

# RELIABILITY DATA ANALYSIS IN THE SAS® SYSTEM

Gordon Johnston, SAS Institute Inc.  
SAS Campus Drive, Cary, NC 27513

## **Key Words: Repair data analysis, Life data analysis, Accelerated life testing**

The RELIABILITY procedure, a recent addition to SAS/QC software, provides tools for reliability and survival data analysis as well as for recurrence data analysis. You can use this procedure to

- construct probability plots and fit life distributions with right and interval censored lifetime data
- fit regression models, including accelerated life test models, to combinations of right, left, 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 or interval censored data. The RELIABILITY procedure provides probability plots of uncensored, right censored, and interval censored data when all the failure data have common interval endpoints.

Features of the RELIABILITY procedure include

- probability plotting and parameter estimation for the common life distributions: Weibull, exponential, extreme value, normal, lognormal, logistic, and loglogistic. The data may 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, loglogistic, 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 repairs and associated confidence intervals from repair data from systems

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 loglogistic 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.

The books by Lawless (1982), Nelson (1990), Nelson (1982), and Trindade and Tobias (1995) provide many examples taken from diverse fields and describe the analyses provided by the RELIABILITY procedure. The books by Nelson emphasize reliability data analysis from an engineering viewpoint. The features of the procedure that deal with analysis of repair data from systems are based on the work of Nelson (1995), Nelson (1988), Doganaksoy and Nelson (1991), and Nelson and Doganaksoy (1989), who provide examples of repair data analysis.

The following examples illustrate some of the analyses that the RELIABILITY procedure 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 using the observed and right censored lifetimes of 70 diesel engine fans given by Nelson (1982, p. 318).

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.

The following statements use the RELIABILITY procedure to produce the graphical output shown in Figure 1:

```
proc reliability;
  distribution weibull;
  probplot lifetime*censor( 1 ) / covb;
  inset / cfill=white;
run;
```

The DISTRIBUTION statement specifies the Weibull distribution as the 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, of the censor-variable (in this case, CENSOR) to indicate censoring times. The COVB option requests the ML parameter estimate covariance matrix. The INSET statement controls the appearance of the inset box on the plot. By default, the inset box is transparent to data points and grid lines. The INSET statement provides a white background in the inset box.

The graphical output, displayed in Figure 1, consists of a probability plot of the data, a 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 placed on the plot by default. The locations of the right censored data values are plotted in an area at the top of the plot.

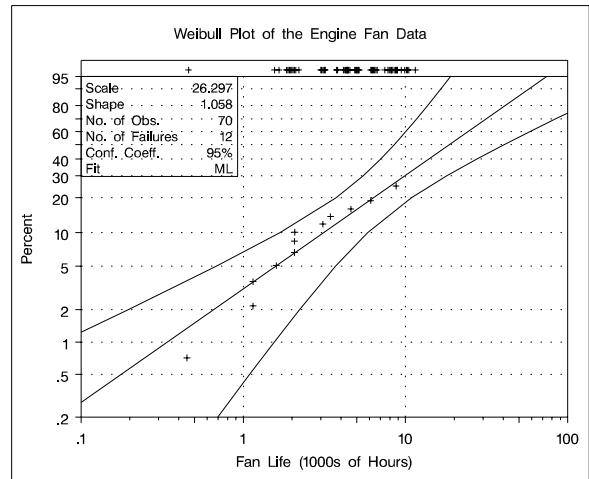


Figure 1. Weibull Probability Plot for the Engine Fan Data

The tabular output produced by the preceding SAS statements is shown in Figure 2. This consists of summary data, fit information, parameter estimates, distribution percentile estimates, standard errors, and confidence intervals for all estimated quantities.

### 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 given by Nelson (1990, p. 129). 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 in Figure 3 for the variable TIME at each level of the *group-variable* VOLTAGE.

```
proc reliability data=fluid;
  distribution weibull;
  probplot time = voltage / overlay noconf;
run;
```

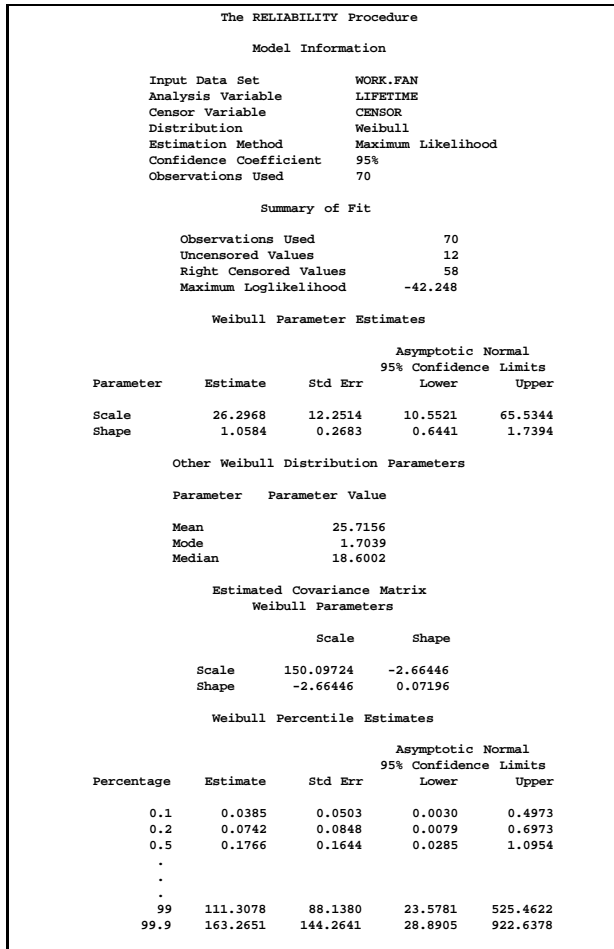


Figure 2. Tabular Output for the Fan Data Analysis

The input data set FLUID is specified by the DATA= option in the PROC RELIABILITY statement. The PROBLOT 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.

A summary table that contains information for all groups is displayed. In addition, information identical to that in Figure 2 is tabulated for each level of voltage. The summary table for all groups and the tables for the 26kV group are shown in Figure 4.

### Analysis of Accelerated Life Test Data

The following example illustrates the analysis of an accelerated life test for Class-B electrical motor insulation using data given by Nelson (1990, p. 243). Forty insulation specimens were tested at four temperatures: 150°, 170°, 190°, and 220°C. The purpose of the test was to

estimate the median life of the insulation at the design operating temperature of 130°C.

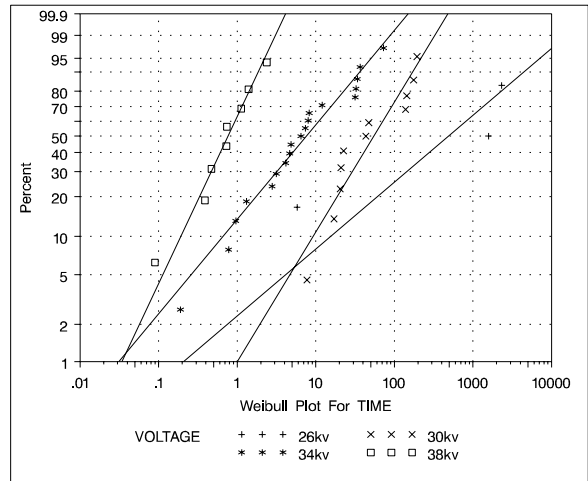


Figure 3. Weibull Probability Plot for the Insulating Fluid Data

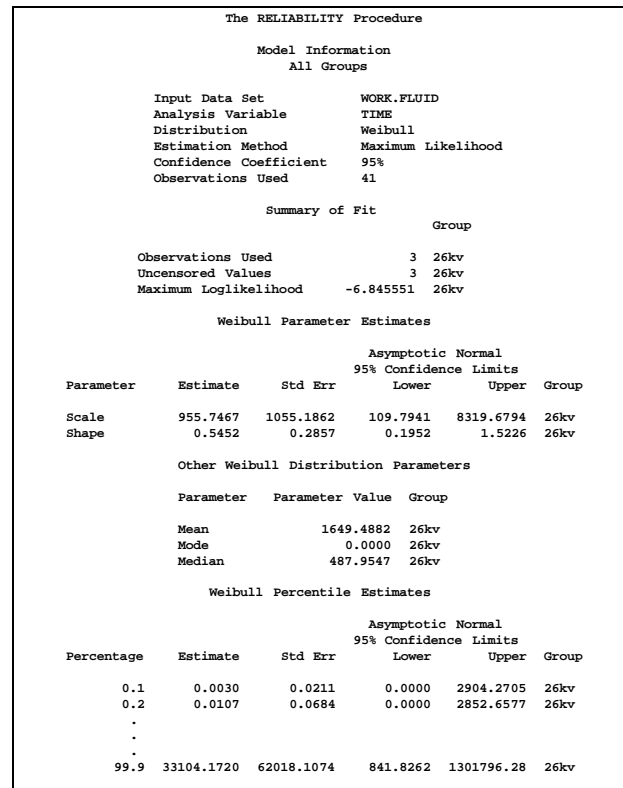


Figure 4. Partial Listing of the Tabular Output for the Insulating Fluid Data

Ten specimens of the insulation were tested at each test temperature. An Arrhenius-lognormal model is fit to the data in this example. In other words, the

fitted model has the lognormal (base 10) distribution, and its location parameter  $\mu$  depends on the centigrade temperature TEMP through the Arrhenius relationship

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

where

$$x = \frac{1000}{TEMP + 273.16}$$

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 display the fitted model distributions side-by-side on the probability plot and the relation plot shown in Figure 5.

```
proc reliability data=classb;
  distribution lognormal10;
  freq count;
  model hours*censor(1) = temp /
    relation=arr
    obstats( q=.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.

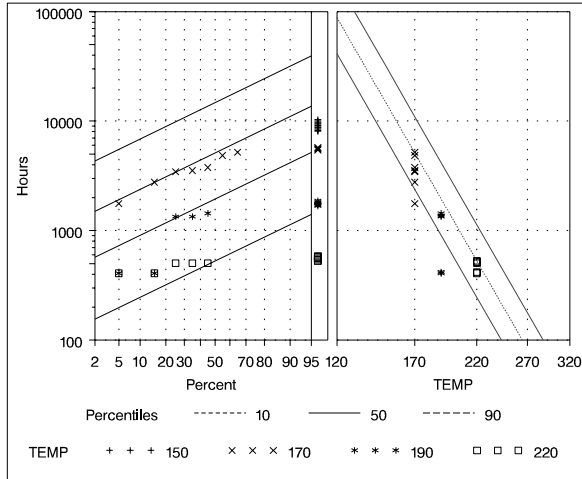
The option OBSTATS requests observation-wise statistics. The options in parentheses following OBSTATS indicate which statistics are to be computed. In this case, QUANTILE = .1 .5 .9 specifies that quantiles of the fitted distribution are to be computed for the value of temperature at each observation. 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 fit 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 PLOT 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). Weibull, extreme value, lognormal (base  $e$ ), normal, loglogistic, 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 10<sup>th</sup>, 50<sup>th</sup>, and 90<sup>th</sup> 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 5. 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.

A partial listing of the tabular output provided is shown in Figure 6.



**Figure 5.** Relation Plot for the Class B Insulation Data

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  $\beta_0$ , TEMP is the estimate of  $\beta_1$ , and Scale is the estimate of the lognormal scale parameter,  $\sigma$ .

The RELIABILITY Procedure					
Lognormal Parameter Estimates					
Parameter	Estimate	Std Err	Asymptotic Normal 95% Confidence Limits		
			Lower	Upper	
INTERCEPT	-6.0184	0.9467	-7.8740	-4.1629	
TEMP	4.3105	0.4366	3.4548	5.1662	
Scale	0.2592	0.0473	0.1812	0.3708	
Observation Statistics					
HOURS	CENSOR	TEMP	COUNT	Prob	Percentile
.	.	130.00000	.	0.1000	21937.5
.	.	130.00000	.	0.5000	47134.8
.	.	130.00000	.	0.9000	101273.6
Observation Statistics					
	Std Error	Lower	Upper		
	6959.1	11780.6	40851.5		
	16125.4	24106.5	92161.2		
	42060.7	44872.1	228568.1		

**Figure 6.** MODEL Statement Output for the Class B Data

The Observation Statistics table provides the estimates of the fitted distribution quantiles, their standard errors, and confidence limits. These are provided 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 is approximately 47,134 hours.

In addition to the MODEL statement output in Figure 6, the RELIABILITY procedure produces tabular output for

each temperature that is identical to the output produced with the PROBLOT statement. This output is not shown here.

### Weibull Analysis with Interval Data

Nelson (1982, p. 415) provides data for 167 identical turbine parts. 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 fit with a Weibull distribution.

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 only known to be greater than 63.48 months.

The RELIABILITY procedure requires that interval data be read from a SAS data set with a special structure. All units must have a common inspection schedule. The SAS data set named CRACKS shown in Figure 7 provides the data with this structure. 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.

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

**Figure 7.** Listing of the Turbine Part Cracking Data

The following statements use the RELIABILITY procedure to produce the probability plot in Figure 8 for the data in CRACKS.

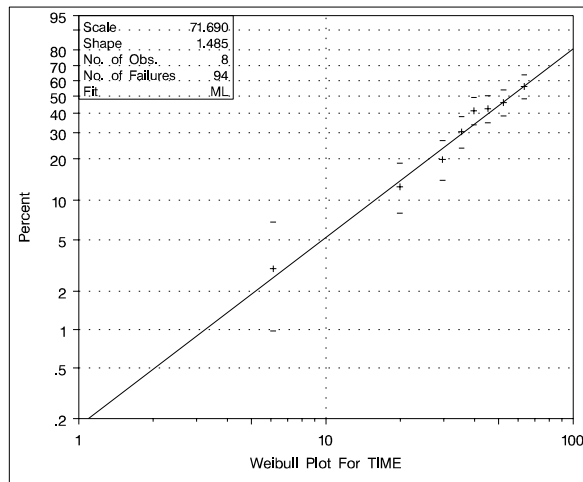
```
proc reliability data=cracks;
  freq fail;
  nenter units;
  distribution weibull;
  probplot time / readout pconfplt noconf;
run;
```

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 is 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 CRACKS are readout (or interval) data. The option PCONFPLT specifies that confidence intervals for the cumulative probability of failure are to 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 8.** Weibull Probability Plot for the Part Cracking Data

A partial listing of the tabular output produced by the preceding SAS statements is shown in Figure 9. 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.

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; that is, they might be right censored. The special structure of the input SAS data set required

for interval data allows the RELIABILITY procedure to analyze this more general case.

The RELIABILITY Procedure					
Readout Data Non-Parametric Cumulative Probability Estimates					
Interval Endpoint	Cum Prob	95% Confidence Limits		Std Err	
		Lower	Upper		
6.12	0.0299	0.0098	0.0685	0.0132	
19.92	0.1257	0.0796	0.1858	0.0257	
.	.	.	.	.	
63.48	0.5629	0.4841	0.6394	0.0384	
Weibull Parameter Estimates					
Parameter	Estimate	Std Err	Asymptotic Normal 95% Confidence Limits		
			Lower	Upper	
Scale	71.6904	5.3335	61.9634	82.9444	
Shape	1.4854	0.1465	1.2242	1.8022	
Weibull Percentile Estimates					
Percentage	Estimate	Std Err	Asymptotic Normal 95% Confidence Limits		
			Lower	Upper	
0.1	0.68534365	0.29999856	0.29060837	1.6162505	
0.2	1.09324644	0.42889769	0.50673205	2.35861884	
.	.	.	.	.	
.	.	.	.	.	
99.9	263.348061	44.7205353	188.791772	367.347584	

**Figure 9.** Partial Listing of the Tabular Output for the Part Cracking Data

### Regression Modeling

This example is an illustration of a Weibull regression model using a load accelerated life test of rolling bearings with data provided by Nelson (1990, p. 305). Bearings were tested at four different loads, and lifetimes in  $10^6$  of revolutions were measured. An outlier identified by Nelson (1990) has been omitted from the analysis.

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

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

where  $\mu_i$  is the location parameter of the extreme value distribution and

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

for the  $i$ th bearing.

Figure 10 shows a partial listing of the bearing data.

The following statements fit the regression model by maximum likelihood using the Weibull distribution.

```

proc reliability data=bearing;
  distribution weibull;
  model life = lload / covb
        corrb
        obstats
        ;
  make 'modobstats' out = RESIDUAL;
run;

```

OBS	LIFE	LOAD	LLOAD
1	1.67	0.87	-0.13926
2	2.20	0.87	-0.13926
3	2.51	0.87	-0.13926
.	.	.	.
.	.	.	.
39	0.456	1.18	0.16551

Figure 10. Partial Listing of the Bearing Data

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 MAKE statement creates a SAS data set named RESIDUAL that contains the table created by the OBSTATS option.

Figure 11 shows a partial listing of 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  $\beta_0$ , LLOAD corresponds to  $\beta_1$ , and SHAPE corresponds to the Weibull shape parameter. Figure 12 shows a partial listing of the output data set RESIDUAL.

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

$$\mathbf{x}_i^T \hat{\beta} = \hat{\beta}_0 + \text{LLOAD} \hat{\beta}_1$$

for each observation. The variable \_SURV\_ contains the value of the reliability function, \_SRESID\_ contains the standardized residual, and \_ARESID\_ contains a residual adjusted for right censored observations.

If the fit regression model is adequate, the standardized residuals will have a standard extreme value distribution. You can check the residuals by creating an extreme value probability plot of the residuals using the RELIABILITY

procedure and the RESIDUAL data set.

The RELIABILITY Procedure				
Model Information				
Input Data Set	WORK.BEARING			
Analysis Variable	LIFE			
Distribution	Weibull			
Weibull Parameter Estimates				
Parameter	Estimate	Std Err	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
Shape	1.2434	0.1545	0.9746	1.5862

Figure 11. Analysis Results for the Bearing Data

OBS	LIFE	LLOAD	_XBETA_	_SURV_	_SRESID_	_ARESID_
1	1.67	-0.13926	2.76147	0.94077	-2.79592	-2.79592
2	2.20	-0.13926	2.76147	0.91758	-2.45320	-2.45320
3	2.51	-0.13926	2.76147	0.90363	-2.28930	-2.28930
.	.	.	.	.	.	.
.	.	.	.	.	.	.
39	0.456	0.16551	-1.46058	0.09871	0.83967	0.83967

Figure 12. Partial Listing of RESIDUAL

The following statements create the plot displayed in Figure 13.

```

proc reliability data=residual;
  distribution ev;
  probplot _sresid_;
run;

```

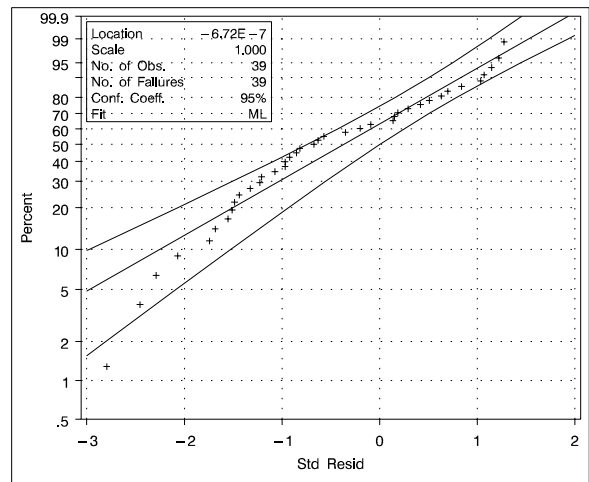


Figure 13. Extreme Value Probability Plot for the Standardized Residuals

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 may be inadequate.

## 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.

Repair histories of 41 diesel engines in a fleet were reported by Nelson (1995). The valve seats in these engines wear out and must be replaced.

The following statements produce the graphical display in Figure 14.

```
proc reliability data=valve;
  unitid id;
  mcfplot days*value( -1 );
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 that identifies the end of history for each engine (system).

In Figure 14, 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 could be based on a fitted line.

A partial listing of the tabular output is shown in Figure 15. It contains a summary of the repair data, estimates of the MCF, the Nelson (1995) standard errors and confidence intervals for the MCF. MCF estimates, standard errors, and confidence limits are shown as missing values (.) at end of history points, since they are not computed at these points.

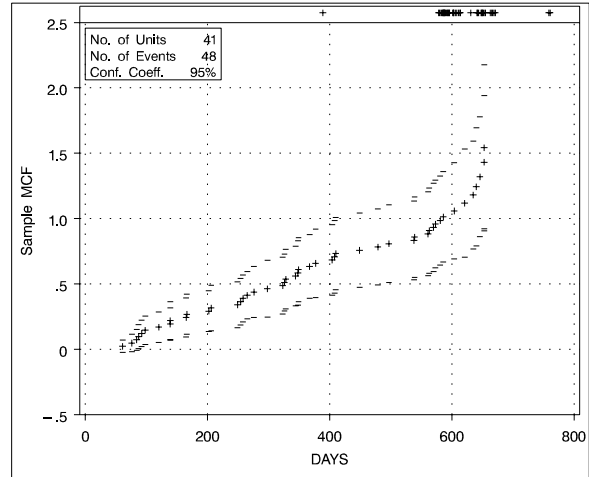


Figure 14. Mean Cumulative Function for the Number of Repairs

The RELIABILITY Procedure					
Repair Data Summary					
Input Data Set	WORK.VALVE				
Observations Used	89				
Number of Units	41				
Number of Events	48				
Repair Data Analysis					
Approximate 95% Confidence Limits					
Age	Sample MCF	Std Error	Lower	Upper	Unit ID
61.00	0.024	0.024	-0.023	0.072	393
76.00	0.049	0.034	-0.018	0.116	395
84.00	0.073	0.041	-0.008	0.154	330
.	.	.	.	.	.
.	.	.	.	.	.
667.00	.	.	.	.	327
759.00	.	.	.	.	252
761.00	.	.	.	.	251

Figure 15. Partial Listing of the Output for the Valve Seat Data

There has been considerable work in parametric modeling of the repair process, which requires more assumptions. Ascher and Feingold (1984) describe parametric models for repair processes. For example, repairs are sometimes modeled as a non-homogeneous Poisson process. The current release of the RELIABILITY procedure does not include this type of parametric modeling, although it is planned for future releases. The MCF plot might be a first step in modeling a repair process, but in many cases, it will provide the required answers without further analysis. An estimate of the MCF for a sample of systems will aid engineers in determining the repair rate at any age and whether the rate increases or decreases with population age. The estimate is also useful for predicting the number of future repairs.

## Comparison of Two Samples of Repair Data

Nelson (1995) and Doganaksoy and Nelson (1991) 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 MCF functions, which can be used to assess whether the difference is statistically significant.

Doganaksoy and Nelson (1991) give an example of two samples of locomotives with braking grids from two different production batches.

The following statements request the Nelson (1995) non-parametric estimate and confidence limits for the difference of the MCF functions shown in Figure 16 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 is to be computed and plotted. Confidence limits for the MCF difference are also computed and plotted. The UNITID statement specifies that the variable ID uniquely identifies each system.

Figure 16 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 partial listing of the tabular output is shown in Figure 17. It contains a summary of the repair data for the two samples, estimates, standard errors, and confidence intervals for the MCF difference.

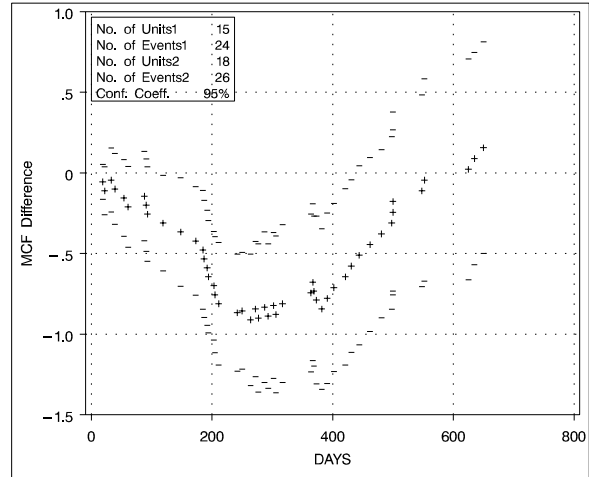


Figure 16. Mean Cumulative Function Difference

The RELIABILITY Procedure						
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				
Sample MCF Differences						
Age	MCF Difference	Std Error	Approximate 95% Confidence Limits			Unit ID
			Lower	Upper		
19.00	-0.06	0.056	-0.16	0.05	S2-16	
22.00	-0.11	0.076	-0.26	0.04	S2-12	
33.00	-0.04	0.101	-0.24	0.15	S1-13	
.	.	.	.	.	.	
.	.	.	.	.	.	
625.00	0.02	0.349	-0.66	0.71	S1-07	
635.00	0.09	0.336	-0.57	0.75	S1-11	
650.00	0.16	0.335	-0.50	0.81	S1-11	

Figure 17. Partial Listing of the Output for the Braking Grids Data

## References

- Ascher, H. and Feingold, H. (1984), *Repairable Systems Reliability*, New York: Marcel Dekker, Inc.
- Doganaksoy, N. and Nelson, W. (1991), "A Method and Computer Program MCFDIFF to Compare Two Samples of Repair Data," GE Research & Development Center TIS Report 91CRD172, P.O. Box 8, Schenectady, NY 12301.
- Lawless, J.E. (1983), *Statistical Models and Methods for Lifetime Data*, New York: John Wiley & Sons, (800) 879-4539.
- Nelson, W. (1982), *Applied Life Data Analysis*, New York: John Wiley & Sons, (800) 879-4539.

- Nelson, W. (1985), "Weibull Analysis of Reliability Data with Few or No Failures," *Journal of Quality Technology*, 17 (3), 140-146.
- Nelson, W. (1988), "Graphical Analysis of System Repair Data," *Journal of Quality Technology*, 20 (1), 24-35.
- Nelson, W. (1990), *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*, New York: John Wiley & Sons, (800) 879-4539.
- Nelson, W. (1995), "Confidence Limits for Recurrence Data--Applied to Cost or Number of Product Repairs," *Technometrics*, 37, 147-157.
- 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," GE Research & Development Center TIS Report 89CRD239, P.O. Box 8, Schenectady, NY 12301.
- Tobias, P.A. and Trindade, D.C. (1995), *Applied Reliability, Second Edition*, New York: Van Nostrand Reinhold.

SAS is a registered trademark of SAS Institute Inc. in the USA and in other countries. ® indicates USA registration.