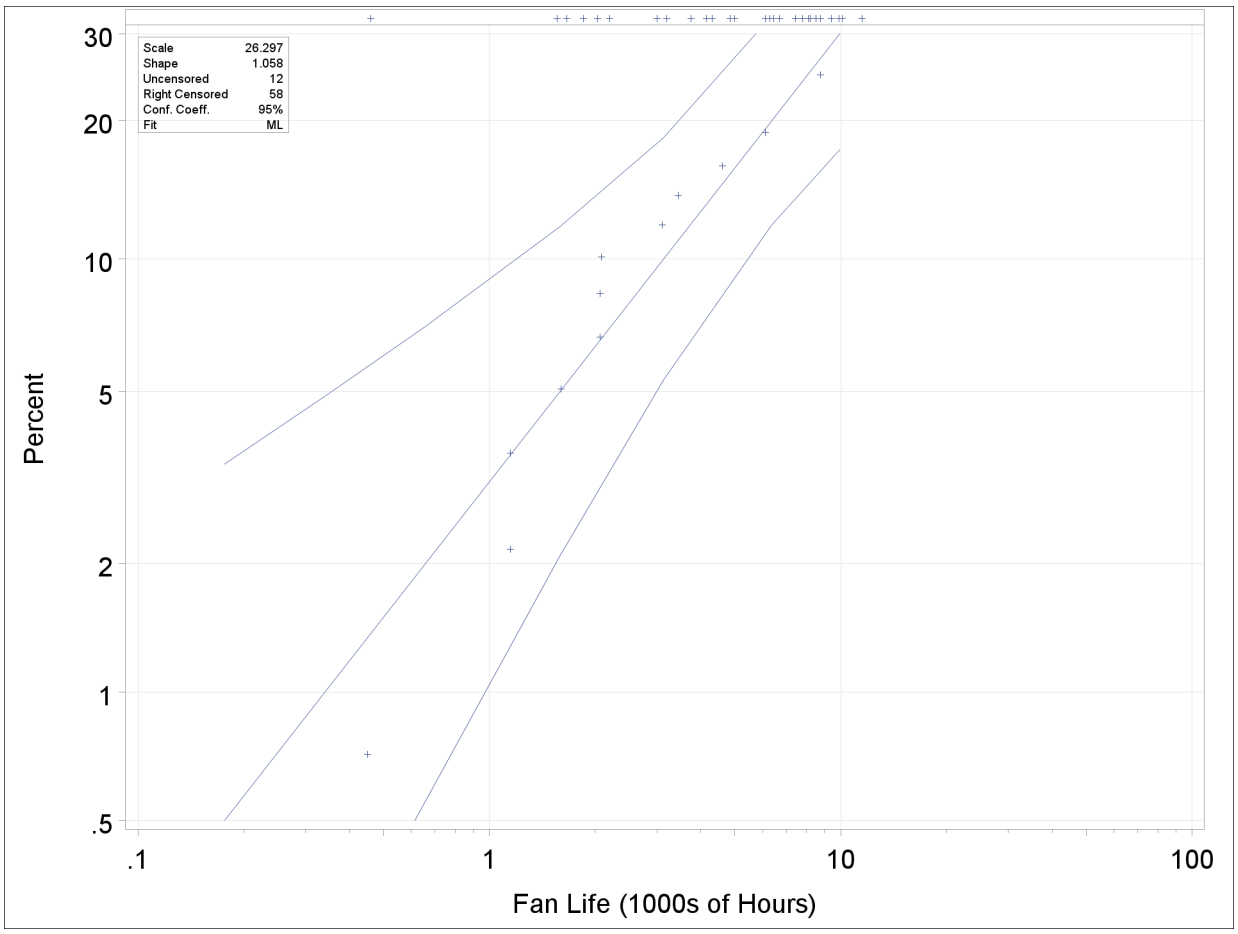
If ODS Graphics is disabled, graphical output is created using traditional graphics; otherwise, ODS Graphics is used. The following statements use the RELIABILITY procedure to produce the traditional graphical output shown in Figure 18.1:

```
ODS Graphics OFF;
proc reliability data=fan;
   distribution Weibull;
   pplot lifetime*censor( 1 ) /  covb ;
run;
ODS Graphics ON;
```

The DISTRIBUTION statement specifies the Weibull distribution for probability plotting and maximum likelihood (ML) parameter estimation. The PROBPLOT statement produces a probability plot for the variable Lifetime and specifies that the value of 1 for the variable Censor denotes censored observations. You can specify any value, or group of values, for the *censor-variable* (in this case, Censor) to indicate censoring times. The option COVB requests the ML parameter estimate covariance matrix.

The graphical output, displayed in Figure 18.1, consists of a probability plot of the data, an ML fitted distribution line, and confidence intervals for the percentile (lifetime) values. An *inset* box containing summary statistics, Weibull scale and shape estimates, and other information is displayed on the plot by default. The locations of the right-censored data values are plotted as plus signs in an area at the top of the plot.

Figure 18.1 Weibull Probability Plot for Engine Fan Data (Traditional Graphics)

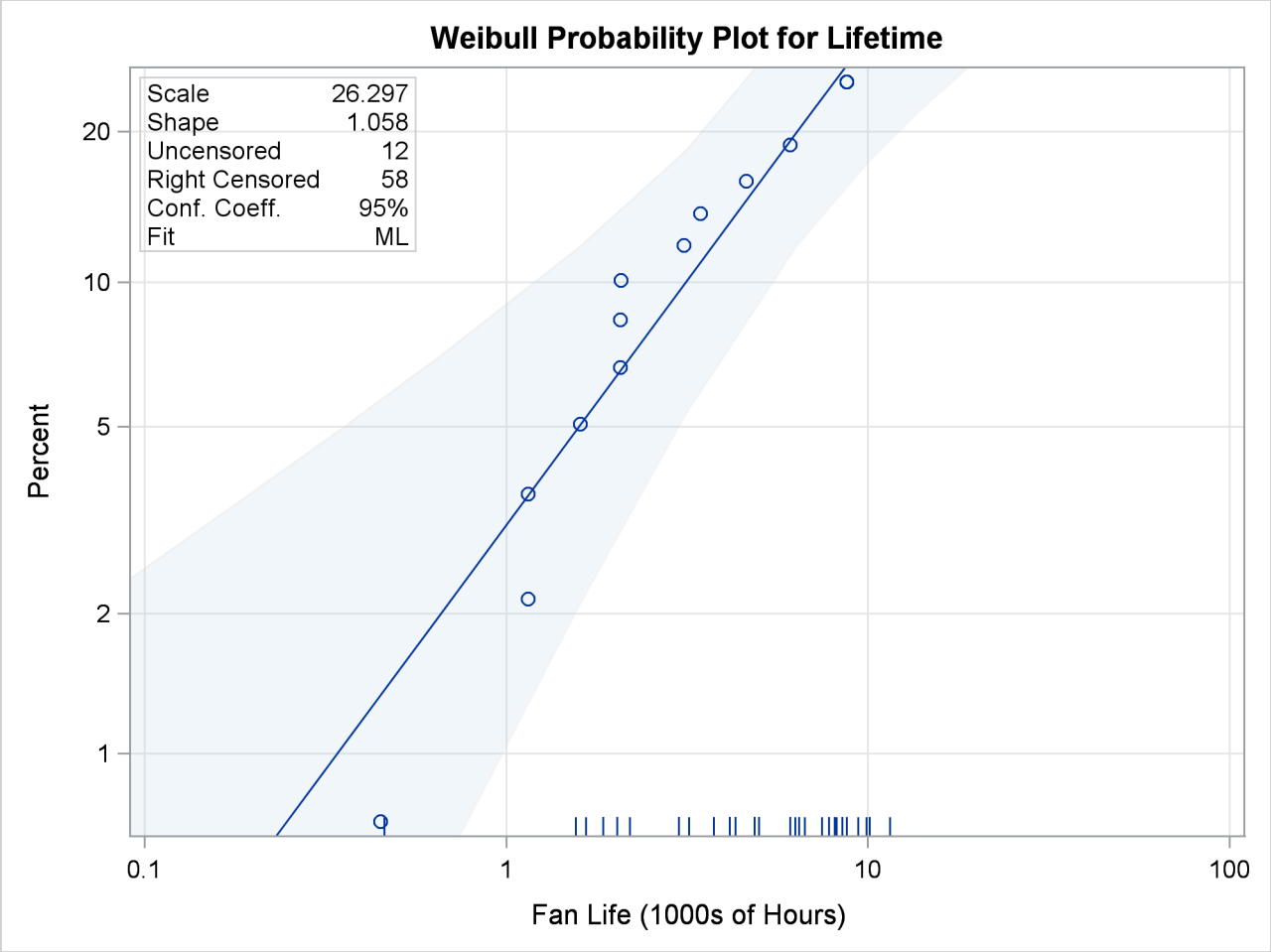


If ODS Graphics is enabled, you can create the probability plot by using ODS Graphics. The following SAS statements use ODS Graphics to create the probability plot shown in Figure 18.1:

```
proc reliability data=fan;
   distribution Weibull;
   pplot lifetime*censor( 1 ) /  covb;
run;
```

The plot is shown in Figure 18.2.

Figure 18.2 Weibull Probability Plot for Engine Fan Data (ODS Graphics)



The tabular output produced by the preceding SAS statements is shown in Figure 18.3 and Figure 18.4. This consists of summary data, fit information, parameter estimates, distribution percentile estimates, standard errors, and confidence intervals for all estimated quantities.

Figure 18.3 Tabular Output for the Fan Data Analysis

The RELIABILITY Procedure

Model Information		
Input Data Set	WORK.FAN	
Analysis Variable	Lifetime	Fan Life (1000s of Hours)
Censor Variable	censor	
Distribution	Weibull	
Estimation Method	Maximum Likelihood	
Confidence Coefficient	95%	
Observations Used	70	

Algorithm converged.

Figure 18.3 *continued*

Summary of Fit	
Observations Used	70
Uncensored Values	12
Right Censored Values	58
Maximum Loglikelihood	-42.248

Weibull Parameter Estimates

Parameter	Estimate	Standard Error	Asymptotic Normal 95% Confidence Limits Lower	Upper
EV Location	3.2694	0.4659	2.3563	4.1826
EV Scale	0.9448	0.2394	0.5749	1.5526
Weibull Scale	26.2968	12.2514	10.5521	65.5344
Weibull Shape	1.0584	0.2683	0.6441	1.7394

Other Weibull Distribution Parameters

Parameter	Value
Mean	25.7156
Mode	1.7039
Median	18.6002
Standard Deviation	24.3066

Estimated Covariance Matrix Weibull Parameters

	EV Location	EV Scale
EV Location	0.21705	0.09044
EV Scale	0.09044	0.05733

Estimated Covariance Matrix Weibull Parameters

	Weibull Scale	Weibull Shape
Weibull Scale	150.09724	-2.66446
Weibull Shape	-2.66446	0.07196

Figure 18.4 Percentile Estimates for the Fan Data

Weibull Percentile Estimates				
			Asymptotic Normal 95% Confidence Limits	
Percent	Estimate	Standard Error	Lower	Upper
0.1	0.03852697	0.05027782	0.002985	0.49726229
0.2	0.07419554	0.08481353	0.00789519	0.69725757
0.5	0.17658807	0.16443381	0.02846732	1.09540855
1	0.34072273	0.2635302	0.07482449	1.55152389
2	0.65900116	0.40845639	0.19556981	2.22060107
5	1.58925244	0.68465855	0.68311002	3.69738878
10	3.13724079	0.99379006	1.68620756	5.83693255
20	6.37467675	1.74261908	3.73051433	10.8930029
30	9.92885165	3.00353842	5.48788931	17.9635721
40	13.9407124	4.85766683	7.04177638	27.5986417
50	18.6002319	7.40416922	8.52475116	40.5840149
60	24.2121441	10.8733301	10.0408557	58.3842593
70	31.3378076	15.750336	11.7018888	83.9230489
80	41.2254517	23.1787018	13.6956839	124.092954
90	57.8253251	36.9266698	16.5405275	202.156081
95	74.1471722	51.6127806	18.9489625	290.137423
99	111.307797	88.1380261	23.5781482	525.462197
99.9	163.265082	144.264145	28.8905203	922.637827

Weibull Analysis Comparing Groups of Data

This example illustrates probability plotting and distribution fitting for data grouped by the levels of a special *group-variable*. The data are from an accelerated life test of an insulating fluid and are the times to electrical breakdown of the fluid under different high voltage levels. Each voltage level defines a subset of data for which a separate analysis and Weibull plot are produced. These data are the 26kV, 30kV, 34kV, and 38kV groups of the data provided by Nelson (1990, p. 129). The following statements create a SAS data set containing the lifetimes and voltages:

```
data fluid;
   input Time voltage $ @@;
   datalines;
5.79    26kv    1579.52 26kv
2323.7  26kv    7.74    30kv
17.05   30kv    20.46   30kv
21.02   30kv    22.66   30kv
43.4    30kv    47.3    30kv
139.07  30kv    144.12  30kv
175.88  30kv    194.90  30kv
0.19     34kv    .78     34kv
0.96     34kv    1.31    34kv
2.78    34kv    3.16    34kv
4.15    34kv    4.67    34kv
4.85    34kv    6.50    34kv
7.35    34kv    8.01    34kv
8.27    34kv    12.06   34kv
31.75   34kv    32.52   34kv
33.91   34kv    36.71   34kv
72.89   34kv    .09     38kv
0.39     38kv    .47     38kv
0.73     38kv    .74     38kv
1.13    38kv    1.40    38kv
2.38    38kv
;
```

The variable Time provides the time to breakdown in minutes, and the variable Voltage provides the voltage level at which the test was conducted. These data are not censored.

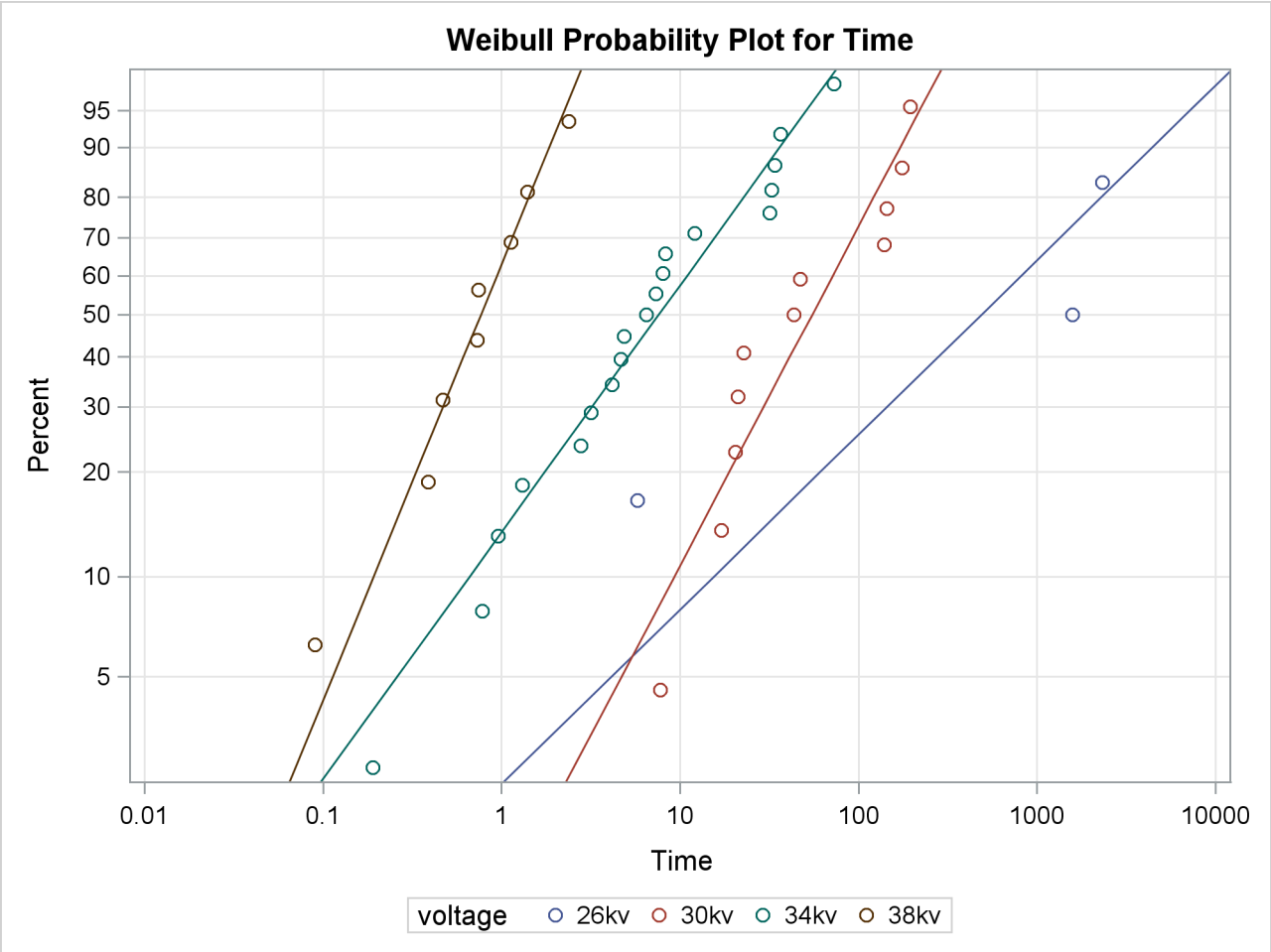
The RELIABILITY procedure plots the data for the different voltage levels on the same Weibull probability plot, fits a separate distribution to the data at each voltage level, and superimposes distribution lines on the plot.

The following statements produce the probability plot shown in Figure 18.5 for the variable Time at each level of the *group-variable* Voltage:

```
proc reliability data=fluid;
   distribution Weibull;
   pplot time=voltage  / overlay
                         noconf;
run;
```

The input data set FLUID is specified by the DATA= option in the PROC RELIABILITY statement. The PROBPLOT statement option OVERLAY specifies that plots for the groups are to be overlaid rather than displayed separately. The option NOCONF specifies that no confidence bands are to be plotted, since these can interfere with one another on overlaid plots; confidence bands are displayed by default.

Figure 18.5 Weibull Probability Plot for the Insulating Fluid Data



A summary table that contains information for all groups is displayed. In addition, information identical to that shown in Figure 18.3 is tabulated for each level of voltage. The summary table for all groups and the tables for the 26kV group are shown in Figure 18.6 and Figure 18.7.

Figure 18.6 Partial Listing of the Tabular Output for the Insulating Fluid Data

The RELIABILITY Procedure

Model Information - All Groups	
Input Data Set	WORK.FLUID
Analysis Variable	Time
Distribution	Weibull
Estimation Method	Maximum Likelihood
Confidence Coefficient	95%
Observations Used	41

The RELIABILITY Procedure

Algorithm converged for group 26kv.

Figure 18.6 *continued*

Summary of Fit		Group
Observations Used	3	26kv
Uncensored Values	3	26kv
Maximum Loglikelihood	-6.845551	26kv

Figure 18.7 Partial Listing of the Tabular Output for the Insulating Fluid Data

The RELIABILITY Procedure

Model Information - All Groups	
Input Data Set	WORK.FLUID
Analysis Variable	Time
Distribution	Weibull
Estimation Method	Maximum Likelihood
Confidence Coefficient	95%
Observations Used	41

Weibull Parameter Estimates

			Asymptotic Normal 95% Confidence Limits		
Parameter	**Estimate**	**Standard Error**	**Lower**	**Upper**	**Group**
EV Location	6.8625	1.1040	4.6986	9.0264	26kv
EV Scale	1.8342	0.9611	0.6568	5.1226	26kv
Weibull Scale	955.7467	1055.1862	109.7941	8319.6794	26kv
Weibull Shape	0.5452	0.2857	0.1952	1.5226	26kv

Other Weibull Distribution Parameters

Parameter	Value	Group
Mean	1649.4882	26kv
Mode	0.0000	26kv
Median	487.9547	26kv
Standard Deviation	3279.0212	26kv

Figure 18.7 *continued*

Weibull Percentile Estimates					
			Asymptotic Normal 95% Confidence Limits		
Percent	Estimate	Standard Error	Lower	Upper	Group
0.1	0.00300636	0.02113841	3.11203E-9	2904.27046	26kv
0.2	0.01072998	0.06838144	4.03597E-8	2852.65767	26kv
0.5	0.0577713	0.31803193	1.19079E-6	2802.78862	26kv
1	0.20695478	1.00385021	0.00001538	2784.16263	26kv
2	0.74484901	3.12705686	0.00019885	2790.0941	26kv
5	4.1142692	13.7388263	0.00591379	2862.3304	26kv
10	15.406565	41.4763373	0.07873508	3014.69497	26kv
20	61.0231127	125.020566	1.10053199	3383.65475	26kv
30	144.246801	242.203982	5.36856883	3875.73303	26kv
40	278.770459	398.048692	16.9761581	4577.77125	26kv
50	487.954708	610.02855	42.0948552	5656.26835	26kv
60	814.147288	920.537706	88.770543	7466.84412	26kv
70	1343.42243	1433.97868	165.818889	10884.0666	26kv
80	2287.87124	2445.52431	281.5628	18590.3635	26kv
90	4412.96962	5148.34986	448.419608	43428.7452	26kv
95	7150.89745	9248.2654	566.892142	90202.9338	26kv
99	15735.8513	24666.0388	728.831025	339745.437	26kv
99.9	33104.172	62018.1074	841.826189	1301796.28	26kv

Analysis of Accelerated Life Test Data

The following example illustrates the analysis of an accelerated life test for Class B electrical motor insulation. The data are provided by Nelson (1990, p. 243). Forty insulation specimens were tested at four temperatures: 150°, 170°, 190°, and 220°C. The purpose of the test is to estimate the median life of the insulation at the design operating temperature of 130°C.

The following SAS program creates the data listed in Figure 18.8. Ten specimens of the insulation were tested at each test temperature. The variable Time provides a specimen time to failure or a censoring time, in hours. The variable Censor is equal to 1 if the value of the variable Time is a right-censoring time and is equal to 0 if the value is a failure time. Some censor times and failure times are identical at some of the temperatures. Rather than repeating identical observations in the input data set, the variable Count provides the number of specimens with identical times and temperatures. The variable Temp provides the test temperature in degrees centigrade. The variable Cntrsl is a control variable specifying that percentiles are to be computed only for the first value of Temp (130°C). The value of Temp in the first observation (130°C) does not correspond to a test temperature. The missing values in the first observation cause the observation to be excluded from the model fit, and the value of 1 for the variable Cntrl causes percentiles corresponding to a temperature of 130°C to be computed.

```
data classb;
   input hours temp count censor;
   if _n_ = 1 then cntrl=1;
   else cntrl=0;
   label hours='Hours';
   datalines;
   . 130 .  .
8064 150 10 1
1764 170 1  0
2772 170 1  0
3444 170 1  0
3542 170 1  0
3780 170 1  0
4860 170 1  0
5196 170 1  0
5448 170 3  1
 408 190 2  0
1344 190 2  0
1440 190 1  0
1680 190 5  1
 408 220 2  0
 504 220 3  0
 528 220 5  1
;
```

Figure 18.8 Listing of the Class B Insulation Data

Obs	hours	temp	count	censor	cntrl
1	.	130	.	.	1
2	8064	150	10	1	0
3	1764	170	1	0	0
4	2772	170	1	0	0
5	3444	170	1	0	0
6	3542	170	1	0	0
7	3780	170	1	0	0
8	4860	170	1	0	0
9	5196	170	1	0	0
10	5448	170	3	1	0
11	408	190	2	0	0
12	1344	190	2	0	0
13	1440	190	1	0	0
14	1680	190	5	1	0
15	408	220	2	0	0
16	504	220	3	0	0
17	528	220	5	1	0

An Arrhenius-lognormal model is fitted to the data in this example. In other words, the failure times follow a lognormal (base 10) distribution, and the lognormal location parameter μ depends on the centigrade temperature Temp through the Arrhenius relationship

$$\mu(x) = \beta_0 + \beta_1 x$$

where

$$x = \frac{1000}{\text{Temp} + 273.15}$$

is 1000 times the reciprocal absolute temperature. The lognormal (base e) distribution is also available.

The following SAS statements fit the Arrhenius-lognormal model, and they display the fitted model distributions side-by-side on the probability and the relation plots shown in Figure 18.9:

```
proc reliability;
   distribution lognormal10;
   freq count;
   model hours*censor(1) = temp /
      relation = arr
      obstats(quantile = .1 .5 .9 control = cntrl);
   rplot hours*censor(1) = temp /
      pplot
      fit = model
      noconf
      relation = arr
      plotdata
      plotfit 10 50 90
      lupper = 1.e5
      slower = 120;
run;
```

The PROC RELIABILITY statement invokes the procedure and specifies CLASSB as the input data set. The DISTRIBUTION statement specifies that the lognormal (base 10) distribution is to be used for maximum likelihood parameter estimation and probability plotting. The FREQ statement specifies that the variable Count is to be used as a frequency variable; that is, if Count=n, then there are n specimens with the time and temperature specified in the observation.

The MODEL statement fits a linear regression equation for the distribution location parameter as a function of independent variables. In this case, the MODEL statement also transforms the independent variable through the Arrhenius relationship. The dependent variable is specified as Time. A value of 1 for the variable Censor indicates that the corresponding value of Time is a right-censored observation; otherwise, the value is a failure time. The temperature variable Temp is specified as the independent variable in the model. The MODEL statement option RELATION=ARR specifies the Arrhenius relationship.

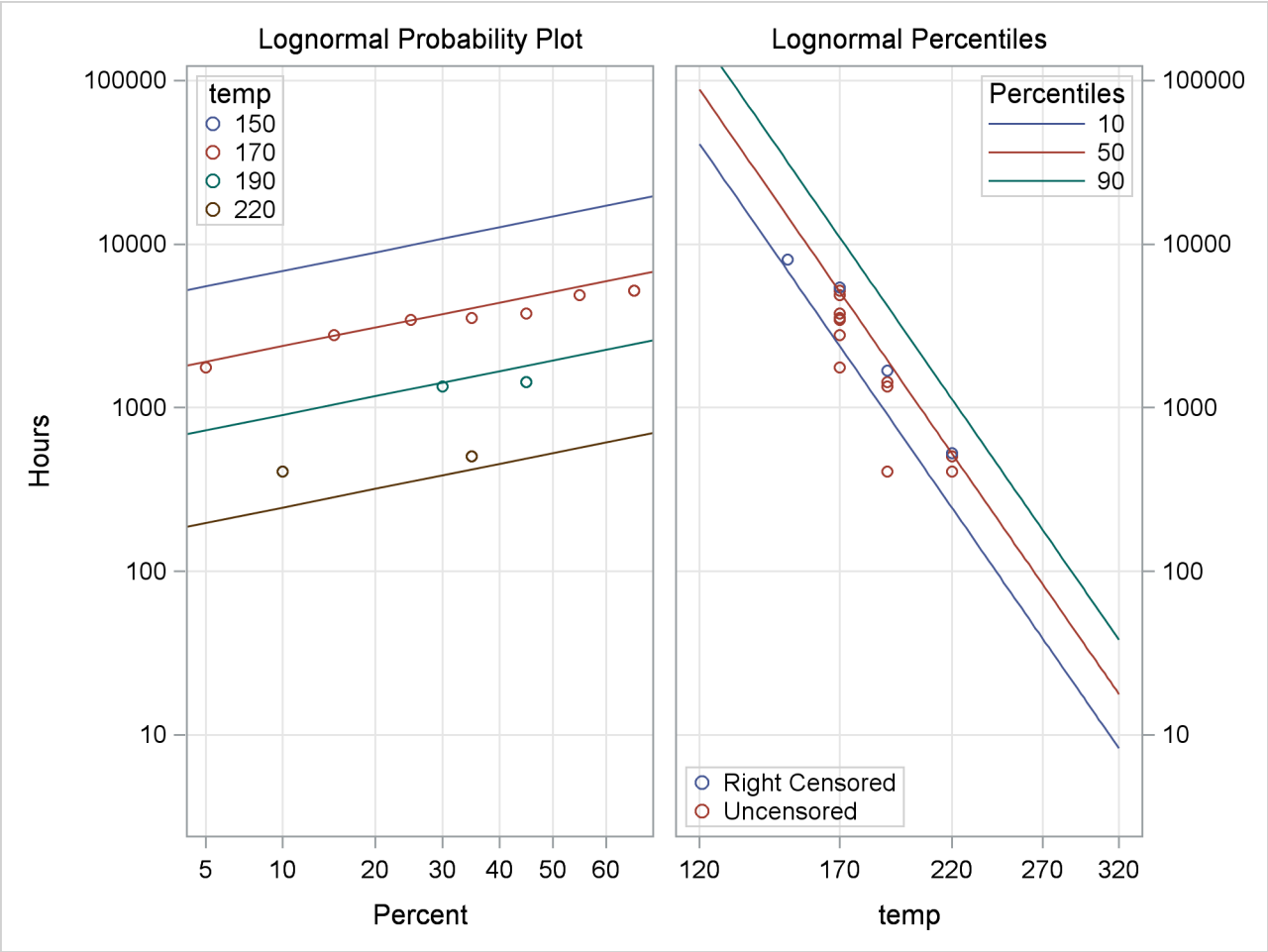
The option OBSTATS requests statistics computed for each observation in the input data set. The options in parentheses following OBSTATS indicate which statistics are to be computed. In this case, QUANTILE=.1 .5 .9 specifies that quantiles of the fitted distribution are to be computed for the value of the variable Temp at each observation. The CONTROL= option requests quantiles only for those observations in which the variable Cntrl has a value of 1. This eliminates unnecessary quantiles in the OBSTATS table since, in this case, only the quantiles at the design temperature of 130°C are of interest.

The RPLOT, or RELATIONPLOT, statement displays a plot of the lifetime data and the fitted model. The dependent variable Time, the independent variable Temp, and the censoring indicator Censor are the same as in the MODEL statement. The option FIT=MODEL specifies that the model fitted with the preceding MODEL statement is to be used for probability plotting and in the relation plot. The option RELATION=ARR specifies an Arrhenius scale for the horizontal axis of the relation plot. The PPLOT option specifies that a probability plot is to be displayed alongside the relation plot. The type of probability plot is determined by the distribution named in the DISTRIBUTION statement, in this case, a lognormal (base 10) distribution. Weibull, extreme value, lognormal (base e), normal, log-logistic, and logistic distributions are also available. The NOCONF option suppresses the default percentile confidence bands on the probability plot. The PLOTDATA option specifies that the failure times are to be plotted on the relation plot. The PLOTFIT option specifies that the 10th, 50th, and 90th percentiles of the fitted relationship are to be plotted on the relation plot. The options LUPPER and SLOWER specify an upper limit on the life axis scale and a lower limit on the stress (temperature) axis scale in the plots.

The plots produced by the preceding statements are shown in Figure 18.9. The plot on the left is an overlaid lognormal probability plot of the data and the fitted model. The plot on the right is a relation plot showing the data and the fitted relation. The fitted straight lines are percentiles of the fitted distribution at each temperature. An Arrhenius relation fitted to the data, plotted on an Arrhenius plot, yields straight percentile lines.

Since all the data at 150°C are right censored, there are no failures corresponding to 150°C on the probability plot. However, the fitted distribution at 150°C is plotted on the probability plot.

Figure 18.9 Probability and Relation Plots for the Class B Insulation Data



The tabular output requested with the MODEL statement is shown in Figure 18.10. The "Model Information" table provides general information about the data and model. The "Summary of Fit" table shows the number of observations used, the number of failures and of censored values (accounting for the frequency count), and the maximum log likelihood for the fitted model.

The "Lognormal Parameter Estimates" table contains the Arrhenius-lognormal model parameter estimates, their standard errors, and confidence interval estimates. In this table, INTERCEPT is the maximum likelihood estimate of β_0, TEMP is the estimate of β_1, and Scale is the estimate of the lognormal scale parameter, σ.

Figure 18.10 MODEL Statement Output for the Class B Data

The RELIABILITY Procedure

Model Information	
Input Data Set	WORK.CLASSB
Analysis Variable	hours Hours
Relation	Arrhenius(temp)
Censor Variable	censor
Frequency Variable	count
Distribution	Lognormal (Base 10)

Figure 18.10 *continued*

Algorithm converged.

Summary of Fit	
Observations Used	16
Uncensored Values	17
Right Censored Values	23
Missing Observations	1
Maximum Loglikelihood	-12.96533

Lognormal Parameter Estimates

Parameter	Estimate	Standard Error	Asymptotic Normal 95% Confidence Limits Lower	Upper
Intercept	-6.0182	0.9467	-7.8737	-4.1628
temp	4.3103	0.4366	3.4546	5.1660
Scale	0.2592	0.0473	0.1812	0.3708

Observation Statistics

Hours	censor	temp	count	Prob	Pcntl	Stderr	Lower	Upper
.	.	130	.	0.1000	21937.658	6959.151	11780.636	40851.857
.	.	130	.	0.5000	47135.132	16125.548	24106.685	92162.016
.	.	130	.	0.9000	101274.29	42061.1	44872.401	228569.92

The "Observation Statistics" table provides the estimates of the fitted distribution quantiles, their standard errors, and the confidence limits. These are given only for the value of 130°C, as specified with the CONTROL= option in the MODEL statement. The predicted median life at 130°C corresponds to a quantile of 0.5, and it is approximately 47,135 hours.

In addition to the MODEL statement output in Figure 18.10, the RELIABILITY procedure produces tabular output for each temperature that is identical to the output produced with the PROBPLOT statement. This output is not shown.

Weibull Analysis of Interval Data with Common Inspection Schedule

Table 18.1 shows data for 167 identical turbine parts provided by Nelson (1982, p. 415). The parts were inspected at certain times to determine which parts had cracked since the last inspection. The times at which parts develop cracks are to be fitted with a Weibull distribution.

Table 18.1 Turbine Part Cracking Data

Inspection (Months)		Number	
Start	**End**	**Cracked**	**Cumulative**
0	6.12	5	5
6.12	19.92	16	21
19.92	29.64	12	33
29.64	35.40	18	51
35.40	39.72	18	69
39.72	45.24	2	71
45.24	52.32	6	77
52.32	63.48	17	94
63.48	Survived	73	167

Table 18.1 shows the time in months of each inspection period and the number of cracked parts found in each period. These data are said to be interval censored since only the time interval in which failures occurred is known, not the exact failure times. Seventy-three parts had not cracked at the last inspection, which took place at 63.48 months. These 73 lifetimes are right censored, since the lifetimes are known only to be greater than 63.48 months.

The interval data in this example are read from a SAS data set with a special structure. All units must have a common inspection schedule. This type of interval data is called *readout data*. The following SAS program creates the SAS data set named CRACKS, shown in Figure 18.11, and provides the data in Table 18.1 with this structure:

```
data cracks;
   input Time units fail;
   datalines;
6.12  167 5
19.92 162 16
29.64 146 12
35.4  134 18
39.72 116 18
45.24 98  2
52.32 96  6
63.48 90 17
;
```

The variable Time is the inspection time—that is, the upper endpoint of each interval. The variable Units is the number of unfailed units at the beginning of each interval, and the variable Fail is the number of units with cracks at the inspection time.

Figure 18.11 Listing of the Turbine Part Cracking Data

Obs	Time	units	fail
1	6.12	167	5
2	19.92	162	16
3	29.64	146	12
4	35.40	134	18
5	39.72	116	18
6	45.24	98	2
7	52.32	96	6
8	63.48	90	17

The following statements use the RELIABILITY procedure to produce the probability plot in Figure 18.12 for the data in the data set CRACKS:

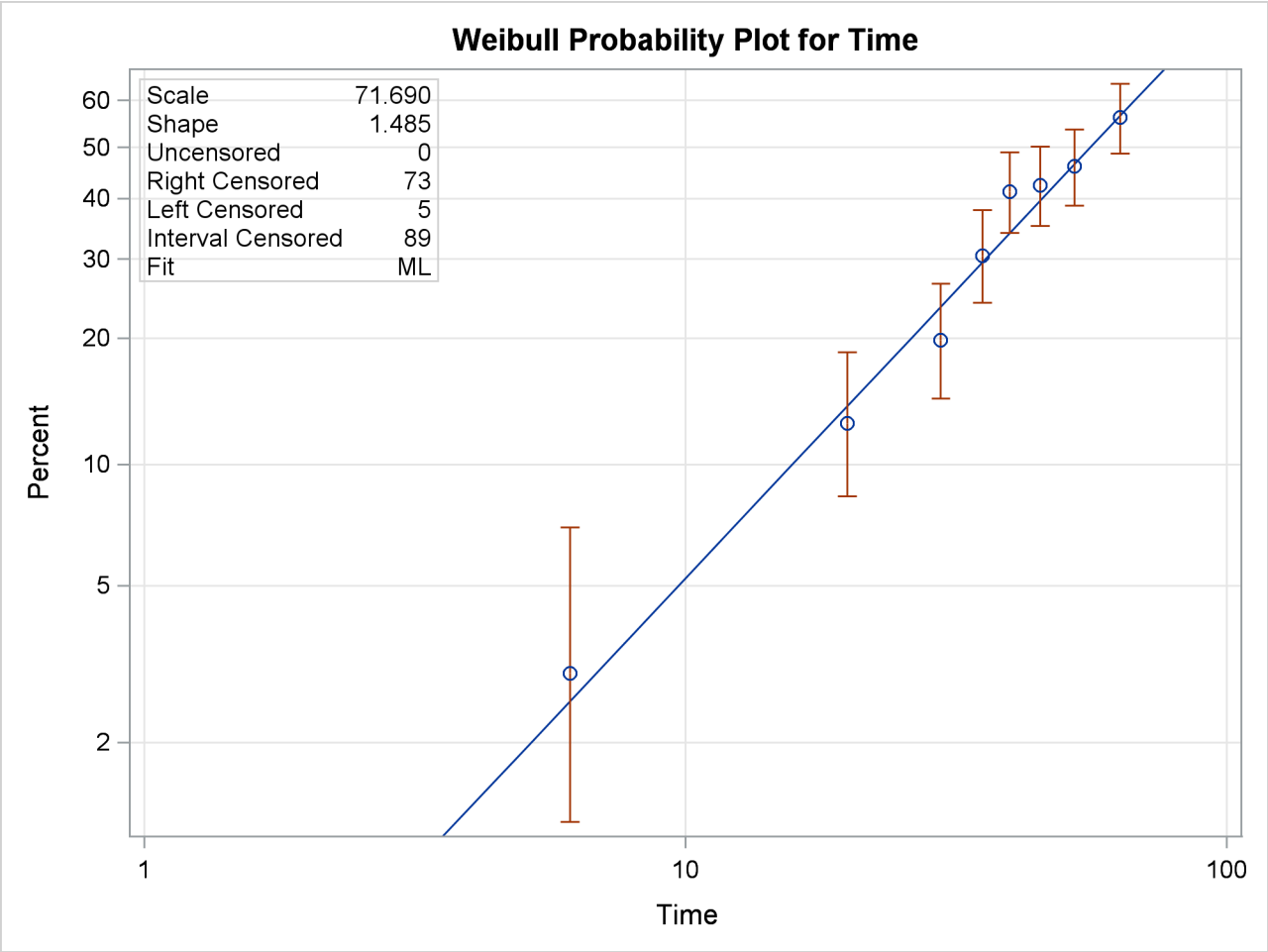
```
proc reliability data=cracks;
   freq fail;
   nenter units;
   distribution Weibull;
   probplot time / readout
                   ppout
                   pconfplt
                   noconf;
run;
```

The FREQ statement specifies that the variable Fail provides the number of failures in each interval. The NENTER statement specifies that the variable Units provides the number of unfailed units at the beginning of each interval. The DISTRIBUTION statement specifies that the Weibull distribution be used for parameter estimation and probability plotting. The PROBPLOT statement requests a probability plot of the data.

The PROBPLOT statement option READOUT indicates that the data in the CRACKS data set are readout (or interval) data. The option PCONFPLT specifies that confidence intervals for the cumulative probability of failure be plotted. The confidence intervals for the cumulative probability are based on the binomial distribution for time intervals until right censoring occurs. For time intervals after right censoring occurs, the binomial distribution is not valid, and a normal approximation is used to compute confidence intervals.

The option NOCONF suppresses the display of confidence intervals for distribution percentiles in the probability plot.

Figure 18.12 Weibull Probability Plot for the Part Cracking Data



A listing of the tabular output produced by the preceding SAS statements is shown in Figure 18.13 and Figure 18.14. By default, the specified Weibull distribution is fitted by maximum likelihood. The line plotted on the probability plot and the tabular output summarize this fit. For interval data, the estimated cumulative probabilities and associated confidence intervals are tabulated. In addition, general fit information, parameter estimates, percentile estimates, standard errors, and confidence intervals are tabulated.

Figure 18.13 Partial Listing of the Tabular Output for the Part Cracking Data

The RELIABILITY Procedure

Model Information	
Input Data Set	WORK.CRACKS
Analysis Variable	Time
Frequency Variable	fail
NENTER Variable	units
Distribution	Weibull
Estimation Method	Maximum Likelihood
Confidence Coefficient	95%
Observations Used	8

Figure 18.13 *continued*

Cumulative Probability Estimates					
			Pointwise 95% Confidence Limits		
Lower Lifetime	Upper Lifetime	Cumulative Probability	Lower	Upper	Standard Error
.	6.12	0.0299	0.0125	0.0699	0.0132
6.12	19.92	0.1257	0.0834	0.1852	0.0257
19.92	29.64	0.1976	0.1440	0.2649	0.0308
29.64	35.4	0.3054	0.2403	0.3793	0.0356
35.4	39.72	0.4132	0.3410	0.4893	0.0381
39.72	45.24	0.4251	0.3524	0.5013	0.0383
45.24	52.32	0.4611	0.3869	0.5370	0.0386
52.32	63.48	0.5629	0.4868	0.6361	0.0384

Algorithm converged.

Summary of Fit	
Observations Used	8
Right Censored Values	73
Left Censored Values	5
Interval Censored Values	89
Maximum Loglikelihood	-309.6684

Figure 18.14 Partial Listing of the Tabular Output for the Part Cracking Data

Weibull Parameter Estimates				
			Asymptotic Normal 95% Confidence Limits	
Parameter	Estimate	Standard Error	Lower	Upper
EV Location	4.2724	0.0744	4.1265	4.4182
EV Scale	0.6732	0.0664	0.5549	0.8168
Weibull Scale	71.6904	5.3335	61.9634	82.9444
Weibull Shape	1.4854	0.1465	1.2242	1.8022

Other Weibull Distribution Parameters	
Parameter	Value
Mean	64.7966
Mode	33.7622
Median	56.0144
Standard Deviation	44.3943

Figure 18.14 *continued*

Weibull Percentile Estimates				
			Asymptotic Normal 95% Confidence Limits	
Percent	Estimate	Standard Error	Lower	Upper
0.1	0.68534385	0.29999861	0.29060848	1.61625083
0.2	1.09324674	0.42889777	0.50673224	2.3586193
0.5	2.02798319	0.67429625	1.05692279	3.8912169
1	3.23938972	0.93123832	1.84401909	5.69063837
2	5.18330703	1.2581604	3.22101028	8.34106988
5	9.70579945	1.78869256	6.76335893	13.9283666
10	15.7577991	2.22445157	11.9491109	20.7804776
20	26.1159906	2.6327383	21.4337103	31.821134
30	35.8126238	2.90557264	30.547517	41.9852137
40	45.6100472	3.27409792	39.6239146	52.5005271
50	56.0143651	3.89410377	48.8792027	64.1910859
60	67.5928125	4.90210777	58.6364803	77.917165
70	81.2334227	6.46932648	69.4938134	94.9562075
80	98.7644937	8.95137184	82.6900902	117.963654
90	125.694556	13.5078386	101.821995	155.164133
95	150.057755	18.2060035	118.300075	190.340791
99	200.437864	29.1957544	150.658574	266.66479
99.9	263.348102	44.7205513	188.791789	367.347666

In this example, the number of unfailed units at the beginning of an interval minus the number failing in the interval is equal to the number of unfailed units entering the next interval. This is not always the case since some unfailed units might be removed from the test at the end of an interval, for reasons unrelated to failure; that is, they might be right censored. The special structure of the input SAS data set required for interval data enables the RELIABILITY procedure to analyze this more general case.

Lognormal Analysis with Arbitrary Censoring

This example illustrates analyzing data that have more general censoring than in the previous example. The data can be a combination of exact failure times, left censored, right censored, and interval censored data. The intervals can be overlapping, unlike in the previous example, where the interval endpoints had to be the same for all units.

Table 18.2 shows data from Nelson (1982, p. 409), analyzed by Meeker and Escobar (1998, p. 135). Each of 435 turbine wheels was inspected once to determine whether a crack had developed in the wheel or not. The inspection time (in 100s of hours), the number inspected at the time that had cracked, and the number not cracked are shown in the table. The quantity of interest is the time for a crack to develop.

Table 18.2 Turbine Wheel Cracking Data

Inspection Time (100 hours)	Number Cracked	Number Not Cracked
4	0	39
10	4	49
14	2	31
18	7	66
22	5	25
26	9	30
30	9	33
34	6	7
38	22	12
42	21	19
46	21	15

These data consist only of left and right censored lifetimes. If a unit exhibits a crack at an inspection time, the unit is left censored at the time; if a unit has not developed a crack, it is right censored at the time. For example, there are 4 left-censored lifetimes and 49 right-censored lifetimes at 1000 hours.

The following statements create a SAS data set named TURBINE that contains the data in the format necessary for analysis by the RELIABILITY procedure:

```
data turbine;
   label t1 = 'Time of Cracking (Hours x 100 )';
   input t1 t2 f;
   datalines;
.   4  0
4   .  39
.  10  4
10  .  49
.  14  2
14  .  31
.  18  7
18  .  66
.  22  5
22  .  25
.  26  9
26  .  30
.  30  9
30  .  33
.  34  6
34  .  7
.  38  22
38  .  12
.  42  21
42  .  19
.  46  21
46  .  15
;
```

The variables T1 and T2 represent the inspection times and determine whether the observation is right or left censored. If T1 is missing (.), then T2 represents a left-censoring time; if T2 is missing, T1 represents a right-censoring time. The variable F is the number of units that were found to be cracked for left-censored observations, or not cracked for right-censored observations at an inspection time.

The following statements use the RELIABILITY procedure to produce the probability plot in Figure 18.15 for the data in the data set TURBINE:

```
proc reliability data = turbine;
   distribution lognormal;
   freq f;
   pplot ( t1 t2 ) / maxitem = 5000
                     ppout;
run;
```

The DISTRIBUTION statement specifies that a lognormal probability plot be created. The FREQ statement identifies the frequency variable F. The option MAXITEM=5000 specifies that the iterative algorithm that computes the points on the probability plot takes a maximum of 5000 iterations. The algorithm does not converge for these data in the default 1000 iterations, so the maximum number of iterations needs to be increased for convergence. The option PPOUT specifies that a table of the cumulative probabilities plotted on the probability plot be printed, along with standard errors and confidence limits.

The tabular output for the maximum likelihood lognormal fit for these data is shown in Figure 18.16. Figure 18.15 shows the resulting lognormal probability plot with the computed cumulative probability estimates and the lognormal fit line.

Figure 18.15 Lognormal Probability Plot for the Turbine Wheel Data

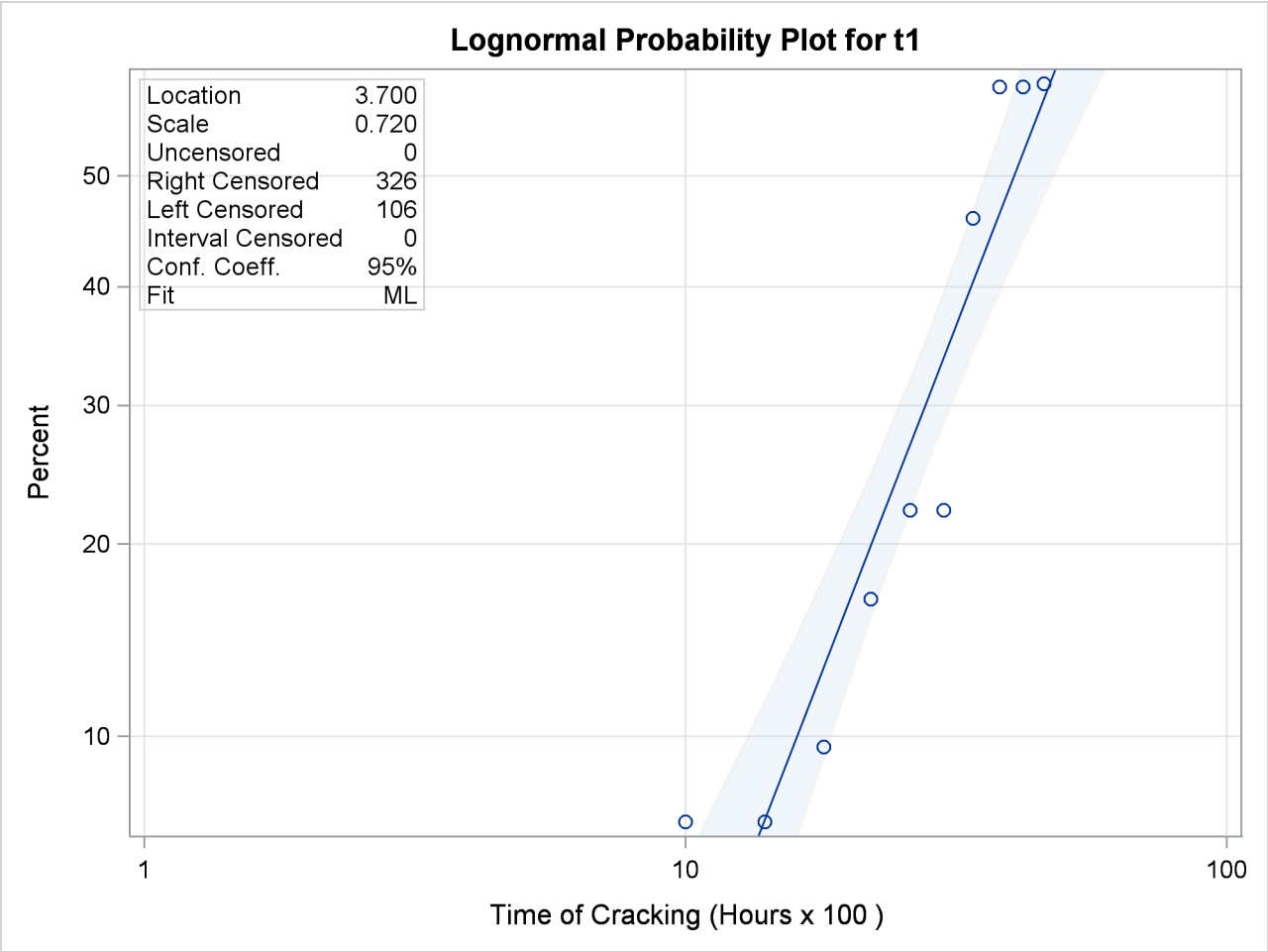


Figure 18.16 Partial Listing of the Tabular Output for the Turbine Wheel Data

The RELIABILITY Procedure

Model Information		
Input Data Set	WORK.TURBINE	
Analysis Variable	t1	Time of Cracking (Hours x 100)
Analysis Variable	t2	
Frequency Variable	f	
Distribution	Lognormal (Base e)	
Estimation Method	Maximum Likelihood	
Confidence Coefficient	95%	
Observations Used	21	

Figure 18.16 *continued*

Cumulative Probability Estimates					
			Pointwise 95% Confidence Limits		
Lower Lifetime	Upper Lifetime	Cumulative Probability	Lower	Upper	Standard Error
.	4	0.0000	0.0000	0.0000	0.0000
10	10	0.0698	0.0264	0.1720	0.0337
14	14	0.0698	0.0177	0.2384	0.0473
18	18	0.0959	0.0464	0.1878	0.0345
22	22	0.1667	0.0711	0.3432	0.0680
26	26	0.2222	0.1195	0.3757	0.0657
30	30	0.2222	0.1203	0.3738	0.0650
34	34	0.4615	0.2236	0.7184	0.1383
38	38	0.5809	0.4085	0.7356	0.0865
42	42	0.5809	0.4280	0.7198	0.0766
46	46	0.5836	0.4195	0.7311	0.0822

Algorithm converged.

Summary of Fit	
Observations Used	21
Uncensored Values	0
Right Censored Values	326
Left Censored Values	106
Maximum Loglikelihood	-190.7315

Lognormal Parameter Estimates				
			Asymptotic Normal 95% Confidence Limits	
Parameter	Estimate	Standard Error	Lower	Upper
Location	3.6999	0.0708	3.5611	3.8387
Scale	0.7199	0.0887	0.5655	0.9165

Other Lognormal Distribution Parameters	
Parameter	Value
Mean	52.4062
Mode	24.0870
Median	40.4436
Standard Deviation	43.1855

Regression Modeling

This example is an illustration of a Weibull regression model that uses a load accelerated life test of rolling bearings, with data provided by Nelson (1990, p. 305). Bearings are tested at four different loads, and lifetimes in 10^6 of revolutions are measured. The data are shown in Table 18.3. An outlier identified by Nelson (1990) is omitted.

Table 18.3 Bearing Lifetime Data

Load	**Life (10^6 Revolutions)**									
0.87	1.67	2.2	2.51	3.00	3.90	4.70	7.53	14.7	27.76	37.4
0.99	0.80	1.0	1.37	2.25	2.95	3.70	6.07	6.65	7.05	7.37
1.09	0.18	0.2	0.24	0.26	0.32	0.32	0.42	0.44	0.88	
1.18	0.073	0.098	0.117	0.135	0.175	0.262	0.270	0.350	0.386	0.456

These data are modeled with a Weibull regression model in which the independent variable is the logarithm of the load. The model is

$$\mu_i = \beta_0 + \beta_1 x_i$$

where μ_i is the location parameter of the extreme value distribution and

$$x_i = \log(\text{load})$$

for the ith bearing. The following statements create a SAS data set containing the loads, log loads, and bearing lifetimes:

```
data bearing;
   input load Life @@;
   lload = log(load);
   datalines;
 .87 1.67    .87 2.2    .87 2.51    .87 3.0    .87 3.9
 .87 4.7     .87 7.53   .87 14.7    .87 27.76  .87 37.4
 .99  .8     .99 1.0    .99 1.37    .99 2.25   .99 2.95
 .99 3.7     .99 6.07   .99 6.65    .99 7.05   .99 7.37
1.09  .18   1.09  .2   1.09  .24   1.09  .26  1.09  .32
1.09  .32   1.09  .42  1.09  .44   1.09  .88  1.18  .073
1.18  .098  1.18  .117 1.18  .135  1.18  .175 1.18  .262
1.18  .270  1.18  .350 1.18  .386  1.18  .456
;
```

Figure 18.17 shows a listing of the bearing data.

Figure 18.17 Listing of the Bearing Data

Obs	load	Life	lload
1	0.87	1.670	-0.13926
2	0.87	2.200	-0.13926
3	0.87	2.510	-0.13926
4	0.87	3.000	-0.13926
5	0.87	3.900	-0.13926
6	0.87	4.700	-0.13926
7	0.87	7.530	-0.13926
8	0.87	14.700	-0.13926
9	0.87	27.760	-0.13926
10	0.87	37.400	-0.13926
11	0.99	0.800	-0.01005
12	0.99	1.000	-0.01005
13	0.99	1.370	-0.01005
14	0.99	2.250	-0.01005
15	0.99	2.950	-0.01005
16	0.99	3.700	-0.01005
17	0.99	6.070	-0.01005
18	0.99	6.650	-0.01005
19	0.99	7.050	-0.01005
20	0.99	7.370	-0.01005
21	1.09	0.180	0.08618
22	1.09	0.200	0.08618
23	1.09	0.240	0.08618
24	1.09	0.260	0.08618
25	1.09	0.320	0.08618
26	1.09	0.320	0.08618
27	1.09	0.420	0.08618
28	1.09	0.440	0.08618
29	1.09	0.880	0.08618
30	1.18	0.073	0.16551
31	1.18	0.098	0.16551
32	1.18	0.117	0.16551
33	1.18	0.135	0.16551
34	1.18	0.175	0.16551
35	1.18	0.262	0.16551
36	1.18	0.270	0.16551
37	1.18	0.350	0.16551
38	1.18	0.386	0.16551
39	1.18	0.456	0.16551

The following statements fit the regression model by maximum likelihood that uses the Weibull distribution:

```
ods output modobstats = Residual;
proc reliability data=bearing;
   distribution Weibull;
   model life = lload /  covb
                         corrb
                         obstats
                         ;
run;
```

The PROC RELIABILITY statement invokes the procedure and identifies BEARING as the input data set. The DISTRIBUTION statement specifies the Weibull distribution for model fitting. The MODEL statement specifies the regression model, identifying Life as the variable that provides the response values (the lifetimes) and Lload as the independent variable (the log loads). The MODEL statement option COVB requests the regression parameter covariance matrix, and the CORRB option requests the correlation matrix. The option OBSTATS requests a table that contains residuals, predicted values, and other statistics. The ODS OUTPUT statement creates a SAS data set named RESIDUAL that contains the table created by the OBSTATS option.

Figure 18.18 shows the tabular output produced by the RELIABILITY procedure. The "Weibull Parameter Estimates" table contains parameter estimates, their standard errors, and 95% confidence intervals. In this table, INTERCEPT corresponds to β_0, LLOAD corresponds to β_1, and SHAPE corresponds to the Weibull shape parameter. Figure 18.19 shows a listing of the output data set RESIDUAL.

Figure 18.18 Analysis Results for the Bearing Data

The RELIABILITY Procedure

Model Information	
Input Data Set	WORK.BEARING
Analysis Variable	Life
Distribution	Weibull

Parameter Information	
Parameter	**Effect**
Prm1	Intercept
Prm2	lload
Prm3	EV Scale

Algorithm converged.

Summary of Fit	
Observations Used	39
Uncensored Values	39
Maximum Loglikelihood	-51.77737

Figure 18.18 *continued*

Weibull Parameter Estimates

Parameter	Estimate	Standard Error	Asymptotic Normal 95% Confidence Limits Lower	Upper
Intercept	0.8323	0.1410	0.5560	1.1086
lload	-13.8529	1.2333	-16.2703	-11.4356
EV Scale	0.8043	0.0999	0.6304	1.0260
Weibull Shape	1.2434	0.1545	0.9746	1.5862

Estimated Covariance Matrix
Weibull Parameters

	Prm1	Prm2	Prm3
Prm1	0.01987	-0.04374	-0.00492
Prm2	-0.04374	1.52113	0.01578
Prm3	-0.00492	0.01578	0.00999

Estimated Correlation Matrix
Weibull Parameters

	Prm1	Prm2	Prm3
Prm1	1.0000	-0.2516	-0.3491
Prm2	-0.2516	1.0000	0.1281
Prm3	-0.3491	0.1281	1.0000

Figure 18.19 Listing of Data Set Residual

Obs	Life	lload	Xbeta	Surv	Resid	SRESID	Aresid
1	1.67	-0.139262	2.7614742	0.9407681	-2.248651	-2.795921	-2.795921
2	2.2	-0.139262	2.7614742	0.9175782	-1.973017	-2.453205	-2.453205
3	2.51	-0.139262	2.7614742	0.9036277	-1.841191	-2.289296	-2.289296
4	3	-0.139262	2.7614742	0.8811799	-1.662862	-2.067565	-2.067565
5	3.9	-0.139262	2.7614742	0.8392186	-1.400498	-1.741347	-1.741347
6	4.7	-0.139262	2.7614742	0.8016738	-1.213912	-1.50935	-1.50935
7	7.53	-0.139262	2.7614742	0.6721971	-0.742579	-0.923306	-0.923306
8	14.7	-0.139262	2.7614742	0.4015113	-0.073627	-0.091546	-0.091546
9	27.76	-0.139262	2.7614742	0.1337746	0.562122	0.6989298	0.6989298
10	37.4	-0.139262	2.7614742	0.0542547	0.8601965	1.069549	1.069549
11	0.8	-0.01005	0.971511	0.7973909	-1.194655	-1.485407	-1.485407
12	1	-0.01005	0.971511	0.741702	-0.971511	-1.207955	-1.207955
13	1.37	-0.01005	0.971511	0.6427726	-0.6567	-0.816526	-0.816526
14	2.25	-0.01005	0.971511	0.4408692	-0.160581	-0.199663	-0.199663
15	2.95	-0.01005	0.971511	0.3175927	0.1102941	0.1371372	0.1371372
16	3.7	-0.01005	0.971511	0.2186832	0.3368218	0.4187966	0.4187966
17	6.07	-0.01005	0.971511	0.0600164	0.8318476	1.0343005	1.0343005
18	6.65	-0.01005	0.971511	0.0428027	0.9231058	1.147769	1.147769
19	7.05	-0.01005	0.971511	0.0337583	0.9815166	1.2203956	1.2203956
20	7.37	-0.01005	0.971511	0.0278531	1.0259067	1.2755892	1.2755892
21	0.18	0.0861777	-0.361531	0.8303684	-1.353268	-1.682623	-1.682623
22	0.2	0.0861777	-0.361531	0.809042	-1.247907	-1.55162	-1.55162
23	0.24	0.0861777	-0.361531	0.7665749	-1.065586	-1.324925	-1.324925
24	0.26	0.0861777	-0.361531	0.7455451	-0.985543	-1.225402	-1.225402
25	0.32	0.0861777	-0.361531	0.6837688	-0.777904	-0.967228	-0.967228
26	0.32	0.0861777	-0.361531	0.6837688	-0.777904	-0.967228	-0.967228
27	0.42	0.0861777	-0.361531	0.5868036	-0.50597	-0.629112	-0.629112
28	0.44	0.0861777	-0.361531	0.5684693	-0.45945	-0.57127	-0.57127
29	0.88	0.0861777	-0.361531	0.2625812	0.2336973	0.290574	0.290574
30	0.073	0.1655144	-1.460578	0.7887184	-1.156718	-1.438237	-1.438237
31	0.098	0.1655144	-1.460578	0.7101313	-0.86221	-1.072052	-1.072052
32	0.117	0.1655144	-1.460578	0.6526714	-0.685003	-0.851717	-0.851717
33	0.135	0.1655144	-1.460578	0.6006317	-0.541902	-0.673789	-0.673789
34	0.175	0.1655144	-1.460578	0.4946523	-0.282391	-0.351119	-0.351119
35	0.262	0.1655144	-1.460578	0.3126729	0.1211675	0.1506569	0.1506569
36	0.27	0.1655144	-1.460578	0.2991233	0.1512449	0.1880546	0.1880546
37	0.35	0.1655144	-1.460578	0.1889073	0.4107561	0.5107249	0.5107249
38	0.386	0.1655144	-1.460578	0.1522503	0.5086604	0.6324568	0.6324568
39	0.456	0.1655144	-1.460578	0.0987061	0.6753158	0.8396724	0.8396724

The value of the lifetime Life and the log load Lload are included in this data set, as well as statistics computed from the fitted model. The variable Xbeta is the value of the linear predictor

$$\mathbf{x}'\hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \mathsf{Lload}\hat{\beta}_1$$

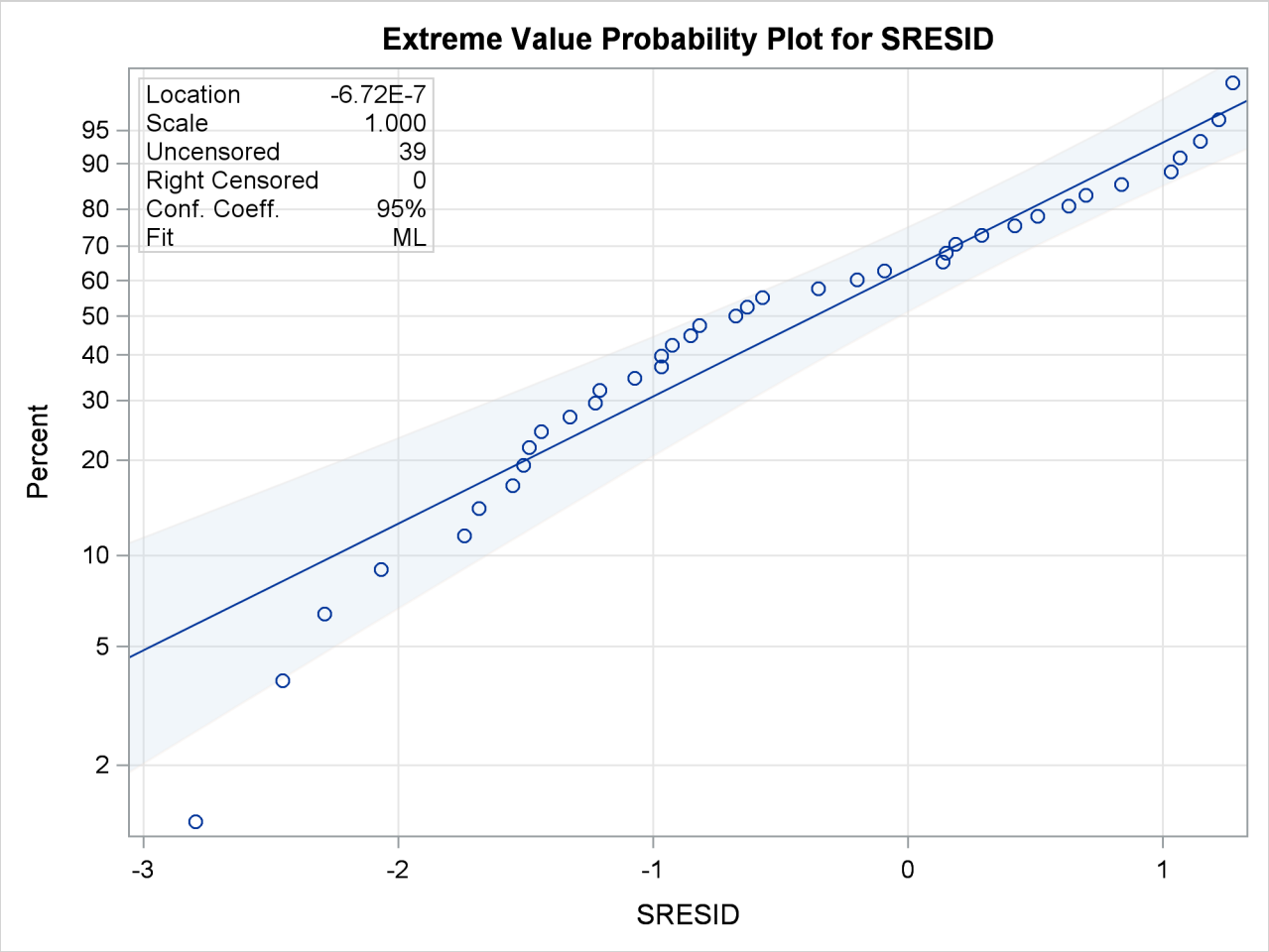
for each observation. The variable Surv contains the value of the reliability function, the variable Sresid contains the standardized residual, and the variable Aresid contains a residual adjusted for right-censored observations. Since there are no censored values in these data, Sresid is equal to Aresid for all the bearings. See Table 18.32 and Table 18.33 for other statistics that are available in the OBSTATS table and data set. See the section "Regression Model Statistics Computed for Each Observation for Lifetime Data" on page 1340 for a description of the residuals and other statistics.

If the fitted regression model is adequate, the standardized residuals have a standard extreme value distribution. You can check the residuals by using the RELIABILITY procedure and the RESIDUAL data set to create an extreme value probability plot of the residuals.

The following statements create the plot in Figure 18.20:

```
proc reliability data=residual;
   distribution ev;
   probplot sresid;
run;
```

Figure 18.20 Extreme Value Probability Plot for the Standardized Residuals



Although the estimated location is near zero and the estimated scale is near one, the plot reveals systematic curvature, indicating that the Weibull regression model might be inadequate.

Regression Model with Nonconstant Scale

Nelson (1990, p. 272) and Meeker and Escobar (1998, p. 439) analyzed data from a strain-controlled fatigue test on 26 specimens of a type of superalloy. The following SAS statements create a SAS data set containing for each specimen the level of pseudo-stress (Pstress), the number of cycles (in thousands) (Kcycles) until failure or removal from the test, and a variable to indicate whether a specimen failed (F) or was right censored (C) (Status):

```
data alloy;
   input pstress kCycles status$ @@;
   cen = ( status = 'C' );
   datalines;
80.3  211.629  F    99.8   43.331  F
80.6  200.027  F   100.1   12.076  F
80.8   57.923  C   100.5   13.181  F
84.3  155.000  F   113.0   18.067  F
85.2   13.949  F   114.8   21.300  F
85.6  112.968  C   116.4   15.616  F
85.8  152.680  F   118.0   13.030  F
86.4  156.725  F   118.4    8.489  F
86.7  138.114  C   118.6   12.434  F
87.2   56.723  F   120.4    9.750  F
87.3  121.075  F   142.5   11.865  F
89.7  122.372  C   144.5    6.705  F
91.3  112.002  F   145.9    5.733  F
;
```

The following statements fit a Weibull regression model with the number of cycles to failure as the response variable:

```
ods output ModObstats = Resids;
proc reliability data = alloy;
   distribution Weibull;
   model kcycles*cen(1) = pstress pstress*pstress / Relation = Pow Obstats;
   logscale pstress;
   rplot kcycles*cen(1) = pstress /  fit=regression
                                     relation = pow
                                     plotfit  10 50 90
                                     slower=60 supper=160
                                     lupper=500;

   label pstress = "Pseudo-Stress";
   label kcycles = "Thousands of Cycles";
run;
```

The data set RESIDS contains standardized residuals created with the ODS OUTPUT statement. The MODEL statement specifies a model quadratic in the log of pseudo-stress for the extreme value location parameter. The quadratic model in pseudo-stress PSTRESS is specified in the MODEL statement, and the RELATION=POW option specifies that the log transformation be applied to Pstress in the MODEL statement and the LOGSCALE statement. The LOGSCALE statement specifies the log of the scale parameter as a linear function of the log of Pstress. The RPLOT statement specifies a plot of the data and the fitted regression model versus the variable Pstress. The FIT=REGRESSION option specifies plotting the regression model

fitted with the preceding MODEL statement. The RELATION=POW option specifies a log stress axis. The PLOTFIT option specifies plotting the 10th, 50th, and 90th percentiles of the regression model at each stress level. The SLOWER, SUPPER, and LUPPER options control limits on the stress and lifetime axes.

Figure 18.21 displays the parameter estimates from the fitted regression model. Parameter estimates for both the model for the location parameter and the scale parameter models are shown. Standard errors and confidence limits for all parameter estimates are included.

Figure 18.21 Parameter Estimates for Fitted Regression Model

The RELIABILITY Procedure

Weibull Parameter Estimates

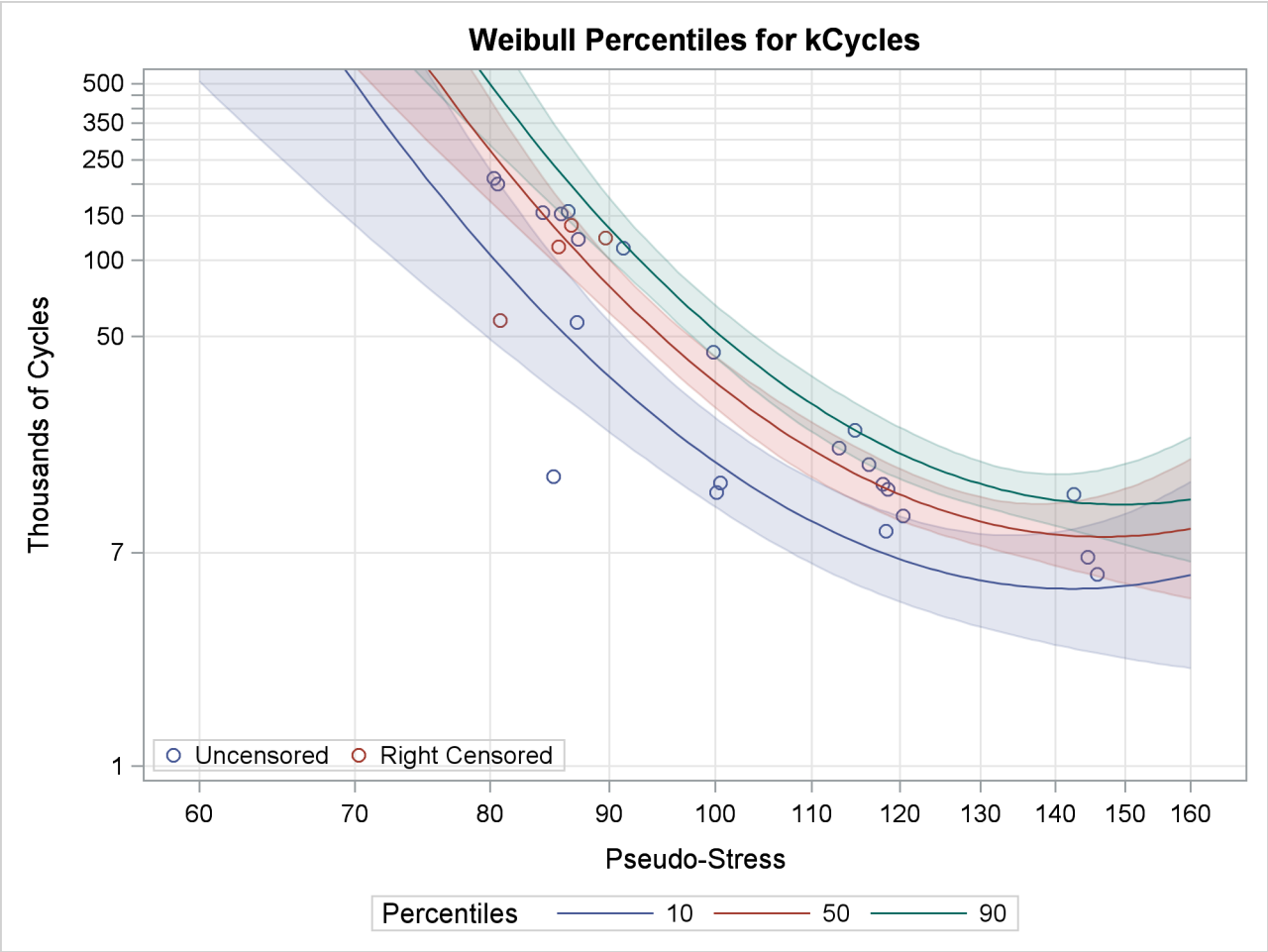
			Asymptotic Normal 95% Confidence Limits	
Parameter	**Estimate**	**Standard Error**	**Lower**	**Upper**
Intercept	243.1680	58.1777	129.1418	357.1943
pstress	-96.5240	24.7558	-145.0445	-48.0035
pstress*pstress	9.6653	2.6299	4.5107	14.8198

Log-Scale Parameter Estimates

			Asymptotic Normal 95% Confidence Limits	
Parameter	**Estimate**	**Standard Error**	**Lower**	**Upper**
Intercept	4.4666	4.1745	-3.7152	12.6484
pstress	-1.1757	0.8950	-2.9299	0.5784

Figure 18.22 displays the plot of the data and fitted regression model.

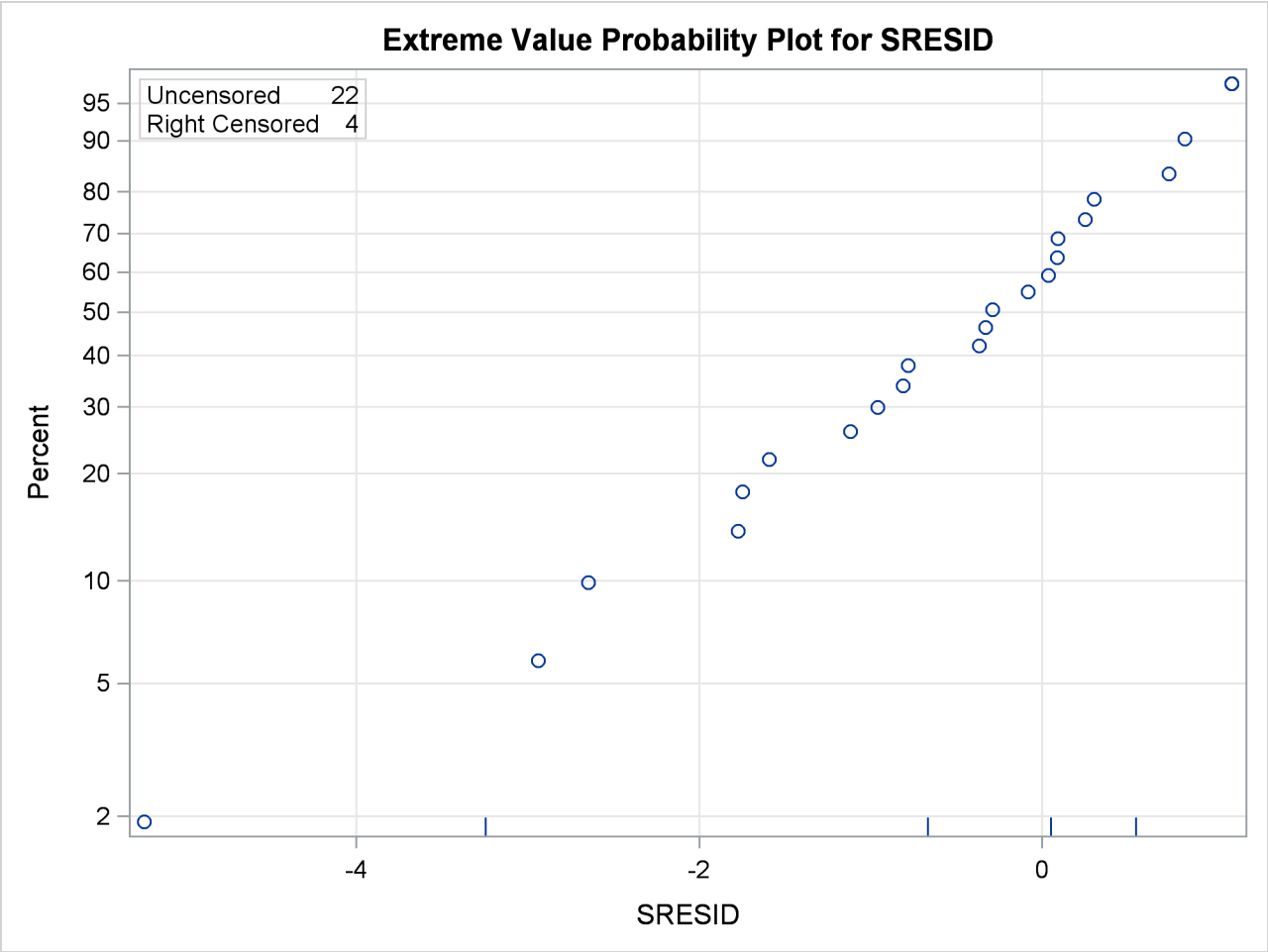
Figure 18.22 Superalloy Fatigue Data with Fitted Regression Model



The following SAS statements create an extreme values probability plot of standardized residuals from the regression model shown in Figure 18.23:

```
proc reliability data = Resids;
   distribution ev;
   pplot sresid*cen(1) / nofit;
run;
```

Figure 18.23 Residuals for Superalloy Fatigue Data Regression Model



Regression Model with Two Independent Variables

Meeker and Escobar (1998, p. 447) analyzed data from an accelerated test on the lifetimes of glass capacitors as a function of operating voltage and temperature. The following SAS statements create a SAS data set containing the data. There are four lifetimes for each of eight combinations and four censored observations after the fourth failure for each combination:

```
data glass;
   input Temp Voltage @;
   do i = 1 to 4;
      cen = 0;
      input Hours @; output;
   end;
   do i = 1 to 4;
      cen = 1;
      output;
   end;
   datalines;
```

```
170 200  439  904  1092  1105
170 250  572  690   904  1090
170 300  315  315   439   628
170 350  258  258   347   588
180 200  959 1065  1065  1087
180 250  216  315   455   473
180 300  241  315   332   380
180 350  241  241   435   455
;
```

The following statements analyze the capacitor data. The MODEL statement fits a regression model with Temp and Voltage as independent variables. Parameter estimates from the fitted regression model are shown in Figure 18.24. An interaction term between Temp and Voltage is included. The PPLOT statement creates a Weibull probability plot shown in Figure 18.25 with all temperature-voltage combinations overlaid on the same plot. The regression model fit is also plotted. The RPLOT statement creates the plot shown in Figure 18.26 of the data and Weibull distribution percentiles from the regression model as a function of voltage for values of temperature of 150, 170, and 180:

```
proc reliability data = glass;
   distribution Weibull;
   model Hours*cen(1) = temp voltage temp * voltage;
   pplot Hours*cen(1) = ( temp voltage ) / fit = model
                                           overlay
                                           noconf
                                           lupper = 2000;
run;

proc reliability data = glass;
   distribution Weibull;
   model Hours*cen(1) = temp voltage temp * voltage;
   rplot Hours*cen(1) = voltage / fit = regression(temp = 150, 170, 180)
                                  plotfit;
run;
```

Figure 18.24 Parameter Estimates for Fitted Regression Model

The RELIABILITY Procedure

Weibull Parameter Estimates

Parameter	Estimate	Standard Error	Asymptotic Normal 95% Confidence Limits Lower	Upper
Intercept	9.4135	10.5402	-11.2449	30.0719
Temp	-0.0062	0.0598	-0.1235	0.1110
Voltage	0.0086	0.0374	-0.0648	0.0820
Temp*Voltage	-0.0001	0.0002	-0.0005	0.0003
EV Scale	0.3624	0.0553	0.2687	0.4887
Weibull Shape	2.7593	0.4210	2.0461	3.7209

Figure 18.25 Probability Plot for Glass Capacitor Regression Model

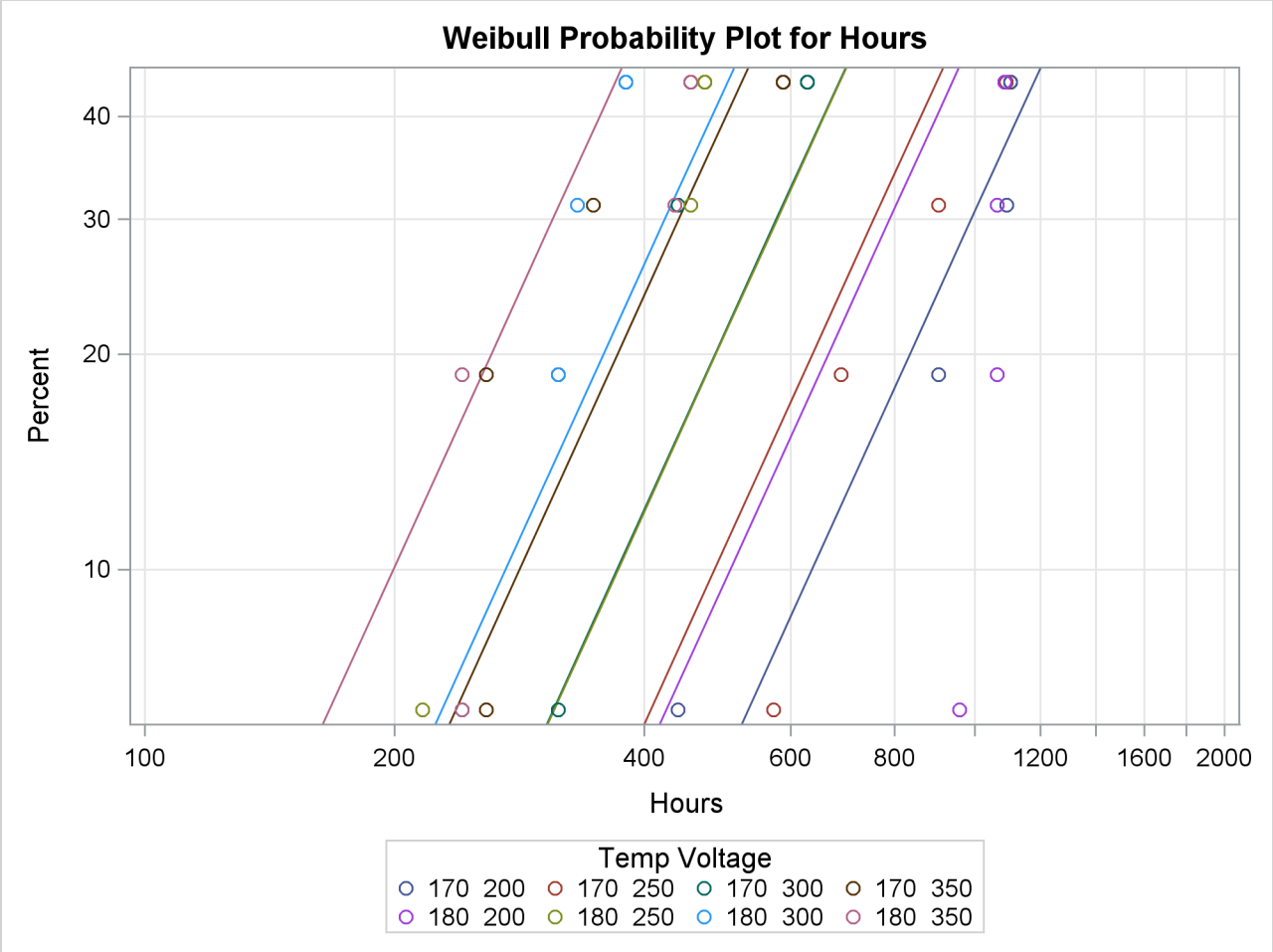
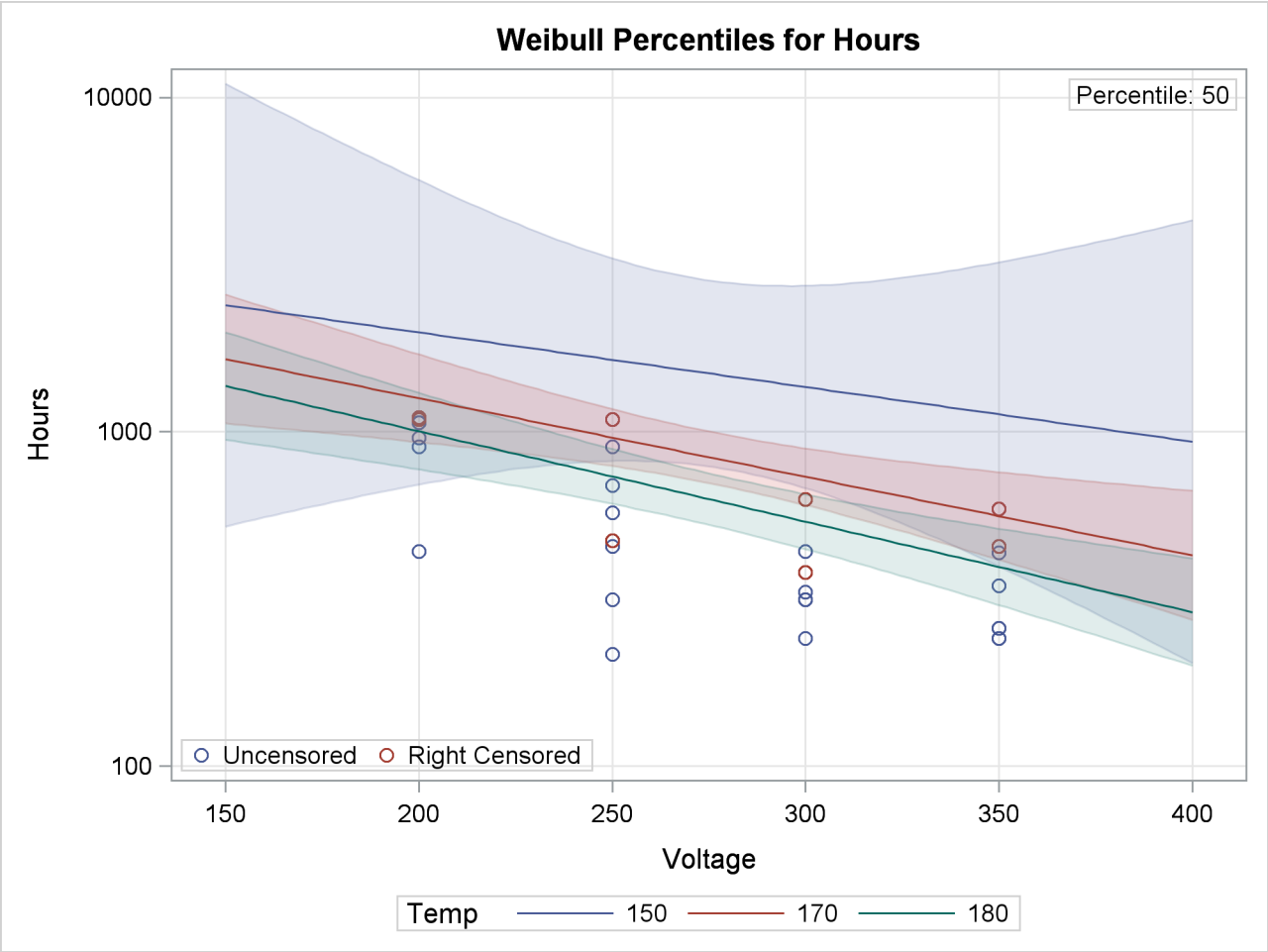


Figure 18.26 Plot of Data and Fitted Weibull Percentiles for Glass Capacitor Regression Model



Weibull Probability Plot for Two Combined Failure Modes

Doganaksoy, Hahn, and Meeker (2002) analyzed failure data for the dielectric insulation of generator armature bars. A sample of 58 segments of bars were subjected to a high voltage stress test. Based on examination of the sample after the test, failures were attributed to one of two modes:

- Mode D (degradation failure): degradation of the organic material. Such failures usually occur later in life.
- Mode E (early failure): insulation defects due to a processing problem. These failures tend to occur early in life.

The following SAS statements create a SAS data set that contains the failure data:

```
data Voltage;
   input Hours Mode$ @@;
   if Mode = 'Cen' then Status = 1;
   else Status = 0;
   datalines;
2   E   3   E   5   E   8   E   13  Cen 21  E
28  E   31  E   31  Cen 52  Cen 53  Cen 64  E
67  Cen 69  E   76  E   78  Cen 104 E   113 Cen
119 E   135 Cen 144 E   157 Cen 160 E   168 D
179 Cen 191 D   203 D   211 D   221 E   226 D
236 E   241 Cen 257 Cen 261 D   264 D   278 D
282 E   284 D   286 D   298 D   303 E   314 D
317 D   318 D   320 D   327 D   328 D   328 D
348 D   348 Cen 350 D   360 D   369 D   377 D
387 D   392 D   412 D   446 D
;
```

The variable Hours represents the number of hours until a failure, or the number of hours on test if the sample unit did not fail. The variable Mode represents the failure mode: D for degradation failure, E for early failures, or Cen if the unit did not fail (i.e., is right-censored). The computed variable Status is a numeric indicator for censored observations.

The following statements fit a Weibull distribution to the individual failure modes (D and E), and compute the failure distribution with both modes acting:

```
proc reliability data=Voltage;
   distribution Weibull;
   pplot Hours*Status(1) / pref(intersect) = 10
                           preflabel = ('10th Percentile')
                           survtime = 100 200 300 400 500 1000
                           lupper = 500;
   fmode combine = Mode( 'D' 'E' );
run;
```

Figure 18.27 contains estimates of the combined failure mode survival function at the times specified with the SURVTIME= option in the PPLOT statement.

Figure 18.27 Survival Function Estimates for Combined Failure Modes

The RELIABILITY Procedure

Combined Failure Modes Weibull Distribution Function Estimates With 95% Asymptotic Normal Confidence Limits						
X	**Pr(<X)**	**Lower**	**Upper**	**Pr(>X)**	**Lower**	**Upper**
100.00	0.1898	0.1172	0.2926	0.8102	0.7074	0.8828
200.00	0.3115	0.2139	0.4292	0.6885	0.5708	0.7861
300.00	0.5866	0.4621	0.7010	0.4134	0.2990	0.5379
400.00	0.9405	0.8476	0.9782	0.0595	0.0218	0.1524
500.00	0.9998	0.9711	1.0000	0.0002	0.0000	0.0289
1000.00	1.0000	0.0000	1.0000	0.0000	0.0000	1.0000

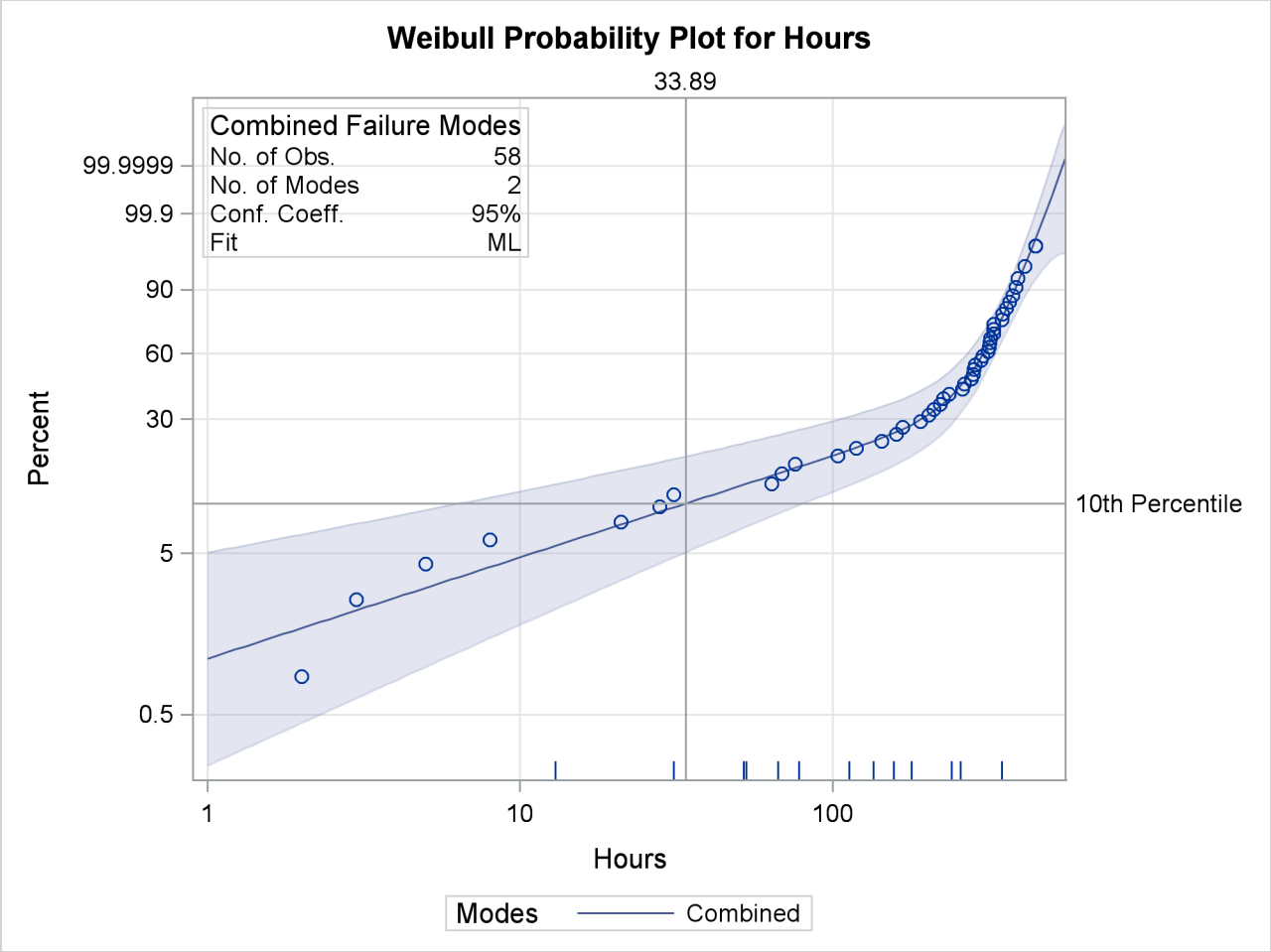
Figure 18.28 shows Weibull parameter estimates for the two individual failure modes.

Figure 18.28 Parameter Estimates for Individual Failure Modes

Individual Failure Mode Weibull Parameter Estimates					
			Asymptotic Normal 95% Confidence Limits		
Parameter	**Estimate**	**Standard Error**	**Lower**	**Upper**	**Mode**
EV Location	5.8415	0.0350	5.7730	5.9100	D
EV Scale	0.1785	0.0254	0.1350	0.2360	D
Weibull Scale	344.2966	12.0394	321.4903	368.7208	D
Weibull Shape	5.6020	0.7985	4.2365	7.4076	D
EV Location	7.0649	0.5109	6.0637	8.0662	E
EV Scale	1.5739	0.3415	1.0287	2.4080	E
Weibull Scale	1170.1832	597.7903	429.9480	3184.8703	E
Weibull Shape	0.6354	0.1379	0.4153	0.9721	E

Figure 18.29 is a Weibull probability plot of the failure probability distribution with the two failure modes combined, along with approximate pointwise 95% confidence limits. A reference line at the 10% point on the probability axis intersecting the distribution curve shows the tenth percentile of lifetimes when both modes act to be about 34 hours.

Figure 18.29 Weibull Plot for Failure Modes D and E

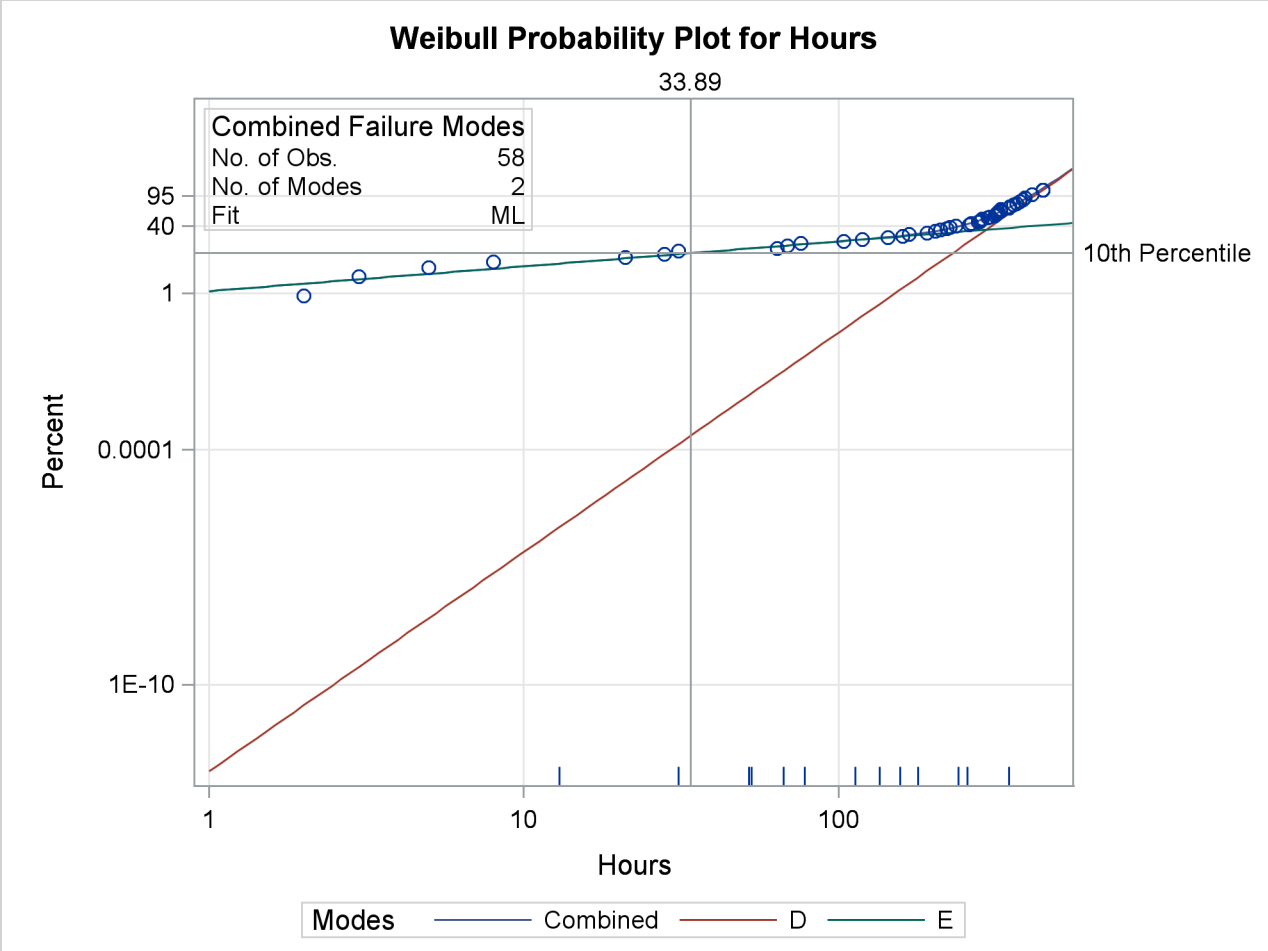


The following SAS statements create the Weibull probability plot in Figure 18.30:

```
proc reliability data=Voltage;
   distribution Weibull;
   pplot Hours*Status(1) / pref(intersect) = 10
                           preflabel = ('10th Percentile')
                           survtime = 100 200 300 400 500 1000
                           noconf
                           lupper = 500;
   fmode combine = Mode( 'D' 'E' ) / plotmodes;
run;
```

The PLOTMODES option in the FMODE statement cause the Weibull fits for the individual failure modes to be included on the probability plot. The combined failure mode curve is almost the same as the fit curve for mode E for lifetimes less than 100 hours, and slightly greater than the fit curve for mode D for lifetimes greater than 100 hours.

Figure 18.30 Weibull Plot for Failure Modes D and E with Individual Modes



Analysis of Recurrence Data on Repairs

This example illustrates analysis of recurrence data from repairable systems. Repair data analysis differs from life data analysis, where units fail only once. As a repairable system ages, it accumulates repairs and costs of repairs. The RELIABILITY procedure provides a nonparametric estimate and plot of the *mean cumulative function* (MCF) for the number or cost of repairs for a population of repairable systems.

The nonparametric estimate of the MCF, the variance of the MCF estimate, and confidence limits for the MCF estimate are based on the work of Nelson (1995). The MCF, also written as $M(t)$, is defined by Nelson (1995) to be the *population mean* of the distribution of the cumulative number or cost of repairs at age t. The method does not assume any underlying structure for the repair process.

The SAS statements that follow create the listing of the SAS data set VALVE shown in Figure 18.31, which contains repair histories of 41 diesel engines in a fleet (Nelson 1995). The valve seats in these engines wear out and must be replaced. The variable Id is a unique identifier for individual engines. The variable Days provides the engine age in days. The value of the variable Value is 1 if the age is a valve seat replacement age or -1 if the age is the end of history, or censoring age, for the engine.

```
data valve;
   input id Days value @@;
   label Days = 'Time of Replacement (Days)';
   datalines;
251 761 -1      252 759 -1      327  98  1      327 667 -1
328 326  1      328 653  1      328 653  1      328 667 -1
329 665 -1      330  84  1      330 667 -1      331  87  1
331 663 -1      389 646  1      389 653 -1      390  92  1
390 653 -1      391 651 -1      392 258  1      392 328  1
392 377  1      392 621  1      392 650 -1      393  61  1
393 539  1      393 648 -1      394 254  1      394 276  1
394 298  1      394 640  1      394 644 -1      395  76  1
395 538  1      395 642 -1      396 635  1      396 641 -1
397 349  1      397 404  1      397 561  1      397 649 -1
398 631 -1      399 596 -1      400 120  1      400 479  1
400 614 -1      401 323  1      401 449  1      401 582 -1
402 139  1      402 139  1      402 589 -1      403 593 -1
404 573  1      404 589 -1      405 165  1      405 408  1
405 604  1      405 606 -1      406 249  1      406 594 -1
407 344  1      407 497  1      407 613 -1      408 265  1
408 586  1      408 595 -1      409 166  1      409 206  1
409 348  1      409 389 -1      410 601 -1      411 410  1
411 581  1      411 601 -1      412 611 -1      413 608 -1
414 587 -1      415 367  1      415 603 -1      416 202  1
416 563  1      416 570  1      416 585 -1      417 587 -1
418 578 -1      419 578 -1      420 586 -1      421 585 -1
422 582 -1
;
```

Figure 18.31 Partial Listing of the Valve Seat Data

Obs	id	Days	value
1	251	761	-1
2	252	759	-1
3	327	98	1
4	327	667	-1
5	328	326	1
6	328	653	1
7	328	653	1
8	328	667	-1
9	329	665	-1
10	330	84	1
11	330	667	-1
12	331	87	1
13	331	663	-1
14	389	646	1
15	389	653	-1
16	390	92	1
17	390	653	-1
18	391	651	-1
19	392	258	1
20	392	328	1

The following statements produce the graphical displays in Figure 18.32 and Figure 18.33.

```
proc reliability;
   unitid id;
   mcfplot Days*value(-1) / nocenprint eventplot;
run;
```

The UNITID statement specifies that the variable Id uniquely identifies each system. The MCFPLOT statement requests a plot of the MCF estimates as a function of the age variable Days, and it specifies -1 as the value of the variable Value, which identifies the end of history for each engine (system). The option NOCENPRINT specifies that only failure times, and not censoring times, be printed in the tabular output. The option EVENTPLOT requests a horizontal plot of failure times for each system.

In Figure 18.32, the MCF estimates and confidence limits are plotted versus system age in days. The end-of-history ages are plotted in an area at the top of the plot. Except for the last few points, the plot is essentially a straight line, suggesting a constant replacement rate. Consequently, the prediction of future replacements of valve seats can be based on a fitted line in this case.

In Figure 18.33, a horizontal line for each system is drawn. Failures are marked by solid circles, and each line terminates at the censoring time for that system.

Figure 18.32 Mean Cumulative Function for the Number of Repairs

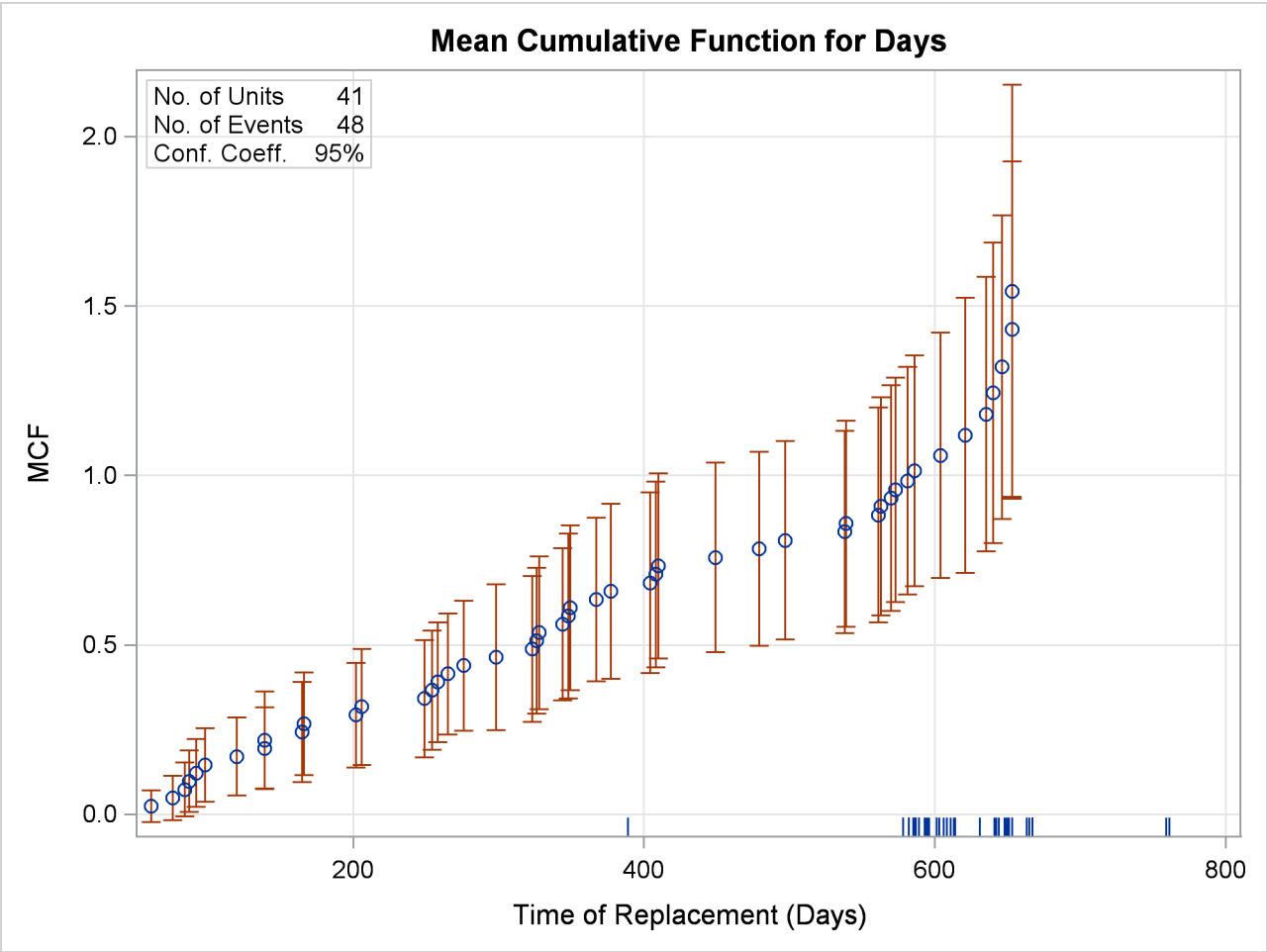
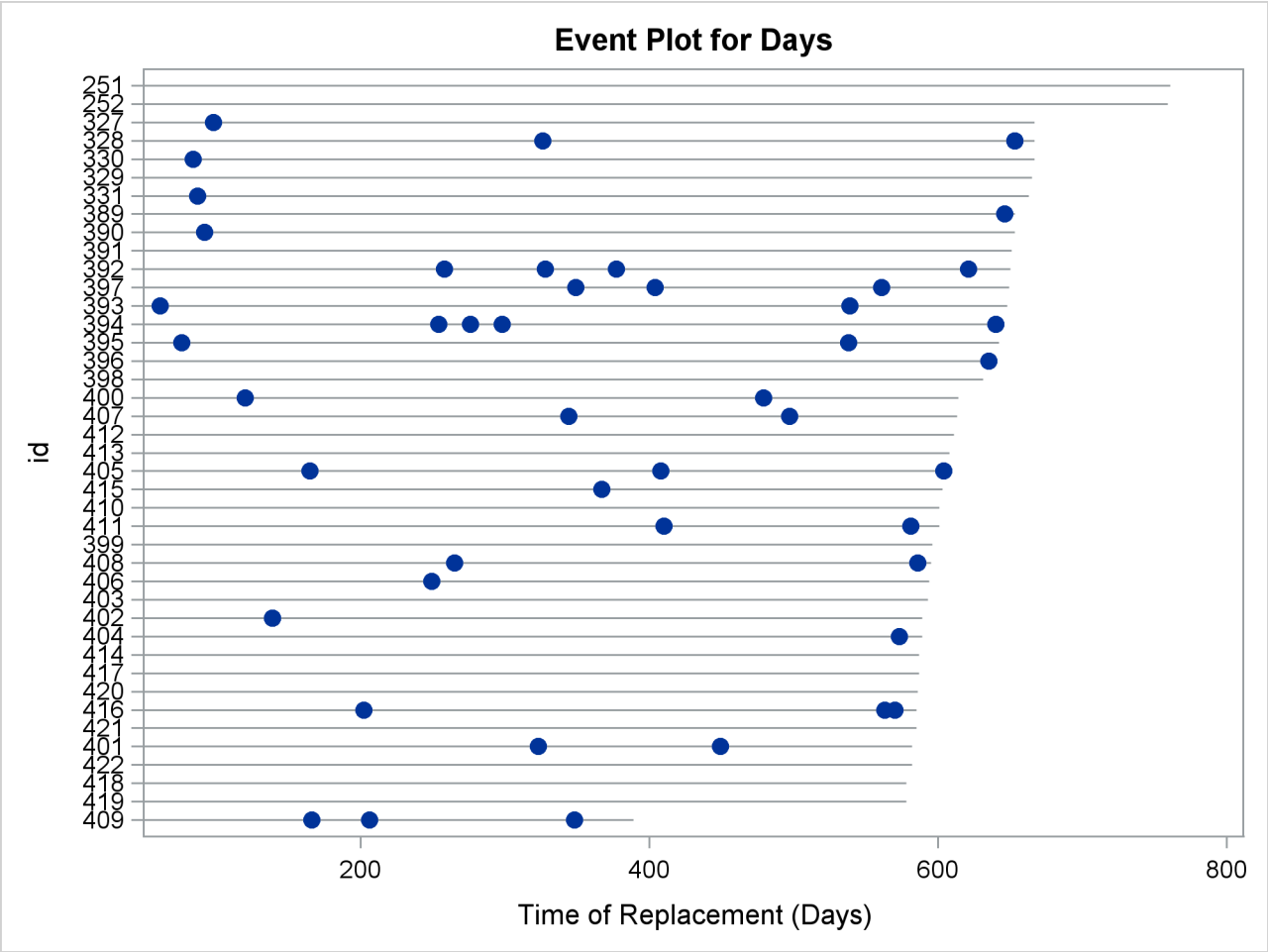


Figure 18.33 Recurrent Events Plot for the Valve Seat Data



A partial listing of the tabular output is shown in Figure 18.34 and Figure 18.35. It contains a summary of the repair data, estimates of the MCF, the Nelson (1995) standard errors, and confidence intervals for the MCF.

Figure 18.34 Partial Listing of the Output for the Valve Seat Data

Recurrence Data Summary	
Input Data Set	WORK.VALVE
Observations Used	89
Number of Units	41
Number of Events	48

Figure 18.35 Partial Listing of the Output for the Valve Seat Data

Recurrence Data Analysis					
			95% Confidence Limits		
Age	Sample MCF	Standard Error	Lower	Upper	Unit ID
61.00	0.024	0.024	-0.023	0.072	393
76.00	0.049	0.034	-0.017	0.115	395
84.00	0.073	0.041	-0.007	0.153	330
87.00	0.098	0.046	0.007	0.188	331
92.00	0.122	0.051	0.022	0.222	390
98.00	0.146	0.055	0.038	0.255	327
120.00	0.171	0.059	0.056	0.286	400
139.00	0.195	0.062	0.074	0.316	402
139.00	0.220	0.073	0.076	0.363	402
165.00	0.244	0.075	0.096	0.392	405
166.00	0.268	0.077	0.117	0.420	409
202.00	0.293	0.079	0.138	0.447	416
206.00	0.317	0.088	0.146	0.489	409
249.00	0.341	0.089	0.168	0.515	406
254.00	0.366	0.090	0.190	0.542	394
258.00	0.390	0.090	0.213	0.568	392
265.00	0.415	0.091	0.236	0.593	408
276.00	0.439	0.098	0.247	0.631	394
298.00	0.463	0.110	0.249	0.678	394
323.00	0.488	0.110	0.273	0.703	401
326.00	0.512	0.110	0.297	0.727	328
328.00	0.537	0.115	0.311	0.762	392
344.00	0.561	0.115	0.336	0.786	407
348.00	0.585	0.124	0.342	0.829	409
349.00	0.610	0.124	0.367	0.852	397
367.00	0.634	0.123	0.393	0.876	415
377.00	0.659	0.132	0.400	0.917	392
404.00	0.684	0.136	0.417	0.950	397
408.00	0.709	0.140	0.435	0.983	405
410.00	0.734	0.139	0.461	1.006	411
449.00	0.759	0.143	0.479	1.038	401
479.00	0.784	0.146	0.497	1.070	400
497.00	0.809	0.149	0.516	1.101	407
538.00	0.834	0.152	0.535	1.132	395
539.00	0.859	0.155	0.554	1.163	393
561.00	0.884	0.162	0.567	1.201	397
563.00	0.909	0.164	0.587	1.230	416
570.00	0.934	0.170	0.600	1.267	416
573.00	0.959	0.169	0.627	1.290	404
581.00	0.985	0.171	0.649	1.320	411
586.00	1.014	0.174	0.674	1.355	408
604.00	1.060	0.185	0.697	1.422	405
621.00	1.119	0.208	0.712	1.525	392
635.00	1.181	0.207	0.776	1.587	396

Figure 18.35 *continued*

Recurrence Data Analysis					
			95% Confidence Limits		
Age	Sample MCF	Standard Error	Lower	Upper	Unit ID
640.00	1.244	0.226	0.800	1.687	394
646.00	1.320	0.229	0.873	1.768	389
653.00	1.432	0.252	0.937	1.926	328
653.00	1.543	0.312	0.932	2.154	328

Parametric modeling of the repair process requires more assumptions than nonparametric modeling, and considerable work has been done in this area. Ascher and Feingold (1984), Tobias and Trindade (1995), Crowder et al. (1991), Meeker and Escobar (1998), Cook and Lawless (2007), Abernethy (2006), and Rigdon and Basu (2000) describe parametric models for repair processes. Repairs are sometimes modeled as a nonhomogeneous Poisson process, and the RELIABILITY procedure provides several forms of Poisson process models for recurrent events data. See the section "Parametric Models for Recurrent Events Data" on page 1351 for details about the Poisson process models that the RELIABILITY procedure provides.

A nonparametric MCF plot might be a first step in modeling a repair process, and, in many cases, provide the required answers without further analysis. An estimate of the MCF for a sample of systems aids engineers in determining the repair rate at any age and the increase or decrease of repair rate with population age. The estimate is also useful for predicting the number of future repairs.

Comparison of Two Samples of Repair Data

Nelson (2002) and Doganaksoy and Nelson (1998) show how the difference of MCFs from two samples can be used to compare the populations from which they are drawn. The RELIABILITY procedure provides Doganaksoy and Nelson's confidence intervals for the pointwise difference of the two MCFs, which can be used to assess whether the difference is statistically significant.

Doganaksoy and Nelson (1998) give an example of two samples of locomotives with braking grids from two different production batches. Figure 18.36 contains a listing of the data. The variable ID is a unique identifier for individual locomotives. The variable Days provides the locomotive age in days. The variable Value is 1 if the age corresponds to a braking grid replacement or -1 if the age corresponds to the locomotive's latest age (the current end of its history). The variable Sample is a group variable that identifies the grid production batch.

```
data Grids;
   if _N_ < 40 then Sample = 'Sample1';
   else Sample = 'Sample2';
   input ID$ Days Value @@;
   datalines;
S1-01 462  1    S1-01 730 -1    S1-02 364  1    S1-02 391  1
S1-02 548  1    S1-02 724 -1    S1-03 302  1    S1-03 444  1
S1-03 500  1    S1-03 730 -1    S1-04 250  1    S1-04 730 -1
S1-05 500  1    S1-05 724 -1    S1-06  88  1    S1-06 724 -1
S1-07 272  1    S1-07 421  1    S1-07 552  1    S1-07 625  1
S1-07 719 -1    S1-08 481  1    S1-08 710 -1    S1-09 431  1
S1-09 710 -1    S1-10 367  1    S1-10 710 -1    S1-11 635  1
S1-11 650  1    S1-11 708 -1    S1-12 402  1    S1-12 700 -1
S1-13  33  1    S1-13 687 -1    S1-14 287  1    S1-14 687 -1
S1-15 317  1    S1-15 498  1    S1-15 657 -1    S2-01 203  1
S2-01 211  1    S2-01 277  1    S2-01 373  1    S2-01 511 -1
S2-02 293  1    S2-02 503 -1    S2-03 173  1    S2-03 470 -1
S2-04 242  1    S2-04 464 -1    S2-05  39  1    S2-05 464 -1
S2-06  91  1    S2-06 462 -1    S2-07 119  1    S2-07 148  1
S2-07 306  1    S2-07 461 -1    S2-08 382  1    S2-08 460 -1
S2-09 250  1    S2-09 434 -1    S2-10 192  1    S2-10 448 -1
S2-11 369  1    S2-11 448 -1    S2-12  22  1    S2-12 447 -1
S2-13  54  1    S2-13 441 -1    S2-14 194  1    S2-14 432 -1
S2-15  61  1    S2-15 419 -1    S2-16  19  1    S2-16 185  1
S2-16 419 -1    S2-17 187  1    S2-17 416 -1    S2-18  93  1
S2-18 205  1    S2-18 264  1    S2-18 415 -1
;
```

Figure 18.36 Partial Listing of the Braking Grids Data

Obs	Sample	ID	Days	Value
1	Sample1	S1-01	462	1
2	Sample1	S1-01	730	-1
3	Sample1	S1-02	364	1
4	Sample1	S1-02	391	1
5	Sample1	S1-02	548	1
6	Sample1	S1-02	724	-1
7	Sample1	S1-03	302	1
8	Sample1	S1-03	444	1
9	Sample1	S1-03	500	1
10	Sample1	S1-03	730	-1
11	Sample1	S1-04	250	1
12	Sample1	S1-04	730	-1
13	Sample1	S1-05	500	1
14	Sample1	S1-05	724	-1
15	Sample1	S1-06	88	1
16	Sample1	S1-06	724	-1
17	Sample1	S1-07	272	1
18	Sample1	S1-07	421	1
19	Sample1	S1-07	552	1
20	Sample1	S1-07	625	1

The following statements request the Nelson (1995) nonparametric estimate and confidence limits for the difference of the MCF functions shown in Figure 18.37 for the braking grids:

```
proc reliability data=Grids;
   unitid ID;
   mcfplot Days*Value(-1) = Sample / mcfdiff;
run;
```

The MCFPLOT statement requests a plot of each MCF estimate as a function of age (provided by Days), and it specifies that the end of history for each system is identified by Value equal to -1. The variable Sample identifies the two samples of braking grids. The option MCFDIFF requests that the difference between the MCFs of the two groups given in the variable Sample be computed and plotted. Confidence limits for the MCF difference are also computed and plotted. The UNITID statement specifies that the variable Id uniquely identify each system.

Figure 18.37 shows the plot of the MCF difference function and pointwise 95% confidence intervals. Since the pointwise confidence limits do not include zero for some system ages, the difference between the two populations is statistically significant. A listing of the tabular output is shown in Figure 18.38. It contains a summary of the repair data for the two samples, estimates, standard errors, and confidence intervals for the MCF difference. A statistical test for different MCFs is also computed and is displayed in the table "Tests for Equality of Mean Functions." The tests also indicate a significant difference between the two samples.

Figure 18.37 Mean Cumulative Function Difference

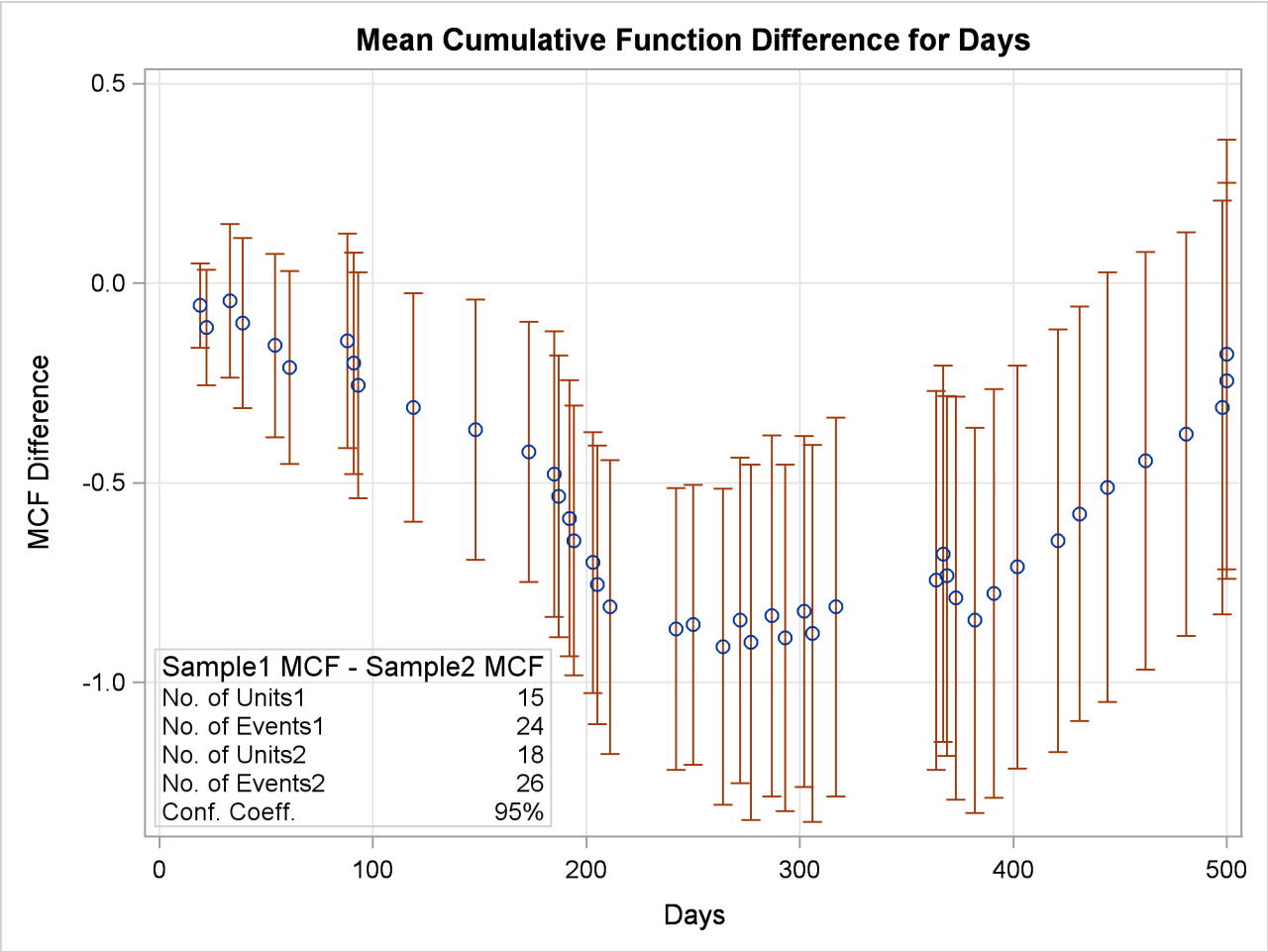


Figure 18.38 Listing of the Output for the Braking Grids Data

MCF Difference Data Summary	
Input Data Set	WORK.GRIDS
Group 1	Sample1
Observations Used	39
Number of Units	15
Number of Events	24
Group 2	Sample2
Observations Used	44
Number of Units	18
Number of Events	26

Figure 18.38 *continued*

Sample MCF Differences					
			95% Confidence Limits		
Age	MCF Difference	Standard Error	Lower	Upper	Unit ID
19.00	-0.056	0.054	-0.161	0.050	S2-16
22.00	-0.111	0.074	-0.256	0.034	S2-12
33.00	-0.044	0.098	-0.237	0.148	S1-13
39.00	-0.100	0.109	-0.313	0.113	S2-05
54.00	-0.156	0.117	-0.385	0.074	S2-13
61.00	-0.211	0.124	-0.453	0.031	S2-15
88.00	-0.144	0.137	-0.414	0.125	S1-06
91.00	-0.200	0.142	-0.478	0.078	S2-06
93.00	-0.256	0.145	-0.539	0.028	S2-18
119.00	-0.311	0.146	-0.598	-0.024	S2-07
148.00	-0.367	0.167	-0.693	-0.040	S2-07
173.00	-0.422	0.166	-0.748	-0.097	S2-03
185.00	-0.478	0.182	-0.835	-0.120	S2-16
187.00	-0.533	0.180	-0.886	-0.181	S2-17
192.00	-0.589	0.177	-0.935	-0.243	S2-10
194.00	-0.644	0.172	-0.982	-0.307	S2-14
203.00	-0.700	0.167	-1.027	-0.373	S2-01
205.00	-0.756	0.178	-1.105	-0.407	S2-18
211.00	-0.811	0.188	-1.179	-0.443	S2-01
242.00	-0.867	0.180	-1.219	-0.514	S2-04
250.00	-0.856	0.179	-1.207	-0.504	S1-04,S2-09
264.00	-0.911	0.202	-1.307	-0.515	S2-18
272.00	-0.844	0.208	-1.252	-0.437	S1-07
277.00	-0.900	0.227	-1.345	-0.455	S2-01
287.00	-0.833	0.231	-1.286	-0.380	S1-14
293.00	-0.889	0.222	-1.323	-0.455	S2-02
302.00	-0.822	0.224	-1.262	-0.383	S1-03
306.00	-0.878	0.241	-1.350	-0.406	S2-07
317.00	-0.811	0.242	-1.286	-0.337	S1-15
364.00	-0.744	0.242	-1.219	-0.270	S1-02
367.00	-0.678	0.241	-1.150	-0.206	S1-10
369.00	-0.733	0.230	-1.185	-0.282	S2-11
373.00	-0.789	0.257	-1.293	-0.284	S2-01
382.00	-0.844	0.246	-1.327	-0.362	S2-08
391.00	-0.778	0.261	-1.290	-0.266	S1-02
402.00	-0.711	0.258	-1.217	-0.206	S1-12
421.00	-0.644	0.270	-1.174	-0.115	S1-07
431.00	-0.578	0.265	-1.097	-0.059	S1-09
444.00	-0.511	0.275	-1.049	0.027	S1-03
462.00	-0.444	0.267	-0.968	0.079	S1-01
481.00	-0.378	0.258	-0.883	0.128	S1-08
498.00	-0.311	0.265	-0.830	0.208	S1-15
500.00	-0.244	0.253	-0.741	0.252	S1-05
500.00	-0.178	0.275	-0.716	0.360	S1-03

Figure 18.38 *continued*

Tests for Equality of Mean Cumulative Functions					
Weight Function	Statistic	Variance	Chi-Square	DF	Pr > Chi Square
Constant	-3.673285	4.556053	2.961560	1	0.0853
Linear	-4.435032	1.424770	13.805393	1	0.0002

You can fit a parametric model that uses Sample as a classification variable. This results in a model with a common shape parameter for the two groups but with different scale parameters. Suppose you want estimates of the parametric mean and intensity functions at values of the time variable 500, 600, 700, 800, 900, and 1,000 days for each of the two groups. The following statements create a new input data set that has observations for the desired prediction times appended to it. The additional observations are not used in the analysis, because the censoring variable Value is set to missing for those observations. Values of the mean and intensity function are computed, however, in the table that is produced by specifying the OBSTATS option in the MODEL statement.

The following statements create the new data set by appending observations to the original Grids data set:

```
data Predict;
   Control=1;
   if _N_ < 7 then Sample = 'Sample1';
   else Sample = 'Sample2';

   input ID$ Days Value;
   datalines;
9999 500  .
9999 600  .
9999 700  .
9999 800  .
9999 900  .
9999 1000 .
9999 500  .
9999 600  .
9999 700  .
9999 800  .
9999 900  .
9999 1000 .
;

data Grids;
   set Predict Grids;
run;
```

The following statements fit a nonhomogeneous Poisson process with a power law mean function that uses Sample as a two-level covariate. The OBSTATS option requests that predicted values be computed for values of the variable Control equal to 1. The MCFPLOT statement plots the fitted model as well as the nonparametric estimates of the MCF. Parametric confidence limits are displayed by default.

```
proc reliability data=Grids;
   unitid ID;
   distribution nhpp(pow);
   class Sample;
   model   Days*Value(-1) = Sample /obstats(control=Control);
   mcfplot Days*Value(-1) = Sample /fit=model overlay;
run;
```

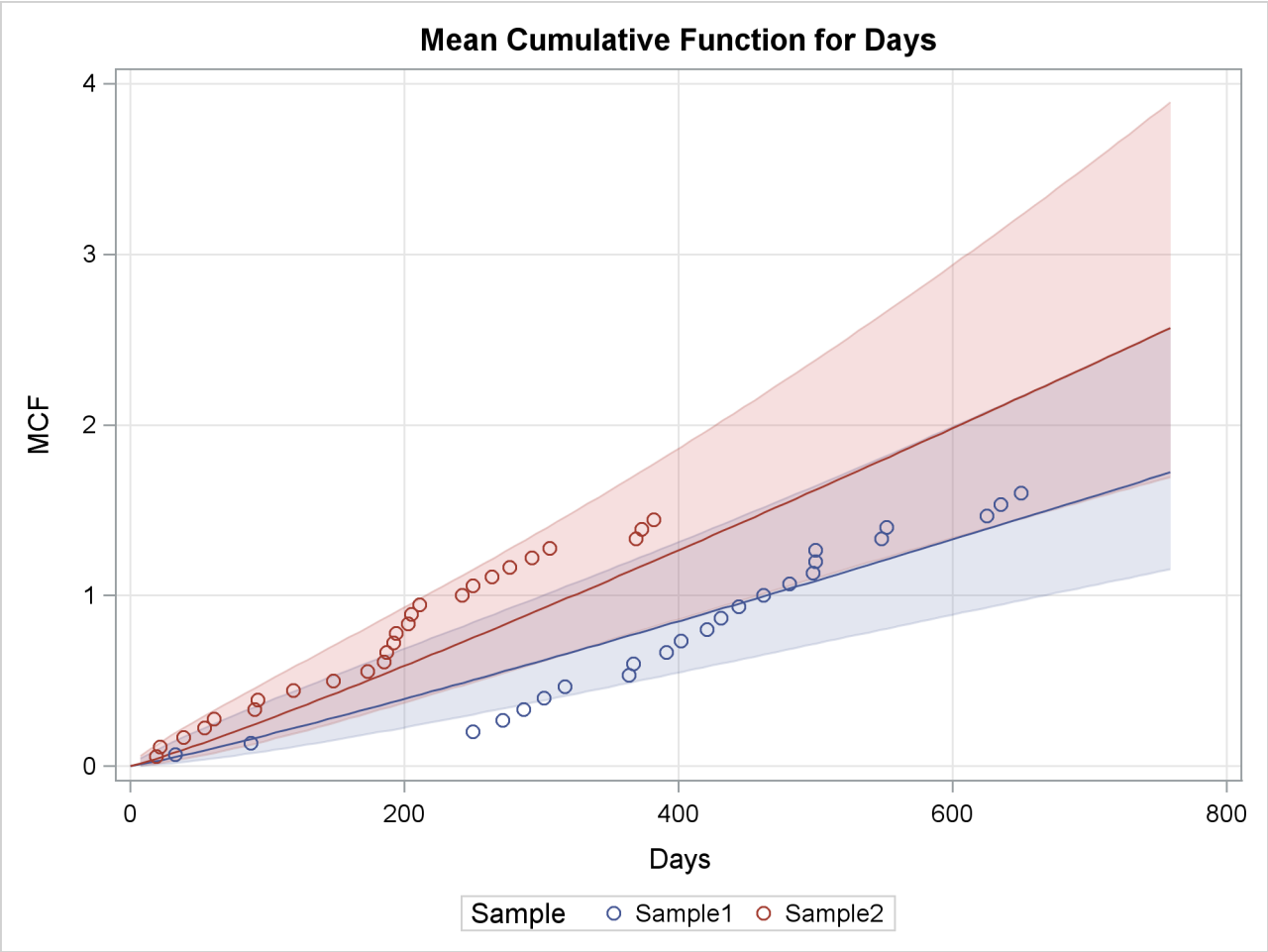
Figure 18.39 Predicted Mean and Intensity Function for the Braking Grids Data

The RELIABILITY Procedure

Observation Statistics								
Days	Value	Sample	ID	Xbeta	Shape	MCF	MCF_Lower	MCF_Upper
500	.	Sample1	9999	464.04648	1.1050556	1.0859585	0.7176451	1.6432995
600	.	Sample1	9999	464.04648	1.1050556	1.3283512	0.8874029	1.9884054
700	.	Sample1	9999	464.04648	1.1050556	1.5750445	1.0556772	2.3499278
800	.	Sample1	9999	464.04648	1.1050556	1.8254803	1.2215469	2.7279987
900	.	Sample1	9999	464.04648	1.1050556	2.0792348	1.3846219	3.1223089
1000	.	Sample1	9999	464.04648	1.1050556	2.3359745	1.5448083	3.5323327
500	.	Sample2	9999	323.23791	1.1050556	1.6193855	1.1012136	2.3813813
600	.	Sample2	9999	323.23791	1.1050556	1.9808425	1.335699	2.9375905
700	.	Sample2	9999	323.23791	1.1050556	2.3487125	1.5633048	3.5287107
800	.	Sample2	9999	323.23791	1.1050556	2.7221634	1.7845232	4.1524669
900	.	Sample2	9999	323.23791	1.1050556	3.100563	2.0000301	4.8066733
1000	.	Sample2	9999	323.23791	1.1050556	3.4834144	2.2104841	5.4893747

Observation Statistics					
Days	MCF_StdErr	Intensity	Int_Lower	Int_Upper	Int_StdErr
500	0.2295199	0.0024001	0.0015543	0.0037061	0.000532
600	0.2733977	0.0024465	0.0015458	0.0038719	0.0005731
700	0.3215248	0.0024864	0.0015325	0.0040343	0.000614
800	0.3741606	0.0025216	0.0015171	0.0041911	0.0006537
900	0.4313142	0.002553	0.0015011	0.0043418	0.0006917
1000	0.4928631	0.0025814	0.0014852	0.0044866	0.000728
500	0.3186232	0.003579	0.0021872	0.0058566	0.0008993
600	0.3982653	0.0036482	0.0021492	0.0061928	0.0009849
700	0.4878045	0.0037078	0.0021124	0.006508	0.0010643
800	0.5864921	0.0037602	0.002078	0.0068042	0.0011378
900	0.6935604	0.003807	0.002046	0.0070837	0.0012061
1000	0.8083117	0.0038494	0.0020164	0.0073484	0.0012699

Figure 18.40 Fitted Model



The predicted values of the mean and intensity functions at the desired values of Days, with standard errors and confidence limits, are shown in Figure 18.39.

A plot of the fitted mean function, along with nonparametric estimates for the two samples, is shown in Figure 18.40.

Analysis of Interval Age Recurrence Data

You can analyze recurrence data when the recurrence ages are grouped into intervals, instead of being exact ages. Figure 18.41 shows a listing of a SAS data set containing field data on replacements of defrost controls in 22,914 refrigerators, whose ages are grouped by months in service. Nelson (2002, problem 5.2, chapter 5) presents these data. Grouping the control data on the 22,914 refrigerators into age intervals enables you to represent the data by 29 data records, instead of requiring a single data record for each refrigerator, as required for exact recurrence data.

The variables Lower and Upper are the lower and upper monthly interval endpoints, Recurrences is the number of defrost control replacements in each month, and Censored is the number of refrigerator histories

censored in each month—that is, the number with current age in the monthly interval. Data are entered as shown in Figure 18.41.

Figure 18.41 Listing of the Defrost Controls Data

Obs	Lower	Upper	Recurrences	Censored
1	0	1	83	0
2	1	2	35	0
3	2	3	23	0
4	3	4	15	0
5	4	5	22	0
6	5	6	16	3
7	6	7	13	36
8	7	8	12	24
9	8	9	15	29
10	9	10	15	37
11	10	11	24	40
12	11	12	12	20041
13	12	13	7	14
14	13	14	11	17
15	14	15	15	13
16	15	16	6	28
17	16	17	8	22
18	17	18	9	27
19	18	19	9	64
20	19	20	5	94
21	20	21	6	119
22	21	22	6	118
23	22	23	6	138
24	23	24	5	1188
25	24	25	7	17
26	25	26	5	28
27	26	27	5	99
28	27	28	6	128
29	28	29	3	590

The following SAS statements create the plot of the sample MCF of defrost control replacement shown in Figure 18.42 and the tabular listing in Figure 18.43:

```
proc reliability data=defrost;
   mcfplot ( interval    = Lower Upper
             recurrences = Recurrences
             censor      = Censored ) / plotsymbol = X
                                        vaxis = 0 to .12 by .04
                                        interpolate = join;
run;
```

Pointwise confidence limits are included on the plot and in the tabular listing. These limits are approximate, and are usually shorter than the correct limits, which have not been developed for interval data.

Here, INTERVAL = LOWER UPPER specifies the input data set variables Lower and Upper as the age interval endpoints. The variable Recurrences identifies the number of recurrences (defrost control replacements) in each time interval, and Censored identifies the number of units censored in each interval (number in an age interval or removed from the sample in an age interval).

Figure 18.42 MCF Plot for the Defrost Controls

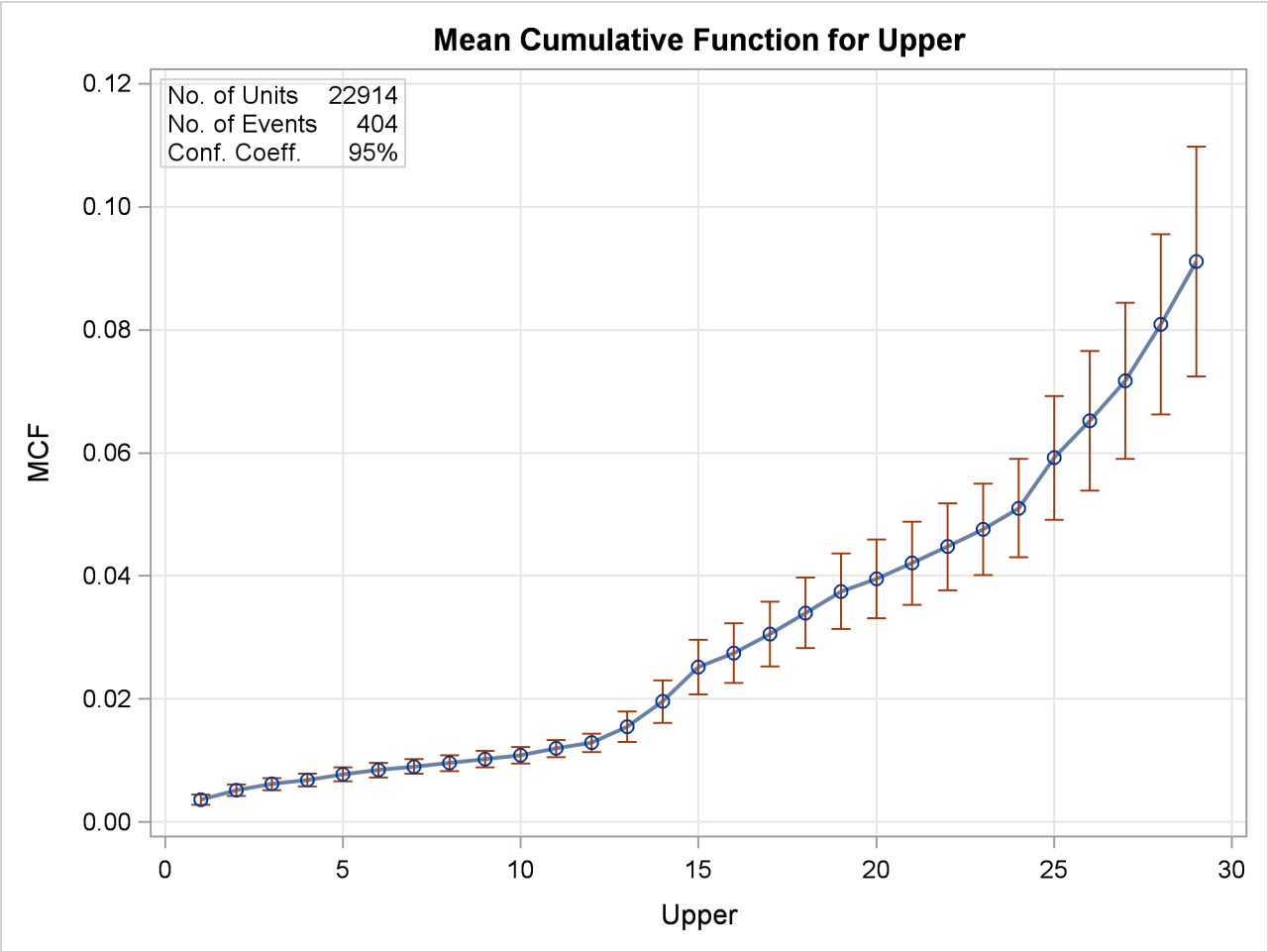


Figure 18.43 Listing of the Output for the Defrost Controls Data

Recurrence Data Summary	
Input Data Set	WORK.DEFROST
Observations Used	29
Number of Units	22914
Number of Events	404

Figure 18.43 *continued*

Recurrence Data Analysis						
	Endpoints				Naive 95% Confidence Limits	
	Lower	Upper	Sample MCF	Standard Error	Lower	Upper
	0.00	**1.00**	0.004	0.000	0.003	0.004
	1.00	**2.00**	0.005	0.000	0.004	0.006
	2.00	**3.00**	0.006	0.001	0.005	0.007
	3.00	**4.00**	0.007	0.001	0.006	0.008
	4.00	**5.00**	0.008	0.001	0.007	0.009
	5.00	**6.00**	0.008	0.001	0.007	0.010
	6.00	**7.00**	0.009	0.001	0.008	0.010
	7.00	**8.00**	0.010	0.001	0.008	0.011
	8.00	**9.00**	0.010	0.001	0.009	0.012
	9.00	**10.00**	0.011	0.001	0.010	0.012
	10.00	**11.00**	0.012	0.001	0.011	0.013
	11.00	**12.00**	0.013	0.001	0.011	0.014
	12.00	**13.00**	0.015	0.001	0.013	0.018
	13.00	**14.00**	0.020	0.002	0.016	0.023
	14.00	**15.00**	0.025	0.002	0.021	0.030
	15.00	**16.00**	0.027	0.002	0.023	0.032
	16.00	**17.00**	0.031	0.003	0.025	0.036
	17.00	**18.00**	0.034	0.003	0.028	0.040
	18.00	**19.00**	0.038	0.003	0.031	0.044
	19.00	**20.00**	0.040	0.003	0.033	0.046
	20.00	**21.00**	0.042	0.003	0.035	0.049
	21.00	**22.00**	0.045	0.004	0.038	0.052
	22.00	**23.00**	0.048	0.004	0.040	0.055
	23.00	**24.00**	0.051	0.004	0.043	0.059
	24.00	**25.00**	0.059	0.005	0.049	0.069
	25.00	**26.00**	0.065	0.006	0.054	0.077
	26.00	**27.00**	0.072	0.006	0.059	0.084
	27.00	**28.00**	0.081	0.007	0.066	0.096
*	**28.00**	**29.00**	0.091	0.010	0.072	0.110

*** The estimate and limits for this interval may not be appropriate.**

The last interval is always marked with a footnote indicating that estimates for the last interval may be biased since censoring ages often are not uniformly spread over that interval.

Analysis of Binomial Data

This example illustrates the analysis of binomial proportions of capacitor failures from nine circuit boards. The data are given by Nelson (1982, p. 451). The following statements create and list a SAS data set named BINEX containing the data:

```
data binex;
   input board sample fail;
   datalines;
1 84 2
2 72 3
3 72 5
4 119 19
5 538 21
6 51 2
7 517 9
8 462 18
9 143 2
;
```

Figure 18.44 displays a listing of the data. The variable Board identifies the circuit board, the variable Sample provides the number of capacitors on the boards, and the variable Fail provides the number of capacitors failing on the boards.

Figure 18.44 Listing of the Capacitor Data

Obs	board	sample	fail
1	1	84	2
2	2	72	3
3	3	72	5
4	4	119	19
5	5	538	21
6	6	51	2
7	7	517	9
8	8	462	18
9	9	143	2

The following statements analyze the proportion of capacitors failing:

```
proc reliability data=Binex;
   distribution binomial;
   analyze fail(sample) = board / predict(1000)
                                  tolerance(.05);
run;
```

The DISTRIBUTION statement specifies the binomial distribution. The analysis requested with the ANALYZE statement consists of tabular output only. Graphical output is not available for the binomial distribution. The variable Fail provides the number of capacitors failing on each board, the variable Sample provides the sample size (number of capacitors) for each board, and the variable Board identifies the individual boards. The statement option PREDICT(1000) requests the predicted number of capacitors failing and prediction limits in a future sample of size 1000. The option TOLERANCE(.05) requests the sample size required to estimate the binomial proportion to within 0.05. Figure 18.45 displays the results of the analysis.

The "Pooled Data Analysis" table displays the estimated binomial probability and exact binomial confidence limits when data from all boards are pooled. The chi-square value and p-value for a test of equality of the binomial probabilities for all of the boards are also shown. In this case, the p-value is less than 0.05, so you reject the test of equality at the 0.05 level.

The "Predicted Values and Limits" table provides the predicted failure count and prediction limits for the number of capacitors that would fail in a future sample of size 1000 for the pooled data, as requested with the PREDICT(1000) option. The "Sample Size for Estimation" table gives the sample size required to estimate the binomial probability to within 0.05 for the pooled data, as requested with the TOLERANCE(.05) option.

The "Estimates by Group" table supplies the estimated binomial probability, confidence limits, and the contribution to the total chi-square for each board. The pooled values are shown in the last line of the table.

The "Predicted Values by Group" table gives the predicted counts in a future sample of size 1000, prediction limits, and the sample size required to estimate the binomial probability to within the tolerance of 0.05 for each board. Values for the pooled data are shown in the last line of the table.

Figure 18.45 Analysis of the Capacitor Data

The RELIABILITY Procedure

Model Information - All Groups	
Input Data Set	WORK.BINEX
Events Variable	fail
Trials Variable	sample
Distribution	Binomial
Confidence Coefficient	95%
Observations Used	9

Binomial Data Analysis	
Pooled Events	81.0000
Pooled Trials	2058.0000
Estimate of Proportion	0.0394
Lower Limit For Proportion	0.0314
Upper Limit For Proportion	0.0487
ChiSquare	56.8504
Pr>ChiSquare	0.0000

Predicted Value and Limits	
Sample Size For Prediction	1000.0000
Predicted Count	39.3586
Lower Prediction Limit	24.8424
Upper Prediction Limit	56.3237

Sample Size For Estimation	
Tolerance	0.0500
Sample Size For Tolerance	58.0975

Figure 18.45 *continued*

Estimates By Group

Group	Events	Trials	Prop	95% Confidence Limits Lower	Upper	X2
1	2	84	0.0238	0.0029	0.0834	0.5371
2	3	72	0.0417	0.0087	0.1170	0.0101
3	5	72	0.0694	0.0229	0.1547	1.7237
4	19	119	0.1597	0.0990	0.2381	45.5528
5	21	538	0.0390	0.0243	0.0590	0.0015
6	2	51	0.0392	0.0048	0.1346	0.0000
7	9	517	0.0174	0.0080	0.0328	6.5884
8	18	462	0.0390	0.0233	0.0609	0.0019
9	2	143	0.0140	0.0017	0.0496	2.4348
Pooled	81	2058	0.0394	0.0314	0.0487	56.8504

Predicted/Tolerance Values By Group

Group	Predicted Count	95% Prediction Limits Lower	Upper	Tolerance Sample Size
1	23.81	1.5476	88.5824	35.71
2	41.67	6.9416	124.6142	61.36
3	69.44	20.4052	165.3499	99.30
4	159.66	91.9722	254.5444	206.17
5	39.03	20.1599	64.7140	57.64
6	39.22	3.3970	144.2494	57.90
7	17.41	5.3506	36.7531	26.28
8	38.96	19.3343	66.3850	57.53
9	13.99	0.3851	53.0715	21.19
Pooled	39.36	24.8424	56.3237	58.10

Three-Parameter Weibull

Meeker and Escobar (1998) give an example of the number of cycles to fatigue failure of specimens of a certain alloy. The first 67 specimens experienced failure, and the last five specimens had no failure at 300,000 cycles. The following statements create a SAS data set named Alloy that contains the number of cycles (in thousands) to failure or end of test for the specimens:

```
data Alloy;
   input kCycles@@;
   Cen = _n_ > 67;
   label kCycles = 'Fatigue Life in Thousands of Cycles';
   datalines;
94  96  99  99  104 108 112 114 117 117
118 121 121 123 129 131 133 135 136 139
139 140 141 141 143 144 149 149 152 153
159 159 159 159 162 168 168 169 170 170
171 172 173 176 177 180 180 184 187 188
189 190 196 197 203 205 211 213 224 226
227 256 257 269 271 274 291 300 300 300
300 300
;
```

The following SAS statements fit a three-parameter Weibull distribution to the specimen lifetimes, in thousands of cycles. The PROFILE option requests a profile likelihood plot for the threshold parameter. ODS Graphics must be enabled to create a profile likelihood plot with the PROFILE option.

```
proc Reliability data=Alloy;
   distribution Weibull3;
   Pplot kCycles*Cen(1) / Profile(noconf range=(50,100)) LifeUpper=500;
run;
```

Figure 18.46 shows the maximum likelihood estimates of the Weibull threshold, shape and scale parameters, and the corresponding extreme value location and scale parameter estimates.

Figure 18.46 Three-Parameter Weibull Parameter Estimates

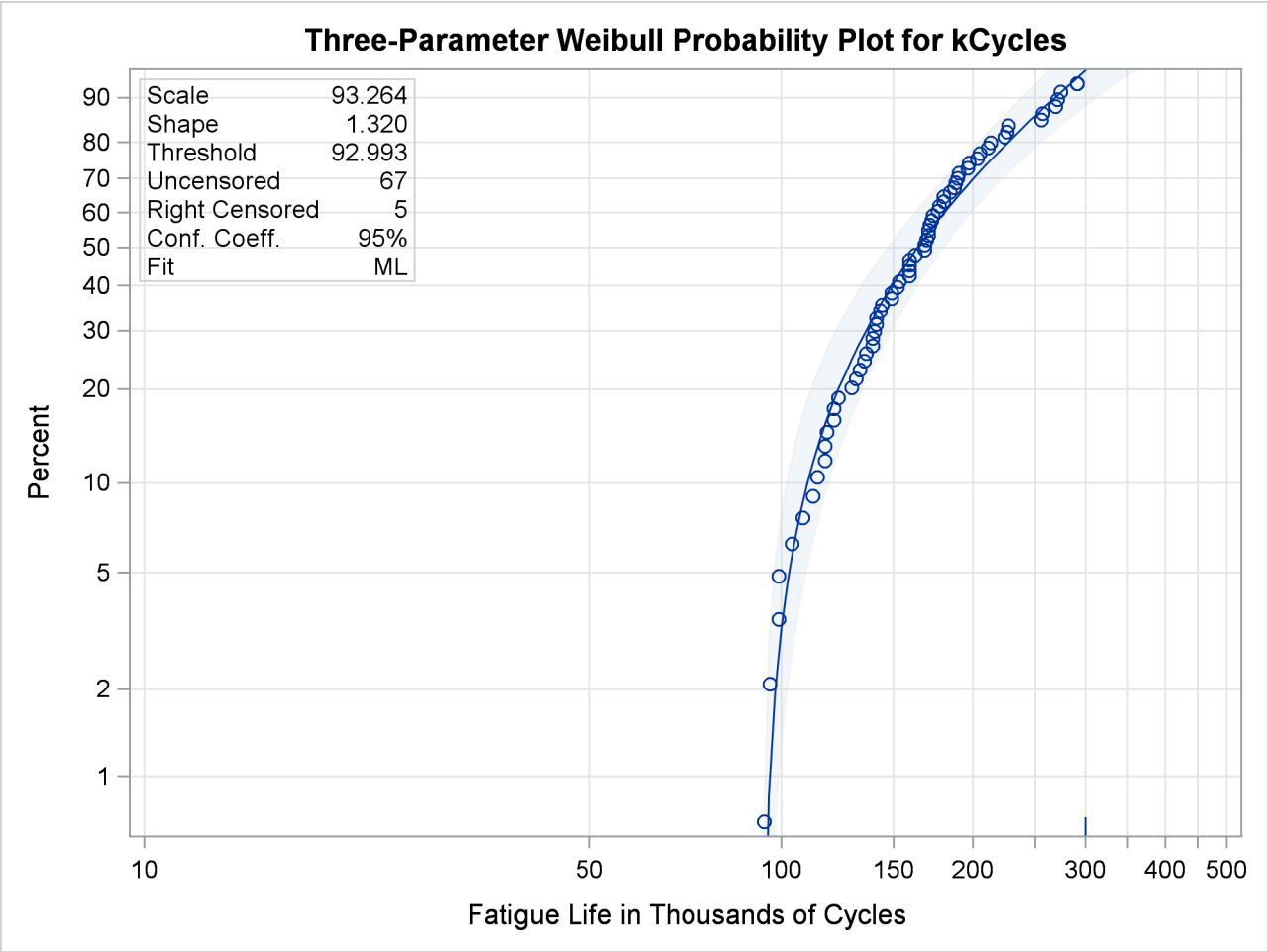
The RELIABILITY Procedure

Three-Parameter Weibull Parameter Estimates

Parameter	Estimate	Standard Error	Asymptotic Normal 95% Confidence Limits Lower	Upper
EV Location	4.5354	0.1009	4.3377	4.7332
EV Scale	0.7575	0.0898	0.6005	0.9556
Weibull Scale	93.2642	9.4082	76.5329	113.6531
Weibull Shape	1.3202	0.1565	1.0465	1.6654
Weibull Threshold	92.9928	1.9516	89.1676	96.8179

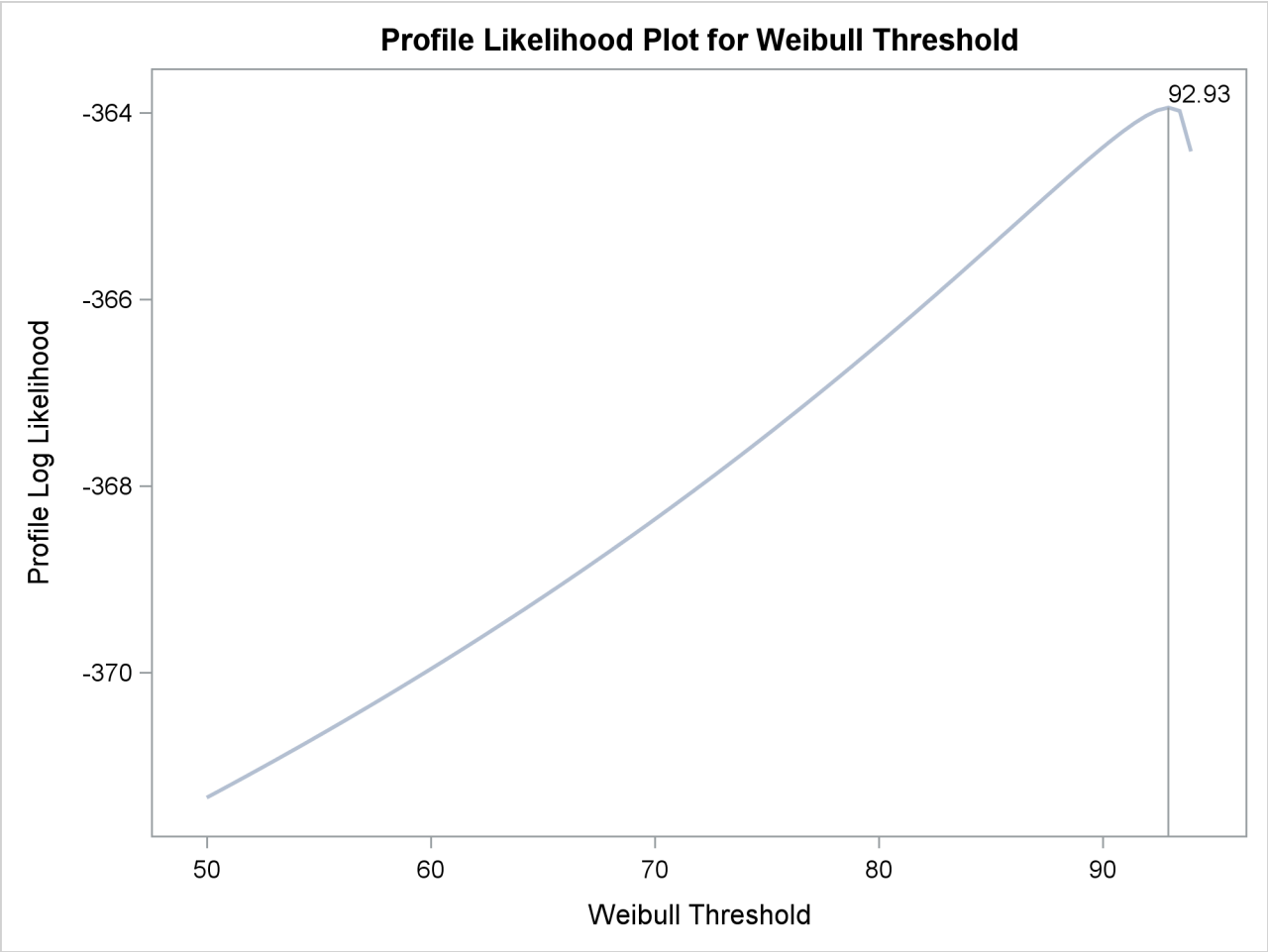
A probability plot of the failure lifetimes and the fitted three-parameter Weibull distribution is shown in Figure 18.47.

Figure 18.47 Three-Parameter Weibull Probability Plot



A profile likelihood plot for the threshold parameter is shown in Figure 18.48. The threshold value at the maximum log likelihood corresponds to the maximum likelihood estimate of the threshold parameter.

Figure 18.48 Profile Likelihood for Three-Parameter Weibull Threshold



Parametric Model for Recurrent Events Data

The following SAS statements fit a non-homogeneous Poisson process with a power intensity function model to the valve seat data described in the section "Analysis of Recurrence Data on Repairs" on page 1215. The FIT=MODEL option in the MCFPLOT statement requests that the fitted model be plotted on the plot with the nonparametric mean cumulative function estimates.

```
proc reliability data=Valve;
   unitid id;
   distribution Nhpp(Pow);
   model   Days*Value(-1);
   mcfplot Days*Value(-1) / Fit=Model Noconf;
run;
```

The model parameter estimates are shown in Figure 18.49.

Figure 18.49 Power Model Parameter Estimates for the Valve Seat Data

The RELIABILITY Procedure

NHPP-Power Parameter Estimates				
			Asymptotic Normal 95% Confidence Limits	
Parameter	Estimate	Standard Error	Lower	Upper
Intercept	553.6430	57.8636	451.0941	679.5048
Shape	1.3996	0.2005	1.0570	1.8533

Figure 18.50 displays a plot of nonparametric estimates of the mean cumulative function and the fitted model mean function. The parametric model matches the data well except at the upper end of the range of repair times, where the parametric model does not capture the rapid increase in the number of replacements of the valve seats. For this reason, the parametric model might not be appropriate for predicting future repairs of the engines.

Figure 18.50 Mean Cumulative Function Plot for the Valve Seat Data

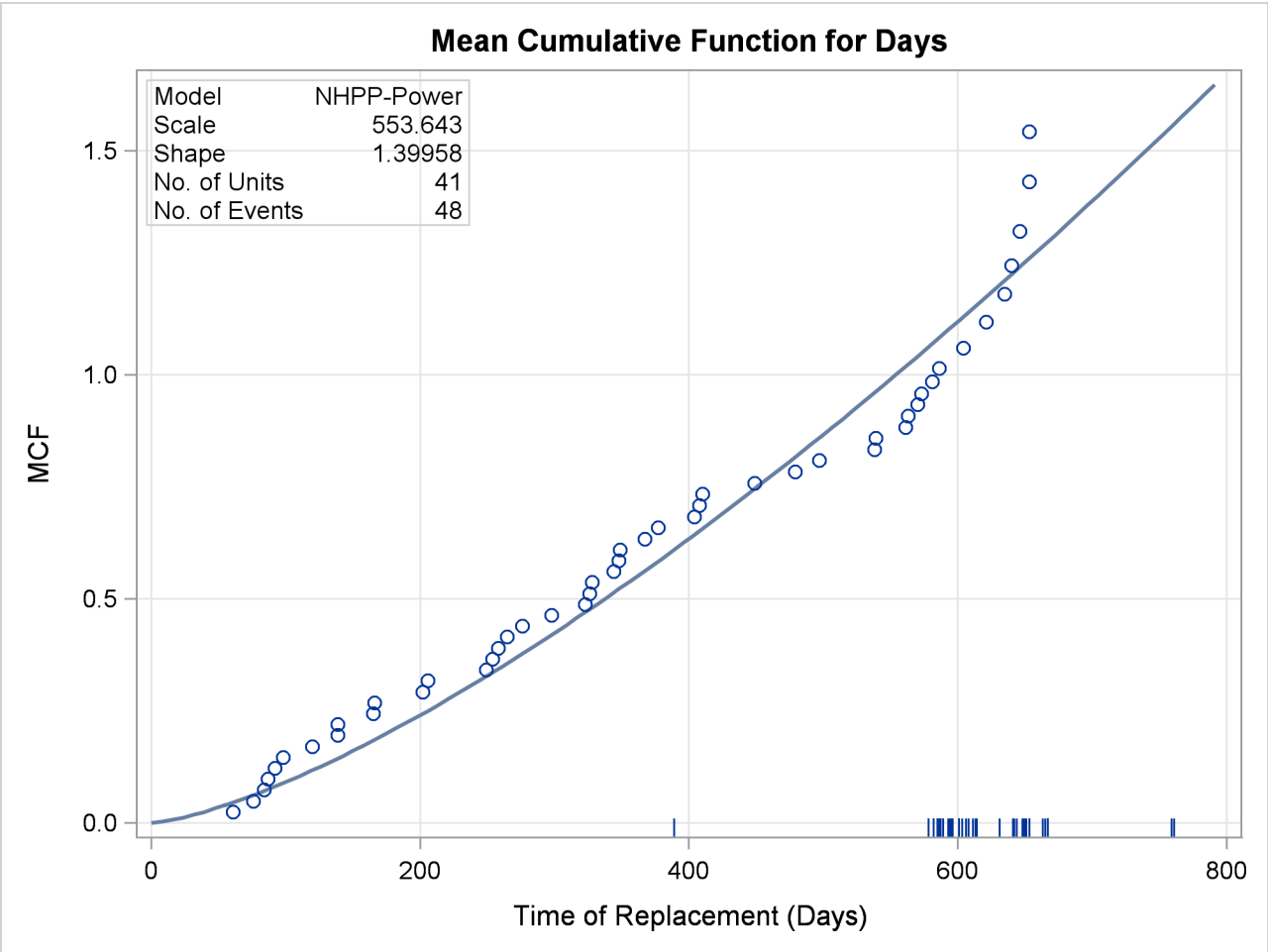
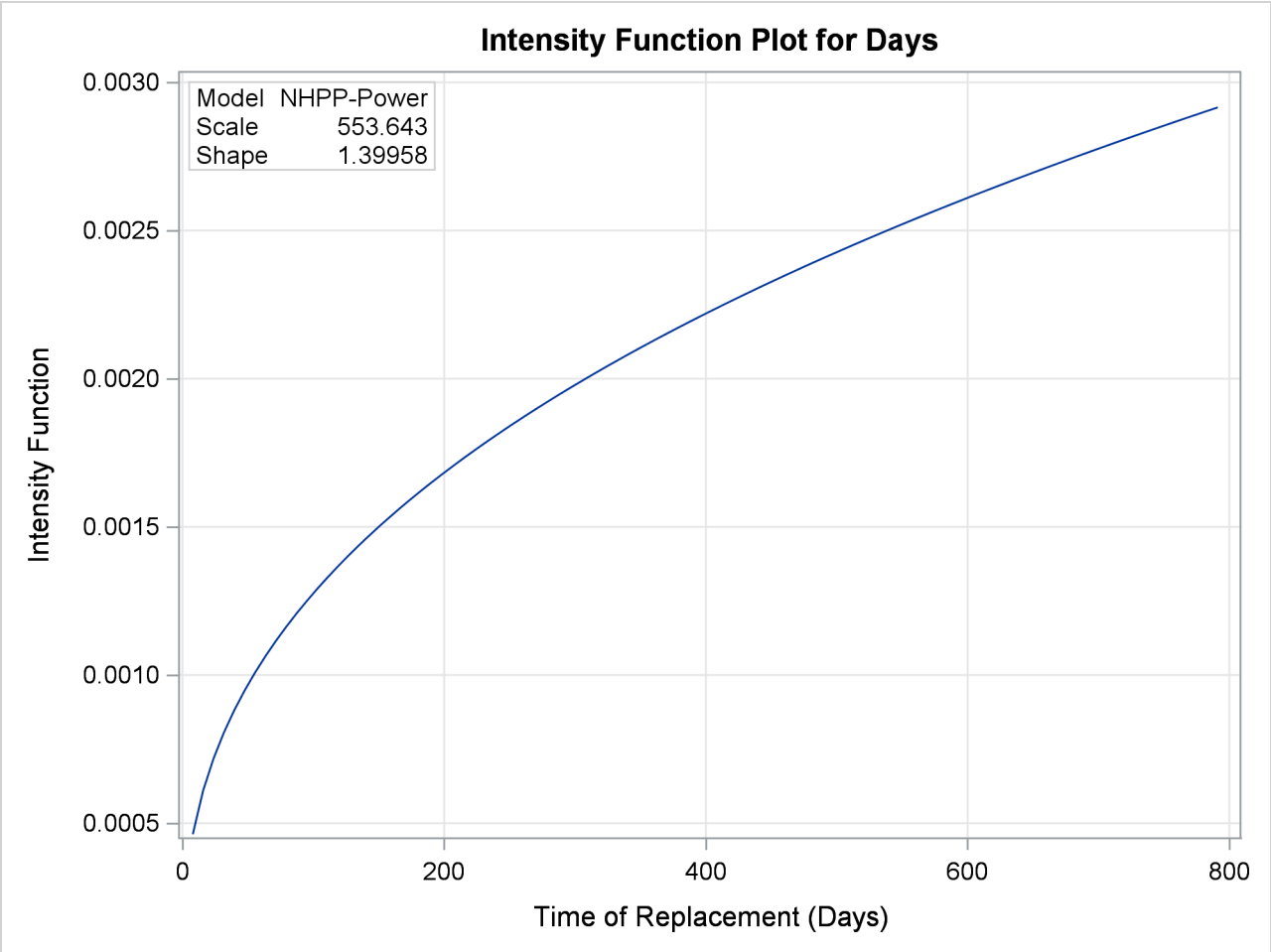


Figure 18.51 shows the parametric model intensity function. The intensity function increases with time, indicating an increasing rate of repairs. This is consistent with the parameter estimates in Figure 18.49, where a shape parameter significantly greater than 1 indicates an increasing failure rate.

Figure 18.51 Intensity Function Plot for the Valve Seat Data



Parametric Model for Interval Recurrent Events Data

Byar (1980) provides data for recurrences of bladder tumors in patients in a clinical trial. Figure 18.52 is a partial listing of data for 86 patients, of which 48 were given a placebo and 38 were treated with the drug Thiotepa. The data are here grouped into one month intervals.

Figure 18.52 Partial Listing of the Bladder Tumor Data

Obs	Group	Age	Age1	N	R
1	Placebo	0	1	48	0
2	Placebo	1	2	47	0
3	Placebo	2	3	46	1
4	Placebo	3	4	46	4
5	Placebo	4	5	46	7
6	Placebo	5	6	45	0
7	Placebo	6	7	45	2
8	Placebo	7	8	45	4
9	Placebo	8	9	44	1
10	Placebo	9	10	44	2
11	Placebo	10	11	44	4
12	Placebo	11	12	42	2
13	Placebo	12	13	42	1
14	Placebo	13	14	42	4
15	Placebo	14	15	42	1
16	Placebo	15	16	41	1
17	Placebo	16	17	41	5
18	Placebo	17	18	41	4
19	Placebo	18	19	41	4
20	Placebo	19	20	38	1

The following SAS statements fit a non-homogeneous Poisson process model with a power intensity function to the interval recurrence data. Some patients were lost to follow-up in each month, so the number of patients observed changes from month to month. The variable N provides the number of patients available at the beginning of each month and assumed to be observed throughout the month. The variable R is the number of recurrences of tumors in each month. Age represents the number of months after randomization into the trial (starting with month 0), and Age1=Age+1 is the end of a month. The variable Group represents the treatment, either Placebo or Thiotepa. The MODEL statement requests a maximum likelihood fit of the model with Group as a classification variable. The MCFPLOT statement requests a plot of the fitted model and nonparametric estimates of the mean cumulative function for each group.

```
proc reliability data=Tumor;
   distribution nhpp(pow);
   freq R;
   nenter N;
   class Group;
   model  (Age Age1) = Group;
   mcfplot(Age Age1) = Group / fit=Model;
run;
```

The resulting maximum likelihood parameter estimates are shown in Figure 18.53.

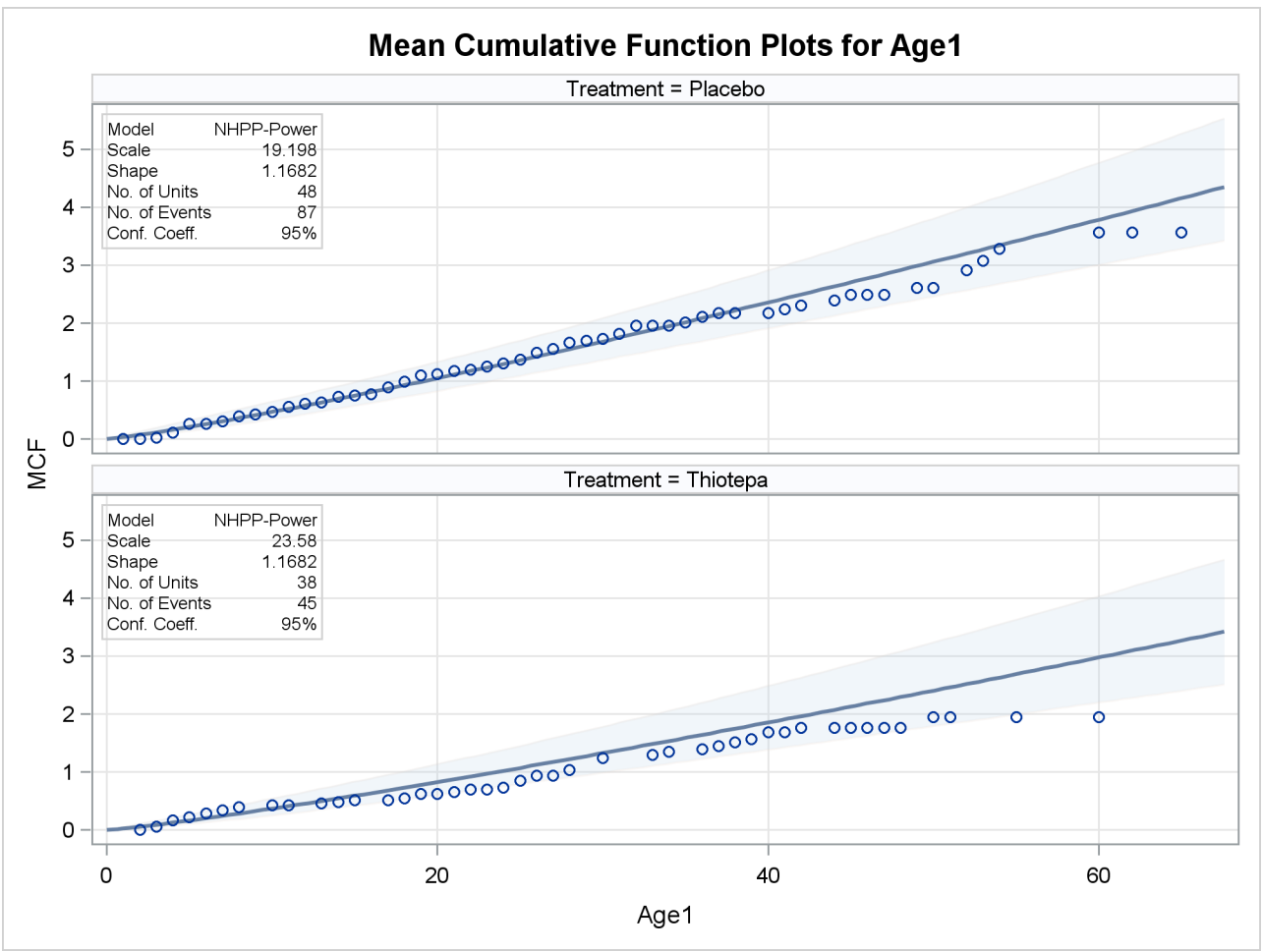
Figure 18.53 Power Model Parameter Estimates for the Bladder Tumor Data

The RELIABILITY Procedure

NHPP-Power Parameter Estimates					
				Asymptotic Normal 95% Confidence Limits	
Parameter		Estimate	Standard Error	Lower	Upper
Intercept		23.5802	3.1567	17.3932	29.7671
Group	Placebo	-4.3826	3.4873	-11.2175	2.4523
Group	Thiotepa	0.0000	0.0000	0.0000	0.0000
Shape		1.1682	0.0960	0.9945	1.3723

Nonparametric estimates of the mean cumulative function are plotted as points, and the fitted model is plotted as the solid line in Figure 18.54. Pointwise parametric confidence intervals are plotted by default when the fit=Model option is used.

Figure 18.54 Mean Cumulative Function Plot for the Bladder Tumor Data



Syntax: RELIABILITY Procedure

Primary Statements

The following are the primary statements that control the RELIABILITY procedure:

```
PROC RELIABILITY <options> ;
    <label:>ANALYZE variable <*censor-variable(values)> <=group-variables>
                    </options> ;
    <label:>MCFPLOT variable * cost/censor-variable(values) <=group-variables>
                    </options> ;
    MODEL variable <*censor-variable(values)> <=independent-variables>
            </options> ;
    <label:>PROBPLOT variable <*censor-variable(values)> <=group-variables>
                    </options> ;
    <label:>RELATIONPLOT variable <*censor-variable(values)> <=group-variables>
                    </options> ;
```

The PROC RELIABILITY statement invokes the procedure.

The plot statements (PROBPLOT, RELATIONPLOT, and MCFPLOT) create graphical displays. Each of the plot statements has options that control the content and appearance of the plots they create. The default settings provide the best plots for many purposes; however, if you want to control specific details of the plots, such as axis limits or background colors, then you need to specify the options.

In addition to graphical output, each plot statement provides analysis results in tabular form. The tabular output also can be controlled with statement options.

The MODEL and ANALYZE statements produce only tabular analysis output, not graphical displays.

You can specify one or more of the plot and ANALYZE statements. If you specify more than one MODEL statement, only the last one specified is used.

Secondary Statements

You can specify the following statements in conjunction with the primary statements listed previously. These statements are used to modify the behavior of the primary statements or to specify additional variables.

```
BY variables ;
CLASS variables ;
DISTRIBUTION distribution-name ;
EFFECTPLOT < plot-type < (plot-definition-options) > > < / options > ;
ESTIMATE < 'label' > estimate-specification < / options > ;
FMODE keyword = variable('value1' . . . 'valuen') ;
FREQ variable ;
INSET keyword-list < / options >  ;
LSMEANS < model-effects > < / options > ;
LSMESTIMATE model-effect < 'label' > values < divisor=n > < , . . . < 'label' > values < divisor=n > >
                < / options > ;
MAKE 'table' OUT=SAS-data-set <  options >  ;
NENTER variable ;
NLOPTIONS < options > ;
SLICE model-effect < / options > ;
STORE < OUT= >item-store-name < / LABEL='label' > ;
TEST < model-effects > < / options > ;
UNITID variable ;
```

The EFFECTPLOT, ESTIMATE, LSMEANS, LSMESTIMATE, SLICE, STORE, and TEST statements are used to provide further analysis of regression models that are fit by using a MODEL statement and are common to many SAS/STAT procedures. Summary descriptions of functionality and syntax for these statements appear after the PROC RELIABILITY statement in alphabetical order, and full documentation about them is available in Chapter 19, "Shared Concepts and Topics" (*SAS/STAT User's Guide*).

You can use the STORE statement to store the results of fitting a regression model with a MODEL statement for later analysis with the SAS/STAT procedure PROC PLM, if you have SAS/STAT software installed at your site.

The BY statement specifies variables in the input data set that are used for BY processing. A separate analysis is performed for each group of observations defined by the levels of the BY variables. The input data set must be sorted in order of the BY variables.

The CLASS statement specifies variables in the input data set that serve as *indicator*, *dummy*, or *classification* variables in the MODEL statement.

The DISTRIBUTION statement specifies a probability distribution name for those statements that require a probability distribution for proper operation (the ANALYZE, PROBPLOT, MODEL, and RELATIONPLOT statements). If you do not specify a distribution with the DISTRIBUTION statement, the normal distribution is used.

The FMODE statement specifies what failure-mode data to include in the analysis of data. Use this statement in conjunction with the ANALYZE, MODEL, PROBPLOT, or RELATIONPLOT statement.

The FREQ statement specifies a variable that provides frequency counts for each observation in the input data set.

The INSET statement specifies what information is printed in the inset box created by the PROBPLOT or MCFPLOT statement. The INSET statement also controls the appearance of the inset box.

The MAKE statement creates a SAS data set from any of the tables produced by the procedure. You specify a table and a SAS data set name for the data set you want to create. There is a unique table name that identifies each table printed; see the tables in the section "MAKE Statement" on page 1261.

The NENTER statement specifies interval-censored data having a special structure; these data are called *readout* data. Use the NENTER statement in conjunction with the FREQ statement.

The NLOPTIONS statement enables you to control aspects of the nonlinear optimizations used for maximum likelihood estimation of the parameters of the three-parameter Weibull distribution in the ANALYZE and PROBPLOT statements, and of parametric models for recurrent events data in the MODEL statement.

The UNITID statement specifies a variable in the input data set that is used to identify each individual unit in an MCFPLOT statement.

Graphical Enhancement Statements

You can use the TITLE, FOOTNOTE, and NOTE statements to enhance printed output. If you are creating plots, you can also use the LEGEND and SYMBOL statements to enhance your plots. For details, see the SAS/GRAPH documentation and the section for the plot statement that you are using.

PROC RELIABILITY Statement

PROC RELIABILITY < *options* > **;**

The PROC RELIABILITY statement invokes the procedure. You can specify the following options.

DATA=*SAS-data-set*
specifies an input data set

GOUT=*graphics-catalog*
specifies a catalog for saving graphical output

NAMELEN=*n*
specifies the length of effect names in tables and output data sets to be *n* characters long, where *n* is a value between 20 and 200 characters. The default length is 20 characters.

ANALYZE Statement

< *label:* >**ANALYZE** *variable* < ∗*censor-variable*(*values*) > < **=***group-variables* > < / *options* > **;**

< *label:* >**ANALYZE** (*variable1 variable2*) < **=***group-variables* > < / *options* > **;**

< *label:* >**ANALYZE** *variable1*(*variable2*) < **=***group-variables* > < / *options* > **;**

You use the ANALYZE statement to estimate the parameters of the probability distribution specified in the DISTRIBUTION statement without producing any graphical output. The ANALYZE statement performs the same analysis as the PROBPLOT statement, but it does not produce any plots. In addition, you can use the ANALYZE statement to analyze data with the binomial and Poisson distributions. The third format for the preceding ANALYZE statement applies only to Poisson and binomial data. You can use any number of ANALYZE statements after a PROC RELIABILITY statement; each ANALYZE statement produces a separate analysis. You can specify an optional *label* to distinguish between multiple ANALYZE statements in the output.

You must specify one *variable*. If your data are right censored, you must specify a *censor-variable* and, in parentheses, the *values* of the *censor-variable* that correspond to censored data values.

If you are using the binomial or Poisson distributions, you must specify *variable1* to represent a binomial or Poisson count and *variable2* to provide an exposure measure for the Poisson distribution or the binomial sample size for the binomial distribution.

You can optionally specify one or two *group-variables*. The ANALYZE statement produces an analysis for each level combination of the *group-variable* values. The observations in a given level are referred to as a *cell*.

The elements of the ANALYZE statement are described as follows.

variable
: represents the data for which an analysis is to be produced. A *variable* must be a numeric variable in the input data set.

censor-variable(values)
: indicates which observations in the input data set are right censored. You specify the values of *censor-variable* that represent censored observations by placing those values in parentheses after the variable name. If your data are not right censored, then you omit the specification of *censor-variable*; otherwise, *censor-variable* must be a numeric variable in the input data set.

(variable1 variable2)
: is another method of specifying the data. You can use this syntax in a situation where uncensored, interval-censored, left-censored, and right-censored values occur in the same set of data. Table 18.31 shows how you use this syntax to specify different types of censoring by using combinations of missing and nonmissing values. See the section "Lognormal Analysis with Arbitrary Censoring" on page 1194 for an example of using this syntax to create a probability plot.

variable1
: represents the count data for which a Poisson or binomial analysis is to be produced. A *variable1* must be a numeric variable in the input data set.

variable2
: provides either an exposure measure for a Poisson analysis or a binomial number of trials for a binomial analysis. A *variable2* must be a numeric variable in the input data set.

group-variables
: are one or two group variables. If no group variables are specified, a single analysis is produced. The *group-variables* can be numeric or character variables in the input data set.

  Note that the parentheses surrounding the textitgroup-variables are needed only if two group variables are specified.

options
: control the features of the analysis. All *options* are specified after a slash (/) in the ANALYZE statement.

Summary of Options

The following tables summarize the options available in the ANALYZE statement. You can specify one or more of these options to control the parameter estimation and provide optional analyses.

Table 18.4 Analysis Options for Distributions Other Than Poisson or Binomial

Option	Option Description
CONFIDENCE=*number*	Specifies the confidence coefficient for all confidence intervals. Specify a *number* between 0 and 1. The default value is 0.95.
CONVERGE=*number*	Specifies the convergence criterion for maximum likelihood fit. See the section "Maximum Likelihood Estimation" on page 1324 for details.
CONVH=*number*	Specifies the convergence criterion for the relative Hessian convergence criterion. See the section "Maximum Likelihood Estimation" on page 1324 for details.
CORRB	Requests the parameter correlation matrix.
COVB	Requests the parameter covariance matrix.
FITTYPE | FIT= *fit-specification*	Specifies the method of estimating distribution parameters. The available *fit-specifications* and their meanings are shown in the following table.

Fit Specification	Definition
LSYX	Least squares fit to the probability plot. The probability axis is the dependent variable.
LSXY	Least squares fit to the probability plot. The lifetime axis is the dependent variable.
MLE	Maximum likelihood (default).
NONE	No fit is computed.
WEIBAYES <(CONFIDENCE | CONF=*number*)>	Weibayes fit. *number* is the confidence coefficient for the Weibayes fit and is between 0 and 1. The default is 0.95.

Table 18.4 Analysis Options for Distributions Other Than Poisson or Binomial (continued)

Option	Option Description
INEST | IN=*SAS-data-set*	Specifies a SAS data set that can contain initial values, equality constraints, upper bounds, or lower bounds for the scale, shape, and threshold parameters in a three-parameter Weibull model for lifetime data. Applies only to three-parameter Weibull models. See the section "INEST Data Set for the Three-Parameter Weibull" on page 1326 for details.
ITPRINT	Requests the iteration history for maximum likelihood fit.
ITPRINTEM	Requests the iteration history for the Turnbull algorithm.
LRCL	Requests likelihood ratio confidence intervals for distribution parameters.
LRCLPER	Requests likelihood ratio confidence intervals for distribution percentiles.
LRCLSURV	Requests likelihood ratio confidence intervals for survival and cumulative distribution functions at times specified with the SURVTIME=*number-list* option.
LOCATION=*number* < LINIT >	Specifies fixed or initial value of location parameter.
MAKEHAM=*number* < MKINIT >	Specifies the fixed or initial value of the Makeham parameter for the three-parameter Gompertz distribution.
MAXIT=*number*	Specifies the maximum number of iterations allowed for maximum likelihood fit.
MAXITEREM | MAXITEM=*number1* < ,*number2* >	*number1* specifies the maximum number of iterations allowed for Turnbull algorithm. Iteration history is printed in increments of *number2* if requested with ITPRINTEM. See the section "Interval-Censored Data" on page 1317 for details.
NOPCTILES	Suppresses computation of percentiles.
NOPOLISH	Suppresses the setting of small interval probabilities to 0 in the Turnbull algorithm. See the section "Interval-Censored Data" on page 1317 for details.

Table 18.4 Analysis Options for Distributions Other Than Poisson or Binomial (continued)

Option	Option Description
PCTLIST=*number-list*	Specifies a list of percentages for which to compute percentile estimates. *number-list* must be a list of numbers separated by blanks or commas. Each number in the list must be between 0 and 100. If this option is not specified, percentiles are computed for a standard list of percentages.
PPOS=*plotting-position*	Specifies the *plotting-position* type used to compute nonparametric estimates of the probability distribution function. See the section "Probability Plotting" on page 1313 for details. The available *plotting-position* types are shown in the following table.

Plotting Position	Type
EXPRANK	Expected ranks
MEDRANK	Median ranks
MEDRANK1	Median ranks (exact formula)
KM	Kaplan-Meier
MKM	Modified Kaplan-Meier (default)
NA | NELSONAALEN	Nelson-Aalen

PPOUT	Requests a table of cumulative probabilities.
PRINTPROBS	Print intervals and associated probabilities for the Turnbull algorithm.
PROBLIST=*number-list*	Specifies a list of initial values for Turnbull algorithm. See the section "Interval-Censored Data" on page 1317 for details.
PSTABLE=*number*	Specifies stable parameterization. The *number* must be between 0 and 1. See the section "Stable Parameters" on page 1329 for further information.
READOUT	Analyzes readout data.
SCALE=*number* < SCINIT >	Specifies the fixed or initial value of scale parameter.
SHAPE=*number* < SHINIT >	Specifies the fixed or initial value of shape parameter.
SINGULAR=*number*	Specifies the singularity criterion for matrix inversion.

Table 18.4 Analysis Options for Distributions Other Than Poisson or Binomial (continued)

Option	Option Description
SURVTIME=*number-list*	Requests that the survival function, cumulative distribution function, and confidence limits be computed for values in *number-list*. See the section "Reliability Function" on page 1334 for details.
THRESHOLD=*number*	Specifies a fixed threshold parameter. See Table 18.57 for the distributions with a threshold parameter.
TOLLIKE=*number*	Specifies the criterion for convergence in the Turnbull algorithm. The default is 10^{-8}. See the section "Interval-Censored Data" on page 1317 for details.
TOLPROB=*number*	Specifies the criterion for setting interval probability to 0 in the Turnbull algorithm. Default is 10^{-6}. See the section "Interval-Censored Data" on page 1317 for details.
WALDCL | NORMALCL	Requests Wald type confidence intervals for distribution parameters. See Table 18.68 and Table 18.74 for details about the computation of Wald confidence intervals. Wald confidence intervals are provided by default, but this option can be combined with LRCL to obtain both types of intervals.

Table 18.5 Analysis Options for Poisson and Binomial Distributions

Option	Option Description
CONFIDENCE=*number*	Specifies the confidence coefficient for all confidence intervals. Specify a *number* between 0 and 1. The default value is 0.95.
PREDICT(*number*)	Requests predicted counts for exposure *number* for Poisson or sample size *number* for binomial.
TOLERANCE(*number*)	Requests exposure for Poisson or sample size for binomial to estimate Poisson rate or binomial probability within *number* with probability given by the CONFIDENCE= option.

BY Statement

BY *variables* **;**

You can specify a BY statement with PROC RELIABILITY to obtain separate analyses of observations in groups that are defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. If you specify more than one BY statement, only the last one specified is used.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data by using the SORT procedure with a similar BY statement.
- Specify the NOTSORTED or DESCENDING option in the BY statement for the RELIABILITY procedure. The NOTSORTED option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the BY variables by using the DATASETS procedure (in Base SAS software).

For more information about BY-group processing, see the discussion in *SAS Language Reference: Concepts*. For more information about the DATASETS procedure, see the discussion in the *Base SAS Procedures Guide*.

CLASS Statement

CLASS *variable-names*< / *options* > **;**

The CLASS statement specifies variables in the input data set that serve as *indicator*, *dummy*, or *classification* variables in the MODEL statement. If a CLASS variable is specified as an independent variable in the MODEL statement, the RELIABILITY procedure automatically generates an indicator variable for each level of the CLASS variable. The indicator variables generated are used as independent variables in the regression model specified in the MODEL statement. An indicator variable for a level of a CLASS variable is a variable equal to 1 for those observations corresponding to the level and equal to 0 for all other observations.

You can specify the following option in the CLASS statement.

TRUNCATE< =*n* >
specifies the length *n* of CLASS variable values to use in determining CLASS variable levels. If you specify TRUNCATE without the length *n*, the first 16 characters of the formatted values are used. When formatted values are longer than 16 characters, you can use this option to revert to the levels as determined in releases before SAS 9. The default is to use the full formatted length of the CLASS variable.

DISTRIBUTION Statement

DISTRIBUTION *probability distribution-name* **;**

The ANALYZE, PROBPLOT, RELATIONPLOT, and MODEL statements require you to specify the probability distribution that describes your data. You can specify a probability distribution for lifetime data by using the DISTRIBUTION statement anywhere after the PROC RELIABILITY statement and before the RUN statement. If you do not specify a distribution in a DISTRIBUTION statement, the normal distribution is assumed. In addition, you can specify a parametric non-homogeneous Poisson process model for recurrent events data in a DISTRIBUTION statement. The probability distribution for lifetime data or the model for recurrent events data specified determines the distribution for which parameters are estimated from your data. The valid distributions and the statements to which they apply are shown in Table 18.6 and Table 18.7.

Table 18.6 Probability Distributions for Lifetime Data

Distribution	Distribution-Name Specified	Statement
Binomial	BINOMIAL	ANALYZE
Exponential	EXPONENTIAL	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
Extreme value	EXTREME | EV	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
Generalized gamma	GAMMA	MODEL
Gompertz	GOMPERTZ | G2	ANALYZE, PROBPLOT
Three-parameter Gompertz	GOMPERTZ3 | G3	ANALYZE, PROBPLOT
Logistic	LOGISTIC | LOGIT	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
Log-logistic	LLOGISTIC | LLOGIT	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
Lognormal (base *e*)	LOGNORMAL | LNORM	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
Lognormal (base 10)	LOGNORMAL10 | LNORM10	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
Normal	NORMAL	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
Poisson	POISSON	ANALYZE
Weibull	WEIBULL | W2	ANALYZE, PROBPLOT, RELATIONPLOT, MODEL
Three-parameter Weibull	WEIBULL3 | W3	ANALYZE, PROBPLOT

Table 18.7 Poisson Process Models for Recurrence Data

NHPP Model	Distribution-Name Specified	Statement
Homogeneous	HPP	MODEL
Crow-AMSAA	NHPP(CA) | NHPP(CROWAMSAA)	MODEL
Log-linear	NHPP(LOG) | NHPP(LOGLINEAR)	MODEL
Power	NHPP(POW) | NHPP(POWER)	MODEL
Proportional intensity	NHPP(PROP) | NHPP(PROPORTIONAL)	MODEL

EFFECTPLOT Statement

EFFECTPLOT < *plot-type* < *(plot-definition-options)* > > < / *options* > **;**

The EFFECTPLOT statement produces a display of the fitted model and provides options for changing and enhancing the displays. Table 18.8 describes the available *plot-types* and their *plot-definition-options*.

Table 18.8 *Plot-Types* and *Plot-Definition-Options*

Plot-Type and Description	Plot-Definition-Options
BOX Displays a box plot of continuous response data at each level of a CLASS effect, with predicted values superimposed and connected by a line. This is an alternative to the INTERACTION *plot-type*.	PLOTBY= variable or CLASS effect X= CLASS variable or effect
CONTOUR Displays a contour plot of predicted values against two continuous covariates.	PLOTBY= variable or CLASS effect X= continuous variable Y= continuous variable
FIT Displays a curve of predicted values versus a continuous variable.	PLOTBY= variable or CLASS effect X= continuous variable
INTERACTION Displays a plot of predicted values (possibly with error bars) versus the levels of a CLASS effect. The predicted values are connected with lines and can be grouped by the levels of another CLASS effect.	PLOTBY= variable or CLASS effect SLICEBY= variable or CLASS effect X= CLASS variable or effect
MOSAIC Displays a mosaic plot of predicted values using up to three CLASS effects.	PLOTBY= variable or CLASS effect X= CLASS effects
SLICEFIT Displays a curve of predicted values versus a continuous variable grouped by the levels of a CLASS effect.	PLOTBY= variable or CLASS effect SLICEBY= variable or CLASS effect X= continuous variable

For full details about the syntax and options of the EFFECTPLOT statement, see the section "EFFECTPLOT Statement" (Chapter 19, *SAS/STAT User's Guide*) in Chapter 19, "Shared Concepts and Topics" (*SAS/STAT User's Guide*).

ESTIMATE Statement

ESTIMATE <'*label*'> *estimate-specification* <(*divisor*=*n*)>
<, . . . <'*label*'> *estimate-specification* <(*divisor*=*n*)>>
</ *options*> ;

The ESTIMATE statement provides a mechanism for obtaining custom hypothesis tests. Estimates are formed as linear estimable functions of the form $\mathbf{L}\boldsymbol{\beta}$. You can perform hypothesis tests for the estimable functions, construct confidence limits, and obtain specific nonlinear transformations.

Table 18.9 summarizes the *options* available in the ESTIMATE statement.

Table 18.9 ESTIMATE Statement Options

Option	Description
Construction and Computation of Estimable Functions	
DIVISOR=	Specifies a list of values to divide the coefficients
NOFILL	Suppresses the automatic fill-in of coefficients for higher-order effects
SINGULAR=	Tunes the estimability checking difference
Degrees of Freedom and *p*-values	
ADJUST=	Determines the method for multiple comparison adjustment of estimates
ALPHA=α	Determines the confidence level $(1-\alpha)$
LOWER	Performs one-sided, lower-tailed inference
STEPDOWN	Adjusts multiplicity-corrected p-values further in a step-down fashion
TESTVALUE=	Specifies values under the null hypothesis for tests
UPPER	Performs one-sided, upper-tailed inference
Statistical Output	
CL	Constructs confidence limits
CORR	Displays the correlation matrix of estimates
COV	Displays the covariance matrix of estimates
E	Prints the **L** matrix
JOINT	Produces a joint F or chi-square test for the estimable functions
PLOTS=	Requests ODS statistical graphics if the analysis is sampling-based
SEED=	Specifies the seed for computations that depend on random numbers
Generalized Linear Modeling	
CATEGORY=	Specifies how to construct estimable functions with multinomial data

Table 18.9 *continued*

Option	Description
EXP	Exponentiates and displays estimates
ILINK	Computes and displays estimates and standard errors on the inverse linked scale

For details about the syntax of the ESTIMATE statement, see the section "ESTIMATE Statement" (Chapter 19, *SAS/STAT User's Guide*) in Chapter 19, "Shared Concepts and Topics" (*SAS/STAT User's Guide*).

FMODE Statement

FMODE keyword=*variable ('value1' . . . 'valuen')* < / *options* > ;

Use the FMODE statement with data that have failures attributable to multiple causes (*failure modes*). You can analyze data by either keeping, eliminating, or combining specific failure modes with the FMODE statement. Use this statement with the KEEP or ELIMINATE keyword in conjunction with the ANALYZE, MODEL, PROBPLOT, or RELATIONPLOT statement. Use this statement with the COMBINE keyword with the ANALYZE or PROBPLOT statement. You can place an FMODE statement anywhere after the PROC RELIABILITY statement and before the RUN statement.

If you specify the keyword KEEP, the life distribution for only the identified failure modes is estimated, with all other failure modes treated as right-censored data. If you specify the keyword ELIMINATE, the life distribution that results if the failure modes identified are completely eliminated is estimated. The keyword ELIMINATE causes the failure modes identified to be treated as right-censored data and causes a single life distribution to be estimated for the remaining data. If you specify the keyword COMBINE, the data are analyzed with all the specified failure modes combined acting. See the section "Weibull Probability Plot for Two Combined Failure Modes" on page 1211 for an example of a Weibull plot of data with two combined failure modes. The failure mode for an observation in the input data set is identified by the value of *variable*, where *variable* is any numeric or character variable in the input data set. You must identify a failure mode for each observation that is not right-censored. You specify failure modes to keep, eliminate, or combine by listing variable values *(value1 . . . valuen)* in parentheses after the failure mode variable name. The list of variable values must have entries separated by blanks or commas. You can specify the following *options* after the slash (/). These options will affect the analysis only when you use the COMBINE keyword.

Table 18.10 FMODE Statement Options

Option	Description
LEGEND=	Specifies a LEGEND statement for individual mode fit lines.
NOLEGEND	Suppresses legend for individual mode fit lines.
PLOTMODES	Plots individual failure distribution lines on probability plot.

FREQ Statement

FREQ *variable-name* **;**

The FREQ statement specifies a variable that provides frequency counts for each observation in the input data set. If *n* is the value of the FREQ variable in the input data set for an observation, then that observation is weighted by *n*. The log-likelihood function for maximum likelihood estimation is multiplied by *n*. If *n* is not an integer, the integer part of *n* is used in creating probability plots.

You can also use the FREQ statement in conjunction with the NENTER statement to specify interval-censored data having a special structure; these data are called *readout* data. The FREQ statement specifies a variable in the input data set that determines the number of units failing in each interval. See the section "Weibull Analysis of Interval Data with Common Inspection Schedule" on page 1189 for an example that uses the FREQ statement with readout data.

You can also use the FREQ statement in conjunction with the NENTER statement to specify recurrent events data when the event times are grouped into intervals, rather than being observed exactly. The FREQ statement specifies a variable in the input data set that determines the number of events in each interval.

You can use the FREQ statement with the MCFPLOT and MODEL statements for exact age data to provide frequency counts for entire recurrence histories. If *n* is the value of the FREQ variable at a censor time, the history of recurrences for the corresponding system is replicated independently *n* times. Values of the FREQ variable at times other than censor times are not used; they can be any value or missing without affecting the analysis.

INSET Statement

INSET *keyword-list* < / *options* > **;**

The box or table of summary information produced on plots made with the PROBPLOT or MCFPLOT statement is called an *inset*. You can use the INSET statement to customize the appearance of the inset box and the information that is printed in the inset box. To supply the information that is displayed in the inset box, you specify a *keyword* that corresponds to the information you want shown. For example, the following statements produce a Weibull plot with the sample size, the number of failures, and the Weibull mean displayed in the inset:

```
proc reliability data=fan;
   distribution Weibull;
   pplot lifetime*censor(1);
   inset n nfail weibull(mean);
run;
```

By default, inset entries are identified with appropriate labels. However, you can provide a customized label by specifying the *keyword* for that entry followed by the equal sign (=) and the label in quotes. For example, the following INSET statement produces an inset that contains the sample size and Weibull mean, labeled "Sample Size" and "Weibull Mean" in the inset:

```
inset n='Sample Size' weibull(mean='Weibull Mean');
```

If you specify a keyword that does not apply to the plot you are creating, then the keyword is ignored.

The *options* control the appearance of the box.

If you specify more than one INSET statement, only the last one is used.

Keywords Used in the INSET Statement

Table 18.11 through Table 18.13 list keywords available in the INSET statement to display summary statistics, distribution parameters, and distribution fitting information.

Table 18.11 Summary Statistics

Keyword	Description
N	Sample size
NFAIL	Number of failures for probability plots.
NEVENTS	Number of events or repairs for MCF plots.
NEVENTS1	Number of events or repairs in the first group for MCF difference plots.
NEVENTS2	Number of events or repairs in the second group for MCF difference plots.
NUNITS	Number of units or systems for MCF plots.
NUNITS1	Number of units or systems in the first group for MCF difference plots.
NUNITS2	Number of units or systems in the second group for MCF difference plots.

Table 18.12 General Information

Keyword	Description
CONFIDENCE	Confidence coefficient for all confidence intervals or for the Weibayes fit
FIT	Method used to estimate distribution parameters for probability plots
RSQUARE	R square for least squares distribution fit to probability plots

Distribution parameters are specified as *distribution-name*(*distribution-parameters*). The following table lists the keywords available.

Table 18.13 Distribution Parameters

Keyword	Secondary Keyword	Description
EXPONENTIAL	SCALE	Scale parameter
	THRESHOLD	Threshold parameter
	MEAN	Expected value
EXTREME | EV	LOCATION	Location parameter
	SCALE	Scale parameter
	MEAN	Expected value
GOMPERTZ | GOMP	SCALE	Scale parameter
	SHAPE	Shape parameter
	MEAN	Expected value
GOMPERTZ3 | GOMP3	SCALE	Scale parameter
	SHAPE	Shape parameter
	MAKEHAM	Makeham mortality component
	MEAN	Expected value
LOGISTIC | LOGIT	LOCATION	Location parameter
	SCALE	Scale parameter
	MEAN	Expected value
LOGLOGISTIC | LLOGIT	LOCATION	Location parameter
	SCALE	Scale parameter
	THRESHOLD	Threshold parameter
	MEAN	Expected value
LOGNORMAL	LOCATION	Location parameter
	SCALE	Scale parameter
	THRESHOLD	Threshold parameter
	MEAN	Expected value
LOGNORMAL10	LOCATION	Location parameter
	SCALE	Scale parameter
	THRESHOLD	Threshold parameter
	MEAN	Expected value
NORMAL	LOCATION	Location parameter
	SCALE	Scale parameter
	MEAN	Expected value
WEIBULL	SCALE	Scale parameter
	SHAPE	Shape parameter
	THRESHOLD	Threshold parameter
	MEAN	Expected value

Options Used in the INSET Statement

Table 18.14 through Table 18.17 list INSET statement options that control the appearance of the inset box.

Table 18.14 lists options that control the appearance of the box when you use traditional graphics.

Table 18.14 General Appearance Options (Traditional Graphics)

Option	Option Description
HEADER=*'quoted-string'*	Specifies text for header or box title.
NOFRAME	Omits frame around box.
POS=*value* < DATA | PERCENT >	Determines the position of the inset. The *value* can be a compass point (N, NE, E, SE, S, SW, W, NW) or a pair of coordinates (x, y) enclosed in parentheses. The coordinates can be specified in axis percent units or axis data units.
REFPOINT=*name*	Specifies the reference point for an inset that is positioned by a pair of coordinates with the POS= option. You use the REFPOINT= option in conjunction with the POS= coordinates. *name* specifies which corner of the inset frame you have specified with coordinates (x, y); it can take the value of BR (bottom right), BL (bottom left), TR (top right), or TL (top left). The default is REFPOINT=BL. If the inset position is specified as a compass point, then the REFPOINT= option is ignored.

Table 18.15 lists options that control the appearance of the box when you use ODS Graphics.

Table 18.15 General Appearance Options (ODS Graphics)

Option	Option Description
HEADER=*'quoted-string'*	Specifies text for header or box title.
NOFRAME	Omits frame around box.
POS=*value*	Determines the position of the inset. The *value* can be a compass point (N, NE, E, SE, S, SW, W, NW).

Table 18.16 lists options that control the appearance of the text within the box when you use traditional graphics. These options are not available if ODS Graphics is enabled.

Table 18.16 Text Enhancement Options (Traditional Graphics)

Option	Option Description
FONT=*font*	Software font for text
HEIGHT=*value*	Height of text

Table 18.17 lists options that control the colors and patterns used in the box when you use traditional graphics. These options are not available if ODS Graphics is enabled.

Table 18.17 Color and Pattern Options (Traditional Graphics)

Option	Option Description
CFILL=*color*	Color for filling box
CFILLH=*color*	Color for filling box header
CFRAME=*color*	Color for frame
CHEADER=*color*	Color for text in header
CTEXT=*color*	Color for text

LOGSCALE Statement

LOGSCALE *effect-list* < */options* > **;**

You use the LOGSCALE statement to model the logarithm of the distribution scale parameter as a function of explanatory variables. A MODEL statement must also be present to specify the model for the distribution location parameter. *effect-list* is a list of variables in the input data set representing the values of the independent variables in the model for each observation, and combinations of variables representing interaction terms. It can contain any variables or combination of variables in the input data set. It can contain the same variables as the MODEL statement, or it can contain different variables. The variables in the *effect-list* can be any combination of indicator variables named in a CLASS statement and continuous variables. The coefficients of the explanatory variables are estimated by maximum likelihood.

Table 18.18 lists the *options* available for the LOGSCALE statement.

Table 18.18 LOGSCALE Statement Options

Option	Option Description
INITIAL=*number-list*	Specifies initial values for log-scale regression parameters other than the intercept term.
INTERCEPT=*number* < INTINIT >	Specifies an initial or fixed value of the intercept parameter, depending on whether INTINIT is present.

LSMEANS Statement

LSMEANS < *model-effects* > < / *options* > **;**

The LSMEANS statement computes and compares least squares means (LS-means) of fixed effects. LS-means are *predicted population margins*—that is, they estimate the marginal means over a balanced population. In a sense, LS-means are to unbalanced designs as class and subclass arithmetic means are to balanced designs.

Table 18.19 summarizes the *options* available in the LSMEANS statement.

Table 18.19 LSMEANS Statement Options

Option	Description
Construction and Computation of LS-Means	
AT	Modifies the covariate value in computing LS-means
BYLEVEL	Computes separate margins
DIFF	Requests differences of LS-means
OM=	Specifies the weighting scheme for LS-means computation as determined by the input data set
SINGULAR=	Tunes estimability checking
Degrees of Freedom and *p*-values	
ADJUST=	Determines the method for multiple-comparison adjustment of LS-means differences
ALPHA=α	Determines the confidence level $(1-\alpha)$
STEPDOWN	Adjusts multiple-comparison *p*-values further in a step-down fashion
Statistical Output	
CL	Constructs confidence limits for means and mean differences
CORR	Displays the correlation matrix of LS-means
COV	Displays the covariance matrix of LS-means
E	Prints the **L** matrix
LINES	Produces a "Lines" display for pairwise LS-means differences
MEANS	Prints the LS-means
PLOTS=	Requests graphs of means and mean comparisons
SEED=	Specifies the seed for computations that depend on random numbers
Generalized Linear Modeling	
EXP	Exponentiates and displays estimates of LS-means or LS-means differences
ILINK	Computes and displays estimates and standard errors of LS-means (but not differences) on the inverse linked scale
ODDSRATIO	Reports (simple) differences of least squares means in terms of odds ratios if permitted by the link function

For details about the syntax of the LSMEANS statement, see the section "LSMEANS Statement" (Chapter 19, *SAS/STAT User's Guide*).

LSMESTIMATE Statement

LSMESTIMATE *model-effect* < '*label*' > *values* < *divisor*=*n* >
< , . . . < '*label*' > *values* < *divisor*=*n* > >
< / *options* > ;

The LSMESTIMATE statement provides a mechanism for obtaining custom hypothesis tests among least squares means.

Table 18.20 summarizes the *options* available in the LSMESTIMATE statement.

Table 18.20 LSMESTIMATE Statement Options

Option	Description
Construction and Computation of LS-Means	
AT	Modifies covariate values in computing LS-means
BYLEVEL	Computes separate margins
DIVISOR=	Specifies a list of values to divide the coefficients
OM=	Specifies the weighting scheme for LS-means computation as determined by a data set
SINGULAR=	Tunes estimability checking
Degrees of Freedom and *p*-values	
ADJUST=	Determines the method for multiple-comparison adjustment of LS-means differences
ALPHA=α	Determines the confidence level $(1-\alpha)$
LOWER	Performs one-sided, lower-tailed inference
STEPDOWN	Adjusts multiple-comparison p-values further in a step-down fashion
TESTVALUE=	Specifies values under the null hypothesis for tests
UPPER	Performs one-sided, upper-tailed inference
Statistical Output	
CL	Constructs confidence limits for means and mean differences
CORR	Displays the correlation matrix of LS-means
COV	Displays the covariance matrix of LS-means
E	Prints the **L** matrix
ELSM	Prints the **K** matrix
JOINT	Produces a joint F or chi-square test for the LS-means and LS-means differences
PLOTS=	Requests graphs of means and mean comparisons
SEED=	Specifies the seed for computations that depend on random numbers
Generalized Linear Modeling	
CATEGORY=	Specifies how to construct estimable functions with multinomial data
EXP	Exponentiates and displays LS-means estimates

Table 18.20 *continued*

Option	Description
ILINK	Computes and displays estimates and standard errors of LS-means (but not differences) on the inverse linked scale

For details about the syntax of the LSMESTIMATE statement, see the section "LSMESTIMATE Statement" (Chapter 19, *SAS/STAT User's Guide*) in Chapter 19, "Shared Concepts and Topics" (*SAS/STAT User's Guide*).

MAKE Statement

MAKE '*table*' **OUT=***SAS-data-set*< *SAS-data-set options* > **;**

The MAKE statement creates a SAS data set from any of the tables produced by the RELIABILITY procedure. You can specify SAS data set options in parentheses after the data set name. You can specify one MAKE statement for each table that you want to save to a SAS data set.

The ODS statement also creates SAS data sets from tables, in addition to providing an extensive and flexible method of controlling output created by the RELIABILITY procedure. The ODS statement is the recommended method of controlling procedure output; however, the MAKE statement is provided for compatibility with earlier releases of the SAS System.

The valid values for *table* are shown in the section "ODS Table Names" on page 1358, organized by the RELIABILITY procedure statement that produces the tabular output. The *table* names are not case sensitive, but they must be enclosed in single quotes.

MCFPLOT Statement

< *label:* >**MCFPLOT** *variable*∗*cost/censor-variable*(*values*) < **=***group-variables* > < / *options* > **;**

< *label:* >**MCFPLOT** (< **INTERVAL=** >*variable1 variable2* < **RECURRENCES=** > *variable3* < **CENSOR=** > *variable4*) < **=***group-variables* > < / *options* > **;**

< *label:* >**MCFPLOT** (*variable1 variable2*) < **=***group-variables* > < / *options* > **;**

You can specify any number of MCFPLOT statements after a PROC RELIABILITY statement. Each MCFPLOT statement creates a separate MCF plot and associated analysis. See the section "Analysis of Recurrence Data on Repairs" on page 1215, the section "Comparison of Two Samples of Repair Data" on page 1220, and the section "Analysis of Interval Age Recurrence Data" on page 1227 for examples that use the MCFPLOT statement. You can specify an optional *label* to distinguish between multiple MCFPLOT statements in the output.

To create a plot of the mean cumulative function for cost or number of repairs with exact age data, you specify a *variable* that represents the times of repairs. You must also specify a *cost/censor-variable* and the *values*, in parentheses, of the *cost/censor-variable* that correspond to end-of-history data values (also referred to as *censored* data values).

To create a plot of the mean cumulative function for cost or number of repairs with interval age data, you specify *variable1 variable2* that represents the age intervals. You must also specify either *variable3* that represents the number of recurrences in the intervals and *variable4* that represents the number censored in the intervals, or a FREQ statement that represents the number of recurrent events in the intervals and a NENTER statement that represents the number of units observed in the intervals.

You can optionally specify one or two *group-variables* (also referred to as *classification variables*). The MCFPLOT statement displays a component plot for each level of the *group-variables*. The observations in a given level are called a *cell*.

For exact data, you must also specify a *unit-identification* variable in conjunction with the MCFPLOT statement to identify the individual unit name for each instance of repair or end of history on the unit. Specify the *unit-identification* variable in the UNITID statement.

Add the EVENTPLOT option to any MCFPLOT statement to obtain a horizontal plot of failure and censoring times for each system.

The elements of the MCFPLOT statement are described as follows.

variable
: represents the time of repair. A *variable* must be a numeric variable in the input data set.

variable1 variable2
: represents time intervals for grouped data. *variable1* and *variable2* must be numeric variables in the input data set.

variable3
: represents the number of recurrences in an interval. A *variable3* must be a numeric variable in the input data set.

variable4
: represents the number censored in an interval. A *variable4* must be a numeric variable in the input data set.

cost/censor-variable(values)
: indicates the cost of each repair or the number of repairs. This variable also indicates which observations in the input data set are end-of-history (censored) data points. You specify the values of *cost/censor-variable* that represent censored observations by placing those values in parentheses after the variable name. A *cost/censor-variable* must be a numeric variable in the input data set.

group-variables
: are one or two group variables. If no group variables are specified, a single plot is produced. The *group-variables* can be any numeric or character variables in the input data set. For exact data, if a single group variable is specified, and the group variable has two levels, then a statistical test for equality of the groups represented by the two levels is computed and displayed in the "Tests for Equality of Mean Functions" table. Refer to "Comparison of Two Groups of Recurrent Events Data" on page 1351 for more details.

  Note that the parentheses surrounding the *group-variables* are needed only if two group variables are specified.

options

control the features of the mean cumulative function plot. All *options* are specified after a slash (/) in the MCFPLOT statement. The "Summary of Options" section, which follows, lists all options by function.

Summary of Options

Table 18.21 lists available analysis options.

Table 18.21 Analysis Options

Option	Option Description
CONFIDENCE=*number*	Specifies the confidence coefficient for all confidence intervals. Specify a *number* between 0 and 1. The default value is 0.95.
EVENTPLOT <(SORT=*sort-order*)>	Specifies a separate horizontal plot of failure and censoring times for each system. The following sort orders are available:

Sort Order	Definition
ASCENDINGTIME	sorts by increasing censoring times (default)
DESCENDINGTIME	sorts by decreasing censoring times
ASCENDINGFORMATTED	sorts alphabetically by system name or label
DESCENDINGFORMATTED	sorts alphabetically in reverse by system name or label

INDINC	Requests variance estimates of the MCF using the Nelson (2002) estimator under the independent increments assumption.
LOGINTERVALS	Requests that confidence intervals be computed based on the asymptotic normality of log(MCF). This is appropriate only when the MCF estimate is positive, so does not apply to MCF differences or when negative costs are specified.
MCFDIFF	Requests a plot of differences of MCFS of two groups specified by a single group variable.
NOVARIANCE	Suppresses MCF variance computation.
VARIANCE=*variance-specification*	Specifies the method of variance calculation. The following methods are available.

Table 18.21 Analysis Options (continued)

Option	Option Description
Method	**Definition**
INDINC	The method of Nelson (2002) assuming independent increments
LAWLESS | VARMETHOD2	The method of Lawless and Nadeau (1995) (the default method)
NELSON	The method of Nelson (2002)
POISSON	Poisson process method

Table 18.22 lists plot layout options that are available when you use traditional graphics.

Table 18.22 Plot Layout Options for Traditional Graphics

Option	Option Description
CENBIN	Plots censored data as frequency counts rather than as individual points.
CENSYMBOL=*symbol* | (*symbol-list*)	Specifies symbols for censored values. *symbol* is one of the symbol names (plus, star, square, diamond, triangle, hash, paw, point, dot, circle) or a letter (A–Z). If you are creating overlaid plots for groups of data, you can specify different symbols for the groups with a list of symbols or letters, separated by blanks, enclosed in parentheses. If no CENSYMBOL option is specified, the symbol used for censored values is the same as for repairs.
HOFFSET=*value*	Specifies offset for horizontal axis.
INBORDER	Requests a border around MCF plots.
INTERPOLATE=JOIN | STEP | NONE	Requests that symbols in an MCF plot be connected with a straight line, step function, or not connected.
INTERTILE=*value*	Specifies the distance between tiles.
MCFLEGEND=*legend-statement-name* | NONE	Identifies a legend statement to specify legend for overlaid MCF plots.
MISSING1	Requests that missing values of first GROUP= variable be treated as a level of the variable.
MISSING2	Requests that missing values of second GROUP= variable be treated as a level of the variable.

Table 18.22 Plot Layout Options (continued)

Option	Option Description
NCOLS=*n*	Specifies the number of columns plotted on a page.
NOCENPLOT	Suppresses plotting of censored data points.
NOCONF	Suppresses plotting of confidence intervals.
NOFRAME	Suppresses the frame around the plotting area.
NOINSET	Suppresses the inset.
NOLEGEND	Suppresses the legend for overlaid MCF plots.
NROWS=*n*	Specifies the number of rows plotted on a page.
ORDER1=DATA | FORMATTED | FREQ | INTERNAL	Specifies display order for values of the first GROUP= variable.
ORDER2=DATA | FORMATTED | FREQ | INTERNAL	Specifies display order for values of the second GROUP= variable.
OVERLAY	Requests that plots with group variables be overlaid on a single page.
PLOTSYMBOL=*symbol* | (*symbol-list*)	Specifies symbols that represent events in an MCF plot.
PLOTCOLOR=*color* | (*color-list*)	Specifies colors of symbols that represent events in an MCF plot.
TURNVLABELS	Vertically strings out characters in labels for vertical axis.
VOFFSET=*value*	Specifies length of offset at upper end of vertical axis.

Table 18.23 lists plot layout options available when you use ODS Graphics.

Table 18.23 Plot Layout Options for ODS Graphics

Option	Option Description
DUANE	Requests that a Duane plot be created in addition to an MCF plot. If you specify the FIT=MODEL option, the fitted parametric model is included in the Duane plot. See the section "Duane Plots" on page 1357 for a description of Duane plots.

Table 18.23 Plot Layout Options (continued)

Option	Option Description
FIT=MODEL	Requests that a parametric cumulative mean function fit with a MODEL statement be plotted on the same plot with nonparametric estimates of the MCF. This option is valid only if the response specification in the MCFPLOT statement matches the response specification in the MODEL statement. If this option is specified, the fit parametric intensity function is plotted on a separate graph.
INTERPOLATE=JOIN | STEP | NONE	Requests that symbols in an MCF plot be connected with a straight line, step function, or not connected.
MISSING1	Requests that missing values of first GROUP= variable be treated as a level of the variable.
MISSING2	Requests that missing values of second GROUP= variable be treated as a level of the variable.
NCOLS=*n*	Specifies the number of columns plotted on a page.
NOCENPLOT	Suppresses plotting of censored data points.
NOCONF	Suppresses plotting of confidence intervals.
NOINSET	Suppresses the inset.
NROWS=*n*	Specifies the number of rows plotted on a page.
ORDER1=DATA | FORMATTED | FREQ | INTERNAL	Specifies the display order for values of the first GROUP= variable.
ORDER2=DATA | FORMATTED | FREQ | INTERNAL	Specifies the display order for values of the second GROUP= variable.
OVERLAY	Requests that plots with group variables be overlaid on a single page.

Table 18.24 lists reference line options that are available when you use traditional graphics.

Table 18.24 Reference Line Options for Traditional Graphics

Option	Option Description
HREF=*value-list*	Specifies reference lines perpendicular to horizontal axis.
HREFLABELS=(*'label1' ... 'labeln'*)	Specifies labels for HREF= lines.

Table 18.24 Reference Line Options (continued)

Option	Option Description
HREFLABPOS=*n*	Specifies the vertical position of labels for HREF= lines. The valid values for *n* and the corresponding label placements are shown in the following table: *n* **Label Placement** 1 Top 2 Staggered from top 3 Bottom 4 Staggered from bottom 5 Alternating from top 6 Alternating from bottom
LHREF=*linetype*	Specifies the line style for HREF= lines.
LVREF=*linetype*	Specifies the line style for VREF= lines.
VREF=*value-list*	Specifies reference lines perpendicular to vertical axis.
VREFLABELS=(*'label1' ... 'labeln'*)	Specifies labels for VREF= lines.
VREFLABPOS=*n*	Specifies the horizontal position of labels for VREF= lines. The valid values for *n* and the corresponding label placements are shown in the following table: *n* **Label Placement** 1 Left 2 Right

Table 18.25 lists reference line options that are available when you use ODS Graphics.

Table 18.25 Reference Line Options for ODS Graphics

Option	Option Description
LREF<(INTERSECT)>=*value-list*	Specifies reference lines perpendicular to the lifetime axis. If (INTERSECT) is specified, a second reference line is drawn perpendicular to the MCF axis and intersects the fit line at the same point as the lifetime axis reference line. If a lifetime axis reference line label is specified, the intersecting MCF axis reference line is labeled with the MCF axis value.
LREFLABELS=(*'label1' ... 'labeln'*)	Specifies labels for LREF= lines.

Table 18.25 Reference Line Options (continued)

Option	Option Description
MREF< (INTERSECT) >=*value-list*	Specifies reference lines perpendicular to the MCF axis. If (INTERSECT) is specified, a second reference line is drawn perpendicular to the lifetime axis and intersects the fit line at the same point as the MCF axis reference line. If an MCF axis reference line label is specified, the intersecting lifetime axis reference line is labeled with the lifetime axis value.
MREFLABELS=(*'label1' . . . 'labeln'*)	Specifies labels for MREF= lines.

Table 18.26 lists the options that control the appearance of the text when you use traditional graphics. These options are not available if ODS Graphics is enabled.

Table 18.26 Text Enhancement Options

Option	Option Description
FONT=*font*	Software font for text.
HEIGHT=*value*	Height of text used outside framed areas.
INFONT=*font*	Software font for text inside framed areas.
INHEIGHT=*value*	Height of text inside framed areas.

Table 18.27 lists options to control the appearance of the axes when you use traditional graphics.

Table 18.27 Axis Options for Traditional Graphics

Option	Option Description

HAXIS=*value1* TO *value2*< BY *value3* >	Specifies tick mark values for the horizontal axis. *value1*, *value2*, and *value3* must be numeric, and *value1* must be less than *value2*. The lower tick mark is *value1*. Tick marks are drawn at increments of *value3*. The last tick mark is the greatest value that does not exceed *value2*. If *value3* is omitted, a value of 1 is used. This method of specification of tick marks is not valid for logarithmic axes. Examples of HAXIS= lists follow:
	`haxis = 0 to 10`
	`haxis = 2 to 10 by 2`
	`haxis = 0 to 200 by 10`

Table 18.27 Axis Options (continued)

Option	Option Description
HLOWER=*number*	Specifies the lower limit on the horizontal axis scale. The HLOWER= option specifies *number* as the lower horizontal axis tick mark. The tick mark interval and the upper axis limit are determined automatically. This option has no effect if the HAXIS option is used.
HUPPER=*number*	Specifies the upper limit on the horizontal axis scale. The HUPPER= option specifies *number* as the upper horizontal axis tick mark. The tick mark interval and the lower axis limit are determined automatically. This option has no effect if the HAXIS= option is used.
LGRID=*number*	Specifies a line style for all grid lines. *number* is between 1 and 46 and specifies a linestyle for grids.
LOGLOG	Requests log scales on both axes.
MINORLOGGRID	Adds a minor grid for log axes.
NOGRID	Suppresses grid lines.
NOHLABEL	Suppresses label for horizontal axis.
NOVLABEL	Suppresses label for vertical axis.
NOVTICK	Suppresses tick marks and tick mark labels for vertical axis.
NOHTICK	Suppresses tick marks and tick mark labels for horizontal axis.
NHTICK=*number*	Specifies the number of tick intervals for the horizontal axis. This option has no effect if the HAXIS= option is used.
NVTICK=*number*	Specifies the number of tick intervals for the vertical axis. This option has no effect if the VAXIS= option is used.

Table 18.27 Axis Options (continued)

Option	Option Description

VAXIS=*value1* TO *value2*< BY *value3* >	Specifies tick mark values for the vertical axis. *value1*, *value2*, and *value3* must be numeric, and *value1* must be less than *value2*. The lower tick mark is *value1*. Tick marks are drawn at increments of *value3*. The last tick mark is the greatest value that does not exceed *value2*. This method of specification of tick marks is not valid for logarithmic axes. If *value3* is omitted, a value of 1 is used.
	`vaxis = 0 to 10`
	`vaxis = 0 to 2 by .1`

VAXISLABEL='*string*'	Specifies a label for the vertical axis
VLOWER=*number*	Specifies the lower limit on the vertical axis scale. The VLOWER= option specifies *number* as the lower vertical axis tick mark. The tick mark interval and the upper axis limit are determined automatically. This option has no effect if the VAXIS= option is used.
VUPPER=*number*	Specifies the upper limit on the vertical axis scale. The VUPPER= option specifies *number* as the upper vertical axis tick mark. The tick mark interval and the lower axis limit are determined automatically. This option has no effect if the VAXIS= option is used.
WAXIS=*n*	Specifies the line thickness for axes and frame.

Table 18.28 lists options that control the appearance of the axes when you use ODS Graphics.

Table 18.28 Axis Options for ODS Graphics

Option	Option Description
HLOWER=*number*	Specifies the lower limit on the horizontal axis scale. The HLOWER= option specifies *number* as the lower horizontal axis tick mark. The tick mark interval and the upper axis limit are determined automatically. This option has no effect if the HAXIS option is used.

Table 18.28 Axis Options (continued)

Option	Option Description
HUPPER=*number*	Specifies the upper limit on the horizontal axis scale. The HUPPER= option specifies *number* as the upper horizontal axis tick mark. The tick mark interval and the lower axis limit are determined automatically. This option has no effect if the HAXIS= option is used.
LOGLOG	Requests log scales on both axes.
MINORLOGGRID	Adds a minor grid for log axes.
NOGRID	Suppresses grid lines.
VLOWER=*number*	Specifies the lower limit on the vertical axis scale. The VLOWER= option specifies *number* as the lower vertical axis tick mark. The tick mark interval and the upper axis limit are determined automatically. This option has no effect if the VAXIS= option is used.
VUPPER=*number*	Specifies the upper limit on the vertical axis scale. The VUPPER= option specifies *number* as the upper vertical axis tick mark. The tick mark interval and the lower axis limit are determined automatically. This option has no effect if the VAXIS= option is used.

Table 18.29 lists options that control colors and patterns used in the graph when you use traditional graphics. These options are not available if ODS Graphics is enabled.

Table 18.29 Color and Pattern Options

Option	Option Description
CAXIS=*color*	Color for axis.
CCENSOR=*color*	Color for filling censor plot area.
CENCOLOR=*color*	Color for censor symbol.
CFRAME=*color*	Color for frame.
CFRAMESIDE=*color*	Color for filling frame for row labels.
CFRAMETOP=*color*	Color for filling frame for column labels.
CGRID=*color*	Color for grid lines.
CHREF=*color*	Color for HREF= lines.
CTEXT=*color*	Color for text.
CVREF=*color*	Color for VREF= lines.

Table 18.30 lists options that control the use of a graphics catalog to store graphs if you use traditional graphics. These options are not available if ODS Graphics is enabled.

Table 18.30 Graphics Catalog Options

Option	Option Description
DESCRIPTION='*string*'	Description for graphics catalog member.
NAME='*string*'	Name for plot in graphics catalog.

MODEL Statement

MODEL *variable* < ∗*censor-variable*(*values*) > < **=***effect-list* > < / *options* > **;**

MODEL (*variable1 variable2*) < **=***effect-list* > < / *options* > **;**

You use the MODEL statement to fit regression models, where life is modeled as a function of explanatory variables.

You can use only one MODEL statement after a PROC RELIABILITY statement. If you specify more than one MODEL statement, only the last is used.

The MODEL statement does not produce any plots, but it enables you to analyze more complicated regression models than the ANALYZE, PROBPLOT, or RELATIONPLOT statement does. The probability distribution specified in the DISTRIBUTION statement is used in the analysis. The following are examples of MODEL statements:

```
model time = temp voltage;
model life*censor(1) = voltage width;
```

See the section "Analysis of Accelerated Life Test Data" on page 1184 and the section "Regression Modeling" on page 1199 for examples that use the MODEL statement to fit regression models.

If your data are right censored lifetime data, you must specify a *censor-variable* and, in parentheses, the *values* of the *censor-variable* that correspond to censored data values.

If your data are recurrent events data with exact event times, you must specify a *censor-variable* and, in parentheses, the *values* of the *censor-variable* that correspond to the end-of-service times for each unit under observation. In this case, you must also specify a UNITID statement to identify the specific unit that corresponds to each observation.

If your lifetime data contain any interval-censored or left-censored values, you must specify *variable1* and *variable2* in parentheses to provide the endpoints of the interval for each observation.

If your data are recurrent events data, and event times are not known exactly, but are known only to have occurred in intervals, you must specify *variable1* and *variable2* in parentheses to provide the endpoints of the interval for each observation. In this case, you must also specify a variable that determines the number of events observed in each interval with a FREQ statement, and a variable that determines the number of units under observation in each interval with a NENTER statement.

The independent variables in your regression model are specified in the *effect-list*. The *effect-list* is any combination of continuous variables, classification variables, and interaction effects.

See the section "Regression Models" on page 1327 for further information on specifying the independent variables.

The elements of the MODEL statement are described as follows.

variable
is the dependent, or response, variable. The *variable* must be a numeric variable in the input data set.

censor-variable(values)
for lifetime data, indicates which observations in the input data set are right censored. You specify the values of *censor-variable* that represent censored observations by placing those values in parentheses after the variable name. If your data are not right censored, then you can omit the specification of a *censor-variable*; otherwise, *censor-variable* must be a numeric variable in the input data set.

If your data are recurrent events data and exact event times are known, then you must specify *censor-variable*. If *censor-variable* is equal to one of the *values*, then the value of *variable* is the end of observation time for a unit. Otherwise, you use *censor-variable* to assign a cost to the event that occurs at the value of *variable*. If all events have unit cost, then *censor-variable* should be set to one for all observations that do not correspond to end of observation times. The *censor-variable* plays the same role as the *cost/censor-variable* in the MCFPLOT statement in this case.

(variable1 variable2)
is another method of specifying the dependent variable in a regression model for lifetime data. You can use this syntax in a situation where uncensored, interval-censored, left-censored, and right-censored values occur in the same set of data. Table 18.31 shows how you use this syntax to specify different types of censoring by using combinations of missing and nonmissing values.

Table 18.31 Specifying Censored Values

Variable1	Variable2	Type of Censoring
Nonmissing	Nonmissing	Uncensored if *variable1* = *variable2*
Nonmissing	Nonmissing	Interval censored if *variable1* < *variable2*
Nonmissing	Missing	Right censored at *variable1*
Missing	Nonmissing	Left censored at *variable2*

For example, if T1 and T2 represent time in hours in the input data set

```
OBS    T1    T2

 1      .     6
 2      6    12
 3     12    24
 4     24     .
 5     24    24
```

then the statement

```
model (t1 t2);
```

specifies a model in which observation 1 is left censored at 6 hours, observation 2 is interval censored in the interval (6, 12), observation 3 is interval censored in (12,24), observation 4 is right censored at 24 hours, and observation 5 is an uncensored lifetime of 24 hours.

You can also use this method to specify a model for recurrent events data when exact recurrence times are not known. In this case, events are observed to have occurred in intervals specified by *(variable1 variable2)*. The values of the variable specified in a FREQ statement determine the number of events that occurred in each interval, and the values of the variable specified in a NENTER statement determine the number of units under observation in each interval.

effect-list
is a list of variables in the input data set representing the values of the independent variables in the model for each observation, and combinations of variables representing interaction terms. If a variable in the *effect-list* is also listed in a CLASS statement, an indicator variable is generated for each level of the variable. An indicator variable for a particular level is equal to 1 for observations with that level, and equal to 0 for all other observations. This type of variable is called a *classification* variable. Classification variables can be either character or numeric. If a variable is not listed in a CLASS statement, it is assumed to be a continuous variable, and it must be numeric.

options
control how the model is fit and what output is produced. All *options* are specified after a slash (/) in the MODEL statement. The "Summary of Options" section, which follows, lists all options by function.

Summary of Options

Table 18.32 Model Statement Options

Option	Option Description
CONFIDENCE=*number*	Specifies the confidence coefficient for all confidence intervals. Specify a *number* between 0 and 1. The default value is 0.95.

Table 18.32 Model Statement Options (continued)

Option	Option Description
CONVERGE=*number*	Specifies the convergence criterion for maximum likelihood fit. See the section "Maximum Likelihood Estimation" on page 1324 for details.
CONVH=*number*	Specifies the convergence criterion for the relative Hessian convergence criterion. See the section "Maximum Likelihood Estimation" on page 1324 for details.
CORRB	Requests parameter correlation matrix.
COVB	Requests parameter covariance matrix.
HPPTEST	Applies only to models for recurrent events data. This option requests a likelihood ratio test for a homogeneous Poisson process.
INEST | IN=*SAS-data-set*	Applies only to models for recurrent events data. This option specifies a SAS data set that can contain initial values, equality constraints, upper bounds, or lower bounds for the intercept and shape parameters in a model for recurrents events data. See the section "INEST Data Set for Recurrent Events Models" on page 1356 for details.
INITIAL=*number list*	Specifies initial values for regression parameters other than the location (intercept) term.
ITPRINT	Requests the iteration history for maximum likelihood fit.
LRCL	Requests likelihood ratio confidence intervals for distribution parameters.
LOCATION=*number* < LINIT >	Specifies the fixed or initial value of the location, or intercept parameter.
MAXIT=*number*	Specifies the maximum number of iterations allowed for maximum likelihood fit.

Table 18.32 Model Statement Options (continued)

Option	Option Description
OBSTATS	Requests a table that contains the XBETA, SURV, SRESID, and ADJRESID statistics in Table 18.33 or the XBETA, MCF, and INTENSITY statistics in Table 18.34. The table also contains the dependent and independent variables in the model. You can use this option to compute statistics such as survival function estimates for lifetime data or mean function estimates for recurrent events data for dependent variable values not included in the analysis. Refer to "Comparison of Two Samples of Repair Data" on page 1220 for an example of computing predicted values for recurrent events data.
OBSTATS(*statistics*)	Requests a table that contains the model variables and the statistics in the specified list of *statistics*. Available statistics are shown in Table 18.33.
ORDER=DATA | FORMATTED | FREQ | INTERNAL	Specifies the sort order for values of the classification variables in the *effect-list*.
PSTABLE=*number*	Specifies stable parameterization. The *number* must be between zero and one. See the section "Stable Parameters" on page 1329 for further information.
READOUT	Analyzes data in readout structure. The FREQ statement must be used to specify the number of units that fail in each interval, and the NENTER statement must be used to specify the number of unfailed units that enter each interval.

Table 18.32 Model Statement Options (continued)

Option	Option Description
RELATION=*transformation-keyword* RELATION=*(transformation-keyword1<,>transformation-keyword2)*	Specifies the type of relationship between independent and dependent variables. In the first form, the transformation specified is applied to the first continuous independent variable in the model. In the second form, the transformations specified within parentheses are applied to the first two continuous independent variables in the model, in the order listed. *transformation-keyword*, *transformation-keyword1*, and *transformation-keyword2* can be any of the transformations listed in the following table. See Table 18.67 for definitions of the transformations.

Transformation Keyword	Type of Transformation
ARRHENIUS	Arrhenius (Nelson parameterization)
ARRHENIUS2	Arrhenius (activation energy)
POWER	Logarithmic
LINEAR	Linear
LOGISTIC	Logistic

SCALE=*number* <SCINIT>	Specifies a fixed or initial value of scale parameter.
SHAPE=*number* <SHINIT>	Specifies a fixed or initial value of shape parameter.
SINGULAR=*number*	Specifies the singularity criterion for matrix inversion.
THRESHOLD=*number*	Specifies a fixed threshold parameter. See Table 18.57 for the distributions with a threshold parameter.
TREND=*trend-test keyword* | *(trend-test keywords)*	Applies only to models for recurrent events data. This option requests one or more tests of trend for a Poisson process. TREND=LRHPP is equivalent to the HPPTEST option. See the section "Tests of Trend" on page 1355 for more information about the tests. The available tests are shown in the following table.

Table 18.32 Model Statement Options (continued)

Option	Option Description
	Trend-Test Keyword / **Description of Test** MH | HDBK | MIL-HDBK — Military handbook test LA | LAPLACE — Laplace's test LR | LEWIS-ROBINSON — Lewis-Robinson test LRHPP | LIKELIHOOD — Likelihood ratio tset ALL — All available tests
WALDCL | NORMALCL	Requests Wald type confidence intervals for distribution parameters. See Table 18.68 and Table 18.74 for details about the computation of Wald confidence intervals. Wald confidence intervals are provided by default, but this option can be combined with LRCL to obtain both types of intervals.

Table 18.33 Available Statistics Computed for Each Observation with the OBSTATS Option for LIfetime Data

Option	Option Description
CENSOR	Is a variable that indicates the type of censoring for each observation in the input data set. The possible values for CENSOR and their interpretations are listed in the following table.

Type of Response	CENSOR Variable Value
Uncensored	0
Right-censored	1
Left-censored	2
Interval-censored	3

CONTROL=*variable*	Specifies a control variable in the input data set that allows the computation of statistics for a subset of observations in the input data set. If the value of *variable* is 1, the statistics are computed for that observation. If the value of the control variable is not equal to 1, the statistics are not computed for that observation.

Table 18.33 Available Statistics Computed for Each Observation with the OBSTATS Option (continued)

Option	Option Description
QUANTILES | QUANTILE | Q=*number-list*	Requests distribution quantiles for each number in *number-list* for each observation. The numbers must be between 0 and 1. Estimated quantile standard errors, and upper and lower confidence limits are also tabulated.
XBETA	Specifies the linear predictor.
SURVIVAL | SURV	Specifies the fitted survival function, evaluated at the value of the dependent variable.
RESID	Specifies the raw residual.
SRESID	Specifies the standardized residual.
GRESID	Specifies the modified Cox-Snell residual.
DRESID	Specifies the deviance residual.
ADJRESID	Specifies the adjusted standardized residuals. These are adjusted for right-censored observations by adding the median of the lifetime greater than the right-censored values to the residuals.
RESIDADJ=*number*	Specifies the adjustment to be added to Cox-Snell residual for right-censored data values. The default of *number* is 1.0, the mean of the standard exponential distribution.
RESIDALPHA | RALPHA=*number*	Specifies that the *number*$\times 100\%$ percentile residual lifetime be used to adjust right-censored standardized residuals. The *number* must be between 0 and 1. The default value is 0.5, which corresponds to the median.

Table 18.34 Available Statistics Computed for Each Observation with the OBSTATS Option for Recurrent Events Data

Option	Option Description
CONTROL=*variable*	Specifies a control variable in the input data set that allows the computation of statistics for a subset of observations in the input data set. If the value of *variable* is 1, the statistics are computed for that observation. If the value of *variable* is not equal to 1, the statistics are not computed for that observation.

Table 18.34 Available Statistics Computed for Each Observation with the OBSTATS Option (continued)

Option	Option Description
MCF	Specifies the mean function, which is evaluated at the value of time for each observation. Standard errors and confidence limits are also computed.
INTENSITY	Specifies the intensity function, which is evaluated at the value of time for each observation. Standard errors and confidence limits are also computed.
XBETA	Specifies the linear predictor.

NENTER Statement

NENTER *variable* **;**

Use the NENTER statement in conjunction with the FREQ statement to specify interval-censored lifetime data having a special structure; these data are called *readout* data. The NENTER statement specifies a *variable* in the input data set that determines the number of unfailed units entering each interval. See the section "Weibull Analysis of Interval Data with Common Inspection Schedule" on page 1189 for an example that uses the NENTER statement with readout data.

You can also use the NENTER statement in conjunction with the FREQ statement to specify recurrent events data when the event times are grouped into intervals, rather than being observed exactly. The NENTER statement specifies a variable in the input data set that determines the number of units observed in each interval.

NLOPTIONS Statement

NLOPTIONS < *options* > **;**

You use the NLOPTIONS statement to control aspects of the optimization system that is used to compute maximum likelihood estimates of the parameters of the three-parameter Weibull distribution with an ANALYZE or PROBPLOT statement, and of the parameters of models for recurrent events data with a MODEL statement. The syntax and options of the NLOPTIONS statement are described in Chapter 19, "Shared Concepts and Topics" (*SAS/STAT User's Guide*).

PROBPLOT Statement

< *label:* >**PROBPLOT** *variable* < ∗*censor-variable*(*values*) > < =*group-variables* > < / *options* > ;

< *label:* >**PROBPLOT** (*variable1 variable2*) < =*group-variables* > < / *options* > ;

You use the PROBPLOT statement to create a probability plot from complete, left-censored, right-censored, or interval-censored data.

You can specify the keyword PPLOT as an alias for PROBPLOT. You can specify any number of PROBPLOT statements after a PROC RELIABILITY statement. Each PROBPLOT statement creates a probability plot and an associated analysis. The probability distribution used in creating the probability plot and performing the analysis is determined by the DISTRIBUTION statement. You can specify an optional *label* to distinguish between multiple PROBPLOT statements in the output.

See the section "Analysis of Right-Censored Data from a Single Population" on page 1176 and the section "Weibull Analysis Comparing Groups of Data" on page 1180 for examples that create probability plots with the PROBPLOT statement.

To create a probability plot, you must specify one *variable*. If your data are right censored, you must specify a *censor-variable* and, in parentheses, the *values* of the *censor-variable* that correspond to censored data values.

You can optionally specify one or two *group-variables* (also referred to as *classification variables*). The PROBPLOT statement displays a probability plot for each level of the *group-variables*. The observations in a given level are called a *cell*.

The elements of the PROBPLOT statement are described as follows.

variable
: represents the data for which a probability plot is to be produced. The *variable* must be a numeric variable in the input data set.

censor-variable(values)
: indicates which observations in the input data set are right censored. You specify the values of *censor-variable* that represent censored observations by placing those values in parentheses after the variable name. If your data are not right censored, then you can omit the specification of *censor-variable*; otherwise, *censor-variable* must be a numeric variable in the input data set.

(variable1 variable2)
: is another method of specifying the data for which a probability plot is to be produced. You can use this syntax in a situation where uncensored, interval-censored, left-censored, and right-censored values occur in the same set of data. Table 18.31 shows how you use this syntax to specify different types of censoring by using combinations of missing and nonmissing values. See the section "Lognormal Analysis with Arbitrary Censoring" on page 1194 for an example that uses this syntax to create a probability plot.

group-variables
: are one or two group variables. If no group variables are specified, a single probability plot is produced. The *group-variables* can be numeric or character variables in the input data set.

  Note that the parentheses surrounding the *group-variables* are needed only if two group variables are specified.

options

control the features of the probability plot. All *options* are specified after the slash (/) in the PROBPLOT statement. See the section "Summary of Options" on page 1282, which follows, for a list of all options by function.

Summary of Options

Table 18.35 lists analysis options that are available when you use either traditional graphics or ODS Graphics.

Table 18.35 Analysis Options

Option	Option Description
CONFIDENCE=*number*	Specifies the confidence coefficient for all confidence intervals. The *number* must be between 0 and 1. The default value is 0.95
CONVERGE=*number*	Specifies the convergence criterion for maximum likelihood fit. See the section "Maximum Likelihood Estimation" on page 1324 for details.
CONVH=*number*	Specifies the convergence criterion for the relative Hessian convergence criterion. See the section "Maximum Likelihood Estimation" on page 1324 for details.
CORRB	Requests the parameter correlation matrix.
COVB	Requests the parameter covariance matrix.
FITTYPE | FIT= *fit-specification*	Specifies the method of estimating distribution parameters. The available *fit-specifications* and their meanings are shown in the following table.

Fit Specification	Definition
LSYX	Least squares fit to the probability plot. The probability axis is the dependent variable.
LSXY	Least squares fit to the probability plot. The lifetime axis is the dependent variable.
MLE	Maximum likelihood (default).
MODEL	Use the fit from the preceding MODEL statement.
NONE	No fit is computed.
WEIBAYES <(CONFIDENCE | CONF=*number*)>	Weibayes fit. *number* is the confidence coefficient for the Weibayes fit and is between 0 and 1. The default is 0.95.

Table 18.35 Analysis Options (continued)

Option	Option Description
INEST | IN=*SAS-data-set*	Specifies a SAS data set that can contain initial values, equality constraints, upper bounds, or lower bounds for the scale, shape, and threshold parameters in a three-parameter Weibull model for lifetime data, and applies only to three-parameter Weibull models. See the section "INEST Data Set for the Three-Parameter Weibull" on page 1326 for details.
ITPRINT	Requests the iteration history for maximum likelihood fit.
ITPRINTEM	Requests the iteration history for the Turnbull algorithm.
LRCL	Requests likelihood ratio confidence intervals for distribution parameters.
LRCLPER	Requests likelihood ratio confidence intervals for distribution percentiles.
LRCLSURV	Requests likelihood ratio confidence intervals for survival and cumulative distribution functions at times specified with the SURVTIME=*number-list* option.
LOCATION=*number* < LINIT >	Specifies a fixed or initial value of location parameter.
MAKEHAM=*number* < MKINIT >	Specifies the fixed or initial value of the Makeham parameter for the three-parameter Gompertz distribution.
MAXIT=*number*	Specifies the maximum number of iterations allowed for a maximum likelihood fit.
MAXITEM=*number1* < ,*number2* >	*number1* Specifies the maximum number of iterations allowed for the Turnbull algorithm. Iteration history will be printed in increments of *number2* if an iteration history is requested with ITPRINTEM. See the section "Interval-Censored Data" on page 1317 for details.
NOPCTILES	Suppresses computation of percentiles for standard list of percentage points.
NOPOLISH	Suppresses the setting of small interval probabilities to 0 in the Turnbull algorithm. See the section "Interval-Censored Data" on page 1317 for details.
NPINTERVALS=*interval-type*	Specifies the type of nonparametric confidence interval displayed in a probability plot. The available types of intervals are listed in the following table.

Table 18.35 Analysis Options (continued)

Option	Option Description
	Interval Type — **Definition**
	POINTWISE | POINT — Pointwise confidence intervals for the CDF. See the section "Pointwise Confidence Intervals" on page 1323 for details.
	SIMULTANEOUS | SIMUL< *number1, number2* > — Simultaneous confidence intervals for the CDF. *number1* and *number2* are constants that control the time interval for which simultaneous intervals are computed. The default time intervals are the lowest and highest times corresponding to failures in the case of right-censored data, or to the lowest and highest intervals for which probabilities are computed for interval-censored data. See the section "Simultaneous Confidence Intervals" on page 1323 for details.
PCTLIST=*number-list*	Specifies a list of percentages for which to compute percentile estimates. The *number-list* must be a list of numbers separated by blanks or commas. Each number in the list must be between 0 and 100. If this option is not specified, percentiles are computed for a standard list of percentages.
PINTERVALS=*interval-type*	Specifies the type of parametric pointwise confidence interval displayed in a probability plot. The available types of intervals are listed in the following table. The default type is PROBABILITY, pointwise confidence intervals on cumulative failure probability.

Table 18.35 Analysis Options (continued)

Option	Option Description

Interval Type	Definition
LIKELIHOOD | LRCI	Likelihood ratio confidence intervals
PERCENTILES | PER	Pointwise parametric confidence intervals for the percentiles of the fitted CDF
PROBABILITY | CDF	Pointwise parametric confidence intervals for the cumulative failure probabilities. See the section "Reliability Function" on page 1334 for details.

PPOS=*plotting-position*

Specifies the *plotting-position* type used to compute nonparametric estimates of the probability distribution function. See the section "Probability Plotting" on page 1313 for details. The plotting position types available are shown in the following table.

Plotting Position	Type
EXPRANK	Expected ranks
MEDRANK	Median ranks
MEDRANK1	Median ranks (exact formula)
KM	Kaplan-Meier
MKM	Modified Kaplan-Meier (default)
NA | NELSONAALEN	Nelson-Aalen

PPOUT

Requests a table of nonparametric cumulative probabilities in the printed output.

PRINTPROBS

Specifies that intervals and associated probabilities for the Turnbull algorithm be printed.

PROBLIST=*number-list*

Specifies a list of initial values for Turnbull algorithm. See the section "Interval-Censored Data" on page 1317 for details.

PSTABLE=*number*

Specifies a stable parameterization. The *number* must be between 0 and 1. See the section "Stable Parameters" on page 1329 for further information.

READOUT

Specifies the data have the readout structure.

SCALE=*number* < SCINIT >

Specifies a fixed or initial value of the scale parameter.

SHAPE=*number* < SHINIT >

Specifies a fixed or initial value of the shape parameter.

SINGULAR=*number*

Specifies the singularity criterion for matrix inversion.

Table 18.35 Analysis Options (continued)

Option	Option Description
SURVTIME=*number-list*	Requests that the survival function, cumulative distribution function, and confidence limits be computed for values in *number-list*. See the section "Reliability Function" on page 1334 for details.
THRESHOLD=*number*	Specifies a fixed threshold parameter. See Table 18.57 for the distributions with a threshold parameter.
TOLLIKE=*number*	Specifies the criterion for convergence in the Turnbull algorithm. The default is 10^{-8}. See the section "Interval-Censored Data" on page 1317 for details.
TOLPROB=*number*	Specifies the criterion for setting interval probabilities to 0 in the Turnbull algorithm. The default is 10^{-6}. See the section "Interval-Censored Data" on page 1317 for details.

Table 18.36 lists analysis options that are available when ODS Graphics is enabled.

Table 18.36 Analysis Options for ODS Graphics

Option	Option Description
PROFILE<(*options*)>	Requests a profile log-likelihood plot of the threshold parameter for a three-parameter Weibull distribution. The *options* listed in the following table are available; they are specified by enclosing them in parentheses after the PROFILE option.

Table 18.36 Analysis Options (continued)

Option	Option Description
	Profile Option — **Option Description**
	NOCONF — Specifies that a reference line on the vertical, log-likelihood axis not be drawn. If this option is not specified, a reference line is drawn at a log-likelihood value that corresponds to the profile likelihood confidence limits on the horizontal axis.
	NPROFILE=*n* — Specifies that the profile log likelihood be computed and plotted at *n* threshold points. If this option is not specified, the profile log likelihood is computed and plotted at 100 points.
	RANGE=(*value1*, *value2*) — Specifies the range of threshold values for which the profile log likelihood is computed and plotted as (*value1*, *value2*). If this option is not specified, the range of threshold values for which the profile log likelihood is computed is from 0 to the minimum failure time.

Table 18.37 lists plot layout options that are available when you use traditional graphics.

Table 18.37 Probability Plot Layout Options for Traditional Graphics

Option	Option Description
CENBIN	Specifies that censored data be plotted as frequency counts rather than as individual points.
CENSYMBOL= *symbol* | (*symbol-list*)	Specifies symbols for censored values. The *symbol* is one of the symbol names (plus, star, square, diamond, triangle, hash, paw, point, dot, circle) or a letter (A–Z). For overlaid plots for groups of data, you can specify different symbols for the groups with a list of symbols or letters, separated by blanks, enclosed in parentheses. If no CENSYMBOL option is specified, the symbol used for censored values is the same as the symbol used for failures.

Table 18.37 Probability Plot Layout Options (continued)

Option	Option Description
HOFFSET=*value*	Specifies the offset for the horizontal axis.
INBORDER	Requests a border around probability plots.
INTERTILE=*value*	Specifies the distance between tiles.
LFIT=*linetype* | (*linetype list*)	Specifies line styles for fit lines and confidence curves in a probability plot. The *linetype list* is a list of numbers from 1 to 46 representing different linetypes; they can be separated by blanks or commas or can be a list in the form n_1 to n_2 <by n_3>.
MISSING1	Requests that missing values of the first GROUP= variable be treated as a level of the variable.
MISSING2	Requests that missing values of the second GROUP= variable be treated as a level of the variable.
NCOLS=*n*	Specifies that *n* columns be plotted on a page.
NOCENPLOT	Suppresses the plotting of censored data points.
NOCONF	Suppresses the plotting of percentile confidence curves.
NOFIT	Suppresses the plotting of fit line and percentile confidence curves.
NOFRAME	Suppresses the frame around the plotting area.
NOINSET	Suppresses the inset.
NOPPLEGEND	Suppresses the legend for overlaid probability plots
NOPPOS	Suppresses plotting of symbols for failures in a probability plot.
NROWS=*n*	Specifies that *n* rows be plotted on a page.
ORDER1=DATA | FORMATTED | FREQ | INTERNAL	Specifies display order for values of the first GROUP= variable.
ORDER2=DATA | FORMATTED | FREQ | INTERNAL	Specifies the display order for values of the second GROUP= variable.
OVERLAY	Requests overlaid plots for group variables.
PCONFPLT	Plots confidence intervals on probabilities for readout data.
PPLEGEND = *legend-statement-name* | NONE	Identifies LEGEND*n* statement to specify a legend for overlaid probability plots.
PPOSSYMBOL=*symbol* | (*symbol-list*)	Specifies symbols to represent failures on a probability plot.
ROTATE	Requests probability plots with the probability scale on the horizontal axis.

Table 18.37 Probability Plot Layout Options (continued)

Option	Option Description
SHOWMULTIPLES	Requests that the count be displayed for multiple overlaying symbols.
TURNVLABELS	Vertically strings out characters in labels for the vertical axis.
VOFFSET=*value*	Specifies *value* as the length of the offset at the upper end of the vertical axis.
WFIT=*n*	Specifies the line width for the fit line and confidence curves.

Table 18.38 lists plot layout options that are available when you use ODS graphics.

Table 18.38 Probability Plot Layout Options for ODS Graphics

Option	Option Description
MISSING1	Requests that missing values of first GROUP= variable be treated as a level of the variable.
MISSING2	Requests that missing values of second GROUP= variable be treated as a level of the variable.
NCOLS=*n*	Specifies that *n* columns be plotted on a page.
NOCENPLOT	Suppresses plotting of censored data points.
NOCONF	Suppresses plotting of percentile confidence curves.
NOFIT	Suppresses plotting of the fit line and percentile confidence curves.
NOINSET	Suppresses the inset.
NOPPLEGEND	Suppresses the legend for overlaid probability plots.
NOPPOS	Suppresses plotting of symbols for failures in a probability plot.
NROWS=*n*	Specifies that *n* rows be plotted on a page.
ORDER1=DATA | FORMATTED | FREQ | INTERNAL	Specifies the display order for values of the first GROUP= variable.
ORDER2=DATA | FORMATTED | FREQ | INTERNAL	Specifies the display order for values of the second GROUP= variable.
OVERLAY	Requests overlaid plots for group variables.
PCONFPLT	Plots confidence intervals on probabilities for readout data.
ROTATE	Requests probability plots with the probability scale on the horizontal axis.

Table 18.39 lists reference line options that are available when you use traditional graphics.

Table 18.39 Reference Line Options for Traditional Graphics

Option	Option Description
HREF <(INTERSECT)>=*value-list*	Requests reference lines perpendicular to horizontal axis. If (INTERSECT) is specified, a second reference line perpendicular to the vertical axis is drawn that intersects the fit line at the same point as the horizontal axis reference line. If a horizontal axis reference line label is specified, the intersecting vertical axis reference line is labeled with the vertical axis value.
HREFLABELS=(*'label1' ... 'labeln'*)	Specifies labels for HREF= lines.
HREFLABPOS=*n*	Specifies vertical position of labels for HREF= lines. The valid values for *n* and the corresponding label placements are shown in the following table: *n* — **Label Placement** 1 — Top 2 — Staggered from top 3 — Bottom 4 — Staggered from bottom 5 — Alternating from top 6 — Alternating from bottom
LHREF=*linetype*	Specifies the line style for HREF= lines.
LVREF=*linetype*	Specifies the line style for VREF= lines.
VREF <(INTERSECT)>=*value-list*	Specifies reference lines perpendicular to vertical axis. If (INTERSECT) is specified, a second reference line perpendicular to the horizontal axis is drawn that intersects the fit line at the same point as the vertical axis reference line. If a vertical axis reference line label is specified, the intersecting horizontal axis reference line is labeled with the horizontal axis value.
VREFLABELS=(*'label1' ... 'labeln'*)	Specifies labels for VREF= lines.
VREFLABPOS=*n*	Specifies horizontal position of labels for VREF= lines. The valid values for *n* and the corresponding label placements are shown in the following table: *n* — **Label Placement** 1 — Left 2 — Right

Table 18.40 lists reference line options that are available when you use ODS graphics.

Table 18.40 Reference Line Options for ODS Graphics

Option	Option Description
LREF < (INTERSECT) >=*value-list*	Requests reference lines perpendicular to the lifetime axis. If (INTERSECT) is specified, a second reference line is drawn perpendicular to the probability axis and intersects the fit line at the same point as the lifetime axis reference line. If a lifetime axis reference line label is specified, the intersecting probability axis reference line is labeled with the probability axis value.
LREFLABELS =(*'label1' ... 'labeln'*)	Specifies labels for LREF= lines.
PREF < (INTERSECT) >=*value-list*	Specifies reference lines perpendicular to the probability axis. If (INTERSECT) is specified, a second reference line is drawn perpendicular to the lifetime axis and intersects the fit line at the same point as the probability axis reference line. If a probability axis reference line label is specified, the intersecting lifetime axis reference line is labeled with the lifetime axis value.
PREFLABELS=(*'label1' ... 'labeln'*)	Specifies labels for PREF= lines.

Table 18.41 lists options that control the appearance of the text when you use traditional graphics. These options are not available if ODS Graphics is enabled.

Table 18.41 Text Enhancement Options

Option	Option Description
FONT=*font*	Specifies a software font for text.
HEIGHT=*value*	Specifies the height of text used outside framed areas.
INFONT=*font*	Specifies a software font for text inside framed areas.
INHEIGHT=*value*	Specifies the height of text inside framed areas.

Table 18.42 lists options that control the appearance of the axes when you use traditional graphics.

Table 18.42 Axis Options for Traditional Graphics

Option	Option Description

LAXIS=*value1 to value2< by value3 >*	Specifies tick mark values for the lifetime axis. *value1*, *value2*, and *value3* must be numeric, and *value1* must be less than *value2*. The lower tick mark is *value1*. Tick marks are drawn at increments of *value3*. The last tick mark is the greatest value that does not exceed *value2*. If *value3* is omitted, a value of 1 is used. This method of specification of tick marks is not valid for logarithmic axes. Examples of LAXIS= lists are
	`laxis = -1 to 10`
	`laxis = 0 to 200 by 10`

LGRID=*number*	Specifies a line style for all grid lines. The *number* is between 1 and 46 and specifies a linestyle for grids.
LIFELOWER | LLOWER=*number*	Specifies the lower limit on the lifetime axis scale. The LLOWER option specifies *number* as the lower lifetime axis tick mark. The tick interval and the upper lifetime axis limit are determined automatically. This option has no effect if the LAXIS option is used.
LIFEUPPER | LUPPER=*number*	Specifies the upper limit on the lifetime axis scale. The LUPPER option specifies *number* as the upper lifetime axis tick mark. The tick interval and the lower lifetime axis limit are determined automatically. This option has no effect if the LAXIS option is used.
MPGRID	Adds a minor grid for the probability axis.
MINORLOGGRID	Adds a minor grid for log axes.
NOGRID	Suppresses grid lines.
NOLLABEL	Suppresses label for life, or analysis variable, axis.
NOLTICK	Suppresses tick marks and tick mark labels for lifetime or analysis variable axis.
NOPLABEL	Suppresses label for probability axis.
NOPTICK	Suppresses tick marks and tick mark labels for the probability axis.
NTICK=*number*	Specifies the number of tick intervals for the lifetime axis. This option has no effect if the LAXIS option is used.

Table 18.42 Axis Options (continued)

Option	Option Description
PCTLOWER | PLOWER=*number*	Specifies the lower limit on probability axis scale.
PCTUPPER | PUPPER=*number*	Specifies the upper limit on probability axis scale.
PAXISLABEL='*string*'	Specifies a label for the probability axis.
WAXIS=*n*	Specifies the line thickness for axes and frame.

Table 18.43 lists options that control the appearance of the axes when you use ODS Graphics.

Table 18.43 Axis Options for ODS Graphics

Option	Option Description

LAXIS=*value1 to value2< by value3 >*	Specifies tick mark values for the lifetime axis. *value1*, *value2*, and *value3* must be numeric, and *value1* must be less than *value2*. The lower tick mark is *value1*. Tick marks are drawn at increments of *value3*. The last tick mark is the greatest value that does not exceed *value2*. If *value3* is omitted, a value of 1 is used. This method of specification of tick marks is not valid for logarithmic axes. Examples of LAXIS= lists are
	`laxis = -1 to 10`
	`laxis = 0 to 200 by 10`

LIFELOWER | LLOWER=*number*	Specifies the lower limit on the lifetime axis scale. The LLOWER option specifies *number* as the lower lifetime axis tick mark. The tick interval and the upper lifetime axis limit are determined automatically. This option has no effect if the LAXIS option is used.
LIFEUPPER | LUPPER=*number*	Specifies the upper limit on the lifetime axis scale. The LUPPER option specifies *number* as the upper lifetime axis tick mark. The tick interval and the lower lifetime axis limit are determined automatically. This option has no effect if the LAXIS option is used.
MINORLOGGRID	Adds a minor grid for log axes.
NOGRID	Suppresses grid lines.
PCTLOWER | PLOWER=*number*	Specifies the lower limit on probability axis scale.

Table 18.43 Axis Options (continued)

Option	Option Description
PCTUPPER | PUPPER=*number*	Specifies the upper limit on probability axis scale.
PAXISLABEL='*string*'	Specifies a label for the probability axis.

Table 18.44 lists options that control colors and patterns used in the graph when you use traditional graphics. These options are not available if ODS Graphics is enabled.

Table 18.44 Color and Pattern Options

Option	Option Description
CAXIS=*color*	Color for axis
CCENSOR=*color*	Color for filling censor plot area
CENCOLOR=*color*	Color for censor symbol
CFIT=*color* | (*color list*)	color for fit lines and confidence curves in a probability plot
CFRAME=*color*	Color for frame
CFRAMESIDE=*color*	Color for filling frame for row labels
CFRAMETOP=*color*	Color for filling frame for column labels
CGRID=*color*	Color for grid lines
CHREF=*color*	Color for HREF= lines
CTEXT=*color*	Color for text
CVREF=*color*	Color for VREF= lines
PPOSCOLOR=*color* | (*color list*)	Colors of symbols that represent failures on a probability plot

Table 18.45 lists options that control the use of a graphics catalog to store graphs if you use traditional graphics. These options are not available if ODS Graphics is enabled.

Table 18.45 Graphics Catalog Options

Option	Option Description
DESCRIPTION='*string*'	Description for graphics catalog member
NAME='*string*'	Name for plot in graphics catalog

RELATIONPLOT Statement

< *label:* >**RELATIONPLOT** *variable* < ∗*censor-variable*(*values*) > < =*group-variable* >
< / *options* > ;

< *label:* >**RELATIONPLOT** (*variable1 variable2*) < =*group-variable* > < / *options* > ;

You use the RELATIONPLOT statement to create life-stress relation plots. A life-stress relation plot is a graphical tool for the analysis of data from accelerated life tests. The plot is a display of the relationship between life and *stress*, such as temperature or voltage. You can also use the RELATIONPLOT statement to display a probability plot alongside the relation plot. See Figure 18.9 for an example of a relation plot.

You can specify the keyword RPLOT as an alias for RELATIONPLOT. You can use any number of RELATIONPLOT statements after a PROC RELIABILITY statement. You can specify an optional *label* to distinguish between multiple RELATIONPLOT statements in the output.

See the section "Analysis of Accelerated Life Test Data" on page 1184 for an example that uses the RELATIONPLOT statement.

To create a life-stress relation plot, you must specify one *variable*. If your data are right censored, you must specify a *censor-variable* and, in parentheses, the *values* of the *censor-variable* that correspond to censored data values. You must specify one *group-variable* to represent the values of stress. The *group-variable* must be a numeric variable.

The RELATIONPLOT statement plots the uncensored values of your data given by *variable* versus the values of the *group-variable*. You can optionally display a boxplot of the values of the data. You can also plot percentiles of the distribution fitted to the data. The RELATIONPLOT statement produces the same tabular output as the PROBPLOT statement, and all the analysis options are the same as for the PROBPLOT statement.

The elements of the RELATIONPLOT statement are described as follows.

variable
: represents the data for which a plot is to be produced. The *variable* must be a numeric variable in the input data set.

censor-variable(values)
: indicates which observations in the input data set are right censored. You specify the values of *censor-variable* that represent censored observations by placing those values in parentheses after the variable name. If your data are not right censored, then you omit the specification of *censor-variable*; otherwise, *censor-variable* must be a numeric variable in the input data set.

(variable1 variable2)
: is another method of specifying the data for which a life-stress plot is to be produced. You can use this syntax in a situation where uncensored, interval-censored, left-censored, and right-censored values occur in the same set of data. Table 18.31 shows how you use this syntax to specify different types of censoring by using combinations of missing and nonmissing values. See the section "Lognormal Analysis with Arbitrary Censoring" on page 1194 for an example that uses this syntax to create a probability plot.

group-variable
is a group variable. The *group-variable* must be a numeric variable in the input data set.

options
control the features of the relation plot. All *options* are specified after the slash (/) in the RELATIONPLOT statement. The "Summary of Options" section, which follows, lists all options by function.

The only type of relation plot currently available for interval data is the type in which percentiles of the fitted distribution are plotted at each stress level.

Summary of Options

Table 18.46 Analysis Options

Option	Option Description
CONFIDENCE=*number*	Specifies the confidence coefficient for all confidence intervals. The *number* must be between 0 and 1. The default value is 0.95.
CONVERGE=*number*	Specifies the convergence criterion for maximum likelihood fit. See the section "Maximum Likelihood Estimation" on page 1324 for details.
CONVH=*number*	Specifies the convergence criterion for the relative Hessian convergence criterion. See the section "Maximum Likelihood Estimation" on page 1324 for details.
CORRB	Requests the parameter correlation matrix.
COVB	Requests the parameter covariance matrix.
FITTYPE | FIT= *fit-specification*	Specifies the method of estimating distribution parameters. The available *fit-specifications* and their meanings are shown in the following table.

Table 18.46 Analysis Options (continued)

Option	Option Description
Fit Specification	**Definition**
LSYX	Least squares fit to the probability plot. The probability axis is the dependent variable.
LSXY	Least squares fit to the probability plot. The lifetime axis is the dependent variable.
MLE	Maximum likelihood (default).
MODEL	Use the fit from the preceding MODEL statement
NONE	No fit is computed.
REGRESSION< (*variable=number-list*) >	Use the fit from the preceding MODEL statement. Nonlinear relations and percentiles from models that use two independent variables can be plotted. *variable* is a second independent variable in the MODEL statement, and *number-list* is a list of values of the second independent variable for which to plot the percentiles as a function of the first independent variable.
WEIBAYES < (CONFIDENCE | CONF=*number*) >	Weibayes fit. *number* is the confidence coefficient for the Weibayes fit and is between 0 and 1. The default is 0.95.
ITPRINT	Requests the iteration history for the maximum likelihood fit.
LRCL	Requests likelihood ratio confidence intervals for distribution parameters.
LRCLPER	Requests likelihood ratio confidence intervals for distribution percentiles.
LOCATION=*number* < LINIT >	Specifies a fixed or initial value of the location parameter.
MAXIT=*number*	Specifies the maximum number of iterations allowed for a maximum likelihood fit.
NOPCTILES	Suppresses computation of percentiles.

Table 18.46 Analysis Options (continued)

Option	Option Description
PCTLIST=*number-list*	Specifies a list of percentages for which to compute percentile estimates. The *number-list* must be a list of numbers separated by blanks or commas. Each number in the list must be between 0 and 100. If this option is not specified, percentiles are computed for a standard list of percentages.
PPOS=*plotting-position*	Specifies the *plotting-position* type used to compute nonparametric estimates of the probability distribution function. See the section "Probability Plotting" on page 1313 for details. The plotting position types available are shown in the following table.

Plotting Position	Type
EXPRANK	Expected ranks
MEDRANK	Median ranks
MEDRANK1	Median ranks (exact formula)
KM	Kaplan-Meier
MKM	Modified Kaplan-Meier (default)
NA | NELSONAALEN	Nelson-Aalen

PPOUT	Requests a table of nonparametric cumulative probabilities in the printed output.
PSTABLE=*number*	Specifies a stable parameterization. The *number* must be between zero and one. See the section "Stable Parameters" on page 1329 for further information.
RELATION=*transformation-keyword*	Specifies the type of relationship between independent (stress) and dependent (lifetime) variables. The transformation specified is applied to the independent (stress) variable in the model. This determines the horizontal scale used in the relation plot. *transformation-keyword* can be any of the transformations listed in the following table. See Table 18.67 for definitions of the transformations.

Transformation Keyword	Type of Transformation
ARRHENIUS	Arrhenius (Nelson parameterization)
ARRHENIUS2	Arrhenius (activation energy)
POWER	Logarithmic
LINEAR	Linear
LOGISTIC	Logistic

Table 18.46 Analysis Options (continued)

Option	Option Description
READOUT	Specifies the data has the readout structure.
SCALE=*number* < SCINIT >	Specifies a fixed or initial value of scale parameter.
SHAPE=*number* < SHINIT >	Specifies a fixed or initial value of shape parameter.
SINGULAR=*number*	Specifies the singularity criterion for matrix inversion.
SURVTIME=*number-list*	Requests that survival function be computed for values in *number-list*. See the section "Reliability Function" on page 1334 for details.
THRESHOLD=*number*	Specifies a fixed threshold parameter. See Table 18.57 for the distributions with a threshold parameter.
variable=*number-list*	Enables creation of plots of percentiles from a regression model when two independent variables are used in a MODEL statement *effect-list*. The FIT=REGRESSION option must be used with this option. Percentile plots are created for each value of the independent *variable* in the *number-list*. *number-list* is a list of numeric values separated by blanks or commas, or in the form of a list n_1 to n_2 < by n_3 >.

Table 18.47 lists plot layout options that are available when you use traditional graphics.

Table 18.47 Plot Layout Options for Traditional Graphics

Option	Option Description
CENSYMBOL=*symbol* | (*symbol-list*)	Specifies symbols for censored values. The *symbol* is one of the symbol names (plus, star, square, diamond, triangle, hash, paw, point, dot, circle) or a letter (A–Z). If you are creating overlaid plots for groups of data, you can specify different symbols for the groups with a list of symbols or letters, separated by blanks, enclosed in parentheses. If no CENSYMBOL option is specified, the symbol used for censored values is the same as for failures.
HOFFSET=*value*	Specifies an offset for horizontal axis.
INBORDER	Requests a border around plots.
LBOXES=*number*	Specifies a line style for boxplots.

Table 18.47 Plot Layout Options (continued)

Option	Option Description
LFIT=*linetype* | (*linetype-list*)	Specifies line styles for fit lines and confidence curves in a probability plot. The *linetype-list* is a list of numbers from 1 to 46 representing different linetypes; the numbers can be separated by blanks or commas or can be a list in the form n_1 to n_2 <by n_3>.
LPLOTFIT=*linetype* | (*linetype-list*)	Specifies line styles for percentile lines. *linetype-list* is a list of numbers that represent different linetypes; the numbers can be separated by blanks or commas or can be a list in the form n_1 to n_2 <by n_3>.
NOCENPLOT	Suppresses plotting of censored data points.
NOCONF	Suppresses plotting of percentile confidence curves.
NOFIT	Suppresses plotting of fit line and percentile confidence curves.
NOFRAME	Suppresses the frame around the plotting area.
NOPPLEGEND	Suppresses the legend for overlaid probability plots.
NOPPOS	Suppresses plotting of symbols for failures in a probability plot.
NORPLEGEND	Suppresses the legend for the relation plot.
PINTERVALS=*interval-type*	Specifies the type of parametric pointwise confidence interval displayed in a probability plot. The available types of intervals are listed in the following table. The default type is PROBABILITY, pointwise confidence intervals on cumulative failure probability.

Interval Type	Definition
LIKELIHOOD | LRCI	Likelihood ratio confidence intervals
PERCENTILES | PER	Pointwise parametric confidence intervals for the percentiles of the fitted CDF
PROBABILITY | CDF	Pointwise parametric confidence intervals for the cumulative failure probabilities. See the section "Reliability Function" on page 1334 for details.

Table 18.47 Plot Layout Options (continued)

Option	Option Description
PLOTDATA <DATA | MEDIANS | BOXES>	Requests that the data be plotted on the relationplot and specifies the representation of the data populations to be plotted.
PLOTFIT < *number-list* >	Specifies that percentiles of the fitted distribution be plotted on the relation plot. The optional *number-list* is a list of percentiles (between 0 and 100); if not specified, the 50th percentile (median) is plotted.
PPLEGEND = *legend-statement-name* | NONE	Identifies a LEGEND*n* statement to specify legend for overlaid probability plots.
PPLOT	Places a probability plot on the same page as the relation plot.
RCENSYMBOL=*symbol* | (*symbol-list*)	Specifies symbols that represent right-censored and left-censored observations in a relation plot. The *symbol* is one of the symbol names (plus, star, square, diamond, triangle, hash, paw, point, dot, circle) or a letter (A–Z).
RPLEGEND = *legend-statement-name* | NONE	Identifies a LEGEND*n* statement to specify legend for the relation plot.
SHOWMULTIPLES	Displays the count for multiple overlaying symbols.
TURNVLABELS	Vertically strings out characters in labels for vertical axis.
VOFFSET=*value*	Specifies length of offset at upper end of vertical axis.
WFIT=*linetype*	Specifies line width for fit line and confidence curves.

Table 18.48 lists plot layout options that are available when you use ODS graphics.

Table 18.48 Plot Layout Options for ODS Graphics

Option	Option Description
NOCENPLOT	Suppresses plotting of censored data points.
NOCONF	Suppresses plotting of percentile confidence curves.
NOFIT	Suppresses plotting of fit line and percentile confidence curves.

Table 18.48 Plot Layout Options (continued)

Option	Option Description
NOPPOS	Suppresses plotting of symbols for failures in a probability plot.
PINTERVALS=*interval-type*	Specifies the type of parametric pointwise confidence interval displayed in a probability plot. The available types of intervals are listed in the following table. The default type is PROBABILITY, pointwise confidence intervals on cumulative failure probability.

Interval Type	Definition
LIKELIHOOD | LRCI	Likelihood ratio confidence intervals
PERCENTILES | PER	Pointwise parametric confidence intervals for the percentiles of the fitted CDF
PROBABILITY | CDF	Pointwise parametric confidence intervals for the cumulative failure probabilities. See the section "Reliability Function" on page 1334 for details.

PLOTDATA	Requests that the data be plotted on the relationplot.
PLOTFIT < *number-list* >	Specifies that percentiles of the fitted distribution be plotted on the relation plot. The optional *number-list* is a list of percentiles (between 0 and 100); if not specified, the 50th percentile (median) is plotted.
PPLOT	Places a probability plot on the same page as the relation plot.

Table 18.49 lists reference line options that are available when you use traditional graphics.

Table 18.49 Reference Line Options for Traditional Graphics

Option	Option Description
HREF < (INTERSECT) >=*value-list*	Requests reference lines perpendicular to horizontal axis. If (INTERSECT) is specified, a second reference line perpendicular to the vertical axis is drawn that intersects the fit line at the same point as the horizontal axis reference line. If a horizontal axis reference line label is specified, the intersecting vertical axis reference line is labeled with the vertical axis value.
HREFLABELS=(*'label1' ... 'labeln'*)	Specifies labels for HREF= lines.
HREFLABPOS=*n*	Specifies the vertical position of labels for HREF= lines. The valid values for *n* and the corresponding label placements are shown in the following table: *n* — **Label Placement** 1 — Top 2 — Staggered from top 3 — Bottom 4 — Staggered from bottom 5 — Alternating from top 6 — Alternating from bottom
LHREF=*linetype*	Specifies a line style for HREF= lines.
LSREF=*linetype*	Specifies a line style for SREF= lines.
LVREF=*linetype*	Specifies a line style for VREF= lines.
SREF=*value-list*	Specifies reference lines perpendicular to horizontal stress axis.
SREFLABELS=(*'label1' ... 'labeln'*)	Specifies labels for SREF= lines
SREFLABPOS=*n*	Specifies horizontal position of labels for SREF= lines. The valid values for *n* and the corresponding label placements are shown in the following table: *n* — **Label Placement** 1 — Top 2 — Staggered from top 3 — Bottom 4 — Staggered from bottom

Table 18.49 Reference Line Options (continued)

Option	Option Description
VREF < (INTERSECT) >=*value-list*	Requests reference lines perpendicular to vertical axis. If (INTERSECT) is specified, a second reference line perpendicular to the horizontal axis is drawn that intersects the fit line at the same point as the vertical axis reference line. If a vertical axis reference line label is specified, the intersecting horizontal axis reference line is labeled with the horizontal axis value.
VREFLABELS=(*'label1' ... 'labeln'*)	Specifies labels for VREF= lines.
VREFLABPOS=*n*	Specifies the horizontal position of labels for VREF= lines. The valid values for *n* and the corresponding label placements are shown in the following table: *n* — **Label Placement** 1 — Left 2 — Right

Table 18.50 lists reference line options that are available when you use ODS graphics.

Table 18.50 Reference Line Options for ODS Graphics

Option	Option Description
LREF < (INTERSECT) >=*value-list*	Requests reference lines perpendicular to lifetime axis. If (INTERSECT) is specified, a second reference line is drawn perpendicular to the stress axis (and the probability axis if applicable), and it intersects the fit line at the same point as the lifetime axis reference line. If a lifetime axis reference line label is specified, the intersecting stress/probability axis reference line is labeled with the stress/probability axis value.
LREFLABELS=(*'label1' ... 'labeln'*)	Specifies labels for LREF= lines.
PREF < (INTERSECT) >=*value-list*	Requests reference lines perpendicular to probability axis, if a probability plot is specified with the PPLOT option. If (INTERSECT) is specified, a second reference line is drawn perpendicular to the lifetime axis and intersects the fit line at the same point as the probability axis reference line. If a probability axis reference line label is specified, the intersecting lifetime axis reference line is labeled with the lifetime axis value.

Table 18.50 Reference Line Options (continued)

Option	Option Description
PREFLABELS=(*'label1' ... 'labeln'*)	Specifies labels for PREF= lines.
SREF <(INTERSECT)>=*value-list*	Specifies reference lines perpendicular to the stress axis. If (INTERSECT) is specified, a second reference line is drawn perpendicular to the lifetime axis and intersects the fit line at the same point as the stress axis reference line. If a stress axis reference line label is specified, the intersecting lifetime axis reference line is labeled with the lifetime axis value.
SREFLABELS=(*'label1' ... 'labeln'*)	Specifies labels for SREF= lines.

Table 18.51 lists options that control the appearance of the text when you use traditional graphics. These options are not available if ODS Graphics is enabled.

Table 18.51 Text Enhancement Options

Option	Option Description
FONT=*font*	Specifies a software font for text.
HEIGHT=*value*	Specifies the height of text used outside framed areas.
INFONT=*font*	Specifies a software font for text inside framed areas.
INHEIGHT=*value*	Specifies the height of text inside framed areas.

Table 18.52 lists options that control the appearance of the axes when you use traditional graphics.

Table 18.52 Axis Options for Traditional Graphics

Option	Option Description

LAXIS=*value1* TO *value2*< BY *value3* >	Specifies tick mark values for the lifetime axis. *value1*, *value2*, and *value3* must be numeric, and *value1* must be less than *value2*. The lower tick mark is *value1*. Tick marks are drawn at increments of *value3*. The last tick mark is the greatest value that does not exceed *value2*. If *value3* is omitted, a value of 1 is used. This method of specification of tick marks is not valid for logarithmic axes. Examples of LAXIS= lists are
	`laxis = -1 to 10`
	`laxis = 0 to 200 by 10`

LGRID=*number*	Specifies a line style for all grid lines. The *number* is between 1 and 46 and specifies a linestyle for grids.
LIFELOWER | LLOWER=*number*	Specifies the lower limit on the lifetime axis scale. The LLOWER option specifies *number* as the lower lifetime axis tick mark. The tick interval and the upper lifetime axis limit are determined automatically. This option has no effect if the LAXIS option is used.
LIFEUPPER | LUPPER=*number*	Specifies the upper limit on the lifetime axis scale. The LUPPER option specifies *number* as the upper lifetime axis tick mark. The tick interval and the lower lifetime axis limit are determined automatically. This option has no effect if the LAXIS option is used.
MPGRID	Adds a minor grid for the probability axis.
MINORLOGGRID	Adds a minor grid for log axes.
NOGRID	Suppresses grid lines.
NOLLABEL	Suppresses the label for the lifetime axis.
NOLTICK	Suppresses tick marks and tick mark labels for lifetime or analysis variable axis.
NOPLABEL	Suppresses the label for the probability axis.
NOPTICK	Suppresses tick marks and tick mark labels for probability axis.
NOSLABEL	Suppresses the label for the stress axis.
NOSTICK	Suppresses tick marks and tick mark labels for stress axis.

Table 18.52 Axis Options (continued)

Option	Option Description
NSTRESSTICK=*number*	Specifies the number of tick intervals for stress axis for relation plot.
NTICK=*number*	Specifies the number of tick intervals for the lifetime axis. This option has no effect if the LAXIS option is used.
PCTLOWER | PLOWER=*number*	Specifies the lower limit on the probability axis scale.
PCTUPPER | PUPPER=*number*	Specifies the upper limit on the probability axis scale.
STRESSLOWER | SLOWER=*number*	Specifies the lower limit on the stress axis scale.
STRESSUPPER | SUPPER=*number*	Specifies the upper limit on the stress axis scale.
PAXISLABEL='*string*'	Specifies a label for the probability axis.
WAXIS=*n*	Specifies the line thickness for axes and frame.

Table 18.53 lists options that control the appearance of the axes when you use ODS Graphics.

Table 18.53 Axis Options for ODS Graphics

Option	Option Description
LIFELOWER | LLOWER=*number*	Specifies the lower limit on the lifetime axis scale. The LLOWER option specifies *number* as the lower lifetime axis tick mark. The tick interval and the upper lifetime axis limit are determined automatically. This option has no effect if the LAXIS option is used.
LIFEUPPER | LUPPER=*number*	Specifies the upper limit on the lifetime axis scale. The LUPPER option specifies *number* as the upper lifetime axis tick mark. The tick interval and the lower lifetime axis limit are determined automatically. This option has no effect if the LAXIS option is used.
NOGRID	Suppresses grid lines.
PCTLOWER | PLOWER=*number*	Specifies lower the limit on the probability axis scale.
PCTUPPER | PUPPER=*number*	Specifies upper the limit on the probability axis scale.
STRESSLOWER | SLOWER=*number*	Specifies the lower limit on the stress axis scale.
STRESSUPPER | SUPPER=*number*	Specifies the upper limit on the stress axis scale.
PAXISLABEL='*string*'	Specifies a label for the probability axis.

Table 18.54 lists options that control the use of a graphics catalog to store graphs if you use traditional graphics. These options are not available if ODS Graphics is enabled.

Table 18.54 Graphics Catalog Options

Option	Option Description
DESCRIPTION='*string*'	Description for graphics catalog member
NAME='*string*'	Name for plot in graphics catalog

Table 18.55 lists options that control colors and patterns used in the graph when you use traditional graphics. These options are not available if ODS Graphics is enabled.

Table 18.55 Color and Pattern Options

Option	Option Description
CAXIS=*color*	Color for axis
CBOXES=*color*	Color for box frame for boxplots
CBOXFILL=*color*	Color for filling boxes for boxplots
CCENSOR=*color*	Color for filling censor plot area
CENCOLOR=*color*	Color for censor symbol
CFIT=*color* | (*color-list*)	color for fit lines and confidence curves in a probability plot
CFRAME=*color*	Color for frame
CGRID=*color*	Color for grid lines
CHREF=*color*	Color for HREF= lines
CPLOTFIT=*color* | (*color-list*)	colors for percentile lines
CSREF=*color*	Color for SREF= lines
CTEXT=*color*	Color for text
CVREF=*color*	Color for VREF= lines
RCENCOLOR=*color* | (*color-list*)	Colors for the symbols representing uncensored, right-censored, and left-censored observations in a relation plot

SLICE Statement

SLICE *model-effect* < / *options* > **;**

The SLICE statement provides a general mechanism for performing a partitioned analysis of the LS-means for an interaction. This analysis is also known as an analysis of simple effects.

The SLICE statement uses the same *options* as the LSMEANS statement, which are summarized in Table 19.21 (*SAS/STAT User's Guide*). For details about the syntax of the SLICE statement, see the section "SLICE Statement" (Chapter 19, *SAS/STAT User's Guide*).

STORE Statement

STORE < **OUT=** >*item-store-name* < / **LABEL=**'*label*' > **;**

The STORE statement requests that the procedure save the context and results of the statistical analysis. The resulting item store has a binary file format that cannot be modified. The contents of the item store can be processed with the PLM procedure. For details about the syntax of the STORE statement, see the section "STORE Statement" (Chapter 19, *SAS/STAT User's Guide*).

TEST Statement

TEST < *model-effects* > < / *options* > **;**

The TEST statement enables you to perform *F* tests for model effects that test Type I, Type II, or Type III hypotheses. See Chapter 15, "The Four Types of Estimable Functions" (*SAS/STAT User's Guide*), for details about the construction of Type I, II, and III estimable functions.

Table 18.56 summarizes the *options* available in the TEST statement.

Table 18.56 TEST Statement Options

Option	Description
CHISQ	Requests chi-square tests
DDF=	Specifies denominator degrees of freedom for fixed effects
E	Requests Type I, Type II, and Type III coefficients
E1	Requests Type I coefficients
E2	Requests Type II coefficients
E3	Requests Type III coefficients
HTYPE=	Indicates the type of hypothesis test to perform
INTERCEPT	Adds a row that corresponds to the overall intercept

For details about the syntax of the TEST statement, see the section "TEST Statement" (Chapter 19, *SAS/STAT User's Guide*) in Chapter 19, "Shared Concepts and Topics" (*SAS/STAT User's Guide*).

UNITID Statement

UNITID *variable* **;**

The UNITID statement names a *variable* in the input data set that is used to identify each individual unit in an MCFPLOT statement. The value of the UNITID variable for an observation corresponds to the name of the unit in the study for which a repair or end of history has occurred. See the section "Analysis of Recurrence Data on Repairs" on page 1215 for an example that uses the UNITID statement with the MCFPLOT statement.

Details: RELIABILITY Procedure

Abbreviations and Notation

The following abbreviations and notation are used in this section:

CDF	Cumulative distribution function: $F(x) = Pr\{X \leq x\}$
log	Base e logarithm
$\log_{10}$	Base 10 logarithm
Reliability or survivor function	$R(x) = Pr\{X > x\}$
x_p	$p \times 100\%$ percentile: $Pr\{X \leq x_p\} = p$

Types of Lifetime Data

This section describes various types of data that you can analyze with the RELIABILITY procedure.

Lifetime data for which the values of all sample units are observed are called *complete* data. This means that the failure times are observed for all units.

Many practical problems in life data analysis involve data for which some units are unfailed. The failure time for an unfailed unit is known only to be greater than the last running time. This type of data is said to be *right censored*, and the censoring time is used in the analysis of the data. Data for which censoring times are intermixed with failure times are sometimes called *multiply censored* or *progressively censored*.

Failure times may be known only to be less than some value. This type of data is called *left censored*.

Another common situation is where the failure times of units are not known exactly, but time intervals that contain the failure times are known. This type of data is called *interval censored*.

Interval-censored data for which all units share common interval endpoints are called *readout*, *inspection*, or *grouped* data.

Arbitrarily censored data can contain a combination of failures, right-, left-, and interval-censored data.

Probability Distributions

This section describes the probability distributions available in the RELIABILITY procedure for probability plotting and parameter estimation.

PROBPLOT and RELATIONPLOT Statements

Probability plots can be constructed for each of the probability distributions in Table 18.57.

For all distributions other than the three-parameter Weibull, estimates of two distribution parameters (*location* and *scale* or *scale* and *shape*) are computed by maximum likelihood or by least squares fitted to points on the probability plot. If one of the parameters is specified as fixed, the other is estimated. In addition, you can

specify a fixed *threshold*, or *shift*, parameter for distributions for which a threshold parameter is indicated in Table 18.57. If you do not specify a threshold parameter, the threshold is set to 0.

For the three-parameter Weibull distribution described in Table 18.57, the scale, shape, and threshold parameters are estimated by maximum likelihood.

You should not interpret the parameters μ and σ as representing the means and standard deviations for all of the distributions in Table 18.57. The normal is the only distribution in Table 18.57 for which this is the case.

Table 18.57 Distributions and Parameters for PROBPLOT and RELATIONPLOT Statements

		Parameters			
Distribution	**Density Function**	**Location**	**Scale**	**Shape**	**Threshold**
Normal	$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ		
Lognormal	$\frac{1}{\sqrt{2\pi}\sigma(x-\theta)}\exp\left(-\frac{(\log(x-\theta)-\mu)^2}{2\sigma^2}\right)$	μ	σ		θ
Lognormal (base 10)	$\frac{\log(10)}{\sqrt{2\pi}\sigma(x-\theta)}\exp\left(-\frac{(\log_{10}(x-\theta)-\mu)^2}{2\sigma^2}\right)$	μ	σ		θ
Extreme value	$\frac{1}{\sigma}\exp\left(\frac{x-\mu}{\sigma}\right)\exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right)$	μ	σ		
Weibull	$\frac{\beta}{\alpha^\beta}(x-\theta)^{\beta-1}\exp\left(-\left(\frac{x-\theta}{\alpha}\right)^\beta\right)$		α	β	θ
Exponential	$\frac{1}{\alpha}\exp\left(-\left(\frac{x-\theta}{\alpha}\right)\right)$		α		θ
Logistic	$\frac{\exp\left(\frac{x-\mu}{\sigma}\right)}{\sigma\left[1+\exp\left(\frac{x-\mu}{\sigma}\right)\right]^2}$	μ	σ		
Log-logistic	$\frac{\exp\left(\frac{\log(x-\theta)-\mu}{\sigma}\right)}{(x-\theta)\sigma\left[1+\exp\left(\frac{\log(x-\theta)-\mu}{\sigma}\right)\right]^2}$	μ	σ		θ
Three-parameter Weibull	$\frac{\beta}{\alpha^\beta}(x-\theta)^{\beta-1}\exp\left(-\left(\frac{x-\theta}{\alpha}\right)^\beta\right)$		α	β	θ
Gompertz	$\frac{\beta}{\alpha}\exp\left\{\frac{t}{\alpha}+\beta-\beta\exp\left(\frac{t}{\alpha}\right)\right\}$		α	β	
Three-parameter Gompertz	$\left\{\frac{\beta}{\alpha}\exp(\frac{t}{\alpha})+\frac{\theta}{\alpha}\right\}$ $\times\exp\left\{-\frac{\theta t}{\alpha}+\beta-\beta\exp(\frac{t}{\alpha})\right\}$		α	β	θ

The threshold parameter for the three-parameter Gompertz distribution is known as the Makeham mortality component and is labeled as such in the output. This component represents an additive risk that is constant over time. The exponential distribution shown in Table 18.57 is a special case of the Weibull distribution with $\beta = 1$. The remaining distributions in Table 18.57 are related to one another as shown in Table 18.58. The threshold parameter, θ, is assumed to be 0 in Table 18.58.

Table 18.58 Relationship among Life Distributions

Distribution of T	Parameters		Distribution of Y=logT	Parameters	
Lognormal	μ	σ	Normal	μ	σ
Weibull	α	β	Extreme value	$\mu = \log\alpha$	$\sigma = \frac{1}{\beta}$
Log-logistic	μ	σ	Logistic	μ	σ

MODEL Statement

All the distributions in Table 18.57 except the three-parameter Weibull and the Gompertz distributions are available for regression model estimation by using the MODEL statement. In addition, you can fit regression models with the generalized gamma distribution with the following probability density function $f(t)$:

$$f(t) = \frac{|\lambda|}{t\sigma\Gamma(\lambda^{-2})}(\lambda^{-2})^{(\lambda^{-2})}\exp\left[\lambda^{-2}\left(\lambda\left(\frac{\log(t)-\mu}{\sigma}\right) - \exp\left(\lambda\left(\frac{\log(t)-\mu}{\sigma}\right)\right)\right)\right]$$

If a lifetime T has the generalized gamma distribution, then the logarithm of the lifetime $X = \log(T)$ has the generalized log-gamma distribution, with the following probability density function $g(x)$. When the gamma distribution is specified, the logarithms of the lifetimes are used as responses, and the generalized log-gamma distribution is used to estimate the parameters by maximum likelihood.

$$g(x) = \frac{|\lambda|}{\sigma\Gamma(\lambda^{-2})}(\lambda^{-2})^{(\lambda^{-2})}\exp\left[\lambda^{-2}\left(\lambda\left(\frac{x-\mu}{\sigma}\right) - \exp\left(\lambda\left(\frac{x-\mu}{\sigma}\right)\right)\right)\right]$$

See Lawless (2003) and Meeker and Escobar (1998) for a description of the generalized gamma and generalized log-gamma distributions.

When $\lambda = 1$, the generalized log-gamma distribution reduces to the extreme value distribution with parameters μ and σ. In this case, the log lifetimes have the extreme value distribution, or, equivalently, the lifetimes have the Weibull distribution with parameters $\alpha = \exp(\mu)$ and $\beta = 1/\sigma$. When $\lambda = 0$, the generalized log-gamma reduces to the normal distribution with parameters μ and σ. In this case, the (unlogged) lifetimes have the lognormal distribution with parameters μ and σ. This chapter uses the notation μ for the *location*, σ for the *scale*, and λ for the *shape* parameters for the generalized log-gamma distribution.

ANALYZE Statement

You can use the ANALYZE statement to compute parameter estimates and other statistics for the distributions in Table 18.57. In addition, you can compute estimates for the binomial and Poisson distributions. The forms of these distributions are shown in Table 18.59.

Table 18.59 Binomial and Poisson Distributions

Distribution	Pr{Y=y}	Parameter	Parameter Name
Binomial	$\binom{n}{y} p^y (1-p)^{n-y}$	p	binomial probability
Poisson	$\frac{\mu^y}{y!} \exp(-\mu)$	μ	Poisson mean

Probability Plotting

Probability plots are useful tools for the display and analysis of lifetime data. See Abernethy (2006) for examples that use probability plots in the analysis of reliability data. Probability plots use a special scale so that a cumulative distribution function (CDF) plots as a straight line. Thus, if lifetime data are a sample from a distribution, the CDF estimated from the data plots approximately as a straight line on a probability plot for the distribution.

You can use the RELIABILITY procedure to construct probability plots for data that are complete, right censored, or interval censored (in readout form) for each of the probability distributions in Table 18.57.

A random variable Y belongs to a *location-scale* family of distributions if its CDF F is of the form

$$Pr\{Y \leq y\} = F(y) = G\left(\frac{y-\mu}{\sigma}\right)$$

where μ is the location parameter, and σ is the scale parameter. Here, G is a CDF that cannot depend on any unknown parameters, and G is the CDF of Y if $\mu = 0$ and $\sigma = 1$. For example, if Y is a normal random variable with mean μ and standard deviation σ,

$$G(u) = \Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

and

$$F(y) = \Phi\left(\frac{y-\mu}{\sigma}\right)$$

Of the distributions in Table 18.57, the normal, extreme value, and logistic distributions are location-scale models. As shown in Table 18.58, if T has a lognormal, Weibull, or log-logistic distribution, then $\log(T)$ has a distribution that is a location-scale model. Probability plots are constructed for lognormal, Weibull, and log-logistic distributions by using $\log(T)$ instead of T in the plots.

Let $y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n)}$ be ordered observations of a random sample with distribution function $F(y)$. A probability plot is a plot of the points $y_{(i)}$ against $m_i = G^{-1}(a_i)$, where $a_i = \hat{F}(y_i)$ is an estimate of the CDF $F(y_{(i)}) = G\left(\frac{y_{(i)}-\mu}{\sigma}\right)$. The points a_i are called *plotting positions*. The axis on which the points m_is are plotted is usually labeled with a probability scale (the scale of a_i).

If F is one of the location-scale distributions, then y is the lifetime; otherwise, the log of the lifetime is used to transform the distribution to a location-scale model.

If the data actually have the stated distribution, then $\hat{F} \approx F$,

$$m_i = G^{-1}(\hat{F}(y_i)) \approx G^{-1}\left(G\left(\frac{y_{(i)} - \mu}{\sigma}\right)\right) = \frac{y_{(i)} - \mu}{\sigma}$$

and points $(y_{(i)}, m_i)$ should fall approximately on a straight line.

There are several ways to compute plotting positions from failure data. These are discussed in the next two sections.

Complete and Right-Censored Data

The censoring times must be taken into account when you compute plotting positions for right-censored data. The RELIABILITY procedure provides several methods for computing plotting positions. These are specified with the PPOS= option in the ANALYZE, PROBPLOT, and RELATIONPLOT statements. All of the methods give similar results, as illustrated in the section "Expected Ranks, Kaplan-Meier, and Modified Kaplan-Meier Methods" on page 1314, the section "Nelson-Aalen" on page 1316, and the section "Median Ranks" on page 1316.

Expected Ranks, Kaplan-Meier, and Modified Kaplan-Meier Methods

Let $y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n)}$ be ordered observations of a random sample including failure times and censor times. Order the data in increasing order. Label all the data with reverse ranks r_i, with $r_1 = n, \ldots, r_n = 1$. For the failure corresponding to reverse rank r_i, compute the reliability, or survivor function estimate

$$R_i = \left[\frac{r_i}{r_i + 1}\right] R_{i-1}$$

with $R_0 = 1$. The expected rank plotting position is computed as $a_i = 1 - R_i$. The option PPOS=EXPRANK specifies the expected rank plotting position.

For the Kaplan-Meier method,

$$R_i = \left[\frac{r_i - 1}{r_i}\right] R_{i-1}$$

The Kaplan-Meier plotting position is then computed as $a'_i = 1 - R_i$. The option PPOS=KM specifies the Kaplan-Meier plotting position.

For the modified Kaplan-Meier method, use

$$R'_i = \frac{R_i + R_{i-1}}{2}$$

where R_i is computed from the Kaplan-Meier formula with $R_0 = 1$. The plotting position is then computed as $a''_i = 1 - R'_i$. The option PPOS=MKM specifies the modified Kaplan-Meier plotting position. If the PPOS option is not specified, the modified Kaplan-Meier plotting position is used as the default method.

For complete samples, $a_i = i/(n+1)$ for the expected rank method, $a'_i = i/n$ for the Kaplan-Meier method, and $a''_i = (i - .5)/n$ for the modified Kaplan-Meier method. If the largest observation is a failure for the Kaplan-Meier estimator, then $F_n = 1$ and the point is not plotted. These three methods are shown for the field winding data in Table 18.60 and Table 18.61.

Table 18.60 Expected Rank Plotting Position Calculations

Ordered Observation	Reverse Rank	$r_i/(r_i+1)$	$\times R_{i-1}$	$= R_i$	$a_i = 1 - R_i$
31.7	16	16/17	1.0000	0.9411	0.0588
39.2	15	15/16	0.9411	0.8824	0.1176
57.5	14	14/15	0.8824	0.8235	0.1765
65.0+	13				
65.8	12	12/13	0.8235	0.7602	0.2398
70.0	11	11/12	0.7602	0.6968	0.3032
75.0+	10				
75.0+	9				
87.5+	8				
88.3+	7				
94.2+	6				
101.7+	5				
105.8	4	4/5	0.6968	0.5575	0.4425
109.2+	3				
110.0	2	2/3	0.5575	0.3716	0.6284
130.0+	1				

+ Censored Times

Table 18.61 Kaplan-Meier and Modified Kaplan-Meier Plotting Position Calculations

Ordered Observation	Reverse Rank	$(r_i - 1)/r_i$	$\times R_{i-1}$	$= R_i$	$a_i' = 1 - R_i$	a_i''
31.7	16	15/16	1.0000	0.9375	0.0625	0.0313
39.2	15	14/15	0.9375	0.8750	0.1250	0.0938
57.5	14	13/14	0.8750	0.8125	0.1875	0.1563
65.0+	13					
65.8	12	11/12	0.8125	0.7448	0.2552	0.2214
70.0	11	10/11	0.7448	0.6771	0.3229	0.2891
75.0+	10					
75.0+	9					
87.5+	8					
88.3+	7					
94.2+	6					
101.7+	5					
105.8	4	3/4	0.6771	0.5078	0.4922	0.4076
109.2+	3					
110.0	2	1/2	0.5078	0.2539	0.7461	0.6192
130.0+	1					

+ Censored Times

Nelson-Aalen

Estimate the cumulative hazard function by

$$H_i = \frac{1}{r_i} + H_{i-1}$$

with $H_0 = 0$. The reliability is $R_i = \exp(-H_i)$, and the plotting position, or CDF, is $a_i''' = 1 - R_i$. You can show that $R_{KM} < R_{NA}$ for all ages. The Nelson-Aalen method is shown for the field winding data in Table 18.62.

Table 18.62 Nelson-Aalen Plotting Position Calculations

Ordered Observation	Reverse Rank	$1/r_i$	$+H_{i-1}$	$= H_i$	$a_i''' = 1 - \exp(-H_i)$
31.7	16	1/16	0.0000	0.0625	0.0606
39.2	15	1/15	0.0625	0.1292	0.1212
57.5	14	1/14	0.1292	0.2006	0.1818
65.0+	13				
65.8	12	1/12	0.2006	0.2839	0.2472
70.0	11	1/11	0.2839	0.3748	0.3126
75.0+	10				
75.0+	9				
87.5+	8				
88.3+	7				
94.2+	6				
101.7+	5				
105.8	4	1/4	0.3748	0.6248	0.4647
109.2+	3				
110.0	2	1/2	0.6248	1.1248	0.6753
130.0+	1				

+ Censored Times

Median Ranks

See Abernethy (2006) for a discussion of the methods described in this section. Let $y_{(1)} \le y_{(2)} \le \ldots \le y_{(n)}$ be ordered observations of a random sample including failure times and censor times. A failure order number j_i is assigned to the ith failure: $j_i = j_{i-1} + \Delta$, where $j_0 = 0$. The increment Δ is initially 1 and is modified when a censoring time is encountered in the ordered sample. The new increment is computed as

$$\Delta = \frac{(n+1) - \text{ previous failure order number}}{1 + \text{ number of items beyond previous censored item}}$$

The plotting position is computed for the ith failure time as

$$a_i = \frac{j_i - .3}{n + .4}$$

For complete samples, the failure order number j_i is equal to i, the order of the failure in the sample. In this case, the preceding equation for a_i is an approximation to the median plotting position computed as the median of the ith-order statistic from the uniform distribution on (0, 1). In the censored case, j_i is not necessarily an integer, but the preceding equation still provides an approximation to the median plotting position. The PPOS=MEDRANK option specifies the median rank plotting position.

For complete data, an alternative method of computing the median rank plotting position for failure i is to compute the exact median of the distribution of the ith order statistic of a sample of size n from the uniform distribution on (0,1). If the data are right censored, the adjusted rank j_i, as defined in the preceding paragraph, is used in place of i in the computation of the median rank. The PPOS=MEDRANK1 option specifies this type of plotting position.

Nelson (1982, p. 148) provides the following example of multiply right-censored failure data for field windings in electrical generators. Table 18.63 shows the data, the intermediate calculations, and the plotting positions calculated by exact (a'_i) and approximate (a_i) median ranks.

Table 18.63 Median Rank Plotting Position Calculations

Ordered Observation	Increment Δ	Failure Order Number j_i	a_i	a'_i
31.7	1.0000	1.0000	0.04268	0.04240
39.2	1.0000	2.0000	0.1037	0.1027
57.5	1.0000	3.0000	0.1646	0.1637
65.0+	1.0769			
65.8	1.0769	4.0769	0.2303	0.2294
70.0	1.0769	5.1538	0.2960	0.2953
75.0+	1.1846			
75.0+	1.3162			
87.5+	1.4808			
88.3+	1.6923			
94.2+	1.9744			
101.7+	2.3692			
105.8	2.3692	7.5231	0.4404	0.4402
109.2+	3.1590			
110.0	3.1590	10.6821	0.6331	0.6335
130.0+	6.3179			

+ Censored Times

Interval-Censored Data

Readout Data

The RELIABILITY procedure can create probability plots for interval-censored data when all units share common interval endpoints. This type of data is called *readout* data in the RELIABILITY procedure. Estimates of the cumulative distribution function are computed at times corresponding to the interval endpoints. Right censoring can also be accommodated if the censor times correspond to interval endpoints. See the section "Weibull Analysis of Interval Data with Common Inspection Schedule" on page 1189 for an example of a Weibull plot and analysis for interval data.

Table 18.64 illustrates the computational scheme used to compute the CDF estimates. The data are failure data for microprocessors (Nelson 1990, p. 147). In Table 18.64, t_i are the interval upper endpoints, in hours, f_i is the number of units failing in interval i, and n_i is the number of unfailed units at the beginning of interval i.

Note that there is right censoring as well as interval censoring in these data. For example, two units fail in the interval (24, 48) hours, and there are 1414 unfailed units at the beginning of the interval, 24 hours. At the beginning of the next interval, (48, 168) hours, there are 573 unfailed units. The number of unfailed units that are removed from the test at 48 hours is $1414 - 2 - 573 = 839$ units. These are right-censored units.

The reliability at the end of interval i is computed recursively as

$$R_i = (1 - (f_i/n_i))R_{i-1}$$

with $R_0 = 1$. The plotting position is $a_i = 1 - R_i$.

Table 18.64 Interval-Censored Plotting Position Calculations

Interval i	**Interval Endpoint** t_i	f_i/n_i	$R_i^{'} = 1 - (f_i/ni)$	$R_i = R_i^{'} R_{i-1}$	$a_i = 1 - R_i$
1	6	6/1423	0.99578	0.99578	.00421
2	12	2/1417	0.99859	0.99438	.00562
3	24	0/1415	1.00000	0.99438	.00562
4	48	2/1414	0.99859	0.99297	.00703
5	168	1/573	0.99825	0.99124	.00876
6	500	1/422	0.99763	0.98889	.01111
7	1000	2/272	0.99265	0.98162	.01838
8	2000	1/123	0.99187	0.97364	.02636

Arbitrarily Censored Data

The RELIABILITY procedure can create probability plots for data that consists of combinations of exact, left-censored, right-censored, and interval-censored lifetimes. Unlike the method in the previous section, failure intervals need not share common endpoints, although if the intervals share common endpoints, the two methods give the same results. The RELIABILITY procedure uses an iterative algorithm developed by Turnbull (1976) to compute a nonparametric maximum likelihood estimate of the cumulative distribution function for the data. Since the technique is maximum likelihood, standard errors of the cumulative probability estimates are computed from the inverse of the associated Fisher information matrix. A technique developed by Gentleman and Geyer (1994) is used to check for convergence to the maximum likelihood estimate. Also see Meeker and Escobar (1998, chap. 3) for more information.

Although this method applies to more general situations, where the intervals may be overlapping, the example of the previous section will be used to illustrate the method. Table 18.65 contains the microprocessor data of the previous section, arranged in intervals. A missing (.) lower endpoint indicates left censoring, and a missing upper endpoint indicates right censoring. These can be thought of as semi-infinite intervals with lower (upper) endpoint of negative (positive) infinity for left (right) censoring.

Table 18.65 Interval-Censored Data

Lower Endpoint	Upper Endpoint	Number Failed
.	6	6
6	12	2
24	48	2
24	.	1
48	168	1
48	.	839
168	500	1
168	.	150
500	1000	2
500	.	149
1000	2000	1
1000	.	147
2000	.	122

The following SAS statements compute the Turnbull estimate and create a lognormal probability plot:

```
data micro;
   input t1 t2 f ;
   datalines;
. 6 6
6 12 2
12 24 0
24 48 2
24 .  1
48 168 1
48 .   839
168 500 1
168 .   150
500 1000 2
500 .    149
1000 2000 1
1000 . 147
2000 . 122
;

proc reliability data=micro;
   distribution lognormal;
   freq f;
   pplot ( t1 t2 ) / itprintem
                     printprobs
                     maxitem = ( 1000, 25 )
                     nofit
                     npintervals = simul
                     ppout;
run;
```

The nonparametric maximum likelihood estimate of the CDF can only increase on certain intervals, and must remain constant between the intervals. The Turnbull algorithm first computes the intervals on which the nonparametric maximum likelihood estimate of the CDF can increase. The algorithm then iteratively estimates the probability associated with each interval. The ITPRINTEM option along with the PRINTPROBS option instructs the procedure to print the intervals on which probability increases can occur and the iterative history of the estimates of the interval probabilities. The PPOUT option requests tabular output of the estimated CDF, standard errors, and confidence limits for each cumulative probability.

Figure 18.55 shows every 25th iteration and the last iteration for the Turnbull estimate of the CDF for the microprocessor data. The initial estimate assigns equal probabilities to each interval. You can specify different initial values with the PROBLIST= option. The algorithm converges in 130 iterations for this data. Convergence is determined if the change in the loglikelihood between two successive iterations less than Δ, where the default value is $\Delta = 10^{-8}$. You can specify a different value for delta with the TOLLIKE= option. This algorithm is an example of an expectation-maximization (EM) algorithm. EM algorithms are known to converge slowly, but the computations within each iteration for the Turnbull algorithm are moderate. Iterations will be terminated if the algorithm does not converge after a fixed number of iterations. The default maximum number of iterations is 1000. Some data may require more iterations for convergence. You can specify the maximum allowed number of iterations with the MAXITEM= option in the PROBPLOT, ANALYZE, or RPLOT statement.

Figure 18.55 Iteration History for Turnbull Estimate

The RELIABILITY Procedure

Iteration History for the Turnbull Estimate of the CDF									
Iteration	**Loglikelihood**	**(., 6)**	**(6, 12)**	**(24, 48)**	**(48, 168)**	**(168, 500)**	**(500, 1000)**	**(1000, 2000)**	**(2000, .)**
0	-1133.4051	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
25	-104.16622	0.00421644	0.00140548	0.00140648	0.00173338	0.00237846	0.00846094	0.04565407	0.93474475
50	-101.15151	0.00421644	0.00140548	0.00140648	0.00173293	0.00234891	0.00727679	0.01174486	0.96986811
75	-101.06641	0.00421644	0.00140548	0.00140648	0.00173293	0.00234891	0.00727127	0.00835638	0.9732621
100	-101.06534	0.00421644	0.00140548	0.00140648	0.00173293	0.00234891	0.00727125	0.00801814	0.97360037
125	-101.06533	0.00421644	0.00140548	0.00140648	0.00173293	0.00234891	0.00727125	0.00798438	0.97363413
130	-101.06533	0.00421644	0.00140548	0.00140648	0.00173293	0.00234891	0.00727125	0.007983	0.97363551

If an interval probability is smaller than a tolerance (10^{-6} by default) after convergence, the probability is set to zero, the interval probabilities are renormalized so that they add to one, and iterations are restarted. Usually the algorithm converges in just a few more iterations. You can change the default value of the tolerance with the TOLPROB= option. You can specify the NOPOLISH option to avoid setting small probabilities to zero and restarting the algorithm.

If you specify the ITPRINTEM option, the table in Figure 18.56 summarizing the Turnbull estimate of the interval probabilities is printed. The columns labeled 'Reduced Gradient' and 'Lagrange Multiplier' are used in checking final convergence to the maximum likelihood estimate. The Lagrange multipliers must all be greater than or equal to zero, or the solution is not maximum likelihood. See Gentleman and Geyer (1994) for more details of the convergence checking.

Figure 18.56 Final Probability Estimates for Turnbull Algorithm

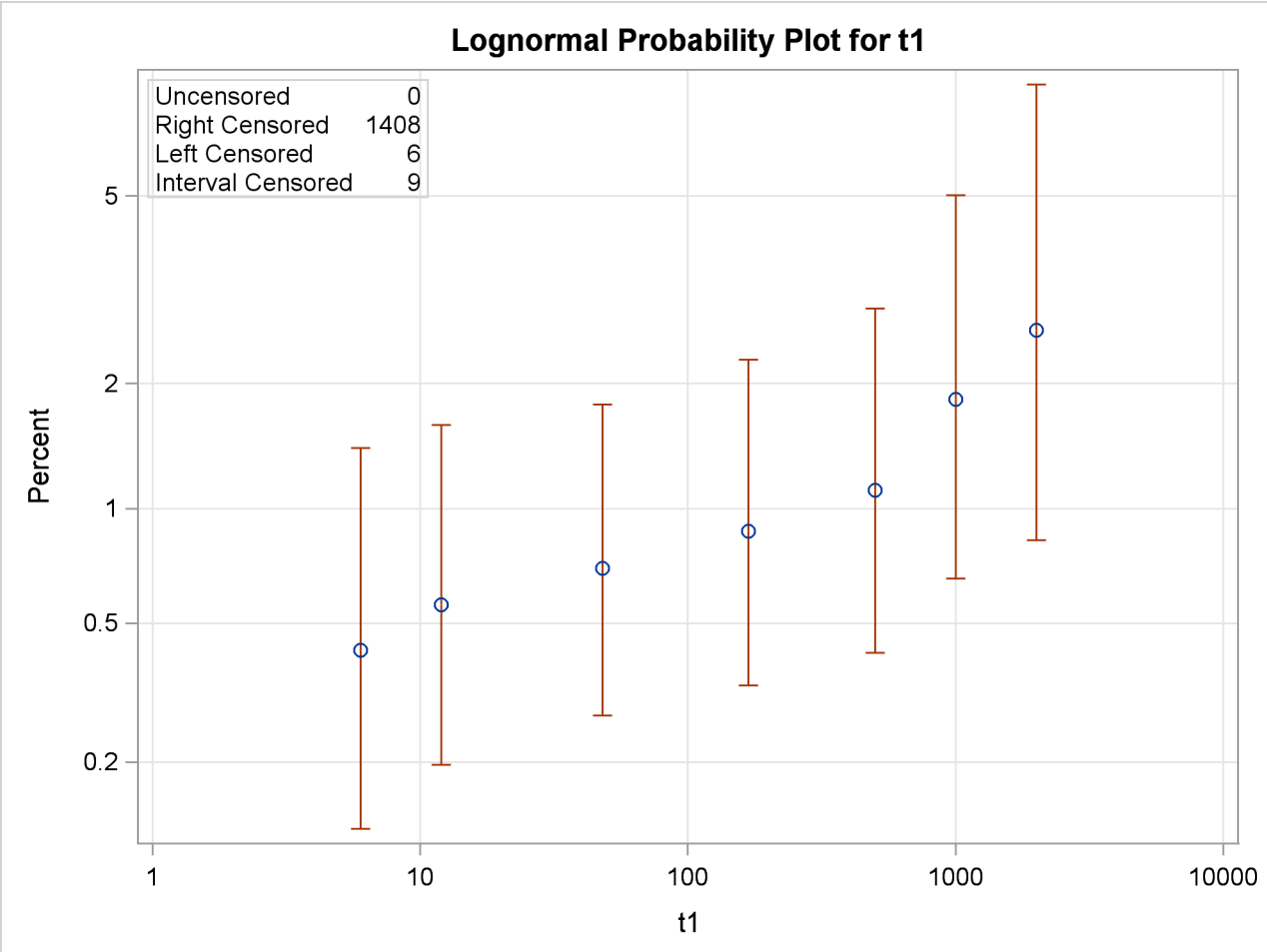
Lower Lifetime	Upper Lifetime	Probability	Reduced Gradient	Lagrange Multiplier
.	6	0.0042	0	0
6	12	0.0014	0	0
24	48	0.0014	0	0
48	168	0.0017	0	0
168	500	0.0023	0	0
500	1000	0.0073	-7.219342E-9	0
1000	2000	0.0080	-0.037063236	0
2000	.	0.9736	0.0003038877	0

Figure 18.57 shows the final estimate of the CDF, along with standard errors and confidence limits. Figure 18.58 shows the CDF and simultaneous confidence limits plotted on a lognormal probability plot.

Figure 18.57 Final CDF Estimates for Turnbull Algorithm

Cumulative Probability Estimates					
			Pointwise 95% Confidence Limits		
Lower Lifetime	Upper Lifetime	Cumulative Probability	Lower	Upper	Standard Error
6	6	0.0042	0.0019	0.0094	0.0017
12	24	0.0056	0.0028	0.0112	0.0020
48	48	0.0070	0.0038	0.0130	0.0022
168	168	0.0088	0.0047	0.0164	0.0028
500	500	0.0111	0.0058	0.0211	0.0037
1000	1000	0.0184	0.0094	0.0357	0.0063
2000	2000	0.0264	0.0124	0.0553	0.0101

Figure 18.58 Lognormal Probability Plot for the Microprocessor Data



Nonparametric Confidence Intervals for Cumulative Failure Probabilities

The method used in the RELIABILITY procedure for computation of approximate pointwise and simultaneous confidence intervals for cumulative failure probabilities relies on the Kaplan-Meier estimator of the cumulative distribution function of failure time and approximate standard deviation of the Kaplan-Meier estimator. For the case of arbitrarily censored data, the Turnbull algorithm, discussed previously, provides an extension of the Kaplan-Meier estimator.

For multiply censored data, the Kaplan-Meier estimator of the cumulative distribution function at failure time t_i is $\hat{F}(t_i) = 1 - \hat{S}(t_i)$, where

$$\hat{S}(t_i) = \prod_{j=1}^{i} (1 - \hat{p}_j),$$

$$\hat{p}_i = \frac{d_i}{n_i},$$

d_i is the number of failures in the interval (t_{i-1}, t_i) , and n_i is the number of unfailed units at the beginning of the interval. This definition of the Kaplan-Meier estimator is equivalent to the one previously given.

An estimator of the variance v_i of the Kaplan-Meier estimator $\hat{F}(t_i)$ is given by

$$\hat{v}_i = [\hat{S}(t_i)]^2 \sum_{j=1}^{i} \frac{\hat{p}_j}{n_j(1-\hat{p}_j)}$$

An estimator of the standard deviation of $\hat{F}(t_i)$ is $\text{se}_{\hat{F}} = \sqrt{\hat{v}_i}$.

For arbitrarily censored data, the Kaplan-Meier estimator is replaced by the nonparametric maximum likelihood estimator computed with the Turnbull algorithm, and the approximate variance of the estimator of $F(t_i)$ is computed from the inverse of the Fisher information matrix.

Pointwise Confidence Intervals

Approximate $(1-\alpha)100\%$ pointwise confidence intervals are computed as in Meeker and Escobar (1998, section 3.6) as

$$[F_L, F_U] = \left[\frac{\hat{F}}{\hat{F} + (1-\hat{F})w}, \quad \frac{\hat{F}}{\hat{F} + (1-\hat{F})/w} \right]$$

where

$$w = \exp\left[\frac{z_{1-\alpha/2}\text{se}_{\hat{F}}}{(\hat{F}(1-\hat{F}))} \right]$$

where z_p is the pth quantile of the standard normal distribution.

Simultaneous Confidence Intervals

Approximate $(1-\alpha)100\%$ simultaneous confidence bands valid over the lifetime interval (t_a, t_b) are computed as the "Equal Precision" case of Nair (1984) and Meeker and Escobar (1998, section 3.8)

$$[F_L, F_U] = \left[\frac{\hat{F}}{\hat{F} + (1-\hat{F})w}, \quad \frac{\hat{F}}{\hat{F} + (1-\hat{F})/w} \right]$$

where

$$w = \exp\left[\frac{e_{a,b,1-\alpha/2}\text{se}_{\hat{F}}}{(\hat{F}(1-\hat{F}))} \right]$$

where the factor $x = e_{a,b,1-\alpha/2}$ is the solution of

$$x \exp(-x^2/2) \log\left[\frac{(1-a)b}{(1-b)a}\right] / \sqrt{8\pi} = \alpha/2$$

The time interval (t_a, t_b) over which the bands are valid depends in a complicated way on the constants a and b defined in Nair (1984), $0 < a < b < 1$. a and b are chosen by default, so that the confidence bands are valid between the lowest and highest times corresponding to failures in the case of multiply censored data, or, to the lowest and highest intervals for which probabilities are computed for arbitrarily censored data. You can optionally specify a and b directly with the NPINTERVALS=SIMULTANEOUS(a,b) option in the PROBPLOT statement.

Parameter Estimation and Confidence Intervals

Maximum Likelihood Estimation

Maximum likelihood estimation of the parameters of a statistical model involves maximizing the likelihood or, equivalently, the log likelihood with respect to the parameters. The parameter values at which the maximum occurs are the maximum likelihood estimates of the model parameters. The likelihood is a function of the parameters and of the data.

Let $x_1, x_2, \ldots, x_n$ be the observations in a random sample, including the failures and censoring times (if the data are censored). Let $f(\boldsymbol{\theta}; x)$ be the probability density of failure time, $S(\boldsymbol{\theta}; x) = Pr\{X \geq x\}$ be the reliability function, and $F(\boldsymbol{\theta}; x) = Pr\{X \leq x\}$ be the cumulative distribution function, where $\boldsymbol{\theta}$ is the vector of parameters to be estimated, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$. The probability density, reliability function, and CDF are determined by the specific distribution selected as a model for the data. The log likelihood is defined as

$$\begin{aligned} L(,,) \quad = \quad & \sum_i \log(f(\boldsymbol{\theta}; x_i)) + \sum_i{}' \log(S(\boldsymbol{\theta}; x_i)) + \\ & \sum_i{}'' \log(F(\boldsymbol{\theta}; x_i)) + \sum_i{}''' [\log(F(\boldsymbol{\theta}; x_{ui}) - F(\boldsymbol{\theta}; x_{li}))] \end{aligned}$$

where

- $\sum$ is the sum over failed units
- $\sum'$ is the sum over right-censored units
- $\sum''$ is the sum over left-censored units
- $\sum'''$ is the sum over interval-censored units

and (x_{li}, x_{ui}) is the interval in which the ith unit is interval censored. Only the sums appropriate to the type of censoring in the data are included when the preceding equation is used.

The RELIABILITY procedure maximizes the log likelihood with respect to the parameters $\boldsymbol{\theta}$ by using a Newton-Raphson algorithm. The Newton-Raphson algorithm is a recursive method for computing the maximum of a function. On the rth iteration, the algorithm updates the parameter vector $\boldsymbol{\theta}_r$ with

$$\boldsymbol{\theta}_{r+1} = \boldsymbol{\theta}_r - \mathbf{H}^{-1}\mathbf{g}$$

where $\mathbf{H}$ is the Hessian (second derivative) matrix, and $\mathbf{g}$ is the gradient (first derivative) vector of the log-likelihood function, both evaluated at the current value of the parameter vector. That is,

$$\mathbf{g} = [g_j] = \left[\frac{\partial L}{\partial \theta_j}\right]\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}$$

and

$$\mathbf{H} = [h_{ij}] = \left[\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}\right]\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}$$

Iteration continues until the parameter estimates converge. The convergence criterion is

$$|\theta_i^{r+1} - \theta_i^r| \le c \quad \text{if} \quad |\theta_i^{r+1}| < 0.01$$

$$\left|\frac{\theta_i^{r+1} - \theta_i^r}{\theta_i^{r+1}}\right| \le c \quad \text{if} \quad |\theta_i^{r+1}| \ge 0.01$$

for all $i = 1, 2, \ldots, p$ where c is the convergence criterion. The default value of c is 0.001, and it can be specified with the CONVERGE= option in the MODEL, PROBPLOT, RELATIONPLOT, and ANALYZE statements.

After convergence by the preceding criterion, the quantity

$$tc = \frac{\mathbf{g}\mathbf{H}^{-1}\mathbf{g}}{L}$$

is computed. If $tc > d$ then a warning is printed that the algorithm did not converge. tc is called the *relative Hessian* convergence criterion. The default value of d is 0.0001. You can specify other values for d with the CONVH= option. The relative Hessian criterion is useful in detecting the occasional case where no progress can be made in increasing the log likelihood, yet the gradient $\mathbf{g}$ is not zero.

A location-scale model has a CDF of the form

$$F(x) = G\left(\frac{x-\mu}{\sigma}\right)$$

where μ is the location parameter, σ is the scale parameter, and G is a standardized form ($\mu = 0, \sigma = 1$) of the cumulative distribution function. The parameter vector is $\boldsymbol{\theta}$=(μ σ). It is more convenient computationally to maximize log likelihoods that arise from location-scale models. If you specify a distribution from Table 18.57 that is not a location-scale model, it is transformed to a location-scale model by taking the natural

(base e) logarithm of the response. If you specify the lognormal base 10 distribution, the logarithm (base 10) of the response is used. The Weibull, lognormal, and log-logistic distributions in Table 18.57 are not location-scale models. Table 18.58 shows the corresponding location-scale models that result from taking the logarithm of the response.

Maximum likelihood is the default method of estimating the location and scale parameters in the MODEL, PROBPLOT, RELATIONPLOT, and ANALYZE statements. If the Weibull distribution is specified, the logarithms of the responses are used to obtain maximum likelihood estimates ($\hat{\mu}$ $\hat{\sigma}$) of the location and scale parameters of the extreme value distribution. The maximum likelihood estimates ($\hat{\alpha}$, $\hat{\beta}$) of the Weibull scale and shape parameters are computed as $\hat{\alpha} = \exp(\hat{\mu})$ and $\hat{\beta} = 1/\hat{\sigma}$.

Maximum likelihood estimates for the Gompertz distributions are obtained by expressing the log-likelihood in terms of $\log(\alpha)$, $\log(\beta)$, and (if applicable) $\log(\theta)$. After the log likelihood is maximized, parameter estimates and their standard errors are transformed from the logarithm metric to the standard metric by using the delta method.

Three-Parameter Weibull

The parameters of the three-parameter Weibull distribution are estimated by maximizing the log likelihood function. The threshold parameter θ must be less than the minimum failure time t_0, unless $\beta = 1$, in which case, θ can be equal to t_0. The RELIABILITY procedure sets a default upper bound of $t_0 - 0.001$ for the threshold in the iterative estimation computations and a default lower bound of 0.0. You can set different bounds by specifying an INEST data set as described in the section "INEST Data Set for the Three-Parameter Weibull" on page 1326.

If the shape parameter β is less than one, then the density function in Table 18.57 has a singularity at $t = \theta$, and the log likelihood is unbounded above as the threshold parameter approaches the minimum failure time t_0. For any fixed $\theta < t_0$, maximum likelihood estimates of the scale and shape parameters α and β exist. If $\hat{\beta} < 1$ in the iterative estimation procedure, the estimate of the threshold θ is set to the upper bound and maximum likelihood estimates of α and β are computed.

INEST Data Set for the Three-Parameter Weibull

You can specify a SAS data set to set lower bounds, upper bounds, equality constraints, or initial values for estimating the parameters of a three-parameter Weibull distribution by using the INEST= option in the ANALYZE or PROBPLOT statement. The data set must contain a variable named _TYPE_ that specifies the action that you want to take in the iterative estimation process, and some combination of variables named _SCALE_, _SHAPE_, and _THRESHOLD_ that represent the distribution parameters. If BY processing is used, the INEST= data set should also include the BY variables, and there must be at least one observation for each BY group.

The possible values of _TYPE_ and corresponding actions are summarized in Table 18.66.

Table 18.66 _TYPE_ Variable Values

Value of _TYPE_	Action
LB	Lower bound
UB	Upper bound
EQ	Equality
PARMS	Initial value

For example, you can use the INEST data set In created by using the following SAS statements to specify bounds for data that contain the BY variable Group with three BY groups: A, B, and D. The data set In specifies a lower bound for the threshold parameter of –100 for groups A, B, and D, and an upper bound of 3 for the threshold parameter for group D. Since the variables _Scale_ and _Shape_ are set to missing, no action is taken for them, and these variables could be omitted from the data set.

```
data In;
   input Group$1 _Type_$ 2-11 _Scale_ _Shape_ _Threshold_;
   datalines;
A    lb    . . -100
B    lb    . . -100
D    lb    . . -100
D    ub    . .  3
;
```

Regression Models

You can specify a regression model by using the MODEL statement. For example, if you want to relate the lifetimes of electronic parts in a test to Arrhenius-transformed operating temperature, then an appropriate model might be

$$\mu_i = \beta_0 + x_i \beta_1$$

where $x_i = 1000/(T_i + 273.15)$, and T_i is the centigrade temperature at which the ith unit is tested. Here, $\mathbf{x}_i' = [\ 1\ x_i\ ]$.

There are two types of explanatory variables: *continuous* variables and *classification* variables. Continuous variables represent physical quantities, such as temperature or voltage, and they must be numeric. Continuous explanatory variables are sometimes called *covariates*.

Classification variables identify classification levels and are declared in the CLASS statement. These are also referred to as *categorical, dummy, qualitative, discrete,* or *nominal* variables. Classification variables can be either character or numeric. The values of classification variables are called *levels*. For example, the classification variable Batch could have levels ‘batch1’ and ‘batch2’ to identify items from two production batches. An indicator (0-1) variable is generated for each level of a classification variable and is used as an explanatory variable. See Nelson (1990, p. 277) for an example that uses an indicator variable in the analysis of accelerated life test data. In a model, an explanatory variable that is not declared in a CLASS statement is assumed to be continuous.

By default, all regression models automatically contain an intercept term; that is, the model is of the form

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \ldots$$

where β_0 does not have an explanatory variable multiplier. The intercept term can be excluded from the model by specifying INTERCEPT=0 as a MODEL statement option.

For numerical stability, continuous explanatory variables are centered and scaled internally to the procedure. This transforms the parameters $\boldsymbol{\beta}$ in the original model to a new set of parameters. The parameter estimates $\boldsymbol{\beta}$ and covariances are transformed back to the original scale before reporting, so that the parameters should be

interpreted in terms of the originally specified model. Covariates that are indicator variables—that is, those specified in a CLASS statement—are not centered and scaled.

Initial values of the regression parameters used in the Newton-Raphson method are computed by ordinary least squares. The parameters $\boldsymbol{\beta}$ and the scale parameter σ are jointly estimated by maximum likelihood, taking a logarithmic transformation of the responses, if necessary, to get a location-scale model.

The generalized gamma distribution is fit using log lifetime as the response variable. The regression parameters $\boldsymbol{\beta}$, the scale parameter σ, and the shape parameter λ are jointly estimated.

The Weibull distribution shape parameter estimate is computed as $\hat{\beta} = 1/\hat{\sigma}$, where σ is the scale parameter from the corresponding extreme value distribution. The Weibull scale parameter $\hat{\alpha}_i = \exp(\mathbf{x}'\hat{\boldsymbol{\beta}})$ is not computed by the procedure. Instead, the regression parameters $\boldsymbol{\beta}$ and the shape β are reported.

In a model with one to three continuous explanatory variables x, you can use the RELATION= option in the MODEL statement to specify a transformation that is applied to the variables before model fitting. Table 18.67 shows the available transformations.

Table 18.67 Variable Transformations

Relation	Transformed variable
ARRHENIUS (Nelson parameterization)	$1000/(x + 273.15)$
ARRHENIUS2 (activation energy parameterization)	$11605/(x + 273.15)$
POWER	$\log(x), \quad x > 0$
LINEAR	x
LOGISTIC	$\log\left(\frac{x}{1-x}\right), \quad 0 < x < 1$

Nonconstant Scale Parameter

In some situations, it is desirable for the scale parameter to change with the values of explanatory variables. For example, Meeker and Escobar (1998, section 17.5) present an analysis of accelerated life test data where the spread of the data is greater at lower levels of the stress. You can use the LOGSCALE statement to specify the scale parameter as a function of explanatory variables. You must also have a MODEL statement to specify the location parameter. Explanatory variables can be continuous variables, indicator variables specified in the CLASS statement, or any interaction combination. The variables can be the same as specified in the MODEL statement, or they can be different variables. Any transformation specified with the RELATION= MODEL statement option will be applied to the same variable appearing in the LOGSCALE statement. See the section "Regression Model with Nonconstant Scale" on page 1205 for an example of fitting a model with nonconstant scale parameter.

The form of the model for the scale parameter is

$$\log(\sigma_i) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$$

where β_0 is the intercept term. The intercept term can be excluded from the model by specifying INTERCEPT=0 as a LOGSCALE statement option.

The parameters $\beta_0, \beta_1, \ldots, \beta_p$ are estimated by maximum likelihood jointly with all the other parameters in the model.

Stable Parameters

The location and scale parameters (μ, σ) are estimated by maximizing the likelihood function by numerical methods, as described previously. An alternative parameterization that is likely to have better numerical properties for heavy censoring is (η, σ), where $\eta = \mu + z_p\sigma$ and z_p is the pth quantile of the standardized distribution. See Meeker and Escobar (1998, p. 90) and Doganaksoy and Schmee (1993) for more details on alternate parameterizations.

By default, RELIABILITY estimates a value of z_p from the data that will improve the numerical properties of the estimation. You can also specify values of p from which the value of z_p will be computed with the PSTABLE= option in the ANALYZE, PROBPLOT, RELATIONPLOT, or MODEL statement. Note that a value of p = 0.632 for the Weibull and extreme value and p = 0.5 for all other distributions will give $z_p = 0$ and the parameterization will then be the usual location-scale parameterization.

All estimates and related statistics are reported in terms of the location and scale parameters (μ, σ). If you specify the ITPRINT option in the ANALYZE, PROBPLOT, or RELATIONPLOT statement, a table showing the values of p, ν, σ, and the last evaluation of the gradient and Hessian for these parameters is produced.

Covariance Matrix

An estimate of the covariance matrix of the maximum likelihood estimators (MLEs) of the parameters $\boldsymbol{\theta}$ is given by the inverse of the negative of the matrix of second derivatives of the log likelihood, evaluated at the final parameter estimates:

$$\boldsymbol{\Sigma} = [\sigma_{ij}] = -\mathbf{H}^{-1} = -\left[\frac{\partial^2 L}{\partial\theta_i\,\partial\theta_j}\right]^{-1}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

The negative of the matrix of second derivatives is called the observed Fisher information matrix. The diagonal term σ_{ii} is an estimate of the variance of $\hat{\theta}_i$. Estimates of standard errors of the MLEs are provided by

$$\mathrm{SE}_{\theta_i} = \sqrt{\sigma_{ii}}$$

An estimator of the correlation matrix is

$$\mathbf{R} = \left[\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}\right]$$

The covariance matrix for the Weibull distribution parameter estimators is computed by a first-order approximation from the covariance matrix of the estimators of the corresponding extreme value parameters (μ, σ) as

$$\begin{aligned}
\mathrm{Var}(\hat{\alpha}) &= [\exp(\hat{\mu})]^2\mathrm{Var}(\hat{\mu}) \\
\mathrm{Var}(\hat{\beta}) &= \frac{\mathrm{Var}(\hat{\sigma})}{\hat{\sigma}^4} \\
\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) &= -\frac{\exp(\hat{\mu})}{\hat{\sigma}^2}\mathrm{Cov}(\hat{\mu}, \hat{\sigma})
\end{aligned}$$

For the regression model, the variance of the Weibull shape parameter estimator $\hat{\beta}$ is computed from the variance of the estimator of the extreme value scale parameter σ as shown previously. The covariance of the regression parameter estimator $\hat{\beta}_i$ and the Weibull shape parameter estimator $\hat{\beta}$ is computed in terms of the covariance between $\hat{\beta}_i$ and $\hat{\sigma}$ as

$$\mathrm{Cov}(\hat{\beta}_i, \hat{\beta}) = -\frac{\mathrm{Cov}(\hat{\beta}_i, \hat{\sigma})}{\hat{\sigma}^2}$$

Confidence Intervals for Distribution Parameters

Table 18.68 shows the method of computation of approximate two-sided $\gamma \times 100\%$ confidence limits for distribution parameters. The default value of confidence is $\gamma = 0.95$. Other values of confidence are specified using the CONFIDENCE= option. In Table 18.68, K_γ represents the $(1 + \gamma)/2 \times 100\%$ percentile of the standard normal distribution, and $\hat{\mu}$ and $\hat{\sigma}$ are the MLEs of the location and scale parameters for the normal, extreme value, and logistic distributions. For the lognormal, Weibull, and log-logistic distributions, $\hat{\mu}$ and $\hat{\sigma}$ represent the MLEs of the corresponding location and scale parameters of the location-scale distribution that results when the logarithm of the lifetime is used as the response. For the Weibull distribution, μ and σ are the location and scale parameters of the extreme value distribution for the logarithm of the lifetime. $\mathrm{SE}_{\hat{\theta}}$ denotes the standard error of the MLE of θ, computed as the square root of the appropriate diagonal element of the inverse of the Fisher information matrix.

For the Gompertz distributions, estimation of all parameters takes place in the logarithm metric. For example, a confidence interval for the logarithm of scale is computed as $\log(\hat{\alpha}) \pm K_\gamma \mathrm{SE}_{\log(\hat{\alpha})}$. The confidence interval in the standard metric is then obtained by taking e to the power equal to each endpoint.

Table 18.68 Confidence Limit Computation

	Parameters		
Distribution	**Location** μ or **Threshold** θ	**Scale**	**Shape**
Normal	$\mu_L = \hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})$ $\mu_U = \hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})$	$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	
Lognormal	$\mu_L = \hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})$ $\mu_U = \hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})$	$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	
Lognormal (base 10)	$\mu_L = \hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})$ $\mu_U = \hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})$	$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	
Extreme value	$\mu_L = \hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})$ $\mu_U = \hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})$	$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	
Weibull		$\alpha_L = \exp[\hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})]$ $\alpha_U = \exp[\hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})]$	$\beta_L = \exp[-K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]/\hat{\sigma}$ $\beta_U = \exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]/\hat{\sigma}$
Exponential		$\alpha_L = \exp[\hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})]$ $\alpha_U = \exp[\hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})]$	
Logistic	$\mu_L = \hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})$ $\mu_U = \hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})$	$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	
Log-logistic	$\mu_L = \hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})$ $\mu_U = \hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})$	$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	
Generalized Gamma		$\sigma_L = \hat{\sigma}/\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$ $\sigma_U = \hat{\sigma}\exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]$	$\mu_L = \hat{\lambda} - K_\gamma(\mathrm{SE}_{\hat{\lambda}})$ $\mu_U = \hat{\lambda} + K_\gamma(\mathrm{SE}_{\hat{\lambda}})$
Three-parameter Weibull	$\theta_L = \hat{\theta} - K_\gamma(\mathrm{SE}_{\hat{\theta}})$ $\theta_U = \hat{\theta} + K_\gamma(\mathrm{SE}_{\hat{\theta}})$	$\alpha_L = \exp[\hat{\mu} - K_\gamma(\mathrm{SE}_{\hat{\mu}})]$ $\alpha_U = \exp[\hat{\mu} + K_\gamma(\mathrm{SE}_{\hat{\mu}})]$	$\beta_L = \exp[-K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]/\hat{\sigma}$ $\beta_U = \exp[K_\gamma(\mathrm{SE}_{\hat{\sigma}})/\hat{\sigma}]/\hat{\sigma}$

Regression Parameters Approximate $\gamma \times 100\%$ confidence limits for the regression parameter β_i are given by

$$\beta_{iL} = \hat{\beta}_i - K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$$

$$\beta_{iU} = \hat{\beta}_i + K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$$

Percentiles

The maximum likelihood estimate of the $p \times 100\%$ percentile x_p for the extreme value, normal, and logistic distributions is given by

$$\hat{x}_p = \hat{\mu} + z_p \hat{\sigma}$$

where $z_p = G^{-1}(p)$, G is the standardized CDF shown in Table 18.69, and $(\hat{\mu}, \hat{\sigma})$ are the maximum likelihood estimates of the location and scale parameters of the distribution. The maximum likelihood estimate of the percentile t_p for the Weibull, lognormal, and log-logistic distributions is given by

$$\hat{t}_p = \exp[\hat{\mu} + z_p \hat{\sigma}]$$

where $z_p = G^{-1}(p)$, and G is the standardized CDF of the location-scale model corresponding to the logarithm of the response. For the lognormal (base 10) distribution,

$$\hat{t}_p = 10^{[\hat{\mu} + z_p \hat{\sigma}]}$$

The maximum likelihood estimate of the percentile t_p for the three-parameter Weibull distribution is computed by

$$\hat{t}_p = \hat{\theta} + \exp[\hat{\mu} + z_p \hat{\sigma}]$$

where $z_p = G^{-1}(p)$, and G is the standardized CDF of extreme value distribution.

The maximum likelihood estimate of the percentile t_p for the standard Gompertz distribution is computed by

$$\hat{t}_p = \hat{\alpha} \log\{1 - \hat{\beta}^{-1} \log(1 - p)\}$$

Because the quantile function depends on Lambert's W function and has no closed form, the percentile for the three-parameter Gompertz distribution is obtained by using a bisection algorithm to solve the following equation:

$$p = F(\hat{t}_p) = 1 - \exp\left\{\hat{\beta} - \frac{\hat{\theta}\hat{t}_p}{\hat{\alpha}} - \hat{\beta} \exp\left(\frac{\hat{t}_p}{\hat{\alpha}}\right)\right\}$$

Table 18.69 Standardized Cumulative Distribution Functions

Distribution	Location-Scale Distribution	Location-Scale CDF
Weibull	Extreme value	$1 - \exp[-\exp(z)]$
Lognormal	Normal	$\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$
Log-logistic	Logistic	$\frac{\exp(z)}{1+\exp(z)}$

Confidence Intervals The variance of the MLE of the $p \times 100\%$ percentile for the normal, extreme value, or logistic distribution is

$$\mathrm{Var}(\hat{x}_p) = \mathrm{Var}(\hat{\mu}) + z_p^2 \mathrm{Var}(\hat{\sigma}) + 2 z_p \mathrm{Cov}(\hat{\mu}, \hat{\sigma})$$

Two-sided approximate $100\gamma\%$ confidence limits for x_p are

$$\begin{aligned} x_{pL} &= \hat{x}_p - K_\gamma \sqrt{\mathrm{Var}(\hat{x}_p)} \\ x_{pU} &= \hat{x}_p + K_\gamma \sqrt{\mathrm{Var}(\hat{x}_p)} \end{aligned}$$

where K_γ represents the $100(1+\gamma)/2 \times 100\%$ percentile of the standard normal distribution.

The limits for the lognormal, Weibull, or log-logistic distributions are

$$\begin{aligned} t_{pL} &= \exp\left(\hat{x}_p - K_\gamma \sqrt{\mathrm{Var}(\hat{x}_p)}\right) \\ t_{pU} &= \exp\left(\hat{x}_p + K_\gamma \sqrt{\mathrm{Var}(\hat{x}_p)}\right) \end{aligned}$$

where x_p refers to the percentile of the corresponding location-scale distribution (normal, extreme value, or logistic) for the logarithm of the lifetime. For the lognormal (base 10) distribution,

$$\begin{aligned} t_{pL} &= 10^{\left(\hat{x}_p - K_\gamma \sqrt{\mathrm{Var}(\hat{x}_p)}\right)} \\ t_{pU} &= 10^{\left(\hat{x}_p + K_\gamma \sqrt{\mathrm{Var}(\hat{x}_p)}\right)} \end{aligned}$$

Approximate limits for the three-parameter Weibull distribution are computed as

$$\begin{aligned} t_{pL} &= \hat{\theta} + \exp\left(\hat{x}_p - K_\gamma \sqrt{\mathrm{Var}(\hat{x}_p)}\right) \\ t_{pU} &= \hat{\theta} + \exp\left(\hat{x}_p + K_\gamma \sqrt{\mathrm{Var}(\hat{x}_p)}\right) \end{aligned}$$

where x_p refers to the percentile of the standard extreme value distribution.

For the Gompertz distributions, confidence limits are computed as

$$\begin{aligned} t_{pL} &= \hat{t}_p - K_\gamma \sqrt{\mathrm{Var}(\hat{t}_p)} \\ t_{pU} &= \hat{t}_p + K_\gamma \sqrt{\mathrm{Var}(\hat{t}_p)} \end{aligned}$$

where $\mathrm{Var}(\hat{t}_p)$ is calculated by the delta method. Because the quantile function has no closed form, the derivatives that are required for the three-parameter Gompertz distribution are obtained by the numerical method of finite differencing.

Reliability Function

For the extreme value, normal, and logistic distributions shown in Table 18.69, the maximum likelihood estimate of the reliability function $R(x) = \Pr\{X > x\}$ is given by

$$\hat{R}(x) = 1 - F\left(\frac{x - \hat{\mu}}{\hat{\sigma}}\right)$$

For the Gompertz distributions, the MLE of the reliability function is

$$\hat{R}(x) = \exp\left\{\hat{\beta} - \frac{\hat{\theta}x}{\hat{\alpha}} - \hat{\beta}\exp\left(\frac{x}{\hat{\alpha}}\right)\right\}$$

where $\hat{\theta} = 0$ for the standard, two-parameter Gompertz distribution.

The MLE of the CDF is $\hat{F}(x) = 1 - \hat{R}(x)$.

Confidence Intervals Let $\hat{u} = \frac{x-\hat{\mu}}{\hat{\sigma}}$. The approximate variance of u is

$$\mathrm{Var}(\hat{u}) \approx \frac{\mathrm{Var}(\hat{\mu}) + \hat{u}^2\mathrm{Var}(\hat{\sigma}) + 2\hat{u}\mathrm{Cov}(\hat{\mu}, \hat{\sigma})}{\hat{\sigma}^2}$$

Two-sided approximate $\gamma \times 100\%$ confidence intervals for $R(x)$ are computed as

$$R_L(x) = \hat{R}(u_2)$$

$$R_U(x) = \hat{R}(u_1)$$

where

$$u_1 = \hat{u} - K_\gamma\sqrt{\mathrm{Var}(\hat{u})}$$

$$u_2 = \hat{u} + K_\gamma\sqrt{\mathrm{Var}(\hat{u})}$$

and K_γ represents the $(1 + \gamma)/2 \times 100\%$ percentile of the standard normal distribution. The corresponding limits for the CDF are

$$F_L(x) = 1 - R_U(x)$$

$$F_U(x) = 1 - R_L(x)$$

For the Gompertz distributions, confidence intervals for $R(x)$ are computed as

$$R_L(x) = \exp\{-\exp(u_2)\}$$

$$R_U(x) = \exp\{-\exp(u_1)\}$$

where $\hat{u} = \log[-\log\{\hat{R}_G(x)\}]$, $R_G(x)$ is the Gompertz reliability function, and

$$u_1 = \hat{u} - K_\gamma\sqrt{\mathrm{Var}(\hat{u})}$$

$$u_2 = \hat{u} + K_\gamma\sqrt{\mathrm{Var}(\hat{u})}$$

The variance term $\mathrm{Var}(\hat{u})$ is obtained by the delta method from the covariance matrix of the original parameter estimates.

As an alternative, you can request that two-sided $\gamma \times 100\%$ likelihood ratio confidence limits for the reliability function and the CDF be computed by specifying the LRCLSURV option in the ANALYZE statement or the LRCLSURV option in the PROBPLOT statement.

Limits for the Weibull, lognormal, and log-logistic reliability function $R(t)$ are the same as those for the corresponding extreme value, normal, or logistic reliability $R(y)$, where $y = \log(t)$. Limits for the three-parameter Weibull use $y = \log(t - \hat{\theta})$ and the extreme value CDF.

You can create a table containing estimates of the reliability function, the CDF, and confidence limits computed as described in this section with the SURVTIME= option in the ANALYZE statement or with the SURVTIME= option in the PROBPLOT statement. You can plot confidence limits for the CDF on probability plots created with the PROBPLOT statement with the PINTERVALS=CDF option in the PROBPLOT statement. PINTERVALS=CDF is the default option for parametric confidence limits on probability plots.

Estimation with the Binomial and Poisson Distributions

In addition to estimating the parameters of the distributions in Table 18.57, you can estimate parameters, compute confidence limits, compute predicted values and prediction limits, and compute chi-square tests for differences in groups for the binomial and Poisson distributions by using the ANALYZE statement. Specify either BINOMIAL or POISSON in the DISTRIBUTION statement to use one of these distributions. The ANALYZE statement options available for the binomial and Poisson distributions are given in Table 18.5. See the section "Analysis of Binomial Data" on page 1230 for an example of an analysis of binomial data.

Binomial Distribution

If r is the number of successes and n is the number of trials in a binomial experiment, then the maximum likelihood estimator of the probability p in the binomial distribution in Table 18.59 is computed as

$$\hat{p} = r/n$$

Two-sided $\gamma \times 100\%$ confidence limits for p are computed as in Johnson, Kotz, and Kemp (1992, p. 130):

$$p_L = \frac{\nu_1 F[(1-\gamma)/2; \nu_1, \nu_2]}{\nu_2 + \nu_1 F[(1-\gamma)/2; \nu_1, \nu_2]}$$

with $\nu_1 = 2r$ and $\nu_2 = 2(n - r + 1)$ and

$$p_U = \frac{\nu_1 F[(1+\gamma)/2; \nu_1, \nu_2]}{\nu_2 + \nu_1 F[(1+\gamma)/2; \nu_1, \nu_2]}$$

with $\nu_1 = 2(r + 1)$ and $\nu_2 = 2(n - r)$, where $F[\gamma; \nu_1, \nu_2]$ is the $\gamma \times 100\%$ percentile of the F distribution with ν_1 degrees of freedom in the numerator and ν_2 degrees of freedom in the denominator.

You can compute a sample size required to estimate p within a specified tolerance w with probability γ. Nelson (1982, p. 206) gives the following formula for the approximate sample size:

$$n \approx \hat{p}(1 - \hat{p}) \left(\frac{K_\gamma}{w} \right)^2$$

where K_γ is the $(1 + \gamma)/2 \times 100\%$ percentile of the standard normal distribution. The formula is based on the normal approximation for the distribution of $\hat{p}$. Nelson recommends using this formula if $np > 10$

and $np(1-p) > 10$. The value of γ used for computing confidence limits is used in the sample size computation. The default value of confidence is $\gamma = 0.95$. Other values of confidence are specified using the CONFIDENCE= option. You specify a tolerance of *number* with the TOLERANCE(*number*) option.

The predicted number of successes X in a future sample of size m, based on the previous estimate of p, is computed as

$$\hat{X} = m(r/n) = m\hat{p}$$

Two-sided approximate $\gamma \times 100)\%$ prediction limits are computed as in Nelson (1982, p. 208). The prediction limits are the solutions X_L and X_U of

$$X_U/m = [(r+1)/n]F[(1+\gamma)/2; 2(r+1), 2X_U]$$

$$m/(X_L+1) = (n/r)F[(1+\gamma)/2; 2(X_L+1), 2r]$$

where $F[\gamma; \nu_1, \nu_2]$ is the $\gamma \times 100\%$ percentile of the F distribution with ν_1 degrees of freedom in the numerator and ν_2 degrees of freedom in the denominator. You request predicted values and prediction limits for a future sample of size *number* with the PREDICT(*number*) option.

You can test groups of binomial data for equality of their binomial probability by using the ANALYZE statement. You specify the K groups to be compared with a group variable having K levels.

Nelson (1982, p. 450) discusses a chi-square test statistic for comparing K binomial proportions for equality. Suppose there are r_i successes in n_i trials for $i = 1, 2, \ldots, K$. The grouped estimate of the binomial probability is

$$\hat{p} = \frac{r_1 + r_2 + \cdots + r_K}{n_1 + n_2 + \cdots + n_K}$$

The chi-square test statistic for testing the hypothesis $p_1 = p_2 = \ldots = p_K$ against $p_i \neq p_j$ for some i and j is

$$Q = \sum_{i=1}^{K} \frac{(r_i - n_i\hat{p})^2}{n_i\hat{p}(1-\hat{p})}$$

The statistic Q has an asymptotic chi-square distribution with $K-1$ degrees of freedom. The RELIABILITY procedure computes the contribution of each group to Q, the value of Q, and the p-value for Q based on the limiting chi-square distribution with $K-1$ degrees of freedom. If you specify the PREDICT option, predicted values and prediction limits are computed for each group, as well as for the pooled group. The p-value is defined as $p_0 = 1 - \chi^2_{K-1}[Q]$, where $\chi^2_{K-1}[x]$ is the chi-square CDF with $K-1$ degrees of freedom, and Q is the observed value. A test of the hypothesis of equal binomial probabilities among the groups with significance level α is

- $p_0 > \alpha$: do not reject the equality hypothesis
- $p_0 \leq \alpha$: reject the equality hypothesis

Poisson Distribution

You can use the ANALYZE statement to model data by using the Poisson distribution. The data consist of a count Y of occurrences in a "length" of observation T. Observation T is typically an *exposure time*, but it can have other units, such as distance. The ANALYZE statement enables you to compute the rate of occurrences, confidence limits, and prediction limits.

An estimate of the rate λ is computed as

$$\hat{\lambda} = Y/T$$

Two-sided $\gamma \times 100\%$ confidence limits for λ are computed as in Nelson (1982, p. 201):

$$\lambda_L = 0.5\chi^2[(1-\gamma)/2; 2Y]/T$$

$$\lambda_U = 0.5\chi^2[(1+\gamma)/2; 2(Y+1)]/T$$

where $\chi^2[\delta; \nu]$ is the $\delta \times 100\%$ percentile of the chi-square distribution with ν degrees of freedom.

You can compute a length T required to estimate λ within a specified tolerance w with probability γ. Nelson (1982, p. 202) provides the following approximate formula:

$$\hat{T} \approx \hat{\lambda}\left(\frac{K_\gamma}{w}\right)^2$$

where K_γ is the $(1+\gamma)/2 \times 100\%$ percentile of the standard normal distribution. The formula is based on the normal approximation for $\hat{\lambda}$ and is more accurate for larger values of λT. Nelson recommends using the formula when $\lambda T > 10$. The value of γ used for computing confidence limits is also used in the length computation. The default value of confidence is $\gamma = 0.95$. Other values of confidence are specified using the CONFIDENCE= option. You specify a tolerance of *number* with the TOLERANCE(*number*) option.

The predicted future number of occurrences in a length S is

$$\hat{X} = (Y/T)S = \hat{\lambda}S$$

Two-sided approximate $\gamma \times 100\%$ prediction limits are computed as in Nelson (1982, p. 203). The prediction limits are the solutions X_L and X_U of

$$X_U/S = [(Y+1)/T]F[(1+\gamma)/2; 2(Y+1), 2X_U]$$

$$S/(X_L+1) = (T/Y)F[(1+\gamma)/2; 2(X_L+1), 2Y]$$

where $F[\gamma; \nu_1, \nu_2]$ is the $\gamma \times 100\%$ percentile of the F distribution with ν_1 degrees of freedom in the numerator and ν_2 degrees of freedom in the denominator. You request predicted values and prediction limits for a future exposure *number* with the PREDICT(*number*) option.

You can compute a chi-square test statistic for comparing K Poisson rates for equality. You specify the K groups to be compared with a group variable having K levels.

See Nelson (1982, p. 444) for more information. Suppose that there are Y_i Poisson counts in lengths T_i for $i = 1, 2, \ldots, K$ and that the Y_i are independent. The grouped estimate of the Poisson rate is

$$\hat{\lambda} = \frac{Y_1 + Y_2 + \cdots + Y_K}{T_1 + T_2 + \cdots + T_K}$$

The chi-square test statistic for testing the hypothesis $\lambda_1 = \lambda_2 = \ldots = \lambda_K$ against $\lambda_i \neq \lambda_j$ for some i and j is

$$Q = \sum_{i=1}^{K} \frac{(Y_i - \hat{\lambda} T_i)^2}{\hat{\lambda} T_i}$$

The statistic Q has an asymptotic chi-square distribution with $K - 1$ degrees of freedom. The RELIABILITY procedure computes the contribution of each group to Q, the value of Q, and the p-value for Q based on the limiting chi-square distribution with $K - 1$ degrees of freedom. If you specify the PREDICT option, predicted values and prediction limits are computed for each group, as well as for the pooled group. The p-value is defined as $p_0 = 1 - \chi^2_{K-1}[Q]$, where $\chi^2_{K-1}[x]$ is the chi-square CDF with $K - 1$ degrees of freedom and Q is the observed value. A test of the hypothesis of equal Poisson rates among the groups with significance level α is

- $p_0 > \alpha$: accept the equality hypothesis
- $p_0 \leq \alpha$: reject the equality hypothesis

Least Squares Fit to the Probability Plot

Fitting to the probability plot by least squares is an alternative to maximum likelihood estimation of the parameters of a life distribution. Only the failure times are used. A least squares fit is computed using points $(x_{(i)}, m_i)$, where $m_i = F^{-1}(a_i)$ and a_i are the plotting positions as defined in the section "Probability Plotting" on page 1313. The x_i are either the lifetimes for the normal, extreme value, or logistic distributions or the log lifetimes for the lognormal, Weibull, or log-logistic distributions. The ANALYZE, PROBPLOT, or RELATIONPLOT statement option FITTYPE=LSXY specifies the $x_{(i)}$ as the dependent variable ('y-coordinate') and the m_i as the independent variable ('x-coordinate'). You can optionally reverse the quantities used as dependent and independent variables by specifying the FITTYPE=LSYX option.

Weibayes Estimation

Weibayes estimation is a method of performing a Weibull analysis when there are few or no failures. The FITTYPE=WEIBAYES option requests this method. The method of Nelson (1985) is used to compute a one-sided confidence interval for the Weibull scale parameter when the Weibull shape parameter is specified. See Abernethy (2006) for more discussion and examples. The Weibull shape parameter β is assumed to be known and is specified to the procedure with the SHAPE=*number* option. Let $T_1, T_2, \ldots, T_n$ be the failure and censoring times, and let $r \geq 0$ be the number of failures in the data. If there are no failures ($r = 0$), a lower $\gamma \times 100\%$ confidence limit for the Weibull scale parameter α is computed as

$$\alpha_L = \{\sum_{i=1}^{n} T_i^{\beta} / [-\log(1 - \gamma)]\}^{1/\beta}$$

The default value of confidence is $\gamma = 0.95$. Other values of confidence are specified using the CONFIDENCE= option.

If $r \geq 1$, the MLE of α is given by

$$\hat{\alpha} = [\sum_{i=1}^{n} T_i^{\beta}/r]^{1/\beta}$$

and a lower $\gamma \times 100\%$ confidence limit for the Weibull scale parameter α is computed as

$$\alpha_L = \hat{\alpha}[2r/\chi^2(\gamma, 2r+2)]^{1/\beta}$$

where $\chi^2(\gamma, 2r+2)$ is the γ percentile of a chi-square distribution with $2r+2$ degrees of freedom. The procedure uses the specified value of β and the computed value of α_L to compute distribution percentiles and the reliability function.

Estimation With Multiple Failure Modes

In many applications, units can experience multiple causes of failure, or *failure modes*. For example, in the section "Weibull Probability Plot for Two Combined Failure Modes" on page 1211, insulation specimens can experience either early failures due to manufacturing defects or degradation failures due to aging. The FMODE statement is used to analyze this type of data. See the section "FMODE Statement" on page 1253 for the syntax of the FMODE statement. This section describes the analysis of data when units experience multiple failure modes.

The assumptions used in the analysis are

- a cause, or mode, can be identified for each failure
- failure modes follow a series-system model; i.e., a unit fails when a failure due to one of the modes occurs
- each failure mode has the specified lifetime distribution with different parameters
- failure modes act statistically independently

Suppose there are m failure modes, with lifetime distribution functions $F_1(t), F_2(t), \ldots, F_m(t)$.

If you wish to estimate the lifetime distribution of a failure mode, say mode i, acting alone, specify the KEEP keyword in the FMODE statement. The failures from all other modes are treated as right-censored observations, and the lifetime distribution is estimated by one of the methods described in other sections, such as maximum likelihood. This lifetime distribution is interpreted as the distribution if the specified failure mode is acting alone, with all other modes eliminated. You can also specify more than one mode to KEEP, but the assumption is that all the specified modes have the same distribution.

If you specify the ELIMINATE keyword, failures due to the specified modes are treated as right censored. The resulting distribution estimate is the failure distribution if the specified modes are eliminated.

If you specify the COMBINE keyword, the failure distribution when all the modes specified in the FMODE statement modes act is estimated. The failure distribution $F_i(t), i = 1, 2, \ldots, m$, from each individual

mode is first estimated by treating all failures from other modes as right censored. The estimated failure distributions are then combined to get an estimate of the lifetime distribution when all modes act,

$$\hat{F}(t) = 1 - \prod_{i=1}^{m} [1 - \hat{F}_i(t)]$$

Pointwise approximate asymptotic normal confidence limits for $F(t)$ can be obtained by the delta method. See Meeker and Escobar (1998, appendix B.2). The delta method variance of $\hat{F}(t)$ is, assuming independence of failure modes,

$$\mathrm{Var}(\hat{F}(t)) = \sum_{i=1}^{m} [S_0(u_1)S_0(u_2)\ldots f_0(u_i)\ldots S_0(u_m)]^2 \mathrm{Var}(u_i)$$

where $u_i = \frac{y-\hat{\mu}_i}{\hat{\sigma}_i}$, y is t for the extreme value, normal, and logistic distributions or $\log(t)$ for the Weibull, lognormal or log-logistic distributions, $\hat{\mu}_i$ and $\hat{\sigma}_i$ are location and scale parameter estimates for mode i, and S_0 and f_0 are the standard ($\mu = 0, \sigma = 1$) survival function and density function for the specified distribution.

Two-sided approximate $(1-\alpha)100\%$ pointwise confidence intervals are computed as in Meeker and Escobar (1998, section 3.6) as

$$[F_L, F_U] = \left[\frac{\hat{F}}{\hat{F} + (1-\hat{F})w}, \quad \frac{\hat{F}}{\hat{F} + (1-\hat{F})/w} \right]$$

where

$$w = \exp\left[\frac{z_{1-\alpha/2}\mathrm{se}_{\hat{F}}}{(\hat{F}(1-\hat{F}))} \right]$$

where $\mathrm{se}_{\hat{F}} = \sqrt{\mathrm{Var}(\hat{F}(t))}$ and z_p is the pth quantile of the standard normal distribution.

Regression Model Statistics Computed for Each Observation for Lifetime Data

This section describes statistics that are computed for each observation when you fit a model for lifetime data. For regression models that are fit using the MODEL statement, you can specify a variety of statistics to be computed for each observation in the input data set. This section describes the method of computation for each statistic. See Table 18.32 and Table 18.33 for the syntax to request these statistics.

Predicted Values

The linear predictor is

$$\hat{\mu}_i = \mathbf{x}'\hat{\boldsymbol{\beta}}$$

where $\mathbf{x}_i$ is the vector of explanatory variables for the ith observation.

Percentiles

An estimator of the $p \times 100\%$ percentile x_p for the ith observation for the extreme value, normal, and logistic distributions is

$$\hat{x}_{i,p} = \mathbf{x}'\hat{\boldsymbol{\beta}} + z_p\hat{\sigma}$$

where $z_p = G^{-1}(p)$, G is the standardized CDF, and σ is the distribution scale parameter.

An estimator of the $p \times 100\%$ percentile t_p for the ith observation for the Weibull, lognormal, and log-logistic distributions is

$$\hat{t}_{i,p} = \exp[\mathbf{x}'\hat{\boldsymbol{\beta}} + z_p\hat{\sigma}]$$

where G is the standardized CDF of the extreme value, normal, or logistic distribution that corresponds to the logarithm of the lifetime, and σ is the distribution scale parameter.

The percentile of the lognormal (base 10) distribution is

$$\hat{t}_{i,p} = 10^{[\mathbf{x}'\hat{\boldsymbol{\beta}}+z_p\hat{\sigma}]}$$

where G is the CDF of the standard normal distribution.

An estimator of the $p \times 100\%$ percentile t_p for the ith observation for the generalized gamma distribution is

$$\hat{t}_{i,p} = \exp[\mathbf{x}'\hat{\boldsymbol{\beta}} + w_{\lambda,p}\hat{\sigma}]$$

where

$$w_{\lambda,p} = \frac{1}{\lambda}\log\left(\frac{\lambda^2}{2}\chi^2_{(2/\lambda^2),p}\right)$$

and $\chi^2_{k,p}$ is the $p \times 100\%$ percentile of the chi-square distribution with k degrees of freedom.

Standard Errors of Percentile Estimator

For the extreme value, normal, and logistic distributions, the standard error of the estimator of the $p \times 100\%$ percentile is computed as

$$\sigma_{i,p} = \sqrt{\mathbf{z}'\boldsymbol{\Sigma}\mathbf{z}}$$

where

$$\mathbf{z} = \begin{bmatrix} \mathbf{x}_i \\ z_p \end{bmatrix}$$

and $\boldsymbol{\Sigma}$ is the covariance matrix of $(\hat{\boldsymbol{\beta}}, \hat{\sigma})$.

For the Weibull, lognormal, and log-logistic distributions, the standard error is computed as

$$\sigma_{i,p} = \exp(x_{i,p})\sqrt{\mathbf{z}'\boldsymbol{\Sigma}\mathbf{z}}$$

where $x_{i,p}$ is the percentile computed from the extreme value, normal, or logistic distribution that corresponds to the logarithm of the lifetime. The standard error for the lognormal (base 10) distribution is computed as

$$\sigma_{i,p} = 10^{x_{i,p}}\sqrt{\mathbf{z}'\boldsymbol{\Sigma}\mathbf{z}}$$

The standard error for the generalized gamma distribution percentile is computed as

$$\sigma_{i,p} = \exp[\mathbf{x}'\hat{\boldsymbol{\beta}} + w_{\lambda,p}\hat{\sigma}]\sqrt{\mathbf{z}'\boldsymbol{\Sigma}\mathbf{z}}$$

where

$$\mathbf{z} = \begin{bmatrix} \mathbf{x}_i \\ w_{\lambda,p} \\ \hat{\sigma}\frac{\partial w_{\lambda,p}}{\partial \lambda} \end{bmatrix}$$

$\boldsymbol{\Sigma}$ is the covariance matrix of $(\hat{\boldsymbol{\beta}}, \hat{\sigma}, \hat{\lambda})$, $\boldsymbol{\beta}$ is the vector of regression parameters, σ is the scale parameter, and λ is the shape parameter.

Confidence Limits for Percentiles

Two-sided approximate $100\gamma\%$ confidence limits for $x_{i,p}$ for the extreme value, normal, and logistic distributions are computed as

$$\begin{aligned} x_L &= \hat{x}_{i,p} - K_\gamma \sigma_{i,p} \\ x_U &= \hat{x}_{i,p} + K_\gamma \sigma_{i,p} \end{aligned}$$

where K_γ represents the $100(1+\gamma)/2 \times 100\%$ percentile of the standard normal distribution.

Limits for the Weibull, lognormal, and log-logistic percentiles are computed as

$$\begin{aligned} t_L &= \exp(x_L) \\ t_U &= \exp(x_U) \end{aligned}$$

where x_L and x_U are computed from the corresponding distributions for the logarithms of the lifetimes. For the lognormal (base 10) distribution,

$$\begin{aligned} t_L &= 10^{x_L} \\ t_U &= 10^{x_U} \end{aligned}$$

Limits for the generalized gamma distribution percentiles are computed as

$$t_L = \exp\left[\mathbf{x}_i'\boldsymbol{\beta} + w_{\lambda,p}\hat{\sigma} - K_\gamma\sqrt{\mathbf{z}'\boldsymbol{\Sigma}\mathbf{z}}\right]$$
$$t_U = \exp\left[\mathbf{x}_i'\boldsymbol{\beta} + w_{\lambda,p}\hat{\sigma} + K_\gamma\sqrt{\mathbf{z}'\boldsymbol{\Sigma}\mathbf{z}}\right]$$

Reliability Function

For the extreme value, normal, and logistic distributions, an estimate of the reliability function evaluated at the response y_i is computed as

$$R(y_i) = 1 - G\left(\frac{y_i - \mathbf{x}'\hat{\boldsymbol{\beta}}}{\hat{\sigma}}\right)$$

where $G(x)$ is the standardized CDF of the distribution from Table 18.69.

Estimates of the reliability function evaluated at the response t_i for the Weibull, lognormal, log-logistic, and generalized gamma distributions are computed as

$$R(t_i) = 1 - G\left(\frac{\log(t_i) - \mathbf{x}'\hat{\boldsymbol{\beta}}}{\hat{\sigma}}\right)$$

where $G(x)$ is the standardized CDF of the corresponding extreme value, normal, logistic, or generalized log-gamma distributions.

Residuals

The RELIABILITY procedure computes several different kinds of residuals. In the following equations, y_i represents the ith response value if the extreme value, normal, or logistic distributions are specified. If t_i is the ith response and if the Weibull, lognormal, log-logistic, or generalized gamma distributions are specified, then y_i represents the logarithm of the response $y_i = \log(t_i)$. If the lognormal (base 10) distribution is specified, then $y_i = \log_{10}(t_i)$.

Raw Residuals

The raw residual is computed as

$$r_{Ri} = y_i - \mathbf{x}'\hat{\boldsymbol{\beta}}$$

Standardized Residuals

The standardized residual is computed as

$$r_{Si} = \frac{y_i - \mathbf{x}'\hat{\boldsymbol{\beta}}}{\hat{\sigma}}$$

Adjusted Residuals

If an observation is right censored, then the standardized residual for that observation is also right censored. Adjusted residuals adjust censored standardized residuals upward by adding a percentile of the residual lifetime distribution, given that the standardized residual exceeds the censoring value. The default percentile is the median (50th percentile), but you can, optionally, specify a $\gamma \times 100\%$ percentile by using the RESIDALPHA=γ option in the MODEL statement. The $\gamma \times 100$ percentile residual life is computed as in Joe and Proschan (1984). The adjusted residual is computed as

$$r_{Ai} = \begin{cases} G^{-1}[1 - (1-\gamma)S(u_i)] & \text{for right-censored observations} \\ u_i & \text{for uncensored observations} \end{cases}$$

where G is the standard CDF,

$$S(u) = 1 - G(u)$$

is the reliability function, and

$$u_i = \frac{y_i - \mathbf{x}'\hat{\boldsymbol{\beta}}}{\hat{\sigma}}$$

If the generalized gamma distribution is specified, the standardized CDF and reliability functions include the estimated shape parameter $\hat{\lambda}$.

Modified Cox-Snell Residuals

Let

$$\delta_i = \begin{cases} 1 & \text{for uncensored observations} \\ 0 & \text{for right-censored observations} \end{cases}$$

The Cox-Snell residual is defined as

$$r_{Ci} = -\log(R(y_i))$$

where

$$R(y) = 1 - G\left(\frac{y - \mathbf{x}'\hat{\boldsymbol{\beta}}}{\hat{\sigma}}\right)$$

is the reliability function. The modified Cox-Snell residual is computed as in Collett (1994, p. 152):

$$r'_{Ci} = r_{Ci} + (1 - \delta_i)\alpha$$

where α is an adjustment factor. If the fitted model is correct, the Cox-Snell residual has approximately a standard exponential distribution for uncensored observations. If an observation is censored, the residual evaluated at the censoring time is not as large as the residual evaluated at the (unknown) failure time. The adjustment factor α adjusts the censored residuals upward to account for the censoring. The default is $\alpha = 1.0$, the mean of the standard exponential distribution. You can, optionally, specify any adjustment factor by using the MODEL statement option RESIDADJ=α. Another commonly used value is the median of the standard exponential distribution, $\alpha = 0.693$.

Deviance Residuals

Deviance residuals are a zero-mean, symmetrized version of modified Cox-Snell residuals. Deviance residuals are computed as in Collett (1994, p. 153):

$$r_{Di} = \text{sgn}(\delta_i - r_{Ci})\{-2[\delta_i - r_{Ci} + \delta_i \log(r_{Ci})]\}^{1/2}$$

where

$$\text{sgn}(u) = \begin{cases} -1 & \text{if } u < 0 \\ 1 & \text{if } u \geq 0 \end{cases}$$

Regression Model Statistics Computed for Each Observation for Recurrent Events Data

This section describes statistics that are computed for each observation when you fit a model for recurrent events data. For regression models that are fit using the MODEL statement, you can specify a variety of statistics to be computed for each observation in the input data set. This section describes the method of computation for each statistic. See Table 18.32 and Table 18.34 for the syntax to request these statistics.

Let t_i be the event time in the ith observation in the input data set. The following statistics use the definitions of the mean function $\mathrm{M}(t;\eta,\beta)$ and intensity function $\lambda(t;\eta,\beta)$ in Table 18.72, where η and β are replaced by their maximum likelihood estimates. The shape parameter β is assumed to be constant for all observations. For regression models, the scale parameter η in Table 18.72 for the ith observation is

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \ldots$$

where $x_{i1}, x_{i2}, \ldots$ are regression coefficients and $\beta_0, \beta_1, \ldots$ are the maximum likelihood estimates of the regression parameters.

Predicted Values of Scale Parameter

The scale parameter that is predicted by the model for the ith observation is

$$\hat{\eta}_i = \mathbf{x}_i' \hat{\boldsymbol{\beta}}$$

where $\mathbf{x}_i$ is the vector of explanatory variables for the ith observation and $\boldsymbol{\beta}$ is the vector of maximum likelihood estimates of the regression parameters.

Mean Function

The predicted mean function is computed as $\mathrm{M}(t_i, \hat{\eta}_i, \hat{\beta})$.

Confidence Limits for the Mean Function

Confidence limits for the estimated $\mathrm{M}(t_i)$ are computed as described in the section "NHPP Model Parameter Confidence Limit Computation" on page 1354, using t_i, $\hat{\eta}_i$, and $\hat{\beta}$.

Standard Error of the Mean Function

The standard error of the estimated $\mathrm{M}(t_i)$ is computed as described in the section "NHPP Model Parameter Confidence Limit Computation" on page 1354, using t_i, $\hat{\eta}_i$, and $\hat{\beta}$.

Intensity Function

The predicted intensity function is computed as $\lambda(t_i, \hat{\eta}_i, \hat{\beta})$.

Confidence Limits for the Intensity Function

Confidence limits for the estimated $\lambda(t_i)$ are computed as described in the section "NHPP Model Parameter Confidence Limit Computation" on page 1354, using t_i, $\hat{\eta}_i$, and $\hat{\beta}$.

Standard Error of the Intensity Function

The standard error of the estimated $\lambda(t_i)$ is computed as described in the section "NHPP Model Parameter Confidence Limit Computation" on page 1354, using t_i, $\hat{\eta}_i$, and $\hat{\beta}$.

Recurrence Data from Repairable Systems

Failures in a system that can be repaired are sometimes modeled as *recurrence data*, or *recurrent events data*. When a repairable system fails, it is repaired and placed back in service. As a repairable system ages, it accumulates a history of repairs and costs of repairs. The mean cumulative function (MCF) $\mathrm{M}(t)$ is defined as the population mean of the cumulative number (or cost) of repairs up until time t. You can use the RELIABILITY procedure to compute and plot nonparametric estimates and plots of the MCF for the number of repairs or the cost of repairs. The Nelson (1995) confidence limits for the MCF are also computed and plotted. You can compute and plot estimates of the difference of two MCFs and confidence intervals. This is useful for comparing the repair performance of two systems.

See Nelson (2002, 1995, 1988), Doganaksoy and Nelson (1998), and Nelson and Doganaksoy (1989) for discussions and examples of nonparametric analysis of recurrence data.

You can also fit a parametric model for recurrent event data and display the resulting model on a plot, along with nonparametric estimates of the MCF.

See Rigdon and Basu (2000), Tobias and Trindade (1995), and Meeker and Escobar (1998) for discussions of parametric models for recurrent events data.

Nonparametric Analysis

Recurrent Events Data with Exact Ages

See the section "Analysis of Recurrence Data on Repairs" on page 1215 and the section "Comparison of Two Samples of Repair Data" on page 1220 for examples of the analysis of recurrence data with exact ages.

Formulas for the MCF estimator $\hat{\mathrm{M}}(t)$ and the variance of the estimator $\mathrm{Var}(\hat{\mathrm{M}}(t))$ are given in Nelson (1995). Table 18.70 shows a set of artificial repair data from Nelson (1988). For each system, the data consist of the system and cost for each repair. If you want to compute the MCF for the number of repairs, rather than cost of repairs, then you should set the cost equal to 1 for each repair. A plus sign (+) in place of a cost indicates that the age is a censoring time. The repair history of each system ends with a censoring time.

Table 18.70 System Repair Histories for Artificial Data

Unit	(Age in Months, Cost in $100)			
6	(5,$3)	(12,$1)	(12,+)	
5	(16,+)			
4	(2,$1)	(8,$1)	(16,$2)	(20,+)
3	(18,$3)	(29,+)		
2	(8,$2)	(14,$1)	(26,$1)	(33,+)
1	(19,$2)	(39,$2)	(42,+)	

Table 18.71 illustrates the calculation of the MCF estimate from the data in Table 18.70. The RELIABILITY procedure uses the following rules for computing the MCF estimates.

1. Order all events (repairs and censoring) by age from smallest to largest.
   - If the event ages of the same or different systems are equal, the corresponding data are sorted from the largest repair cost to the smallest. Censoring events always sort as smaller than repair events with equal ages.
   - When event ages and values of more than one system coincide, the corresponding data are sorted from the largest system identifier to the smallest. The system IDs can be numeric or character, but they are always sorted in ASCII order.
2. Compute the number of systems I in service at the current age as the number in service at the last repair time minus the number of censored units in the intervening times.
3. For each repair, compute the mean cost as the cost of the current repair divided by the number in service I.
4. Compute the MCF for each repair as the previous MCF plus the mean cost for the current repair.

Table 18.71 Calculation of MCF for Artificial Data

Event	(Age,Cost)	Number *I* in Service	Mean Cost	MCF
1	(2,$1)	6	$1/6=0.17	0.17
2	(5,$3)	6	$3/6=0.50	0.67
3	(8,$2)	6	$2/6=0.33	1.00
4	(8,$1)	6	$1/6=0.17	1.17
5	(12,$1)	6	$1/6=0.17	1.33
6	(12,+)	5		
7	(14,$1)	5	$1/5=0.20	1.53
8	(16,$2)	5	$2/5=0.40	1.93
9	(16,+)	4		
10	(18,$3)	4	$3/4=0.75	2.68
11	(19,$2)	4	$2/4=0.50	3.18
12	(20,+)	3		
13	(26,$1)	3	$1/3=0.33	3.52
14	(29,+)	2		
15	(33,+)	1		
16	(39,$2)	1	$2/1=2.00	5.52
17	(42,+)	0		

If you specify the VARIANCE=NELSON option, the variance of the estimator of the MCF $\mathrm{Var}(\hat{\mathrm{M}}(t))$ is computed as in Nelson (1995). If the VARIANCE=LAWLESS or VARMETHOD2 option is specified, the method of Lawless and Nadeau (1995) is used to compute the variance of the estimator of the MCF. This method is recommended if the number of systems or events is large or if a FREQ statement is used to specify

a frequency variable. If you do not specify a variance computation method, the method of Lawless and Nadeau (1995) is used.

Default approximate two-sided $\gamma \times 100\%$ pointwise confidence limits for $\mathrm{M}(t)$ are computed as

$$\mathrm{M}_L(t) = \hat{\mathrm{M}}(t) - K_\gamma \sqrt{\mathrm{Var}(\hat{\mathrm{M}}(t))}$$

$$\mathrm{M}_U(t) = \hat{\mathrm{M}}(t) + K_\gamma \sqrt{\mathrm{Var}(\hat{\mathrm{M}}(t))}$$

where K_γ represents the $100(1+\gamma)/2$ percentile of the standard normal distribution.

If you specify the LOGINTERVALS option in the MCFPLOT statement, alternative confidence intervals based on the asymptotic normality of $\log(\hat{\mathrm{M}}(t))$, rather than of $\hat{\mathrm{M}}(t)$, are computed. Let

$$w = \exp\left[\frac{K_\gamma \sqrt{\mathrm{Var}(\hat{\mathrm{M}}(t))}}{\hat{\mathrm{M}}(t)}\right]$$

Then the limits are computed as

$$\mathrm{M}_L(t) = \frac{\hat{\mathrm{M}}(t)}{w}$$

$$\mathrm{M}_U(t) = \hat{\mathrm{M}}(t) \times w$$

These alternative limits are always positive, and can provide better coverage than the default limits when the MCF is known to be positive, such as for counts or for positive costs. They are not appropriate for MCF differences, and are not computed in this case.

The following SAS statements create the tabular output shown in Figure 18.59 and the plot shown in Figure 18.60:

```
data Art;
   input Sysid $ Time Cost;
   datalines;
sys1 19 2
sys1 39 2
sys1 42 -1
sys2 8  2
sys2 14 1
sys2 26 1
sys2 33 -1
sys3 18 3
sys3 29 -1
```

```
sys4 16 2
sys4 2  1
sys4 20  -1
sys4 8  1
sys5 16 -1
sys6 5 3
sys6 12 1
sys6 12 -1
;

proc reliability data=Art;
   unitid Sysid;
   mcfplot Time*Cost(-1) ;
run;
```

The first table in Figure 18.59 displays the input data set, the number of observations used in the analysis, the number of systems (units), and the number of repair events. The second table displays the system age, MCF estimate, standard error, approximate confidence limits, and system ID for each event.

Figure 18.59 PROC RELIABILITY Output for the Artificial Data

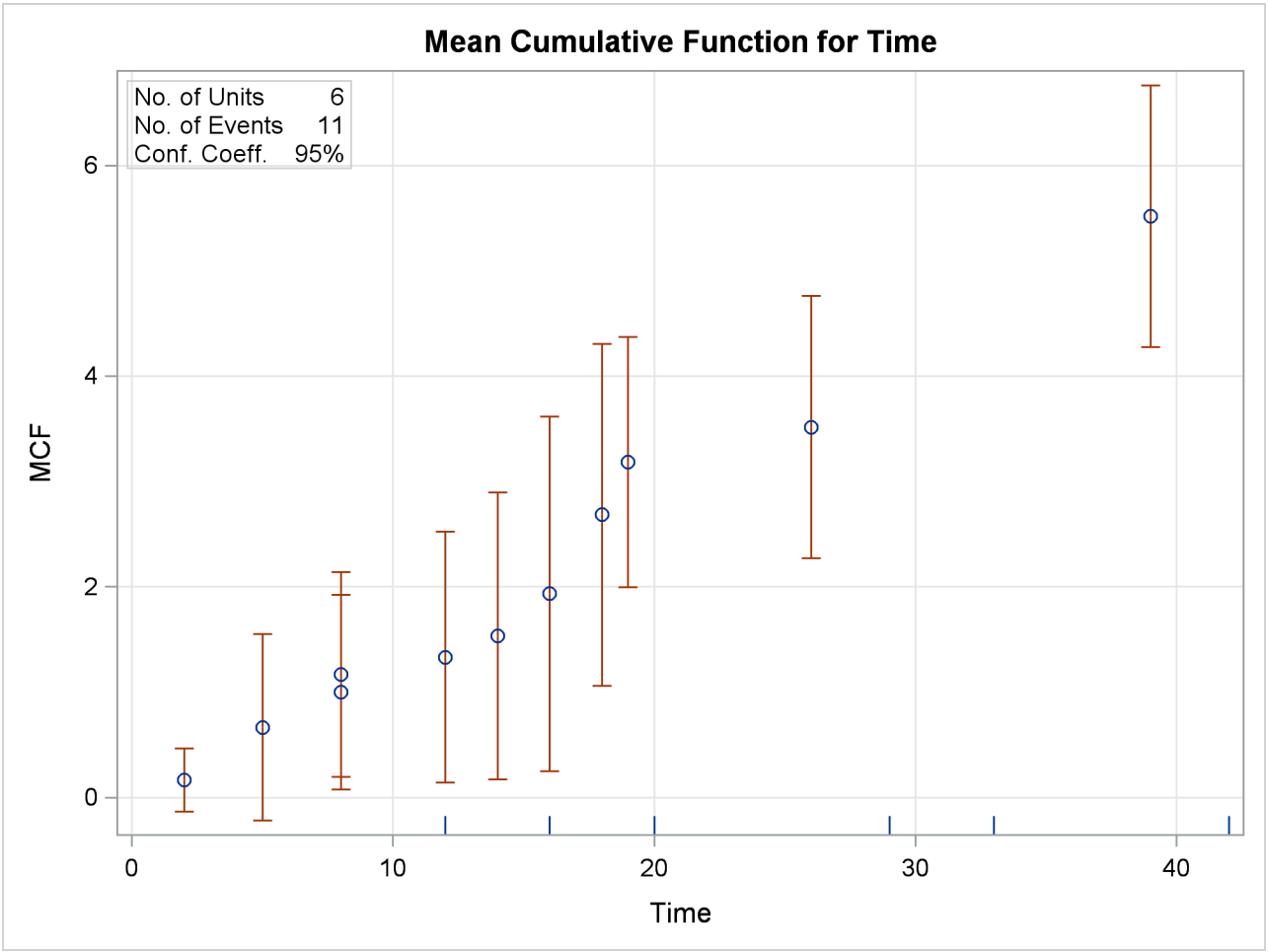
The RELIABILITY Procedure

Recurrence Data Summary	
Input Data Set	WORK.ART
Observations Used	17
Number of Units	6
Number of Events	11

Recurrence Data Analysis

Age	Sample MCF	Standard Error	95% Confidence Limits Lower	Upper	Unit ID
2.00	0.167	0.152	-0.132	0.465	sys4
5.00	0.667	0.451	-0.218	1.551	sys6
8.00	1.000	0.471	0.076	1.924	sys2
8.00	1.167	0.495	0.196	2.138	sys4
12.00	1.333	0.609	0.141	2.526	sys6
14.00	1.533	0.695	0.172	2.895	sys2
16.00	1.933	0.859	0.249	3.618	sys4
18.00	2.683	0.828	1.061	4.306	sys3
19.00	3.183	0.607	1.993	4.373	sys1
26.00	3.517	0.634	2.274	4.759	sys2
39.00	5.517	0.634	4.274	6.759	sys1

Figure 18.60 MCF Plot for the Artificial Data



Recurrent Events Data with Ages Grouped into Intervals

Recurrence data are sometimes grouped into time intervals for convenience, or to reduce the number of data records to be stored and analyzed. Interval recurrence data consist of the number of recurrences and the number of censored units in each time interval.

You can use PROC RELIABILITY to compute and plot MCFs and MCF differences for interval data. Formulas for the MCF estimator $\hat{\mathrm{M}}(t)$ and the variance of the estimator $\mathrm{Var}(\hat{\mathrm{M}}(t))$ for interval data, as well as examples and interpretations, are given in Nelson (2002, chapter 5). These calculations apply only to the number of recurrences, and not to cost.

Let N_0 be the total number of units, R_i the number of recurrences in interval i, $i = 1, \ldots, n$, and C_i the number of units censored into interval i. Then $\mathrm{N}_0 = \sum_{i=1}^{n} \mathrm{C}_i$ and the number entering interval i is $\mathrm{N}_i = \mathrm{N}_{i-1} - \mathrm{C}_{i-1}$ with $\mathrm{C}_0 = 0$. The MCF estimate for interval i is $\mathrm{M}_0 = 0$,

$$\mathrm{M}_i = \mathrm{M}_{i-1} + \frac{\mathrm{R}_i}{\mathrm{N}_i - 0.5\mathrm{C}_i}$$

The denominator $\mathrm{N}_i - 0.5\mathrm{C}_i$ approximates the number at risk in interval i, and treats the censored units as if they were censored halfway through the interval. Since no censored units are likely to have ages lasting

through the entire last interval, the MCF estimate for the last interval is likely to be biased. A footnote is printed in the tabular output as a reminder of this bias for the last interval.

See the section "Analysis of Interval Age Recurrence Data" on page 1227 for an example of interval recurrence data analysis.

Comparison of Two Groups of Recurrent Events Data

If you specify a group variable in an MCFPLOT statement, and the group variable has only two levels representing two groups of data, then there are two ways to compare the MCFs of the two groups for equality.

If you specify the MCFDIFF option in the MCFPLOT statement, estimates of the difference between two MCFs $\mathrm{MDIFF}(t) = \mathrm{M}_1(t) - \mathrm{M}_2(t)$ and the variance of the estimator are computed and plotted as in Doganaksoy and Nelson (1998). Confidence limits for the MCF difference function are computed in the same way as for the MCF, by using the variance of the MCF difference function estimator. If the confidence limits do not enclose zero at any time point, then the statistical hypothesis that the two MCFs are equal at all times is rejected.

Cook and Lawless (2007, section 3.7.5) describe statistical tests based on weighted sums of sample differences in the MCFs of the two groups. These tests, similar to log-rank tests for survival data, are computed and displayed in the "Tests for Equality of Mean Functions" table. Two cases are computed, corresponding to different weight functions in the test statistic. The "constant" case is powerful in cases where the two MCFs are approximately proportional. The "linear" case is more powerful for cases where the MCFs are not proportional, but do not cross.

These methods are not available for grouped data, as described in "Recurrent Events Data with Ages Grouped into Intervals" on page 1350.

Parametric Models for Recurrent Events Data

The parametric models used for recurrent events data in PROC RELIABILITY are called *Poisson process models*. Some important features of these models are summarized below. See, for example, Rigdon and Basu (2000) and Meeker and Escobar (1998) for a full mathematical description of these models. See Cook and Lawless (2007) for a general discussion of maximum likelihood estimation in Poisson processes. Abernethy (2006) and US Army (2000) provide examples of the application of Poisson process models to system reliability.

Let $\mathrm{N}(t)$ be the number of events up to time t, and let $\mathrm{N}(a, b)$ be the number of events in the interval $(a, b]$. Then, for a Poisson process,

- $\mathrm{N}(0) = 0$.
- $\mathrm{N}(a, b)$ and $\mathrm{N}(c, d)$ are statistically independent if $a < b \leq c < d$.
- $\mathrm{N}(a, b)$ is a Poisson random variable with mean $\mathrm{M}(a, b) = \mathrm{M}(b) - \mathrm{M}(a)$ where $\mathrm{M}(t)$ is the mean number of failures up to time t. $\mathrm{M}(0) = 0$.

Poisson processes are characterized by their cumulative mean function $\mathrm{M}(t)$, or equivalently by their *intensity*, *rate*, or *rate of occurrence of failure* (ROCOF) function $\lambda(t) = \frac{d}{dt}\mathrm{M}(t)$, so that

$$\mathrm{M}(a, b) = \mathrm{M}(b) - \mathrm{M}(a) = \int_a^b \lambda(t)dt$$

Poisson processes are parameterized through their mean and rate functions. The RELIABILITY procedure provides the Poisson process models shown in Table 18.72.

Table 18.72 Models and Parameters for Recurrent Events Data

Model	Intensity Function	Mean Function
Crow-AMSAA	$\beta\eta t^{\beta-1}$	ηt^{β}
Homogeneous	$\exp(\eta)$	$\exp(\eta)t$
Log-linear	$\exp(\eta+\beta t)$	$\frac{\exp(\eta)}{\beta}[\exp(\beta t)-1]$
Power	$\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}$	$\left(\frac{t}{\eta}\right)^{\beta}$
Proportional intensity	$\exp(\eta)\beta t^{\beta-1}$	$\exp(\eta)t^{\beta}$

For a homogeneous Poisson process, the intensity function is a constant; that is, the rate of failures does not change with time. For the other models, the rate function can change with time, so that a rate of failures that increases or decreases with time can be modeled. These models are called *non-homogeneous* Poisson processes.

In the RELIABILITY procedure, you specify a Poisson model with a DISTRIBUTION statement and a MODEL statement. You must also specify additional statements, depending on whether failure times are observed exactly, or observed to occur in time intervals. These statements are explained in the sections "Recurrent Events Data with Exact Event Ages" on page 1356, "Recurrent Events Data with Interval Event Ages" on page 1357, and "MODEL Statement" on page 1272. The DISTRIBUTION statement specifications for the models described in Table 18.72 are summarized in Table 18.73.

Table 18.73 DISTRIBUTION Statement Specification for Recurrent Events Data Models

Model	DISTRIBUTION Statement Value
Crow-AMSAA	NHPP(CA)
Homogeneous	HPP
Log-linear	NHPP(LOG)
Power	NHPP(POW)
Proportional intensity	NHPP(PROP)

For each of the models, you can specify a regression model for the parameter η in Table 18.72 for the ith observation as

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \ldots$$

where $x_{i1}, x_{i2}, \ldots$ are regression coefficients specified as described in the section "MODEL Statement" on page 1272. The parameter β_0 is labeled Intercept in the printed output, and the parameter β in Table 18.72 is labeled Shape. If no regression coefficients are specified, Intercept represents the parameter η in Table 18.72.

Model parameters are estimated by maximizing the log-likelihood function, which is equivalent to maximizing the likelihood function. The sections "Recurrent Events Data with Exact Event Ages" on page 1356 and "Recurrent Events Data with Interval Event Ages" on page 1357 contain descriptions of the form of the log likelihoods for the different models. An estimate of the covariance matrix of the maximum likelihood estimators (MLEs) of the parameters $\boldsymbol{\theta}$ is given by the inverse of the negative of the matrix of second derivatives of the log likelihood, evaluated at the final parameter estimates:

$$\boldsymbol{\Sigma} = [\sigma_{ij}] = -\mathbf{H}^{-1} = -\left[\frac{\partial^2 LL}{\partial\theta_i\,\partial\theta_j}\right]^{-1}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

The negative of the matrix of second derivatives is called the observed Fisher information matrix. The diagonal term σ_{ii} is an estimate of the variance of $\hat{\theta}_i$. Estimates of standard errors of the MLEs are provided by

$$\mathrm{SE}_{\theta_i} = \sqrt{\sigma_{ii}}$$

An estimator of the correlation matrix is

$$\mathbf{R} = \left[\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}\right]$$

Wald-type confidence intervals are computed for the model parameters as described in Table 18.74. Wald intervals use asymptotic normality of maximum likelihood estimates to compute approximate confidence intervals. If a parameter must be greater than zero, then an approximation based on the asymptotic normality of the logarithm of the parameter estimate is often more accurate, and the lower endpoint is strictly positive. The intercept term β_0 in an intercept-only power NHPP model with no other regression parameters represents η in Table 18.72, and is a model parameter that must be strictly positive. Also, the shape parameter for the power and proportional intensity models, represented by β in Table 18.72, must be strictly positive. In these cases, formula 7.11 of Meeker and Escobar (1998, p. 163) is used in Table 18.74 to compute confidence limits.

Table 18.74 shows the method of computation of approximate two-sided $\gamma \times 100\%$ confidence limits for model parameters. The default value of confidence is $\gamma = 0.95$. Other values of confidence are specified using the CONFIDENCE= option. In Table 18.74, K_γ represents the $(1+\gamma)/2 \times 100\%$ percentile of the standard normal distribution, and $w(\hat{\theta}) = \exp[K_\gamma(\mathrm{SE}_{\hat{\theta}})/\hat{\theta}]$.

Table 18.74 NHPP Model Parameter Confidence Limit Computation

Model		Parameters: Intercept	Regression Parameters	Shape
Crow-AMSAA				
Intercept-only model				
	Lower	$\hat{\beta}_0/w(\hat{\beta}_0)$		$\hat{\beta}/w(\hat{\beta})$
	Upper	$\hat{\beta}_0 w(\hat{\beta}_0)$		$\hat{\beta} w(\hat{\beta})$
Regression model				
	Lower	$\hat{\beta}_0 - K_\gamma(\mathrm{SE}_{\hat{\beta}_0})$	$\hat{\beta}_i - K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$	$\hat{\beta}/w(\hat{\beta})$
	Upper	$\hat{\beta}_0 + K_\gamma(\mathrm{SE}_{\hat{\beta}_0})$	$\hat{\beta}_i + K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$	$\hat{\beta} w(\hat{\beta})$
Homogeneous				
	Lower	$\hat{\beta}_0 - K_\gamma(\mathrm{SE}_{\hat{\beta}_0})$	$\hat{\beta}_i - K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$	
	Upper	$\hat{\beta}_0 + K_\gamma(\mathrm{SE}_{\hat{\beta}_0})$	$\hat{\beta}_i + K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$	
Log-linear				
	Lower	$\hat{\beta}_0 - K_\gamma(\mathrm{SE}_{\hat{\beta}_0})$	$\hat{\beta}_i - K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$	$\hat{\beta} - K_\gamma(\mathrm{SE}_{\hat{\beta}})$
	Upper	$\hat{\beta}_0 + K_\gamma(\mathrm{SE}_{\hat{\beta}_0})$	$\hat{\beta}_i + K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$	$\hat{\beta} + K_\gamma(\mathrm{SE}_{\hat{\beta}})$
Power				
Intercept-only model				
	Lower	$\hat{\beta}_0/w(\hat{\beta}_0)$		$\hat{\beta}/w(\hat{\beta})$
	Upper	$\hat{\beta}_0 w(\hat{\beta}_0)$		$\hat{\beta} w(\hat{\beta})$
Regression model				
	Lower	$\hat{\beta}_0 - K_\gamma(\mathrm{SE}_{\hat{\beta}_0})$	$\hat{\beta}_i - K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$	$\hat{\beta}/w(\hat{\beta})$
	Upper	$\hat{\beta}_0 + K_\gamma(\mathrm{SE}_{\hat{\beta}_0})$	$\hat{\beta}_i + K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$	$\hat{\beta} w(\hat{\beta})$
Proportional intensity				
	Lower	$\hat{\beta}_0 - K_\gamma(\mathrm{SE}_{\hat{\beta}_0})$	$\hat{\beta}_i - K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$	$\hat{\beta}/w(\hat{\beta})$
	Upper	$\hat{\beta}_0 + K_\gamma(\mathrm{SE}_{\hat{\beta}_0})$	$\hat{\beta}_i + K_\gamma(\mathrm{SE}_{\hat{\beta}_i})$	$\hat{\beta} w(\hat{\beta})$

You can request that profile likelihood confidence intervals for model parameters be computed instead of Wald intervals with the LRCL option in the MODEL statement.

Confidence limits for the mean and intensity functions for plots that are created with the MCFPLOT statement and for the table that is created with the OBSTATS option in the MODEL statement are computed by using the delta method. See Meeker and Escobar (1998, Appendix B) for a full explanation of this method. If $\mathbf{\Sigma}$ represents the covariance matrix of the estimates of the parameters η and β in Table 18.72, then the variance of the mean or intensity function estimate is given by

$$\mathrm{V} = [\frac{\partial g}{\partial \eta}\frac{\partial g}{\partial \beta}]\mathbf{\Sigma}[\frac{\partial g}{\partial \eta}\frac{\partial g}{\partial \beta}]'$$

where $g = g(t; \eta, \beta)$ represents either the mean function or the intensity function in Table 18.72. Since both of these functions must be positive, formula 7.11 of Meeker and Escobar (1998, p. 163) is used to compute

confidence limits, using the standard error $\sigma = \sqrt{\mathrm{V}}$. For regression models, the full covariance matrix of all the regression parameter estimates and the parameter β is used to compute $\mathbf{\Sigma}$.

Tests of Trend

For the nonhomogeneous models in Table 18.72, you can request a test for a homogeneous Poisson process by specifying the HPPTEST option in the MODEL statement. In this case the test is a likelihood ratio test for $\beta = 1$ for the power, Crow-AMSAA, and proportional intensity models, and $\beta = 0$ for the log-linear model.

You can request other tests of trend by using the TREND= option in the MODEL statement. These tests are not available for the kind of grouped data that are described in the section "Recurrent Events Data with Interval Event Ages" on page 1357. See Lindqvist and Doksum (2003), Kvaloy and Lindqvist (1998), and Meeker and Escobar (1998) for a discussion of these kinds of tests.

Let there be m independent systems observed, and let $t_{i1}, t_{i2}, \ldots t_{in_i}$ be the times of observed events for system i. Let the last time of observation of system i be T_i, with $t_{in_i} \leq T_i$. The following test statistics can be computed. These are extended versions of trend tests for a single system, and they allow valid tests for HPP versus NHPP even if the intensities vary from system to system.

- Military Handbook (MH)

  $$\mathrm{MH} = 2\sum_{i=1}^{m}\sum_{j=1}^{n_i}\log\left(\frac{T_i}{t_{ij}}\right)$$

  The asymptotic distribution of MH is the chi-square with $2\sum_{i=1}^{m} n_i$ degrees of freedom. This test statistic is powerful for testing HPP versus NHPP in the power law model.

- Laplace (LA)

  $$\mathrm{LA} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i}(t_{ij} - \frac{1}{2}T_i)}{\sqrt{\frac{1}{12}\sum_{i=1}^{m} n_i T_i^2}}$$

  The asymptotic distribution of LA is the standard normal. This test statistic is powerful for testing HPP versus NHPP in a log-linear model.

- Lewis-Robinson (LR)

  Let LA_i be the Laplace test statistic for system i,

  $$\mathrm{LA}_i = \frac{\sum_{j=1}^{n_i}(t_{ij} - \frac{1}{2}T_i)}{\sqrt{\frac{1}{12}n_i}}$$

  The extended Lewis-Robinson test statistic is defined as

  $$\mathrm{LR} = \sum_{i=1}^{m}\frac{\bar{X}_i}{\sigma_i}\mathrm{LA}_i$$

  where $\bar{X}_i$ and σ_i are the estimated mean and standard deviation of the event interarrival times for system i. The asymptotic distribution of LR is the standard normal. This test statistic is powerful for testing HPP versus NHPP in a log-linear model.

INEST Data Set for Recurrent Events Models

You can specify a SAS data set to set lower bounds, upper bounds, equality constraints, or initial values for estimating the intercept and shape parameters of the models in Table 18.72 by using the INEST= option in the MODEL statement. The data set must contain a variable named _TYPE_ that specifies the action that you want to take in the iterative estimation process, and some combination of variables named _INTERCEPT_ and _SHAPE_ that represent the distribution parameters. If BY processing is used, the INEST= data set should also include the BY variables, and there must be at least one observation for each BY group.

The possible values of _TYPE_ and corresponding actions are summarized in Table 18.75.

Table 18.75 _TYPE_ Variable Values

Value of _TYPE_	Action
LB	Lower bound
UB	Upper bound
EQ	Equality
PARMS	Initial value

For example, you can use the INEST data set In created by the following SAS statements to specify an equality constraint for the shape parameter. The data set In specifies that the shape parameter be constrained to be 1.5. Since the variable _Intercept_ is set to missing, no action is taken for it, and this variable could be omitted from the data set.

```
data In ;
   input _Type_$ 1-5  _Intercept_  _Shape_;
   datalines;
eq    . 1.5
;
```

Recurrent Events Data with Exact Event Ages

Let there be m independent systems observed, and let $t_{i1}, t_{i2}, \ldots t_{in_i}$ be the times of observed events for system i. Let the last time of observation of system i be T_i, with $t_{in_i} \leq T_i$.

If there are no regression parameters in the model, or there are regression parameters and they are constant for each system, then the log-likelihood function is

$$LL = \sum_{i=1}^{m}\{\sum_{j=1}^{n_i}[\log(\lambda(t_{ij})] - \mathrm{M}(T_i)\}$$

If there are regression parameters that can change over time for individual systems, the RELIABILITY procedure uses the convention that a covariate value specified at a given event time takes effect immediately after the event time; that is, the value of a covariate used at an event time is the value specified at the previous event time. The value used at the first event time is the value specified at that event time. You can establish a different value for the first event time by specifying a zero cost event previous to the first actual event. The zero cost event is not used in the analysis, but it is used to establish a covariate value for the next event time. The covariate value used at the end time T_i is the value established at the last event time.

With these conventions, the log likelihood is

$$LL = \sum_{i=1}^{m}\{\sum_{j=1}^{n_i}[\log(\lambda(t_{ij}) - (\mathrm{M}(t_{ij}) - \mathrm{M}(t_{i,j-1})] - (\mathrm{M}(T_i) - \mathrm{M}(t_{in_i})\}$$

with $t_{i0} = 0$ for each $i = 1, 2, \ldots, m$. Note that this log likelihood reduces to the previous log likelihood if covariate values do not change over time for each system.

In order to specify a parametric model for recurrence data with exact event times, you specify the event times, end of observation times, and regression model, if any, with a MODEL statement, as described in the section "MODEL Statement" on page 1272. In addition, you specify a variable that uniquely identifies each system with a UNITID statement. See the section "Parametric Model for Recurrent Events Data" on page 1236 for an example of fitting a parametric recurrent events model to data with exact recurrence times.

Recurrent Events Data with Interval Event Ages

If n independent and statistically identical systems are observed in the time interval $(t_a, t_b]$, then the number r of events that occur in the interval is a Poisson random variable with mean $n\mathrm{M}(t_a, t_b) = n[\mathrm{M}(t_a) - \mathrm{M}(t_b)]$, where $\mathrm{M}(t)$ is the cumulative mean function for an individual system.

Let $(t_0, t_1], (t_1, t_2], \ldots, (t_{m-1}, t_m]$ be nonoverlapping time intervals for which r_i events are observed among the n_i systems observed in time interval $(t_{i-1}, t_i]$. The parameters in the mean function $\mathrm{M}(t)$ are estimated by maximizing the log likelihood

$$LL = \sum_{i=1}^{m}[r_i \log(n_i) + r_i \log(\mathrm{M}(t_{i-1}, t_i)) - n_i \mathrm{M}(t_{i-1}, t_i) - \log(r_i!)]$$

The time intervals do not have to be of the same length, and they do not have to be adjacent, although the preceding formula shows them as adjacent.

If you have data from groups of systems to which you are fitting a regression model (for example, to model the effects of different manufacturing lines or different vendors), the time intervals in the different groups do not have to coincide. The only requirement is that the data in the different groups be independent; for example, you cannot have data from the same systems in two different groups.

In order to specify a parametric model for interval recurrence data, you specify the time intervals and regression model, if any, with a MODEL statement, as described in the section "MODEL Statement" on page 1272. In addition, you specify a variable that contains the number n_i of systems under observation in time interval i with an NENTER statement, and the number of events r_i observed with a FREQ statement. See the section "Parametric Model for Interval Recurrent Events Data" on page 1238 for an example of fitting a parametric recurrent events model to data with interval recurrence times.

Duane Plots

A Duane plot is defined as a graph of the quantity $\mathrm{H}(t) = \mathrm{M}(t)/t$ versus t, where $\mathrm{M}(t)$ is the MCF. The graph axes are usually both on the log scale, so that if $\mathrm{M}(t)$ is the power law type in Table 18.72, a linear graph is produced. Duane plots are traditionally used as a visual assessment of the goodness of fit of a power law model. You should exercise caution in using Duane plots, because even if the underlying model is a power law process, a nonlinear Duane plot can result. See Rigdon and Basu (2000, section 4.1.1) for a discussion of Duane plots. You can create a Duane plot by specifying the DUANE option in the MCFPLOT statement. A scatter plot of nonparametric estimates of $\hat{\mathrm{H}}(t_i) = \hat{\mathrm{M}}(t_i)/t_i$ versus t_i is created on a log-log scale, where

$\hat{\mathrm{M}}(t_i)$ are the nonparametric estimates of the MCF that are described in the section "Nonparametric Analysis" on page 1346. If you specify a parametric model with the FIT=MODEL option in the MCFPLOT statement, the corresponding parametric estimate of $\mathrm{H}(t)$ is plotted on the same graph as the scatter plot.

ODS Table Names

The following tables contain the ODS table names created by the RELIABILITY Procedure, organized by the statements that produce them.

Table 18.76 Tables Produced with the ANALYZE Statement

Table Name	Description
ConvergenceStatus	Convergence status
CorrMat	Parameter correlation matrix
CovMat	Parameter covariance matrix
DatSum	Summary of fit
GradHess	Last evaluation of parameters, gradient, and Hessian
IterEM	Iteration history for Turnbull algorithm
IterLRParm	Iteration history for likelihood ratio confidence intervals for parameters
IterLRPer	Iteration history for likelihood ratio confidence intervals for percentiles
IterParms	Iteration history for parameter estimates
Lagrange	Lagrange multiplier statistics
NObs	Observations summary
PBEst	Poisson/binomial estimates by group
PBPred	Poisson/binomial predicted values
PBPredTol	Poisson/binomial predicted values by group
PBSum	Poisson/binomial analysis summary
PBTol	Poisson/binomial tolerance estimates
PctEst	Percentile estimates
ParmEst	Parameter estimates
ParmOther	Fitted distribution mean, median, mode
PGradHess	Last evaluation of parameters, gradient, and Hessian in terms of stable parameters
ProbabilityEstimates	Nonparametric cumulative distribution function estimates
RelInfo	Model information
SurvEst	Survival function estimates
TurnbullGrad	Interval probabilities, reduced gradient, Lagrange multipliers for Turnbull algorithm
WCorrMat	Parameter correlation matrix for Weibull distribution
WCovMat	Parameter covariance matrix for Weibull distribution

Table 18.77 Tables Produced with the MCFPLOT Statement

Table Name	Description
McfDEst	MCF difference estimates

Table 18.77 Tables Produced with the MCFPLOT Statement (continued)

Table Name	Description
McfDSum	MCF difference data summary
McfEst	MCF estimates
McfLogRank	Tests of difference between two MCFs
McfSum	MCF data summary

Table 18.78 Tables Produced with the MODEL Statement

Table Name	Description
MConvergenceStatus	Convergence status
ClassLevels	Class level information
ModCorMat	Parameter correlation matrix
ModCovMat	Parameter covariance matrix
ModFitSum	Summary of fit
ModInfo	Model information
ModIterLRparm	Iteration history for likelihood ratio confidence intervals for parameters
ModIterParms	Iteration history for parameter estimates
ModLagr	Lagrange multiplier statistics
ModLastGradHess	Last evaluation of the gradient and Hessian
ModNObs	Observations summary
ModObstats	Observation statistics
ModParmInfo	Parameter information
ModPrmEst	Parameter estimates
RecurGoodFit	Test for homogeneous Poisson process

Table 18.79 Tables Produced with PROBPLOT and RELATIONPLOT Statements

Table Name	Description
ConvergenceStatus	Convergence status
CorrMat	Parameter correlation matrix
CovMat	Parameter covariance matrix
DatSum	Summary of fit
GradHess	Last evaluation of parameters, gradient, and Hessian
IterEM	Iteration history for Turnbull algorithm
IterLRParm	Iteration history for likelihood ratio confidence intervals for parameters
IterLRPer	Iteration history for likelihood ratio confidence intervals for percentiles
IterParms	Iteration history for parameter estimates
Lagrange	Lagrange multiplier statistics
NObs	Observations summary
PctEst	Percentile estimates

Table 18.79 Tables Produced with PROBPLOT and RELATIONPLOT Statements (continued)

Table Name	Description
ParmEst	Parameter estimates
ParmOther	Fitted distribution mean, median, mode
PGradHess	Last evaluation of parameters, gradient, and Hessian in terms of stable parameters
ProbabilityEstimates	Nonparametric cumulative distribution function estimates
RelInfo	Model information
SurvEst	Survival function estimates
TurnbullGrad	Interval probabilities, reduced gradient, Lagrange multipliers for Turnbull algorithm
WCorrMat	Parameter correlation matrix for Weibull distribution
WCovMat	Parameter covariance matrix for Weibull distribution

ODS Graphics

SAS/QC procedures use ODS Graphics functionality to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, "Statistical Graphics Using ODS" (*SAS/STAT User's Guide*).

Before you create graphs, ODS Graphics must be enabled. For example:

```
ods graphics on;

proc reliability;
   probplot y;
run;

ods graphics off;
```

For more information about enabling and disabling ODS Graphics, see Chapter 21, "Statistical Graphics Using ODS" (*SAS/STAT User's Guide*).

See Chapter 4, "SAS/QC Graphics," for alternative methods of creating graphics with PROC RELIABILITY. See the section "Analysis of Right-Censored Data from a Single Population" on page 1176 for an example that uses ODS Graphics in PROC RELIABILITY to create a probability plot and the section "ODS Graph Names" on page 1360 for ODS Graphics table names.

ODS Graph Names

If ODS Graphics is enabled (for example, with the ODS GRAPHICS ON statement), PROC RELIABILITY creates graphs by using ODS Graphics. You can reference every graph produced through ODS Graphics with a name. The names of the graphs that PROC RELIABILITY generates are listed in Table 18.80, along with the required statements and options.

Table 18.80 Graphs Produced by PROC RELIABILITY

ODS Graph Name	Description	Statement	Option
IntensityPlot	Plot of intensity function for NHPP model	MCFPLOT	Fit=MODEL
IntensityPlots	Plot of intensity functions for NHPP model for multiple groups	MCFPLOT	Fit=MODEL
MCFDiffPlot	Plot of mean cumulative function differences	MCFPLOT	MCFDIFF
MCFPlot	Plot of mean cumulative function plot for single population	MCFPLOT	Default
MCFPlotPanel	Plot of mean cumulative function plots for multiple groups	MCFPLOT	Group variable
ProbabilityPlot	Probability plot for single population	PROBPLOT	Default
ProbabilityPlotFM	Probability plot with failure modes	PROBPLOT	FMODE
ProbabilityPlotPanel	Probability plots for multiple groups	PROBPLOT	Group variable
PercentilePlot	Plot of model percentiles	RELATIONPLOT	Default
RecurrentEventsPlot	Plot of recurrent event times	MCFPLOT	EVENTPLOT
RecurrentEventsPlotPanel	Plots of recurrent events for multiple groups	MCFPLOT	EVENTPLOT
RelationPlot	Plot of model percentiles with probability plot	RELATIONPLOT	PPLOT

References

Abernethy, R. B. (2006). *The New Weibull Handbook*. 5th ed. North Palm Beach, FL: Robert B. Abernethy.

Ascher, H., and Feingold, H. (1984). *Repairable Systems Reliability*. New York: Marcel Dekker.

Byar, D. P. (1980). *The Veterans Administration Study of Chemoprophylaxis for Recurrent Stage 1 Bladder Tumors: Comparisons of Placebo, Pyridoxine, and Topical Thiotepa.* Edited by M. Pavone-Macaluso, P. H. Smith, and F. Edsmyr. New York: Plenum.

Collett, D. (1994). *Modelling Survival Data in Medical Research*. London: Chapman & Hall.

Cook, R. J., and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. New York: Springer.

Crowder, M. J., Kimber, A. C., Smith, R. L., and Sweeting, T. J. (1991). *Statistical Analysis of Reliability Data*. New York: Chapman & Hall.

Doganaksoy, N., Hahn, G. J., and Meeker, W. Q. (2002). "Reliability Analysis by Failure Mode." *Quality Progress* 35:47–52.

Doganaksoy, N., and Nelson, W. (1998). "A Method to Compare Two Samples of Recurrence Data." *Lifetime Data Analysis* 4:51–63.

Doganaksoy, N., and Schmee, J. (1993). "Orthogonal Parameters with Censored Data." *Communications in Statistics—Theory and Methods* 22:669–685.

Gentleman, R., and Geyer, C. J. (1994). "Maximum Likelihood for Interval Censored Data: Consistency and Computation." *Biometrika* 81:618–623.

Joe, H., and Proschan, F. (1984). "Percentile Residual Life Functions." *Operations Research* 32:668–678.

Kvaloy, J. T., and Lindqvist, B. H. (1998). "TTT-Based Tests for Trend in Repairable Systems Data." *Reliability Engineering and Systems Safety* 60:13–28.

Lawless, J. F. (2003). *Statistical Model and Methods for Lifetime Data*. 2nd ed. New York: John Wiley & Sons.

Lawless, J. F., and Nadeau, C. (1995). "Some Simple Robust Methods for the Analysis of Recurrent Events." *Technometrics* 37:158–168.

Lindqvist, B. H., and Doksum, K. A., eds. (2003). *Mathematical and Statistical Methods in Reliability.* Vol. 7 of Series on Quality, Reliability, and Engineering Statistics. Singapore: World Scientific.

Meeker, W. Q., and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. New York: John Wiley & Sons.

Nair, V. N. (1984). "Confidence Bands for Survival Functions with Censored Data: A Comparative Study." *Technometrics* 26:265–275.

Nelson, W. (1982). *Applied Life Data Analysis*. New York: John Wiley & Sons.

Nelson, W. (1985). "Weibull Analysis of Reliability Data with Few or No Failures." *Journal of Quality Technology* 17:140–146.

Nelson, W. (1988). "Graphical Analysis of System Repair Data." *Journal of Quality Technology* 20:24–35.

Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*. New York: John Wiley & Sons.

Nelson, W. (1995). "Confidence Limits for Recurrence Data—Applied to Cost or Number of Product Repairs." *Technometrics* 37:147–157.

Nelson, W. (2002). *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications*. Philadelphia: SIAM.

Nelson, W., and Doganaksoy, N. (1989). *A Computer Program for an Estimate and Confidence Limits for the Mean Cumulative Function for Cost or Number of Repairs of Repairable Products*. Technical Report 89CRD239, GE Research & Development Center, Schenectady, NY.

Rigdon, S. E., and Basu, A. P. (2000). *Statistical Methods for the Reliability of Repairable Systems*. New York: John Wiley & Sons.

Tobias, P. A., and Trindade, D. C. (1995). *Applied Reliability*. 2nd ed. New York: Van Nostrand Reinhold.

Turnbull, B. W. (1976). "The Empirical Distribution Function with Arbitrarily Grouped, Censored, and Truncated Data." *Journal of the Royal Statistical Society, Series B* 38:290–295.

US Army (2000). *AMSAA Reliability Growth Guide*. Technical Report TR-652, US Army Materiel Systems Analysis Activity, Aberdeen Proving Ground, MD.

Subject Index

ODS tables
  RELIABILITY procedure, 1358, 1359
options summary
  ESTIMATE statement, 1252

reliability analysis
  analyzing accelerated life test data, 1184–1189
  analyzing arbitrarily censored data, 1194
  analyzing binomial data, 1230, 1231
  analyzing combined failure modes, 1211
  analyzing groups of data, 1180, 1181, 1184
  analyzing interval-censored data, 1189, 1191, 1192, 1197
  analyzing regression models, 1199, 1201, 1202, 1204
  analyzing repair data, 1215, 1217, 1220
  analyzing right-censored data, 1176, 1177, 1180
  analyzing two groups of repair data, 1220, 1222, 1223, 1227
  arbitrarily censored data, 1318
  binomial parameter estimation, 1335, 1336
  classification variables, 1249
  comparison of groups of recurrence data, 1351
  confidence intervals for parameters, 1330, 1331
  covariance matrix of parameters, 1329
  creating life-stress relation plots, 1295, 1296, 1299, 1302, 1305, 1307, 1308
  creating output data sets, 1261, 1359
  creating probability plots, 1281, 1282, 1287, 1289, 1291, 1294
  details, 1310
  Duane plots, 1357
  estimating distribution parameters, 1243, 1245, 1248
  examples, 1176
  failure modes, 1211, 1253, 1339
  fitting regression models, 1272, 1274, 1278
  frequency variables, 1254
  insets, 1254, 1255, 1257
  least squares estimation, 1338
  log-scale regression model parameters, 1328
  maximum likelihood estimation, 1324
  mean cumulative function plots, 1261, 1263, 1266, 1268, 1272, 1309
  observation-wise percentiles, 1341, 1342
  observation-wise predicted values, 1340
  observation-wise predicted values, recurrent events, 1345
  observation-wise reliability function estimates, 1343
  observation-wise statistics, 1340–1344
  observation-wise statistics, recurrent events, 1345
  ODS Graphics, 1360
  optimization options recurrent events, 1280
  optimization options three-parameter Weibull, 1280
  overview, 1174, 1175
  parameter estimation, 1324, 1327–1339
  parametric model for recurrent events data, 1236, 1238
  parametric model for two groups of repair data, 1227
  percentile estimation, 1332, 1333
  Poisson parameter estimation, 1337, 1338
  predicted values for recurrent events data, 1227
  probability distributions, 1310–1312
  probability plots, 1313–1315, 1317
  readout data, 1280
  recurrence data, 1346, 1347, 1351
  recurrent events data, 1236, 1238
  regression model parameters, 1327, 1328
  reliability function estimation, 1334, 1335
  residuals, 1343, 1344
  specifying failure modes, 1253
  specifying probability distributions, 1250
  syntax, 1241
  three-parameter Weibull, 1233
  Turnbull algorithm, 1318
  types of lifetime data, 1310
  Weibayes estimation, 1338

Syntax Index

BY statement
RELIABILITY procedure, 1249

EFFECTPLOT statement
RELIABILITY procedure, 1251
ESTIMATE statement
RELIABILITY procedure, 1252

LSMEANS statement
RELIABILITY procedure, 1258
LSMESTIMATE statement
RELIABILITY procedure, 1260

PPLOT statement, *see* RELIABILITY procedure, PROBPLOT statement
PROBPLOT statement, *see* RELIABILITY procedure, PROBPLOT statement

reliability, 1174
RELIABILITY procedure, 1241
details, 1310
examples, 1176
overview, 1174
syntax, 1241
RELIABILITY procedure, ANALYZE statement, 1231, 1243, 1245, 1248
CONVERGE= option, 1325
PPOS= option, 1314, 1315, 1317
PREDICT option, 1231
summary of options, 1245, 1248
TOLERANCE option, 1231
RELIABILITY procedure, BY statement, 1241, 1242, 1249
RELIABILITY procedure, CLASS statement, 1241, 1242, 1249
TRUNCATE option, 1249
RELIABILITY procedure, DISTRIBUTION statement, 1176, 1177, 1186, 1187, 1191, 1192, 1201, 1231, 1241, 1242, 1250
RELIABILITY procedure, EFFECTPLOT statement, 1251
RELIABILITY procedure, ESTIMATE statement, 1252
RELIABILITY procedure, FMODE statement, 1241, 1242, 1253
RELIABILITY procedure, FREQ statement, 1186, 1187, 1191, 1192, 1241, 1242, 1254
RELIABILITY procedure, INSET statement, 1241, 1242, 1254, 1255, 1257
keywords, 1255
summary of options, 1257
RELIABILITY procedure, LSMEANS statement, 1258
RELIABILITY procedure, LSMESTIMATE statement, 1260
RELIABILITY procedure, MAKE statement, 1201, 1241, 1242, 1261
NOPRINT option, 1261
table keywords, 1261
RELIABILITY procedure, MCFPLOT statement, 1217, 1220, 1222, 1241, 1261, 1263, 1266, 1268, 1272
MCFDIFF option, 1222
summary of options, 1263, 1266, 1268, 1272
RELIABILITY procedure, MODEL statement, 1186, 1187, 1201, 1241, 1272, 1274, 1278
CONVERGE= option, 1325
CORRB option, 1201
COVB option, 1201
No intercept option, 1327
OBSTATS option, 1186, 1187, 1201, 1278
RELATION= option, 1186, 1187
summary of options, 1274, 1278
RELIABILITY procedure, NENTER statement, 1191, 1192, 1241, 1242, 1280
RELIABILITY procedure, NLOPTIONS statement, 1280
RELIABILITY procedure, ODS
table keywords, 1359
RELIABILITY procedure, ODS table names, 1359
RELIABILITY procedure, PPLOT statement, *see* RELIABILITY procedure, PROBPLOT statement
RELIABILITY procedure, PROBPLOT statement, 1176, 1177, 1181, 1191, 1192, 1241, 1281, 1282, 1287, 1289, 1291, 1294
CONVERGE= option, 1325
COVB option, 1176, 1177
NOCONF option, 1181, 1191, 1192
OVERLAY option, 1181
PCONFPLT option, 1191, 1192
PPOS= option, 1314, 1315, 1317
READOUT option, 1191, 1192
summary of options, 1282, 1287, 1289, 1291, 1294
RELIABILITY procedure, RELATIONPLOT statement, 1186, 1187, 1241, 1295, 1296,

1299, 1302, 1305, 1307, 1308
CONVERGE= option, 1325
FIT= option, 1186, 1187
LUPPER= option, 1186, 1187
NOCONF option, 1186, 1187
PLOTDATA option, 1186, 1187
PLOTFIT option, 1186, 1187
PPLOT option, 1186, 1187
PPOS= option, 1314, 1315, 1317
RELATION= option, 1186, 1187
SLOWER= option, 1186, 1187
summary of options, 1296, 1299, 1302, 1305, 1307, 1308
RELIABILITY procedure, RPLOT statement, *see* RELIABILITY procedure, RELATIONPLOT statement
RELIABILITY procedure, SLICE statement, 1308
RELIABILITY procedure, STORE statement, 1309
RELIABILITY procedure, TEST statement, 1309
RELIABILITY procedure, UNITID statement, 1217, 1222, 1241, 1242, 1309

SLICE statement
RELIABILITY procedure, 1308
STORE statement
RELIABILITY procedure, 1309

TEST statement
RELIABILITY procedure, 1309
TRUNCATE option
CLASS statement (RELIABILITY), 1249