Chapter 37
The LIFETEST Procedure

Chapter Table of Contents

Chapter 37
The LIFETEST Procedure

Overview

A common feature of lifetime or survival data is the presence of right-censored observations due either to withdrawal of experimental units or to termination of the experiment. For such observations, you know only that the lifetime exceeded a given value; the exact lifetime remains unknown. Such data cannot be analyzed by ignoring the censored observations because, among other considerations, the longer-lived units are generally more likely to be censored. The analysis methodology must correctly use the censored observations as well as the noncensored observations. Several texts that discuss the survival analysis methodology are Collett (1994), Cox and Oakes (1984), Kalbfleish and Prentice (1980), Lawless (1982), and Lee (1992).

Usually, a first step in the analysis of survival data is the estimation of the distribution of the survival times. Survival times are often called *failure* times, and *event* times are uncensored survival times. The survival distribution function (SDF), also known as the survivor function, is used to describe the lifetimes of the population of interest. The SDF evaluated at t is the probability that an experimental unit from the population will have a lifetime exceeding t, that is

$$S(t) = \Pr(T > t)$$

where $S(t)$ denotes the survivor function and T is the lifetime of a randomly selected experimental unit. The LIFETEST procedure can be used to compute nonparametric estimates of the survivor function either by the product-limit method (also called the Kaplan-Meier method) or by the life table method.

Some functions closely related to the SDF are the cumulative distribution function (CDF), the probability density function (PDF), and the hazard function. The CDF, denoted $F(t)$, is defined as $1 - S(t)$ and is the probability that a lifetime does not exceed t. The PDF, denoted $f(t)$, is defined as the derivative of $F(t)$, and the hazard function, denoted $h(t)$, is defined as $f(t)/S(t)$. If the life table method is chosen, the estimates of the probability density function and the hazard function can also be computed. Plots of these estimates can be produced by a graphical or line printer device.

An important task in the analysis of survival data is the comparison of survival curves. It is of interest to determine whether two or more samples have arisen from identical survivor functions. PROC LIFETEST provides two rank tests and a likelihood ratio test for testing the homogeneity of survival functions across strata. The rank tests are censored-data generalizations of the Savage (exponential scores) test and the Wilcoxon test. The generalized Savage test is also known as the log-rank test, while the generalized Wilcoxon test is simply referred to as the Wilcoxon test. The likeli-

hood ratio test is based on an underlying exponential model, whereas the rank tests are not.

Often there are prognostic variables called covariates that are thought to be related to the failure time. These variables can be used to define strata, and the resulting SDF estimates can be compared visually or by using the tests of homogeneity of strata. The variables can also be used to construct statistics to test for association between the covariates and the lifetime variable. PROC LIFETEST can compute two such test statistics: censored data linear rank statistics based on the exponential scores and the Wilcoxon scores. The corresponding tests are known as the log-rank test and the Wilcoxon test, respectively. These tests are computed by pooling over any defined strata, thus adjusting for the stratum variables. Except for a difference in the treatment of ties, these two rank tests are the same as those used to test for homogeneity over strata.

Getting Started

You can use the LIFETEST procedure to compute nonparametric estimates of the survivor function and to compute rank tests for association of the response variable with other variables.

For simple analyses, only the PROC LIFETEST and TIME statements are required. Consider a sample of survival data. Suppose that the time variable is t and the censoring variable is c with value 1 indicating censored observations. The following statements compute the product-limit estimate for the sample:

```
proc lifetest;
   time t*c(1);
run;
```

You can use the STRATA statement to divide the data into various strata. A separate survivor function is then estimated for each stratum, and tests of the homogeneity of strata are performed. You can specify covariates in the TEST statement. PROC LIFETEST computes linear rank statistics to test the effects of these covariates on survival.

For example, consider the results of a small randomized trial on rats. Suppose you assign forty rats exposed to a carcinogen into two treatment groups. The event of interest is death from cancer induced by the carcinogen. The response is the time from randomization to death. Four rats died of other causes; their survival times are regarded as censored observations. Interest lies in whether the survival distributions differ between the two treatments.

The data set Exposed contains four variables: Days (survival time in days from treatment to death), Status (censoring indicator variable: 0 if censored and 1 if not censored), Treatment (treatment indicator), and Sex (gender: F if female and M if male).

```
data Exposed;
  input Days Status Treatment Sex $ @@;
  datalines;
179   1   1   F   378   0   1   M
256   1   1   F   355   1   1   M
262   1   1   M   319   1   1   M
256   1   1   F   256   1   1   M
255   1   1   M   171   1   1   F
224   0   1   F   325   1   1   M
225   1   1   F   325   1   1   M
287   1   1   M   217   1   1   F
319   1   1   M   255   1   1   F
264   1   1   M   256   1   1   F
237   0   2   F   291   1   2   M
156   1   2   F   323   1   2   M
270   1   2   M   253   1   2   M
257   1   2   M   206   1   2   F
242   1   2   M   206   1   2   F
157   1   2   F   237   1   2   M
249   1   2   M   211   1   2   F
180   1   2   F   229   1   2   F
226   1   2   F   234   1   2   F
268   0   2   M   209   1   2   F
;
```

PROC LIFETEST is invoked to compute the product-limit estimate of the survivor function for each treatment and to compare the survivor functions between the two treatments. In the TIME statement, the survival time variable, Days, is crossed with the censoring variable, Status, with the value 0 indicating censoring. That is, the values of Days are considered censored if the corresponding values of Status are 0; otherwise, they are considered as event times. In the STRATA statement, the variable Treatment is specified, which indicates that the data are to be divided into strata based on the values of Treatment. PROC LIFETEST computes the product-limit estimate for each stratum and tests whether the survivor functions are identical across strata.

```
symbol1 c=blue; symbol2 c=orange;
proc lifetest data=Exposed plots=(s,ls,lls);
   time Days*Status(0);
   strata Treatment;
run;
```

The PLOTS= option in the PROC LIFETEST statement is used to request a plot of the estimated survivor function against time (by specifying S), a plot of the negative log of the estimated survivor function against time (by specifying LS), and a plot of the log of the negative log of the estimated survivor function against log time (by specifying LLS). The LS and LLS plots provide an empirical check of the appropriateness of the exponential model and the Weibull model, respectively, for the survival data (Kalbfleisch and Prentice 1980, Chapter 2).

If the exponential model is appropriate, the LS curve should be approximately linear through the origin. If the Weibull model is appropriate, the LLS curve should be approximately linear. Since there are more than one stratum, the LLS plot may also be used to check the proportional hazards model assumption. Under this assumption, the LLS curves should be approximately parallel across strata.

The results of the analysis are displayed in the following figures.

Figure 37.1 displays the product-limit survival estimate for the first stratum (Treatment=1). The figure lists, for each observed time, the survival estimate, failure rate, standard error of the estimate, number of failures, and number of subjects remaining in the study.

The SAS System

The LIFETEST Procedure

Stratum 1: Treatment = 1

Product-Limit Survival Estimates

Days	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.000	1.0000	0	0	0	20
171.000	0.9500	0.0500	0.0487	1	19
179.000	0.9000	0.1000	0.0671	2	18
217.000	0.8500	0.1500	0.0798	3	17
224.000*	.	.	.	3	16
225.000	0.7969	0.2031	0.0908	4	15
255.000	.	.	.	5	14
255.000	0.6906	0.3094	0.1053	6	13
256.000	.	.	.	7	12
256.000	.	.	.	8	11
256.000	.	.	.	9	10
256.000	0.4781	0.5219	0.1146	10	9
262.000	0.4250	0.5750	0.1135	11	8
264.000	0.3719	0.6281	0.1111	12	7
287.000	0.3188	0.6813	0.1071	13	6
319.000	.	.	.	14	5
319.000	0.2125	0.7875	0.0942	15	4
325.000	.	.	.	16	3
325.000	0.1063	0.8938	0.0710	17	2
355.000	0.0531	0.9469	0.0517	18	1
378.000*	.	.	.	18	0

NOTE: The marked survival times are censored observations.

Figure 37.1. Product-Limit Survivor Function Estimate for Treatment=1

Figure 37.2 displays summary statistics of survival times for Treatment=1. It contains estimates of the 25th, 50th, and 75th percentiles and the corresponding 95% confidence limits.

The median survival time for rats in this treatment is 256 days. The mean and standard error are also displayed; however, it is noted that these values are underestimated because the largest observed time is censored and the estimation is restricted to the largest event time.

The LIFETEST Procedure

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval [Lower	Upper)
75	319.000	262.000	325.000
50	256.000	255.000	319.000
25	255.000	217.000	256.000

Mean	Standard Error
271.131	11.877

NOTE: The mean survival time and its standard error were underestimated because the largest observation was censored and the estimation was restricted to the largest event time.

Figure 37.2. Summary Statistics of Survival Times for Treatment=1

The LIFETEST Procedure

Stratum 2: Treatment = 2

Product-Limit Survival Estimates

Days	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.000	1.0000	0	0	0	20
156.000	0.9500	0.0500	0.0487	1	19
157.000	0.9000	0.1000	0.0671	2	18
180.000	0.8500	0.1500	0.0798	3	17
206.000	.	.	.	4	16
206.000	0.7500	0.2500	0.0968	5	15
209.000	0.7000	0.3000	0.1025	6	14
211.000	0.6500	0.3500	0.1067	7	13
226.000	0.6000	0.4000	0.1095	8	12
229.000	0.5500	0.4500	0.1112	9	11
234.000	0.5000	0.5000	0.1118	10	10
237.000	0.4500	0.5500	0.1112	11	9
237.000*	.	.	.	11	8
242.000	0.3938	0.6063	0.1106	12	7
249.000	0.3375	0.6625	0.1082	13	6
253.000	0.2813	0.7188	0.1038	14	5
257.000	0.2250	0.7750	0.0971	15	4
268.000*	.	.	.	15	3
270.000	0.1500	0.8500	0.0891	16	2
291.000	0.0750	0.9250	0.0693	17	1
323.000	0	1.0000	0	18	0

NOTE: The marked survival times are censored observations.

Figure 37.3. Product-Limit Survivor Function Estimate for Treatment=2

Figure 37.3 and Figure 37.4 display the survival estimates and the summary statistics of the survival times for Treatment=2. The median survival time for rats in this treatment is 235 days.

The LIFETEST Procedure

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval [Lower	Upper)
75	257.000	237.000	291.000
50	235.500	209.000	253.000
25	207.500	180.000	234.000

Mean	Standard Error
235.156	10.211

Figure 37.4. Survival Times Summary for Treatment=2

A summary of the number of censored and event observations is shown in Figure 37.5. The figure lists, for each stratum, the number of event and censored observations, and the percentage of censored observations.

The LIFETEST Procedure

Summary of the Number of Censored and Uncensored Values

Stratum	Treatment	Total	Failed	Censored	Percent Censored
1	1	20	18	2	10.00
2	2	20	18	2	10.00
Total		40	36	4	10.00

Figure 37.5. Summary of Censored and Uncensored Values

Figure 37.6 displays the graph of the product-limit survivor function estimates versus survival time. The two treatments differ primarily at larger survival times.

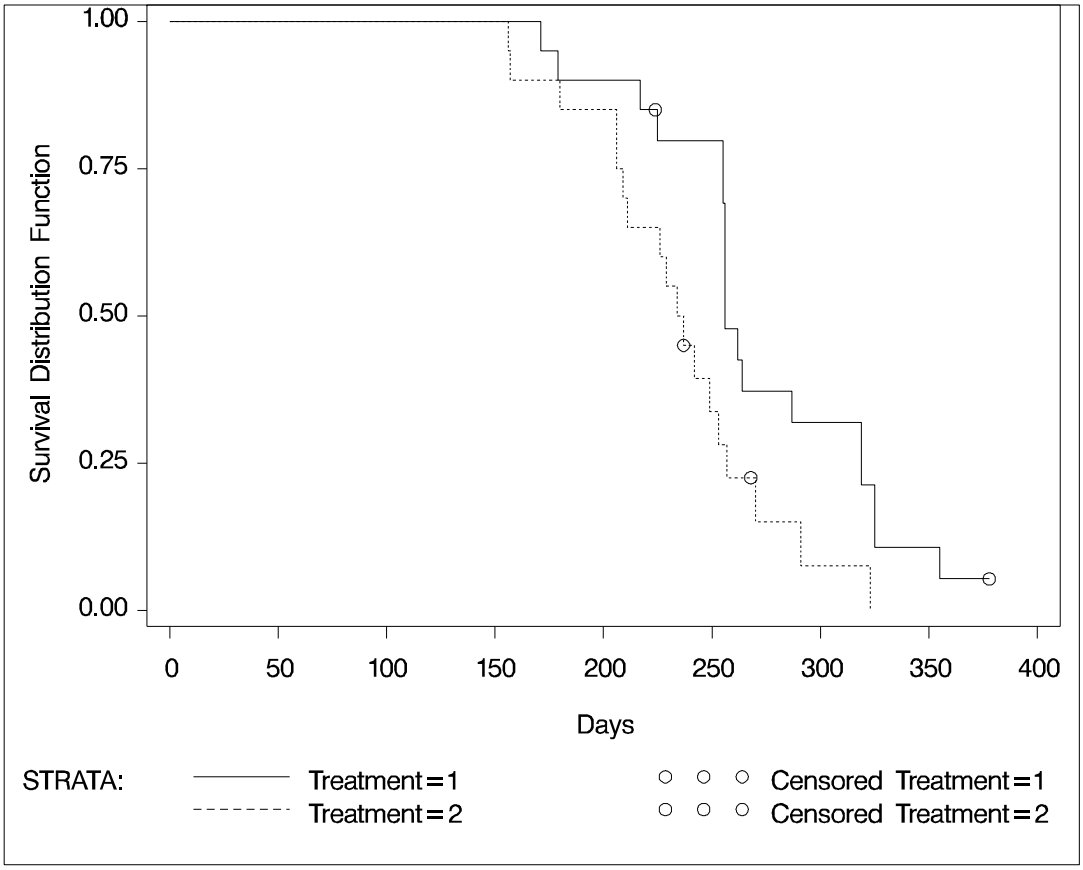


Figure 37.6. Product-Limit Survivor Functions

Figure 37.7 displays the graph of the log survival function estimates versus survival time for the two treatments. Neither curve approximates a straight line through the origin; therefore, the exponential model is not appropriate for the survival data.

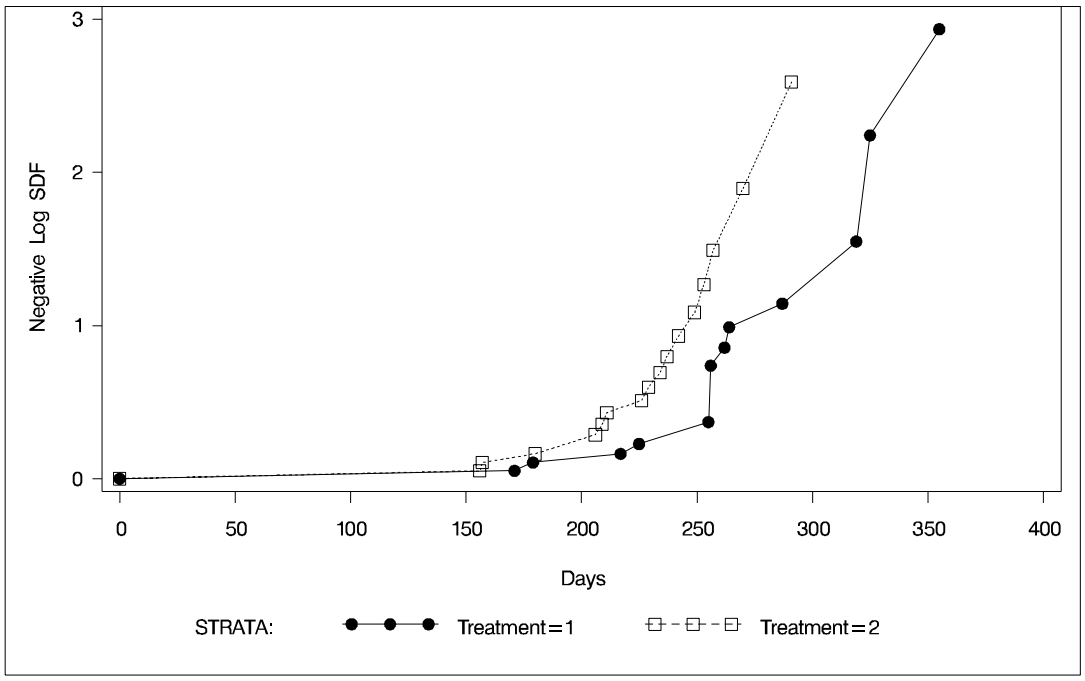


Figure 37.7. Log Survivor Function Estimates

Figure 37.8 displays the graph of the negative log-log survivor function estimates versus log time for the two treatments.

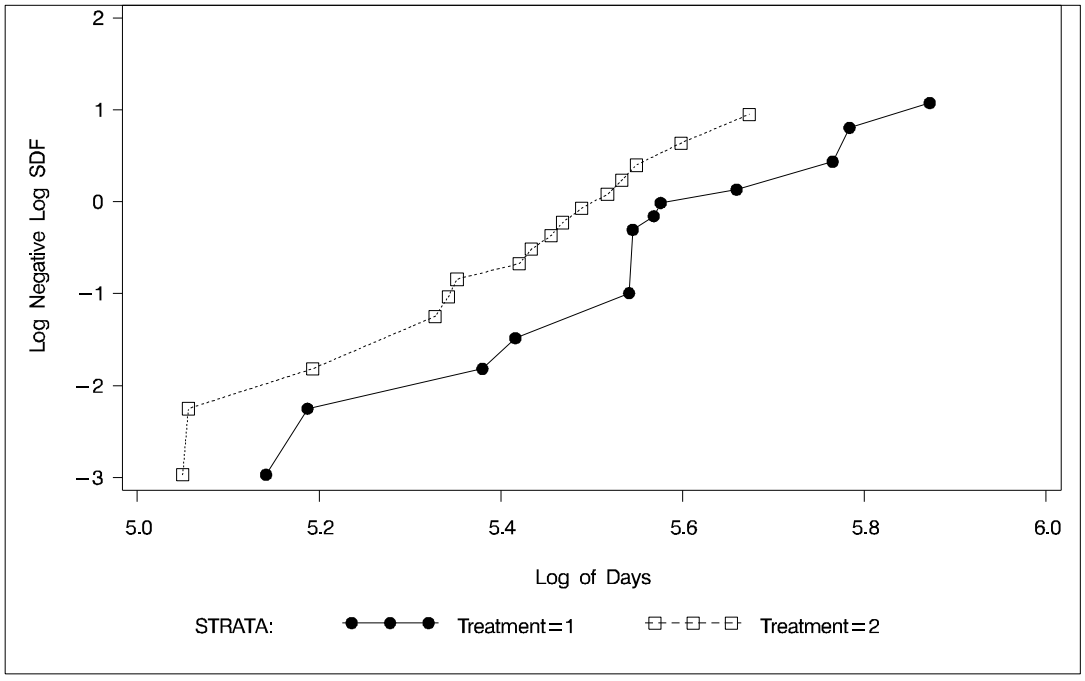


Figure 37.8. Log of Negative Log Survivor Function Estimates

The LIFETEST Procedure

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	5.6485	1	0.0175
Wilcoxon	5.0312	1	0.0249
-2Log(LR)	0.1983	1	0.6561

Figure 37.9. Tests for Strata Homogeneity

Results of the comparison of survival curves between the two treatments are shown in Figure 37.9. The rank tests for homogeneity indicate a significant difference between the treatments (p=0.0175 for the log-rank test and p=0.0249 for the Wilcoxon test). Rats in Treatment=1 live significantly longer than those in Treatment=2. The log-rank test, which places more weight on larger survival times, is more significant than the Wilcoxon test, which places more weight on early survival times. As noted earlier, the exponential model is not appropriate for the given survival data; consequently, the result of the likelihood ratio test should be ignored.

Next, suppose that gender is thought to be related to survival time, and you want to study the treatment effect while adjusting for the gender of the rats. By specifying the variable Sex in the STRATA statement and by specifying the variable Treatment in the TEST statement, you can test the effect of Treatment while adjusting for the effect of Sex. The log-rank and Wilcoxon linear rank statistics are computed by pooling over the strata defined by the values of Sex, thus adjusting for the effect of Sex.

The NOTABLE option is added to the PROC LIFETEST statement to avoid estimating a survival curve for each gender.

```
proc lifetest data=Exposed notable;
   time Days*Status(0);
   strata Sex;
   test Treatment;
run;
```

Results of the linear rank tests are shown in Figure 37.10. The treatment effect is statistically significant for both the Wilcoxon test (p=0.0147) and the log-rank test (p=0.0075). As compared to the results of the homogenity test in Figure 37.9, the significance of the treatment effect has been sharpened by controlling for the effect of the gender of the subjects.

The LIFETEST Procedure

Univariate Chi-Squares for the Wilcoxon Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
Treatment	-4.2372	1.7371	5.9503	0.0147

Univariate Chi-Squares for the Log-Rank Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
Treatment	-6.8021	2.5419	7.1609	0.0075

Figure 37.10. Tests for Association of Time with Covariates

Syntax

The following statements are available in PROC LIFETEST:

PROC LIFETEST < *options* > **;**
TIME *variable* < *****censor(list)* > **;**
BY *variables* **;**
FREQ *variable* **;**
ID *variables* **;**
STRATA *variable* < *(list)* > < . . . *variable* < *(list)* > > **;**
TEST *variables* **;**

The simplest use of PROC LIFETEST is to request the nonparametric estimates of the survivor function for a sample of survival times. In such a case, only the PROC LIFETEST statement and the TIME statement are required. You can use the STRATA statement to divide the data into various strata. A separate survivor function is then estimated for each stratum, and tests of the homogeneity of strata are performed. You can specify covariates in the TEST statement. PROC LIFETEST computes linear rank statistics to test the effects of these covariates on survival.

The PROC LIFETEST statement invokes the procedure. All statements except the TIME statement are optional, and there is no required order for the statements following the PROC LIFETEST statement. The TIME statement is used to specify the variables that define the survival time and censoring indicator. The STRATA statement specifies a variable or set of variables defining the strata for the analysis. The TEST statement specifies a list of numeric covariates to be tested for their association with the response survival time. Each variable is tested individually, and a joint test statistic is also computed. The ID statement provides a list of variables whose values are used to identify observations in the product-limit estimates of the survival function. When only the TIME statement appears, no strata are defined and no tests of homogeneity are performed.

PROC LIFETEST Statement

PROC LIFETEST < *options* > **;**

The PROC LIFETEST statement invokes the procedure. The following options can appear in the PROC LIFETEST statement and are described in alphabetic order. If no options are requested, PROC LIFETEST computes and displays product-limit estimates of the survival distribution within each stratum and tests the equality of the survival functions across strata.

Task	Options	Description
Specify Data Set	DATA=	specifies the input SAS data set
	OUTSURV=	names an output data set to contain survival estimates and confidence limits
	OUTTEST=	names an output data set to contain rank test statistics for association of survival time with covariates limits
Specify Model	ALPHA=	sets confidence level for survival estimates
	ALPHAQT=	sets confidence level for survival time quartiles
	INTERVALS=	specifies interval endpoints for life table calculations
	MAXTIME=	sets maximum value of time variable for plot
	METHOD=	specifies method to compute survivor function
	MISSING	allows missing values to be a stratum level
	NINTERVAL=	specifies number of intervals for life table estimates
	SINGULAR=	sets tolerance for testing singularity of covariance matrix of rank statistics
	TIMELIM=	specifies the time limit used to estimate the mean survival time and its standard error
	WIDTH=	specifies width of intervals for life table estimates
Control Output	CENSOREDSYMBOL=	defines symbol used for censored observations in plots
	EVENTSYMBOL=	specifies symbol used for event observations in plots
	FORMCHAR(1,2,7,9)=	defines characters used for line printer plot axes
	LINEPRINTER	specifies that plots are produced by line printer
	NOCENSPLOT	suppresses the plot of censored observations
	NOPRINT	suppresses display of output
	NOTABLE	suppresses display of survival function estimates

Table 37.0. (continued)

Task	Options	Description
	PLOTS=	plots survival estimates
	REDUCEOUT	specifies that only INTERVAL= or TIMELIST= observations are listed in the OUTSURV= data set
	TIMELIST=	specifies a list of time points at which the Kaplan-Meier estimates are displayed
Enhance Graphical Output	ANNOTATE=	specifies an annotate data set that adds features to plots
	DESCRIPTION=	specifies string that appears in the description field of the PROC GREPLAY master menu for the plots
	GOUT=	specifies graphics catalog name for saving graphics output
	LANNOTATE=	specifies an input data set that contains variables for local annotation

ALPHA=*value*

specifies a number between 0.0001 and 0.9999 that sets the confidence level for the confidence intervals for the survivor function. The confidence level for the interval is 1 - ALPHA. For example, the option ALPHA=0.05 requests a 95% confidence interval for the SDF at each time point. The default value is 0.05.

ALPHAQT=*value*

specifies a number between 0.0001 and 0.9999 that sets the level for the confidence intervals for the quartiles of the survival time. The confidence level for the interval is 1 - ALPHAQT. For example, the option ALPHAQT=0.05 requests a 95% confidence interval for the quantiles of the survival time. The default value is 0.05.

ANNOTATE=*SAS-data-set*
ANNO=*SAS-data-set*

specifies an input data set that contains appropriate variables for annotation. The ANNOTATE= option enables you to add features (for example, labels explaining extreme observations) to plots produced on graphics devices. The ANNOTATE= option cannot be used if the LINEPRINTER option is specified. The data set specified must be an ANNOTATE= type data set, as described in *SAS/GRAPH Software: Reference*.

The data set specified with the ANNOTATE= option in the PROC LIFETEST statement is "global" in the sense that the information in this data set is displayed on every plot produced by a single invocation of PROC LIFETEST.

CENSOREDSYMBOL=*name* **|** *'string'*
CS=*name* **|** *'string'*

specifies the symbol value for the censored observations. The value, *name* or *'string'*, is the symbol value specification allowed in SAS/GRAPH software. The default is CS=CIRCLE. If you want to omit plotting the censored observations, specify CS=NONE. The CENSOREDSYMBOL= option cannot be used if the LINEPRINTER option is specified.

DATA=*SAS-data-set*

names the SAS data set used by PROC LIFETEST. By default, the most recently created SAS data set is used.

DESCRIPTION='*string***'**
DES='*string***'**

specifies a descriptive string of up to 40 characters that appears in the "Description" field of the graphics catalog. The description does not appear on the plots. By default, PROC LIFETEST assigns a description of the form PLOT OF *vname* vs *hname*, where *vname* and *hname* are the names of the *y* variable and the *x* variable, respectively. The DESCRIPTION= option cannot be used if the LINEPRINTER option is specified.

EVENTSYMBOL=*name* **|** *'string'*
ES=*name* **|** *'string'*

specifies the symbol value for the event observations. The value, *name* or *'string'*, is the symbol value specification allowed in SAS/GRAPH software. The default is ES=NONE. The EVENTSYMBOL= option cannot be used if the LINEPRINTER option is specified.

FORMCHAR(1,2,7,9)=*'string'*

defines the characters used for constructing the vertical and horizontal axes of the line printer plots. The string should be four characters. The first and second characters define the vertical and horizontal bars, respectively, which are also used in drawing the *steps* of the product-limit survival function. The third character defines the tick mark for the axes, and the fourth character defines the lower left corner of the plot. If the FORMCHAR option in PROC LIFETEST is not specified, the value supplied, if any, with the system option FORMCHAR= is used. The default is FORMCHAR(1,2,7,9)='|-+-'. Any character or hexadecimal string can be used to customize the plot appearance. To send the plot output to a printer with the IBM graphics character set (1 or 2) or display it directly on your PC screen, you can use the following hexadecimal representation

```
formchar(1,2,7,9)='B3C4C5C0'x
```

or system option

```
formchar='B3C4DAC2BFC3C5B4C0C1D9'x
```

Refer to the chapter titled "The PLOT Procedure," in the *SAS Procedures Guide* or the section "System Options" in *SAS Language Reference: Dictionary* for further information.

GOUT=*graphics-catalog*

specifies the graphics catalog for saving graphics output from PROC LIFETEST. The default is WORK.GSEG. The GOUT= option cannot be used if the LINEPRINTER option is specified. For more information, refer to the chapter titled "The GREPLAY Procedure" in *SAS/GRAPH Software: Reference*.

INTERVALS=*values*

specifies a list of interval endpoints for the life table calculations. These endpoints must all be nonnegative numbers. The initial interval is assumed to start at zero whether or not zero is specified in the list. Each interval contains its lower endpoint but does not contain its upper endpoint. When this option is used with the product-limit method, it reduces the number of survival estimates displayed by displaying only the estimates for the smallest time within each specified interval. The INTERVALS= option can be specified in any of the following ways:

list separated by blanks	**intervals=1 3 5 7**
list separated by commas	**intervals=1,3,5,7**
x to y	**intervals=1 to 7**
x to y by z	**intervals=1 to 7 by 1**
combination of the above	**intervals=1,3 to 5,7**

For example, the specification

```
intervals=5,10 to 30 by 10
```

produces the set of intervals

$$\{[0, 5), [5, 10), [10, 20), [20, 30), [30, \infty)\}$$

LANNOTATE=*SAS-data-set*
LANN=*SAS-data-set*

specifies an input data set that contains variables for local annotation. You can use the LANNOTATE= option to specify a different annotation for each BY group, in which case the BY variables must be included in the LANNOTATE= data set. The LANNOTATE= option cannot be used if the LINEPRINTER option is specified. The data set specified must be an ANNOTATE= type data set, as described in *SAS/GRAPH Software: Reference*.

If there is no BY-group processing, the ANNOTATE= and LANNOTATE= options have the same effects.

LINEPRINTER
LS

specifies that plots are produced by a line printer instead of by a graphical device.

MAXTIME=*value*

specifies the maximum value of the time variable allowed on the plots so that outlying points do not determine the scale of the time axis of the plots. This parameter only affects the displayed plots and has no effect on the calculations.

METHOD=*type*

specifies the method used to compute the survival function estimates. Valid values for *type* are as follows.

PL | KM specifies that product-limit (PL) or Kaplan-Meier (KM) estimates are computed.

ACT | LIFE | LT specifies that life table (or actuarial) estimates are computed.

By default, METHOD=PL.

MISSING

allows missing values for numeric variables and blank values for character variables as valid stratum levels. See the section "Missing Values" on page 1818 for details.

By default, PROC LIFETEST does not use observations with missing values for any stratum variables.

NINTERVAL=*value*

specifies the number of intervals used to compute the life table estimates of the survivor function. This parameter is overridden by the WIDTH= option or the INTERVALS= option. When you specify the NINTERVAL= option, PROC LIFETEST tries to find an interval that results in round numbers for the endpoints. Consequently, the number of intervals may be different from the number requested. Use the INTERVALS= option to control the interval endpoints. The default is NINTERVAL=10.

NOCENSPLOT

NOCENS

requests that the plot of censored observations be suppressed when the PLOTS= option is specified. This option is not needed when the life table method is used to compute the survival estimates, since the plot of censored observations is not produced.

NOPRINT

suppresses the display of output. This option is useful when only an output data set is needed. Note that this option temporarily disables the Output Delivery System (ODS). For more information, see Chapter 15, "Using the Output Delivery System."

NOTABLE

suppresses the display of survival function estimates. Only the number of censored and event times, plots, and test results are displayed.

OUTSURV=*SAS-data-set*

OUTS=*SAS-data-set*

creates an output SAS data set to contain the estimates of the survival function and corresponding confidence limits for all strata. See the section "Output Data Sets" on page 1825 for more information on the contents of the OUTSURV= SAS data set.

OUTTEST=*SAS-data-set*

OUTT=*SAS-data-set*

creates an output SAS data set to contain the overall chi-square test statistic for association with failure time for the variables in the TEST statement, the values of the univariate rank test statistics for each variable in the TEST statement, and the estimated covariance matrix of the univariate rank test statistics. See the section "Output Data Sets" on page 1825 for more information on the contents of the OUTTEST= SAS data set.

PLOTS= *(type <(NAME=name)> <, ..., type <(NAME=name)> >)*

creates plots of survival estimates or censored observations, where *type* is the type of plot and *name* is a catalog entry name of up to eight characters. Valid values of *type* are as follows:

CENSORED | C	specifies a plot of censored observations by strata.
SURVIVAL | S	specifies a plot of the estimated SDF versus time.
LOGSURV | LS	specifies a plot of the $-\log(\text{estimated SDF})$ versus time.
LOGLOGS | LLS	specifies a plot of the $\log(-\log(\text{estimated SDF})$ versus $\log(\text{time})$.
HAZARD | H	specifies a plot of the estimated hazard function versus time.
PDF | P	specifies a plot of the estimated probability density function versus time.

Parentheses are required in specifying the plots. For example,

```
plots = (s)
```

requests a plot of the estimated survivor function versus time, and

```
plots = (s(name=Surv2), h(name=Haz2))
```

requests a plot of the estimated survivor function versus time and a plot of the estimated hazard function versus time, with Surv2 and Haz2 as their catalog names, respectively.

REDUCEOUT

specifies that the OUTSURV= data set contains only those observations that are included in the INTERVALS= or TIMELIST= option. This option has no effect if the OUTSURV= option is not specified. It also has no effect if neither the INTERVALS= option nor the TIMELIST= option is specified.

SINGULAR=*value*

specifies the tolerance for testing singularity of the covariance matrix for the rank test statistics. The test requires that a pivot for sweeping a covariance matrix be at least this number times a norm of the matrix. The default value is 1E-12.

TIMELIM=*time-limit*

specifies the time limit used in the estimation of the mean survival time and its standard error. The mean survival time can be shown to be the area under the Kaplan-Meier survival curve. However, if the largest observed time in the data is censored, the area under the survival curve is not a closed area. In such a situation, you can choose a time limit L and estimate the mean survival curve limited to a time L (Lee 1992, pp. 72–76). This option is ignored if the largest observed time is an event time. Valid *time-limit* values are as follows.

EVENT | LET specifies that the time limit L is the largest event time in the data. TIMELIM=EVENT is the default.

OBSERVED | LOT specifies that the time limit L is the largest observed time in the data.

number specifies that the time limit L is the given *number*. The *number* must be positive and at least as large as the largest event time in the data.

TIMELIST=*number-list*
specifies a list of time points at which the Kaplan-Meier estimates are displayed. The time points are listed in the column labeled as _TIME_. Since the Kaplan-Meier survival curve is a decreasing step function, each given time point falls in an interval that has a constant survival estimate. The event time that corresponds to the beginning of the time interval is displayed along with its survival estimate.

WIDTH=*value*
sets the width of the intervals used in the life table calculation of the survival function. This parameter is overridden by the INTERVALS= option.

BY Statement

BY *variables* **;**

You can specify a BY statement with PROC LIFETEST to obtain separate analyses on observations in groups defined by the BY variables.

The BY statement is more efficient than the STRATA statement for defining strata in large data sets. However, if you use the BY statement to define strata, PROC LIFETEST does not pool over strata for testing the association of survival time with covariates nor does it test for homogeneity across the BY groups.

Interval size is computed separately for each BY group. When intervals are determined by default, they may be different for each BY group. To make intervals the same for each BY group, use the INTERVALS= option in the PROC LIFETEST statement.

When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data using the SORT procedure with a similar BY statement.
- Specify the BY statement option NOTSORTED or DESCENDING in the BY statement for the LIFETEST 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 using the DATASETS procedure.

For more information on the BY statement, refer to the discussion in *SAS Language Reference: Concepts*. For more information on the DATASETS procedure, refer to the discussion in the *SAS Procedures Guide*.

FREQ Statement

FREQ *variable* **;**

The *variable* in the FREQ statement identifies a variable containing the frequency of occurrence of each observation. PROC LIFETEST treats each observation as if it appeared n times, where n is the value of the FREQ variable for the observation. The FREQ statement is useful for producing life tables when the data are already in the form of a summary data set. If not an integer, the frequency value is truncated to an integer. If the frequency value is less than one, the observation is not used.

ID Statement

ID *variables* **;**

The ID variable values are used to label the observations of the product-limit survival function estimates. SAS format statements can be used to format the values of the ID variables.

STRATA Statement

STRATA *variable* $<$ *(list)* $>$ $<$ $\ldots$ *variable* $<$ *(list)* $>$ $>$ **;**

The STRATA statement indicates which variables determine strata levels for the computations. The strata are formed according to the nonmissing values of the designated strata variables. The MISSING option can be used to allow missing values as a valid stratum level.

In the preceding syntax, *variable* is a variable whose values determine the stratum levels and *list* is a list of endpoints for a numeric variable. The values for *variable* can be formatted or unformatted. If the variable is a character variable, or if the variable is numeric and no list appears, then the strata are defined by the unique values of the strata variable. More than one variable can be specified in the STRATA statement, and each numeric variable can be followed by a list. Each interval contains its lower endpoint but does not contain its upper endpoint. The corresponding strata are formed by the combination of levels. If a variable is numeric and is followed by a list, then the levels for that variable correspond to the intervals defined by the list. The initial interval is assumed to start at $-\infty$ and the final interval is assumed to end at ∞.

The STRATA statement can have any of the following forms:

list separated by blanks	**strata age(5 10 20 30)**
list separated by commas	**strata age(5,10,20,30)**
x to y	**strata age(5 to 10)**
x to y by z	**strata age(5 to 30 by 10)**
combination of the above	**strata age(5,10 to 50 by 10)**

For example, the specification

```
strata age(5,20 to 50 by 10) sex;
```

indicates the following levels for the Age variable

$$\{(-\infty, 5), [5, 20), [20, 30), [30, 40), [40, 50), [50, \infty)\}$$

This statement also specifies that the age strata is further subdivided by values of the variable Sex. In this example, there are 6 age groups by 2 sex groups, forming a total of 12 strata.

The specification of several variables (for example, A B C) is equivalent to the A*B*C... syntax of the TABLES statement in the FREQ procedure. The number of strata levels usually grows very rapidly with the number of strata variables, so you must be cautious when specifying the STRATA list.

TEST Statement

TEST *variables* **;**

The TEST statement specifies a list of numeric (continuous) covariates that you want tested for association with the failure time.

Two sets of rank statistics are computed. These rank statistics and their variances are pooled over all strata. Univariate (marginal) test statistics are displayed for each of the covariates.

Additionally, a sequence of test statistics for joint effects of covariates is displayed. The first element of the sequence is the largest univariate test statistic. Other variables are then added on the basis of the largest increase in the joint test statistic. The process continues until all the variables have been added or until the remaining variables are linearly dependent on the previously added variables. See the section "Computational Formulas" on page 1818 for more information.

TIME Statement

TIME *variable* < ***censor(list)*** > **;**

The TIME statement is required. It is used to indicate the failure time variable, where *variable* is the name of the failure time variable that can be optionally followed by an asterisk, the name of the censoring variable, and a parenthetical list of values that correspond to right censoring. The censoring values should be numeric, nonmissing values. For example, the statement

```
time t*flag(1,2);
```

identifies the variable T as containing the values of the event or censored time. If the variable Flag has value 1 or 2, the corresponding value of T is a right-censored value and not an observed failure time.

Details

Missing Values

Observations with a missing value for either the failure time or the censoring variable are not used in the analysis. If a stratum variable value is missing, survival function estimates are computed for the strata labeled by the missing value, but these data are not used in any rank tests. However, the MISSING option can be used to request that missing values be treated as valid stratum values. If any variable specified in the TEST statement has a missing value, that observation is not used in the calculation of the rank statistics.

Computational Formulas

Product-Limit Method

Let $t_1 < t_2 < \cdots < t_k$ represent the distinct event times. For each $i = 1, \ldots, k$, let n_i be the number of surviving units, the size of the risk set, just prior to t_i. Let d_i be the number of units that fail at t_i, and let $s_i = n_i - d_i$.

The product-limit estimate of the SDF at t_i is the cumulative product

$$\hat{S}(t_i) = \prod_{j=1}^{i} \left(1 - \frac{d_j}{n_j}\right)$$

Notice that the estimator is defined to be right continuous; that is, the events at t_i are included in the estimate of $S(t_i)$. The corresponding estimate of the standard error is computed using Greenwood's formula (Kalbfleish and Prentice 1980) as

$$\hat{\sigma}\left(\hat{S}(t_i)\right) = \hat{S}(t_i)\sqrt{\sum_{j=1}^{i} \frac{d_j}{n_j s_j}}$$

The first sample quartile of the survival time distribution is given by

$$q_{0.25} = \frac{1}{2}(\inf\left\{t : 1 - \hat{S}(t) \geq 0.25\right\} + \sup\left\{t : 1 - \hat{S}(t) \leq 0.25\right\})$$

Confidence intervals for the quartiles are based on the sign test (Brookmeyer and Crowley 1982). The $100(1-\alpha)\%$ confidence interval for the first quartile is given by

$$I_{0.25} = \left\{t : (1 - \hat{S}(t) - 0.25)^2 \leq c_\alpha \hat{\sigma}^2 \left(\hat{S}(t)\right)\right\}$$

where c_α is the upper α percentile of a central chi-squared distribution with 1 degree of freedom. The second and third sample quartiles and the corresponding confidence intervals are calculated by replacing the 0.25 in the last two equations by 0.50 and 0.75, respectively.

The estimated mean survival time is

$$\hat{\mu} = \sum_{i=1}^{k} \hat{S}(t_{i-1})(t_i - t_{i-1})$$

where t_0 is defined to be zero. If the last observation is censored, this sum underestimates the mean. The standard error of $\hat{\mu}$ is estimated as

$$\hat{\sigma}(\hat{\mu}) = \sqrt{\frac{m}{m-1}\sum_{i=1}^{k-1}\frac{A_i^2}{n_i s_i}}$$

where

$$\begin{aligned} A_i &= \sum_{j=i}^{k-1} \hat{S}(t_j)(t_{j+1} - t_j) \\ m &= \sum_{j=1}^{k} d_j \end{aligned}$$

Life Table Method

The life table estimates are computed by counting the numbers of censored and uncensored observations that fall into each of the time intervals $[t_{i-1}, t_i)$, $i = 1, 2, \ldots, k+1$, where $t_0 = 0$ and $t_{k+1} = \infty$. Let n_i be the number of units entering the interval $[t_{i-1}, t_i)$, and let d_i be the number of events occurring in the interval. Let $b_i = t_i - t_{i-1}$, and let $n_i' = n_i - w_i/2$, where w_i is the number of units censored in the interval. The *effective sample size* of the interval $[t_{i-1}, t_i)$ is denoted by n_i'. Let t_{mi} denote the midpoint of $[t_{i-1}, t_i)$.

The conditional probability of an event in $[t_{i-1}, t_i)$ is estimated by

$$\hat{q}_i = \frac{d_i}{n'_i}$$

and its estimated standard error is

$$\hat{\sigma}\left(\hat{q}_i\right) = \sqrt{\frac{\hat{q}_i \hat{p}_i}{n'_i}}$$

where $\hat{p}_i = 1 - \hat{q}_i$.

The estimate of the survival function at t_i is

$$\hat{S}(t_i) = \begin{cases} 1 & i = 0 \\ \hat{S}(t_{i-1}) p_{i-1} & i > 0 \end{cases}$$

and its estimated standard error is

$$\hat{\sigma}\left(\hat{S}(t_i)\right) = \hat{S}(t_i) \sqrt{\sum_{j=1}^{i-1} \frac{\hat{q}_j}{n'_j \hat{p}_j}}$$

The density function at t_{mi} is estimated by

$$\hat{f}(t_{mi}) = \frac{\hat{S}(t_i)\hat{q}_i}{b_i}$$

and its estimated standard error is

$$\hat{\sigma}\left(\hat{f}(t_{mi})\right) = \hat{f}(t_{mi}) \sqrt{\sum_{j=1}^{i-1} \frac{\hat{q}_j}{n'_j \hat{p}_j} + \frac{\hat{p}_i}{n'_i \hat{q}_i}}$$

The estimated hazard function at t_{mi} is

$$\hat{h}(t_{mi}) = \frac{2\hat{q}_i}{b_i(1 + \hat{p}_i)}$$

and its estimated standard error is

$$\hat{\sigma}\left(\hat{h}(t_{mi})\right) = \hat{h}(t_{mi}) \sqrt{\frac{1 - (b_i \hat{h}(t_{mi})/2)^2}{n'_i \hat{q}_i}}$$

Let $[t_{j-1}, t_j)$ be the interval in which $\hat{S}(t_{j-1}) \geq \hat{S}(t_i)/2 > \hat{S}(t_j)$. The median residual lifetime at t_i is estimated by

$$\hat{M}_i = t_{j-1} - t_i + b_j \frac{\hat{S}(t_{j-1}) - \hat{S}(t_i)/2}{\hat{S}(t_{j-1}) - \hat{S}(t_j)}$$

and the corresponding standard error is estimated by

$$\hat{\sigma}(\hat{M}_i) = \frac{\hat{S}(t_i)}{2\hat{f}(t_{mj})\sqrt{n_i'}}$$

Interval Determination

If you want to determine the intervals exactly, use the INTERVALS= option in the PROC LIFETEST statement to specify the interval endpoints. Use the WIDTH= option to specify the width of the intervals, thus indirectly determining the number of intervals. If neither the INTERVALS= option nor the WIDTH= option is specified in the life table estimation, the number of intervals is determined by the NINTERVAL= option. The width of the time intervals is 2, 5, or 10 times an integer (possibly a negative integer) power of 10. Let $c = \log_{10}$(maximum event or censored time/number of intervals), and let b be the largest integer not exceeding c. Let $d = 10^{c-b}$ and let

$$a = 2 \times I(d \leq 2) + 5 \times I(2 < d \leq 5) + 10 \times I(d > 5)$$

with I being the indicator function. The width is then given by

$$\text{width} = a \times 10^b$$

By default, NINTERVAL=10.

Confidence Limits Added to the Output Data Set

The upper confidence limits (UCL) and the lower confidence limits (LCL) for the distribution estimates for both the product-limit and life table methods are computed as

$$\begin{aligned} \text{UCL} &= \hat{\lambda} + z_{\alpha/2}\hat{\sigma} \\ \text{LCL} &= \hat{\lambda} - z_{\alpha/2}\hat{\sigma} \end{aligned}$$

where $\hat{\lambda}$ is the estimate (either the survival function, the density, or the hazard function), $\hat{\sigma}$ is the corresponding estimate of the standard error, and $z_{\alpha/2}$ is the critical value for the normal distribution. That is, $\Phi(-z_{\alpha/2}) = \alpha/2$, where Φ is the cumulative distribution function for the standard normal distribution.

The value of α can be specified with the ALPHA= option.

Tests for Equality of Survival Curves across Strata

Log-Rank Test and Wilcoxon Test

The rank statistics used to test homogeneity between the strata (Kalbfleish and Prentice 1980) have the form of a $c \times 1$ vector $\mathbf{v} = (v_1, v_2, \ldots, v_c)'$ with

$$v_j = \sum_{i=1}^{k} w_i \left(d_{ij} - \frac{n_{ij} d_i}{n_i} \right)$$

where c is the number of strata, and the estimated covariance matrix, $\mathbf{V} = (V_{jl})$, is given by

$$V_{jl} = \sum_{i=1}^{k} \frac{w_i^2 d_i s_i (n_i n_{il} \delta_{jl} - n_{ij} n_{il})}{n_i^2 (n_i - 1)}$$

where i labels the distinct event times, δ_{jl} is 1 if $j = l$ and 0 otherwise, n_{ij} is the size of the risk set in the jth stratum at the ith event time, d_{ij} is the number of events in the jth stratum at the ith time, and

$$\begin{aligned} n_i &= \sum_{j=1}^{c} n_{ij} \\ d_i &= \sum_{j=1}^{c} d_{ij} \\ s_i &= n_i - d_i \end{aligned}$$

The term v_j can be interpreted as a weighted sum of observed minus expected numbers of failure under the null hypothesis of identical survival curves. The weight w_i is 1 for the log-rank test and n_i for the Wilcoxon test. The overall test statistic for homogeneity is $\mathbf{v}'\mathbf{V}^{-}\mathbf{v}$, where $\mathbf{V}^{-}$ denotes a generalized inverse of $\mathbf{V}$. This statistic is treated as having a chi-square distribution with degrees of freedom equal to the rank of $\mathbf{V}$ for the purposes of computing an approximate probability level.

Likelihood Ratio Test

The likelihood ratio test statistic (Lawless 1982) for homogeneity assumes that the data in the various strata are exponentially distributed and tests that the scale parameters are equal. The test statistic is computed as

$$Z = 2N \log\left(\frac{T}{N}\right) - 2 \sum_{j=1}^{c} N_j \log\left(\frac{T_j}{N_j}\right)$$

where N_j is the total number of events in the jth stratum, $N = \sum_{j=1}^{c} N_j$, T_j is the total time on test in the jth stratum, and $T = \sum_{j=1}^{c} T_j$. The approximate probability value is computed by treating Z as having a chi-square distribution with $c-1$ degrees of freedom.

Rank Tests for the Association of Survival Time with Covariates

The rank tests for the association of covariates are more general cases of the rank tests for homogeneity. A good discussion of these tests can be found in Kalbfleisch and Prentice (1980). In this section, the index α is used to label all observations, $\alpha = 1, 2, \ldots, n$, and the indices i, j range only over the observations that correspond to events, $i, j = 1, 2, \ldots, k$. The ordered event times are denoted as $t_{(i)}$, the

corresponding vectors of covariates are denoted as $\mathbf{z}_{(i)}$, and the ordered times, both censored and event times, are denoted as t_α.

The rank test statistics have the form

$$\mathbf{v} = \sum_{\alpha=1}^{n} c_{\alpha,\delta_\alpha} \mathbf{z}_\alpha$$

where n is the total number of observations, c_{α,δ_α} are rank scores, which can be either log-rank or Wilcoxon rank scores, δ_α is 1 if the observation is an event and 0 if the observation is censored, and $\mathbf{z}_\alpha$ is the vector of covariates in the TEST statement for the αth observation. Notice that the scores, c_{α,δ_α}, depend on the censoring pattern and that the summation is over all observations.

The log-rank scores are

$$c_{\alpha,\delta_\alpha} = \sum_{(j:t_{(j)} \le t_\alpha)} \left(\frac{1}{n_j} - \delta_\alpha \right)$$

and the Wilcoxon scores are

$$c_{\alpha,\delta_\alpha} = 1 - (1 + \delta_\alpha) \prod_{(j:t_{(j)} \le t_\alpha)} \frac{n_j}{n_j + 1}$$

where n_j is the number at risk just prior to $t_{(j)}$.

The estimates used for the covariance matrix of the log-rank statistics are

$$\mathbf{V} = \sum_{i=1}^{k} \frac{\mathbf{V}_i}{n_i}$$

where $\mathbf{V}_i$ is the corrected sum of squares and crossproducts matrix for the risk set at time $t_{(i)}$; that is,

$$\mathbf{V}_i = \sum_{(\alpha:t_\alpha \ge t_{(i)})} (\mathbf{z}_\alpha - \bar{\mathbf{z}}_i)'(\mathbf{z}_\alpha - \bar{\mathbf{z}}_i)$$

where

$$\bar{\mathbf{z}}_i = \sum_{(\alpha:t_\alpha \ge t_{(i)})} \frac{\mathbf{z}_\alpha}{n_i}$$

The estimate used for the covariance matrix of the Wilcoxon statistics is

$$\mathbf{V} = \sum_{i=1}^{k} \left[a_i(1 - a_i^*)(2\mathbf{z}_{(i)}\mathbf{z}'_{(i)} + \mathbf{S}_i) - (a_i^* - a_i) \left(a_i \mathbf{x}_i \mathbf{x}_i' + \sum_{j=i+1}^{k} a_j (\mathbf{x}_i \mathbf{x}_j' + \mathbf{x}_j \mathbf{x}_i') \right) \right]$$

where

$$
\begin{aligned}
a_i &= \prod_{j=1}^{i} \frac{n_j}{n_j + 1} \\
a_i^* &= \prod_{j=1}^{i} \frac{n_j + 1}{n_j + 2} \\
\mathbf{S}_i &= \sum_{(\alpha : t_{(i+1)} > t_\alpha > t_{(i)})} \mathbf{z}_\alpha \mathbf{z}_\alpha' \\
\mathbf{x}_i &= 2\mathbf{z}_{(i)} + \sum_{(\alpha : t_{(i+1)} > t_\alpha > t_{(i)})} \mathbf{z}_\alpha
\end{aligned}
$$

In the case of tied failure times, the statistics $\mathbf{v}$ are averaged over the possible orderings of the tied failure times. The covariance matrices are also averaged over the tied failure times. Averaging the covariance matrices over the tied orderings produces functions with appropriate symmetries for the tied observations; however, the actual variances of the $\mathbf{v}$ statistics would be smaller than the preceding estimates. Unless the proportion of ties is large, it is unlikely that this will be a problem.

The univariate tests for each covariate are formed from each component of $\mathbf{v}$ and the corresponding diagonal element of $\mathbf{V}$ as v_i^2 / V_{ii}. These statistics are treated as coming from a chi-square distribution for calculation of probability values.

The statistic $\mathbf{v}' \mathbf{V}^- \mathbf{v}$ is computed by sweeping each pivot of the $\mathbf{V}$ matrix in the order of greatest increase to the statistic. The corresponding sequence of partial statistics is tabulated. Sequential increments for including a given covariate and the corresponding probabilities are also included in the same table. These probabilities are calculated as the tail probabilities of a chi-square distribution with one degree of freedom. Because of the selection process, these probabilities should not be interpreted as p-values.

If desired for data screening purposes, the output data set requested by the OUTTEST= option can be treated as a sum of squares and crossproducts matrix and processed by the REG procedure using the option METHOD=RSQUARE. Then the sets of variables of a given size can be found that give the largest test statistics. Example 37.1 illustrates this process.

Output Data Sets

OUTSURV= Data Set

The OUTSURV= option in the LIFETEST statement creates an output data set containing survival estimates. It contains

- any specified BY variables
- any specified STRATA variables, their values coming from either their original values or the midpoints of the stratum intervals if endpoints are used to define strata (semi-infinite intervals are labeled by their finite endpoint)
- _STRTUM_, a numeric variable that numbers the strata
- the time variable as given in the TIME statement. In the case of the product-limit estimates, it contains the observed failure or censored times. For the life table estimates, it contains the lower endpoints of the time intervals.
- SURVIVAL, a variable containing the survival function estimates
- SDF_LCL, a variable containing the lower endpoint of the survival confidence interval
- SDF_UCL, a variable containing the upper endpoint of the survival confidence interval

If the estimation uses the product-limit method, then the data set also contains

- _CENSOR_, an indicator variable that has a value 1 for a censored observation and a value 0 for an event observation

If the estimation uses the life table method, then the data set also contains

- MIDPOINT, a variable containing the value of the midpoint of the time interval
- PDF, a variable containing the density function estimates
- PDF_LCL, a variable containing the lower endpoint of the PDF confidence interval
- PDF_UCL, a variable containing the upper endpoint of the PDF confidence interval
- HAZARD, a variable containing the hazard estimates
- HAZ_LCL, a variable containing the lower endpoint of the hazard confidence interval
- HAZ_UCL, a variable containing the upper endpoint of the hazard confidence interval

Each survival function contains an initial observation with the value 1 for the SDF and the value 0 for the time. The output data set contains an observation for each distinct failure time if the product-limit method is used or an observation for each time interval if the life table method is used. The product-limit survival estimates are defined to be right continuous; that is, the estimates at a given time include the factor for the failure events that occur at that time.

Labels are assigned to all the variables in the output data set except the BY variable and the STRATA variable.

OUTTEST= Data Set

The OUTTEST= option in the LIFETEST statement creates an output data set containing the rank statistics for testing the association of failure time with covariates. It contains

- any specified BY variables
- _TYPE_, a character variable of length 8 that labels the type of rank test, either "LOG-RANK" or "WILCOXON"
- _NAME_, a character variable of length 8 that labels the rows of the covariance matrix and the test statistics
- the TIME variable, containing the overall test statistic in the observation that has _NAME_ equal to the name of the time variable and the univariate test statistics under their respective covariates.
- all variables listed in the TEST statement

The output is in the form of a symmetric matrix formed by the covariance matrix of the rank statistics bordered by the rank statistics and the overall chi-square statistic. If the value of _NAME_ is the name of a variable in the TEST statement, the observation contains a row of the covariance matrix and the value of the rank statistic in the time variable. If the value of _NAME_ is the name of the TIME variable, the observation contains the values of the rank statistics in the variables from the TEST list and the value of the overall chi-square test statistic in the TIME variable.

Two complete sets of statistics labeled by the _TYPE_ variable are produced, one for the log-rank test and one for the Wilcoxon test.

Computer Resources

The data are first read and sorted into strata. If the data are originally sorted by failure time and censoring state, with smaller failure times coming first and event values preceding censored values in cases of ties, the data can be processed by strata without additional sorting. Otherwise, the data are read into memory by strata and sorted.

Memory Requirements

For a given BY group, define

N	the total number of observations
V	the number of STRATA variables
C	the number of covariates listed on the TEST statement
L	total length of the ID variables in bytes
S	number of strata
n	maximum number of observations within strata
b	$12 + 8C + L$
$m1$	$(112 + 16V) \times S$
$m2$	$50 \times b \times S$
$m3$	$(50 + n) \times (b + 4)$
$m4$	$8(C + 4)^2$
$m5$	$20N + 8S \times (S + 4)$

The memory, in bytes, required to process the BY-group is at least

$$m1 + \max(m2, m3) + m4$$

The test of equality of survival functions across strata requires additional memory ($m5$ bytes). However, if this additional memory is not available, PROC LIFETEST skips the test for equality of survival functions and finishes the other computations. Additional memory is required for the PLOTS= option. Temporary storage of $16n$ bytes is required to store the product-limit estimates for plotting.

Displayed Output

For each stratum, the LIFETEST procedure displays

the values of the stratum variables, if you specify the STRATA statement.

The following items are displayed when you request product-limit estimates:

- the observed event or censored times
- the estimate of the survival function
- the estimate of the cumulative distribution function of the failure time
- the standard error estimate of the estimated survival function
- the number of event times that have been observed
- the number of event or censored times which remain to be observed

- the frequency of the observed event or censored times if you specify the FREQ statement
- the values of the ID variables if you specify the ID statement
- the sample quartiles of the survival times
- the estimated mean survival time
- the estimated standard error of the estimated mean

The following items are displayed when you request life table estimates:

- time intervals into which the failure and censored times are distributed; each interval is from the lower limit, up to but not including the upper limit. If the upper limit is infinity, the missing value is printed.
- the number of events that occur in the interval
- the number of censored observations that fall into the interval
- the effective sample size for the interval
- the estimate of conditional probability of events (failures) in the interval
- the standard error of the estimated conditional probability of events
- the estimate of the survival function at the beginning of the interval
- the estimate of the cumulative distribution function of the failure time at the beginning of the interval
- the standard error estimate of the estimated survival function
- the estimate of the median residual lifetime which is the amount of time elapsed before reducing the number of at-risk units to one-half. This is also known as the it median future lifetime in Johnson and Johnson (1980).
- the estimated standard error of the estimated median residual lifetime
- the density function estimated at the midpoint of the interval
- the standard error estimate of the estimated density
- the hazard rate estimated at the midpoint of the interval
- the standard error estimate of the estimated hazard

The following results, summarized over all strata, are displayed:

- a summary of the number of censored and event times
- a table of rank statistics for testing homogeneity over strata. For each stratum, the log rank statistic can be interpreted as the difference between the observed number of failures and the expected numbers of failures under the null hypothesis of identical survival function.
- the covariance matrix for the log rank statistics for testing homogeneity over strata

- the covariance matrix for the Wilcoxon statistics for testing homogeneity over strata
- the approximate chi-square statistic for the log rank test, computed as a quadratic form of the log rank statistics (see **Computational Formulas**)
- the approximate chi-square statistic for the Wilcoxon test
- the likelihood ratio test for homogeneity over strata based on the exponential distribution

You can generate plots for

- the estimated SURVIVAL FUNCTION against FAILURE TIME
- the −log(estimated SURVIVAL FUNCTION) against FAILURE TIME
- the log(−log(estimated SURVIVAL FUNCTION)) against log(FAILURE TIME)
- censored observations for each stratum if the product-limit estimation method was used.

If you request the life table estimation method, you can also generate plots for the estimated HAZARD against FAILURE TIME and the estimated DENSITY against FAILURE TIME.

If you specify the TEST statement, the following statistics are printed:

- the univariate Wilcoxon statistics
- the standard deviations of the Wilcoxon statistics
- the corresponding approximate chi-square statistics
- the approximate probability values of the univariate chi-square statistics
- the covariance matrix for the Wilcoxon statistics
- the sequence of partial chi-square statistics for the Wilcoxon test in the order of the greatest increase to the overall test statistic
- the approximate probability values of the partial chi-square statistics
- the chi-square increments for including the given covariate
- the probability values of the chi-square increments. See **Computational Formulas** earlier in this chapter for a warning concerning these probabilities.
- the univariate log rank statistics
- the standard deviations of the log rank statistics
- the corresponding approximate chi-square statistics
- the approximate probability values of the univariate chi-square statistics
- the covariance matrix for the log rank statistics
- the sequence of partial chi-square statistics for the log rank test in the order of the greatest increase to the overall test statistic

- the approximate probability values of the partial chi-square statistics
- the chi-square increments for including the given covariate
- the probability values of the chi-square increments. See **Computational Formulas** earlier in this chapter for a warning concerning these probabilities

ODS Table Names

PROC LIFETEST assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in the following table. For more information on ODS, see Chapter 15, "Using the Output Delivery System."

Table 37.1. ODS Tables Produced in PROC LIFETEST

ODS Table Name	Description	Statement	Option
CensorPlot	Line-printer plot of censored observations	PROC	PLOT=(C) and METHOD=PL and LINEPRINTER
CensoredSummary	Number of event and censored observations	PROC	METHOD=PL (default)
DensityPlot	Line-printer plot of the density	PROC	PLOT=(D) and METHOD=LT and LINEPRINTER
HazardPlot	Line-printer plot of the hazards function	PROC	PLOT=(H) and METHOD=LT and LINEPRINTER
HomStats	Rank statistics for testing strata homogeneity	STRATA	
HomTests	Tests for strata homogeneity	STRATA	
LifetableEstimates	Lifetable survival estimates	PROC	METHOD=LT
LogForStepSeq	Forward stepwise sequence for the log-rank statistics for association	TEST	
LogHomCov	Covariance matrix for the log-rank statistics for strata homogeneity	STRATA	
LogLogSurvivalPlot	Line-printer plot of the log of the negative log survivor function	PROC	PLOT=(LLS) and LINEPRINTER
LogSurvivalPLot	Line-printer plot of the log survivor function	PROC	PLOT=(LS) and LINEPRINTER
LogTestCov	Covariance matrix for log-rank statistics for association	TEST	
LogUniChisq	Univariate chi-squares for log-rank statistic for association	TEST	
Means	Mean and Standard Error of survival times	PROC	METHOD=PL (default)
ProductLimitEstimates	Product-limit survival estimates	PROC	METHOD=PL (default)

Table 37.1. (continued)

ODS Table Name	Description	Statement	Option
Quartiles	Quartiles of the survival distribution	PROC	METHOD=PL (default)
SurvivalPlot	Line-printer plot of the survivor function	PROC	PLOT=(S) and LINEPRINTER
WilForStepSeq	Forward stepwise sequence for the log-rank statistics for association	TEST	
WilHomCov	Covariance matrix for the Wilcoxon statistics for strata homogeneity	STRATA	
WilTestCov	Covariance matrix for log-rank statistics for association	TEST	
WilUniChiSq	Univariate chi-squares for Wilcoxon statistic for association	TEST	

Examples

Example 37.1. Product-Limit Estimates and Tests of Association for the VA Lung Cancer Data

This example uses the data presented in Appendix I of Kalbfleisch and Prentice (1980). The response variable, SurvTime, is the survival time in days of a lung cancer patient. Negative values of SurvTime are censored values. The covariates are Cell (type of cancer cell), Therapy (type of therapy: standard or test), Prior (prior therapy: 0=no, 10=yes), Age (age in years), DiagTime (time in months from diagnosis to entry into the trial), and Kps (performance status). A censoring indicator variable Censor is created from the data, with value 1 indicating a censored time and value 0 an event time. Since there are only two types of therapy, an indicator variable, Treatment, is constructed for therapy type, with value 0 for standard therapy and value 1 for test therapy.

```
options ls=120;
data VALung;
   drop check m;
   retain Therapy Cell;
   infile cards column=column;
   length Check $ 1;
   label SurvTime='failure or censoring time'
      Kps='karnofsky index'
      DiagTime='months till randomization'
      Age='age in years'
      Prior='prior treatment?'
      Cell='cell type'
```

```
      Therapy='type of treatment'
      Treatment='treatment indicator';
   M=Column;
   input Check $ @@;
   if M>Column then M=1;
   if Check='s'|Check='t' then input @M Therapy $ Cell $ ;
   else input @M SurvTime Kps DiagTime Age Prior @@;
   if SurvTime > .;
   censor=(SurvTime<0);
   SurvTime=abs(SurvTime);
   Treatment=(Therapy='test');
   datalines;
standard squamous
 72 60  7 69  0   411 70  5 64 10   228 60  3 38  0   126 60  9 63 10
118 70 11 65 10    10 20  5 49  0    82 40 10 69 10   110 80 29 68  0
314 50 18 43  0  -100 70  6 70  0    42 60  4 81  0     8 40 58 63 10
144 30  4 63  0   -25 80  9 52 10    11 70 11 48 10
standard small
 30 60  3 61  0   384 60  9 42  0     4 40  2 35  0    54 80  4 63 10
 13 60  4 56  0  -123 40  3 55  0   -97 60  5 67  0   153 60 14 63 10
 59 30  2 65  0   117 80  3 46  0    16 30  4 53 10   151 50 12 69  0
 22 60  4 68  0    56 80 12 43 10    21 40  2 55 10    18 20 15 42  0
139 80  2 64  0    20 30  5 65  0    31 75  3 65  0    52 70  2 55  0
287 60 25 66 10    18 30  4 60  0    51 60  1 67  0   122 80 28 53  0
 27 60  8 62  0    54 70  1 67  0     7 50  7 72  0    63 50 11 48  0
392 40  4 68  0    10 40 23 67 10
standard adeno
  8 20 19 61 10    92 70 10 60  0    35 40  6 62  0   117 80  2 38  0
132 80  5 50  0    12 50  4 63 10   162 80  5 64  0     3 30  3 43  0
 95 80  4 34  0
standard large
177 50 16 66 10   162 80  5 62  0   216 50 15 52  0   553 70  2 47  0
278 60 12 63  0    12 40 12 68 10   260 80  5 45  0   200 80 12 41 10
156 70  2 66  0  -182 90  2 62  0   143 90  8 60  0   105 80 11 66  0
103 80  5 38  0   250 70  8 53 10   100 60 13 37 10
test squamous
999 90 12 54 10   112 80  6 60  0   -87 80  3 48  0  -231 50  8 52 10
242 50  1 70  0   991 70  7 50 10   111 70  3 62  0     1 20 21 65 10
587 60  3 58  0   389 90  2 62  0    33 30  6 64  0    25 20 36 63  0
357 70 13 58  0   467 90  2 64  0   201 80 28 52 10     1 50  7 35  0
 30 70 11 63  0    44 60 13 70 10   283 90  2 51  0    15 50 13 40 10
test small
 25 30  2 69  0  -103 70 22 36 10    21 20  4 71  0    13 30  2 62  0
 87 60  2 60  0     2 40 36 44 10    20 30  9 54 10     7 20 11 66  0
 24 60  8 49  0    99 70  3 72  0     8 80  2 68  0    99 85  4 62  0
 61 70  2 71  0    25 70  2 70  0    95 70  1 61  0    80 50 17 71  0
 51 30 87 59 10    29 40  8 67  0
test adeno
 24 40  2 60  0    18 40  5 69 10   -83 99  3 57  0    31 80  3 39  0
 51 60  5 62  0    90 60 22 50 10    52 60  3 43  0    73 60  3 70  0
  8 50  5 66  0    36 70  8 61  0    48 10  4 81  0     7 40  4 58  0
140 70  3 63  0   186 90  3 60  0    84 80  4 62 10    19 50 10 42  0
 45 40  3 69  0    80 40  4 63  0
test large
 52 60  4 45  0   164 70 15 68 10    19 30  4 39 10    53 60 12 66  0
 15 30  5 63  0    43 60 11 49 10   340 80 10 64 10   133 75  1 65  0
111 60  5 64  0   231 70 18 67 10   378 80  4 65  0    49 30  3 37  0
;
```

PROC LIFETEST is invoked to compute the product-limit estimate of the survivor function for each type of cancer cell and to analyze the effects of the variables Age, Prior, DiagTime, Kps, and Treatment on the survival of the patients. These prognostic factors are specified in the TEST statement, and the variable Cell is specified in the STRATA statement. Graphs of the product-limit estimates, the log estimates, and the negative log-log estimates are requested through the PLOTS= option in the PROC LIFETEST statement. Because of a few large survival times, a MAXTIME of 600 is used to set the scale of the time axis; that is, the time scale extends from 0 to a maximum of 600 days in the plots. The variable Therapy is specified in the ID statement to identify the type of therapy for each observation in the product-limit estimates. The OUTTEST option specifies the creation of an output data set named Test to contain the rank test matrices for the covariates.

```
title 'VA Lung Cancer Data';
symbol1 c=blue ; symbol2 c=orange; symbol3 c=green;
symbol4 c=red; symbol5 c=cyan; symbol6 c=black;
proc lifetest plots=(s,ls,lls) outtest=Test maxtime=600;
   time SurvTime*Censor(1);
   id Therapy;
   strata Cell;
   test Age Prior DiagTime Kps Treatment;
run;
```

Output 37.1.1 through Output 37.1.5 display the product-limit estimates of the survivor functions for the four cell types. Summary statistics of the survival times are also shown. The median survival times are 51 days, 156 days, 51 days, and 118 days for patients with adeno cells, large cells, small cells, and squamous cells, respectively.

Output 37.1.1. Product-Limit Survival Estimate for Cell=adeno

VA Lung Cancer Data

The LIFETEST Procedure

Stratum 1: Cell = adeno

Product-Limit Survival Estimates

SurvTime	Survival	Failure	Survival Standard Error	Number Failed	Number Left	Therapy
0.000	1.0000	0	0	0	27	
3.000	0.9630	0.0370	0.0363	1	26	standard
7.000	0.9259	0.0741	0.0504	2	25	test
8.000	.	.	.	3	24	standard
8.000	0.8519	0.1481	0.0684	4	23	test
12.000	0.8148	0.1852	0.0748	5	22	standard
18.000	0.7778	0.2222	0.0800	6	21	test
19.000	0.7407	0.2593	0.0843	7	20	test
24.000	0.7037	0.2963	0.0879	8	19	test
31.000	0.6667	0.3333	0.0907	9	18	test
35.000	0.6296	0.3704	0.0929	10	17	standard
36.000	0.5926	0.4074	0.0946	11	16	test
45.000	0.5556	0.4444	0.0956	12	15	test
48.000	0.5185	0.4815	0.0962	13	14	test
51.000	0.4815	0.5185	0.0962	14	13	test
52.000	0.4444	0.5556	0.0956	15	12	test
73.000	0.4074	0.5926	0.0946	16	11	test
80.000	0.3704	0.6296	0.0929	17	10	test
83.000*	.	.	.	17	9	test
84.000	0.3292	0.6708	0.0913	18	8	test
90.000	0.2881	0.7119	0.0887	19	7	test
92.000	0.2469	0.7531	0.0850	20	6	standard
95.000	0.2058	0.7942	0.0802	21	5	standard
117.000	0.1646	0.8354	0.0740	22	4	standard
132.000	0.1235	0.8765	0.0659	23	3	standard
140.000	0.0823	0.9177	0.0553	24	2	test
162.000	0.0412	0.9588	0.0401	25	1	standard
186.000	0	1.0000	0	26	0	test

NOTE: The marked survival times are censored observations.

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval [Lower	Upper)
75	92.000	73.000	140.000
50	51.000	31.000	90.000
25	19.000	8.000	45.000

Mean	Standard Error
65.556	10.127

Output 37.1.2. Product-Limit Survival Estimate for Cell=large

VA Lung Cancer Data

The LIFETEST Procedure

Stratum 2: Cell = large

Product-Limit Survival Estimates

SurvTime	Survival	Failure	Survival Standard Error	Number Failed	Number Left	Therapy
0.000	1.0000	0	0	0	27	
12.000	0.9630	0.0370	0.0363	1	26	standard
15.000	0.9259	0.0741	0.0504	2	25	test
19.000	0.8889	0.1111	0.0605	3	24	test
43.000	0.8519	0.1481	0.0684	4	23	test
49.000	0.8148	0.1852	0.0748	5	22	test
52.000	0.7778	0.2222	0.0800	6	21	test
53.000	0.7407	0.2593	0.0843	7	20	test
100.000	0.7037	0.2963	0.0879	8	19	standard
103.000	0.6667	0.3333	0.0907	9	18	standard
105.000	0.6296	0.3704	0.0929	10	17	standard
111.000	0.5926	0.4074	0.0946	11	16	test
133.000	0.5556	0.4444	0.0956	12	15	test
143.000	0.5185	0.4815	0.0962	13	14	standard
156.000	0.4815	0.5185	0.0962	14	13	standard
162.000	0.4444	0.5556	0.0956	15	12	standard
164.000	0.4074	0.5926	0.0946	16	11	test
177.000	0.3704	0.6296	0.0929	17	10	standard
182.000*	.	.	.	17	9	standard
200.000	0.3292	0.6708	0.0913	18	8	standard
216.000	0.2881	0.7119	0.0887	19	7	standard
231.000	0.2469	0.7531	0.0850	20	6	test
250.000	0.2058	0.7942	0.0802	21	5	standard
260.000	0.1646	0.8354	0.0740	22	4	standard
278.000	0.1235	0.8765	0.0659	23	3	standard
340.000	0.0823	0.9177	0.0553	24	2	test
378.000	0.0412	0.9588	0.0401	25	1	test
553.000	0	1.0000	0	26	0	standard

NOTE: The marked survival times are censored observations.

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval [Lower	Upper)
75	231.000	164.000	340.000
50	156.000	103.000	216.000
25	53.000	43.000	133.000

Mean	Standard Error
170.506	25.098

Output 37.1.3. Product-Limit Survival Estimate for Cell=small

VA Lung Cancer Data

The LIFETEST Procedure

Stratum 3: Cell = small

Product-Limit Survival Estimates

SurvTime	Survival	Failure	Survival Standard Error	Number Failed	Number Left	Therapy
0.000	1.0000	0	0	0	48	
2.000	0.9792	0.0208	0.0206	1	47	test
4.000	0.9583	0.0417	0.0288	2	46	standard
7.000	.	.	.	3	45	standard
7.000	0.9167	0.0833	0.0399	4	44	test
8.000	0.8958	0.1042	0.0441	5	43	test
10.000	0.8750	0.1250	0.0477	6	42	standard
13.000	.	.	.	7	41	standard
13.000	0.8333	0.1667	0.0538	8	40	test
16.000	0.8125	0.1875	0.0563	9	39	standard
18.000	.	.	.	10	38	standard
18.000	0.7708	0.2292	0.0607	11	37	standard
20.000	.	.	.	12	36	standard
20.000	0.7292	0.2708	0.0641	13	35	test
21.000	.	.	.	14	34	standard
21.000	0.6875	0.3125	0.0669	15	33	test
22.000	0.6667	0.3333	0.0680	16	32	standard
24.000	0.6458	0.3542	0.0690	17	31	test
25.000	.	.	.	18	30	test
25.000	0.6042	0.3958	0.0706	19	29	test
27.000	0.5833	0.4167	0.0712	20	28	standard
29.000	0.5625	0.4375	0.0716	21	27	test
30.000	0.5417	0.4583	0.0719	22	26	standard
31.000	0.5208	0.4792	0.0721	23	25	standard
51.000	.	.	.	24	24	standard
51.000	0.4792	0.5208	0.0721	25	23	test
52.000	0.4583	0.5417	0.0719	26	22	standard
54.000	.	.	.	27	21	standard
54.000	0.4167	0.5833	0.0712	28	20	standard
56.000	0.3958	0.6042	0.0706	29	19	standard
59.000	0.3750	0.6250	0.0699	30	18	standard
61.000	0.3542	0.6458	0.0690	31	17	test
63.000	0.3333	0.6667	0.0680	32	16	standard
80.000	0.3125	0.6875	0.0669	33	15	test
87.000	0.2917	0.7083	0.0656	34	14	test
95.000	0.2708	0.7292	0.0641	35	13	test
97.000*	.	.	.	35	12	standard
99.000	.	.	.	36	11	test
99.000	0.2257	0.7743	0.0609	37	10	test
103.000*	.	.	.	37	9	test
117.000	0.2006	0.7994	0.0591	38	8	standard
122.000	0.1755	0.8245	0.0567	39	7	standard
123.000*	.	.	.	39	6	standard
139.000	0.1463	0.8537	0.0543	40	5	standard
151.000	0.1170	0.8830	0.0507	41	4	standard
153.000	0.0878	0.9122	0.0457	42	3	standard
287.000	0.0585	0.9415	0.0387	43	2	standard
384.000	0.0293	0.9707	0.0283	44	1	standard
392.000	0	1.0000	0	45	0	standard

NOTE: The marked survival times are censored observations.

VA Lung Cancer Data

The LIFETEST Procedure

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval [Lower	Upper)
75	99.000	59.000	151.000
50	51.000	25.000	61.000
25	20.000	13.000	25.000

Mean	Standard Error
78.981	14.837

Output 37.1.4. Product-Limit Survival Estimate for Cell=squamous

VA Lung Cancer Data

The LIFETEST Procedure

Stratum 4: Cell = squamous

Product-Limit Survival Estimates

SurvTime	Survival	Failure	Survival Standard Error	Number Failed	Number Left	Therapy
0.000	1.0000	0	0	0	35	
1.000	.	.	.	1	34	test
1.000	0.9429	0.0571	0.0392	2	33	test
8.000	0.9143	0.0857	0.0473	3	32	standard
10.000	0.8857	0.1143	0.0538	4	31	standard
11.000	0.8571	0.1429	0.0591	5	30	standard
15.000	0.8286	0.1714	0.0637	6	29	test
25.000	0.8000	0.2000	0.0676	7	28	test
25.000*	.	.	.	7	27	standard
30.000	0.7704	0.2296	0.0713	8	26	test
33.000	0.7407	0.2593	0.0745	9	25	test
42.000	0.7111	0.2889	0.0772	10	24	standard
44.000	0.6815	0.3185	0.0794	11	23	test
72.000	0.6519	0.3481	0.0813	12	22	standard
82.000	0.6222	0.3778	0.0828	13	21	standard
87.000*	.	.	.	13	20	test
100.000*	.	.	.	13	19	standard
110.000	0.5895	0.4105	0.0847	14	18	standard
111.000	0.5567	0.4433	0.0861	15	17	test
112.000	0.5240	0.4760	0.0870	16	16	test
118.000	0.4912	0.5088	0.0875	17	15	standard
126.000	0.4585	0.5415	0.0876	18	14	standard
144.000	0.4257	0.5743	0.0873	19	13	standard
201.000	0.3930	0.6070	0.0865	20	12	test
228.000	0.3602	0.6398	0.0852	21	11	standard
231.000*	.	.	.	21	10	test
242.000	0.3242	0.6758	0.0840	22	9	test
283.000	0.2882	0.7118	0.0820	23	8	test
314.000	0.2522	0.7478	0.0793	24	7	standard
357.000	0.2161	0.7839	0.0757	25	6	test
389.000	0.1801	0.8199	0.0711	26	5	test
411.000	0.1441	0.8559	0.0654	27	4	standard
467.000	0.1081	0.8919	0.0581	28	3	test
587.000	0.0720	0.9280	0.0487	29	2	test
991.000	0.0360	0.9640	0.0352	30	1	test
999.000	0	1.0000	0	31	0	test

NOTE: The marked survival times are censored observations.

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval [Lower	Upper)
75	357.000	201.000	467.000
50	118.000	72.000	242.000
25	33.000	11.000	111.000

Mean	Standard Error
230.225	48.475

Output 37.1.5. Summary of Censored and Uncensored Values

VA Lung Cancer Data

The LIFETEST Procedure

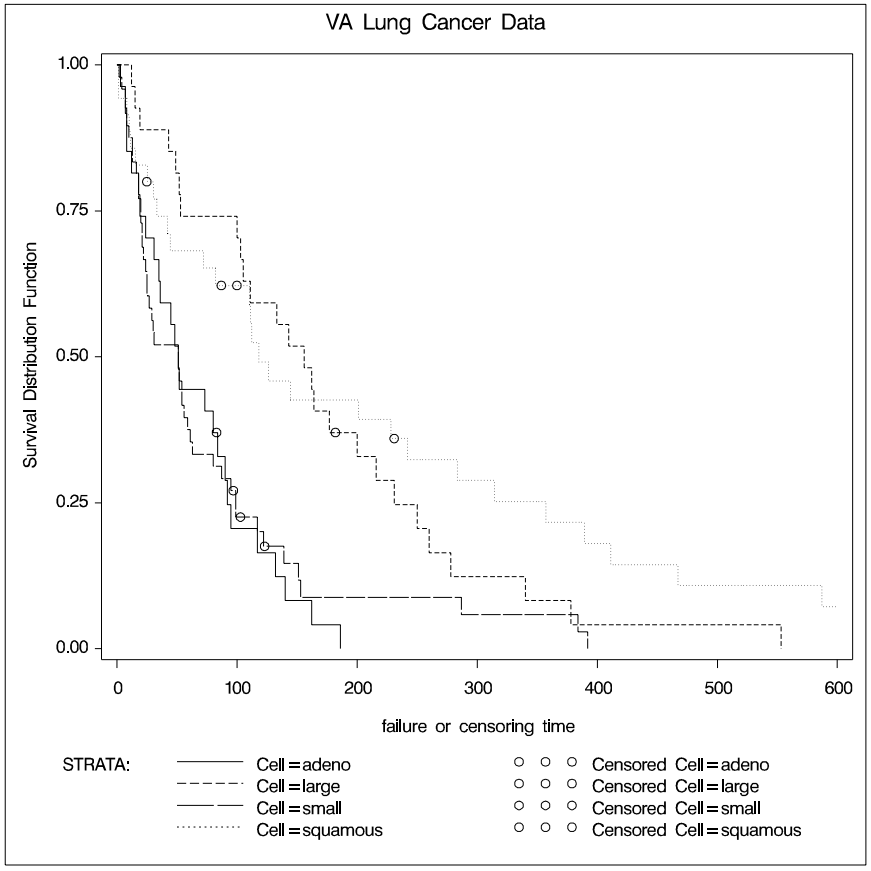
Summary of the Number of Censored and Uncensored Values

Stratum	Cell	Total	Failed	Censored	Percent Censored
1	adeno	27	26	1	3.70
2	large	27	26	1	3.70
3	small	48	45	3	6.25
4	squamous	35	31	4	11.43
Total		137	128	9	6.57

Output 37.1.5 displays a summary of the number of censored and event observations by cell type.

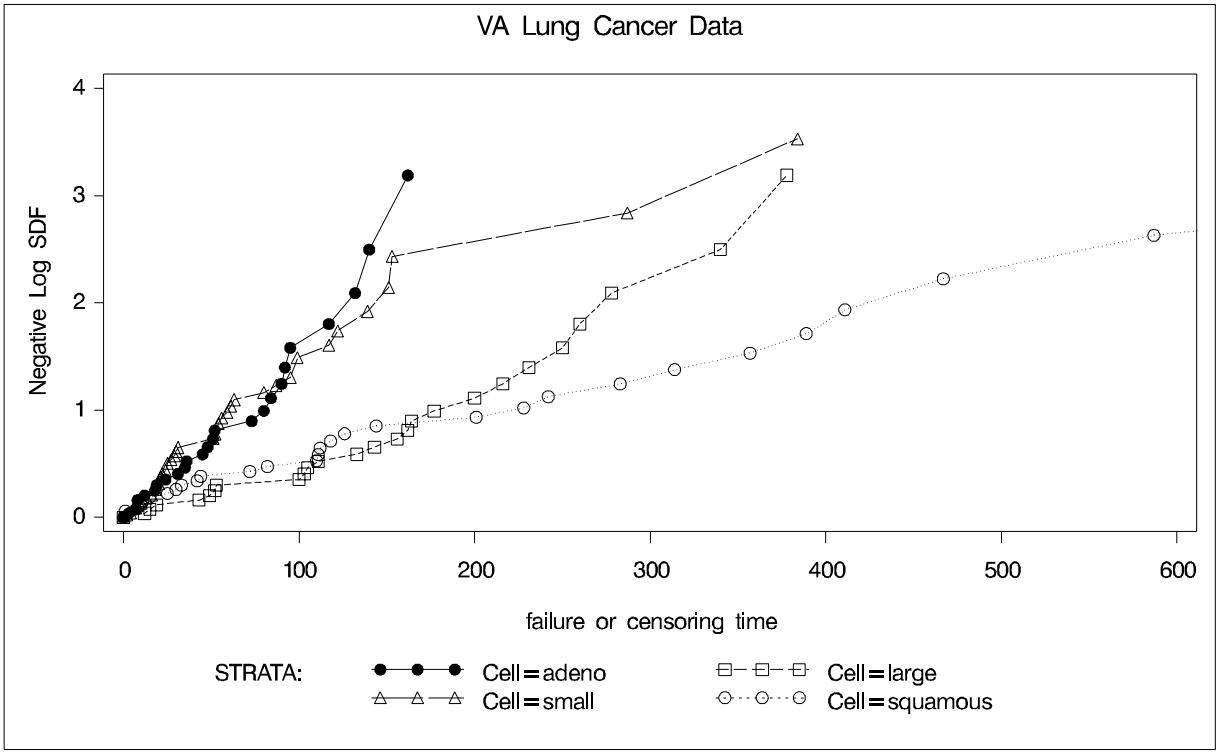
The graph of the estimated survivor functions is shown in Output 37.1.6. The adeno cell curve and the small cell curve are much closer to each other than to the large cell curve or the squamous cell curve. The survival rates of the adeno cell patients and the small cell patients decrease rapidly to approximately 29% in 90 days. Shapes of the large cell curve and the squamous cell curve are quite different, although both decrease less rapidly than those of the adeno and small cells. The squamous cell curve decreases more rapidly initially than the large cell curve, but the role is reversed in the later period.

Output 37.1.6. Graph of the Estimated Survivor Functions

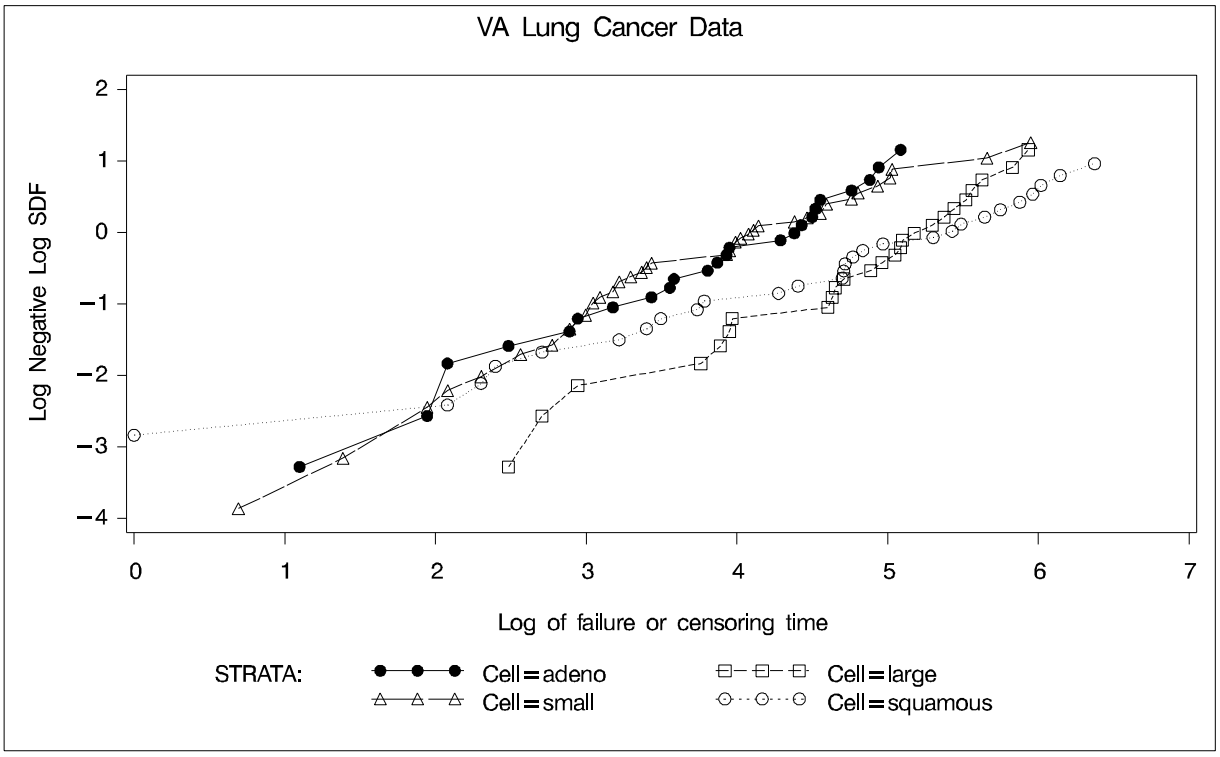


Output 37.1.7 displays the graph of the log of the estimated survivor functions and Output 37.1.8 displays the log of the negative log of the estimated survivor functions.

Output 37.1.7. Graph of the Log of the Estimated Survivor Functions



Output 37.1.8. Graph of the Negative Log-Log of the Estimated Survivor Functions



Output 37.1.9. Homogeneity Tests Across Strata

VA Lung Cancer Data

The LIFETEST Procedure

Rank Statistics

Cell	Log-Rank	Wilcoxon
adeno	10.306	697.0
large	-8.549	-1085.0
small	14.898	1278.0
squamous	-16.655	-890.0

Covariance Matrix for the Log-Rank Statistics

Cell	adeno	large	small	squamous
adeno	12.9662	-4.0701	-4.4087	-4.4873
large	-4.0701	24.1990	-7.8117	-12.3172
small	-4.4087	-7.8117	21.7543	-9.5339
squamous	-4.4873	-12.3172	-9.5339	26.3384

Covariance Matrix for the Wilcoxon Statistics

Cell	adeno	large	small	squamous
adeno	121188	-34718	-46639	-39831
large	-34718	151241	-59948	-56576
small	-46639	-59948	175590	-69002
squamous	-39831	-56576	-69002	165410

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	25.4037	3	<.0001
Wilcoxon	19.4331	3	0.0002
-2Log(LR)	33.9343	3	<.0001

Results of the homogeneity tests across cell types are given in Output 37.1.9. The log-rank and Wilcoxon statistics and their corresponding covariance matrices are displayed. Also given is a table that consists of the approximate chi-square statistics, degrees of freedom, and p-values for the log-rank, Wilcoxon, and likelihood ratio tests. All three tests indicate strong evidence of a significant difference among the survival curves for the four types of cancer cells ($p < 0.001$).

Output 37.1.10. Log-Rank Rank Test of the Prognostic Factors

VA Lung Cancer Data

The LIFETEST Procedure

Univariate Chi-Squares for the Log-Rank Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square	Label
Age	-40.7383	105.7	0.1485	0.7000	age in years
Prior	-19.9435	46.9836	0.1802	0.6712	prior treatment?
DiagTime	-115.9	97.8708	1.4013	0.2365	months till randomization
Kps	1123.1	170.3	43.4747	<.0001	karnofsky index
Treatment	-4.2076	5.0407	0.6967	0.4039	treatment indicator

Covariance Matrix for the Log-Rank Statistics

Variable	Age	Prior	DiagTime	Kps	Treatment
Age	11175.4	-301.2	-892.2	-2948.4	119.3
Prior	-301.2	2207.5	2010.9	78.6	13.9
DiagTime	-892.2	2010.9	9578.7	-2295.3	21.9
Kps	-2948.4	78.6	-2295.3	29015.6	61.9
Treatment	119.3	13.9	21.9	61.9	25.4

Forward Stepwise Sequence of Chi-Squares for the Log-Rank Test

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment	Label
Kps	1	43.4747	<.0001	43.4747	<.0001	karnofsky index
Treatment	2	45.2008	<.0001	1.7261	0.1889	treatment indicator
Age	3	46.3012	<.0001	1.1004	0.2942	age in years
Prior	4	46.4134	<.0001	0.1122	0.7377	prior treatment?
DiagTime	5	46.4200	<.0001	0.00665	0.9350	months till randomization

Results of the log-rank test of the prognostic variables are shown in Output 37.1.10. The univariate test results correspond to testing each prognostic factor marginally. The joint covariance matrix of these univariate test statistics is also displayed. In computing the overall chi-square statistic, the partial chi-square statistics following a forward stepwise entry approach are tabulated.

Consider the log-rank test in Output 37.1.10. Since the univariate test for Kps has the largest chi-square (43.4747) among all the covariates, Kps is entered first. At this stage, the partial chi-square and the chi-square increment for Kps are the same as the univariate chi-square. Among all the covariates not in the model (Age, Prior, DiagTime, Treatment), Treatment has the largest approximate chi-square increment (1.7261) and is entered next. The approximate chi-square for the model containing Kps and Treatment is 43.4747+1.7261=45.2008 with 2 degrees of freedom. The third covariate entered is Age. The fourth is Prior, and the fifth is DiagTime . The overall chi-square statistic on the last line of output is the partial chi-square for including all the covariates. It has a value of 46.4200 with 5 degrees of freedom, which is highly significant ($p < 0.0001$).

You can establish this forward stepwise entry of prognostic factors by passing the matrix corresponding to the log-rank test to the RSQUARE method in the REG procedure. PROC REG finds the sets of variables that yield the largest chi-square statistics.

```
data RSq;
   set Test;
   if _type_='LOG RANK';
      _type_='cov';

proc print data=RSq;

proc reg data=RSq(type=cov);
   model SurvTime=Age Prior DiagTime Kps Treatment
         / selection=rsquare;
   title 'All Possible Subsets of Covariables for the
          log-rank Test';
run;
```

Output 37.1.11 displays the univariate statistics and their covariance matrix. Results of the best subset regression are shown in Output 37.1.12. The variable Kps generates the largest univariate test statistic among all the covariates, the pair Kps and Age generate the largest test statistic among any other pairs of covariates, and so on. The entry order of covariates is identical to that of PROC LIFETEST.

Output 37.1.11. Log-Rank Statistics and Covariance Matrix

Obs	_TYPE_	_NAME_	SurvTime	Age	Prior	DiagTime	Kps	Treatment
1	cov	SurvTime	46.42	-40.74	-19.94	-115.86	1123.14	-4.208
2	cov	Age	-40.74	11175.44	-301.23	-892.24	-2948.45	119.297
3	cov	Prior	-19.94	-301.23	2207.46	2010.85	78.64	13.875
4	cov	DiagTime	-115.86	-892.24	2010.85	9578.69	-2295.32	21.859
5	cov	Kps	1123.14	-2948.45	78.64	-2295.32	29015.62	61.945
6	cov	Treatment	-4.21	119.30	13.87	21.86	61.95	25.409

Output 37.1.12. Best Subset Regression from the REG Procedure

All Possible Subsets of Covariables for the log-rank Test

The REG Procedure
Model: MODEL1
Dependent Variable: SurvTime

R-Square Selection Method

Number in Model	R-Square	Variables in Model
1	0.9366	Kps
1	0.0302	DiagTime
1	0.0150	Treatment
1	0.0039	Prior
1	0.0032	Age
2	0.9737	Kps Treatment
2	0.9472	Age Kps
2	0.9417	Prior Kps
2	0.9382	DiagTime Kps
2	0.0434	DiagTime Treatment
2	0.0353	Age DiagTime
2	0.0304	Prior DiagTime
2	0.0181	Prior Treatment
2	0.0159	Age Treatment
2	0.0075	Age Prior
3	0.9974	Age Kps Treatment
3	0.9774	Prior Kps Treatment
3	0.9747	DiagTime Kps Treatment
3	0.9515	Age Prior Kps
3	0.9481	Age DiagTime Kps
3	0.9418	Prior DiagTime Kps
3	0.0456	Age DiagTime Treatment
3	0.0438	Prior DiagTime Treatment
3	0.0355	Age Prior DiagTime
3	0.0192	Age Prior Treatment
4	0.9999	Age Prior Kps Treatment
4	0.9976	Age DiagTime Kps Treatment
4	0.9774	Prior DiagTime Kps Treatment
4	0.9515	Age Prior DiagTime Kps
4	0.0459	Age Prior DiagTime Treatment
5	1.0000	Age Prior DiagTime Kps Treatment

Example 37.2. Life Table Estimates for Males with Angina Pectoris

The data in this example come from Lee (1992, p. 91) and represent the survival rate of males with angina pectoris. Survival time is measured as years from the time of diagnosis. The data are read as number of events and number of withdrawals in each one-year time interval for 16 intervals. Three variables are constructed from the data: Years (an artificial time variable with values that are the midpoints of the time intervals), Censored (a censoring indicator variable with value 1 indicating censored observations and value 0 indicating event observations), and Freq (the frequency variable). Two observations are created for each interval, one representing the event observations and the other representing the censored observations.

```
title 'Survival of Males with Angina Pectoris';
data males;
   keep Freq Years Censored;
   retain Years -.5;
   input fail withdraw @@;
   Years + 1;
   Censored=0;
   Freq=fail;
   output;
   Censored=1;
   Freq=withdraw;
   output;
   datalines;
456   0 226  39 152  22 171  23 135 24 125 107
 83 133  74 102  51  68  42  64  43 45  34  53
 18  33   9  27   6  23   0  30
;
```

PROC LIFETEST is invoked to compute the various life table survival estimates, the median residual time, and their standard errors. The life table method of computing estimates is requested by specifying METHOD=LT. The intervals are specified by the INTERVAL= option. Graphs of the life table estimate, log of the estimate, negative log-log of the estimate, estimated density function, and estimated hazard function are requested by the PLOTS= option. No tests for homogeneity are carried out because the data are not stratified.

```
symbol1 c=blue;
proc lifetest data=males  method=lt intervals=(0 to 15 by 1)
              plots=(s,ls,lls,h,p);
   time Years*Censored(1);
   freq Freq;
run;
```

Output 37.2.1. Life Table Survival Estimates

Survival of Males with Angina Pectoris

The LIFETEST Procedure

Life Table Survival Estimates

Interval [Lower,	Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime
0	1	456	0	2418.0	0.1886	0.00796	1.0000	0	0	5.3313
1	2	226	39	1942.5	0.1163	0.00728	0.8114	0.1886	0.00796	6.2499
2	3	152	22	1686.0	0.0902	0.00698	0.7170	0.2830	0.00918	6.3432
3	4	171	23	1511.5	0.1131	0.00815	0.6524	0.3476	0.00973	6.2262
4	5	135	24	1317.0	0.1025	0.00836	0.5786	0.4214	0.0101	6.2185
5	6	125	107	1116.5	0.1120	0.00944	0.5193	0.4807	0.0103	5.9077
6	7	83	133	871.5	0.0952	0.00994	0.4611	0.5389	0.0104	5.5962
7	8	74	102	671.0	0.1103	0.0121	0.4172	0.5828	0.0105	5.1671
8	9	51	68	512.0	0.0996	0.0132	0.3712	0.6288	0.0106	4.9421
9	10	42	64	395.0	0.1063	0.0155	0.3342	0.6658	0.0107	4.8258
10	11	43	45	298.5	0.1441	0.0203	0.2987	0.7013	0.0109	4.6888
11	12	34	53	206.5	0.1646	0.0258	0.2557	0.7443	0.0111	.
12	13	18	33	129.5	0.1390	0.0304	0.2136	0.7864	0.0114	.
13	14	9	27	81.5	0.1104	0.0347	0.1839	0.8161	0.0118	.
14	15	6	23	47.5	0.1263	0.0482	0.1636	0.8364	0.0123	.
15	.	0	30	15.0	0	0	0.1429	0.8571	0.0133	.

Evaluated at the Midpoint of the Interval

Interval [Lower,	Upper)	Median Standard Error	PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	1	0.1749	0.1886	0.00796	0.208219	0.009698
1	2	0.2001	0.0944	0.00598	0.123531	0.008201
2	3	0.2361	0.0646	0.00507	0.09441	0.007649
3	4	0.2361	0.0738	0.00543	0.119916	0.009154
4	5	0.1853	0.0593	0.00495	0.108043	0.009285
5	6	0.1806	0.0581	0.00503	0.118596	0.010589
6	7	0.1855	0.0439	0.00469	0.1	0.010963
7	8	0.2713	0.0460	0.00518	0.116719	0.013545
8	9	0.2763	0.0370	0.00502	0.10483	0.014659
9	10	0.4141	0.0355	0.00531	0.112299	0.017301
10	11	0.4183	0.0430	0.00627	0.155235	0.023602
11	12	.	0.0421	0.00685	0.17942	0.030646
12	13	.	0.0297	0.00668	0.149378	0.03511
13	14	.	0.0203	0.00651	0.116883	0.038894
14	15	.	0.0207	0.00804	0.134831	0.054919
15	.	.	.	.	.	.

Results of the life table estimation are shown in Output 37.2.1. The five-year survival rate is 0.5193 with a standard error of 0.0103. The estimated median residual lifetime, which is 5.33 years initially, has reached a maximum of 6.34 years at the beginning of the second year and decreases gradually to a value lower than the initial 5.33 years at the beginning of the seventh year.

Output 37.2.2. Summary of Censored and Event Observations

The LIFETEST Procedure

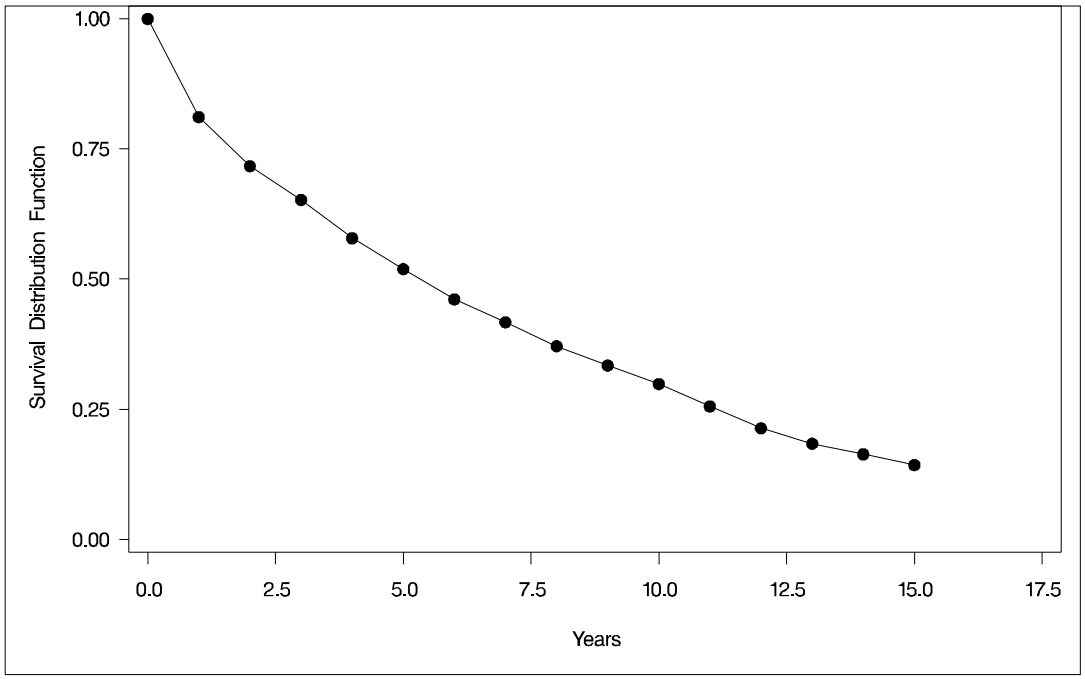
Summary of the Number of Censored and Uncensored Values

Total	Failed	Censored	Percent Censored
2418	1625	793	32.80

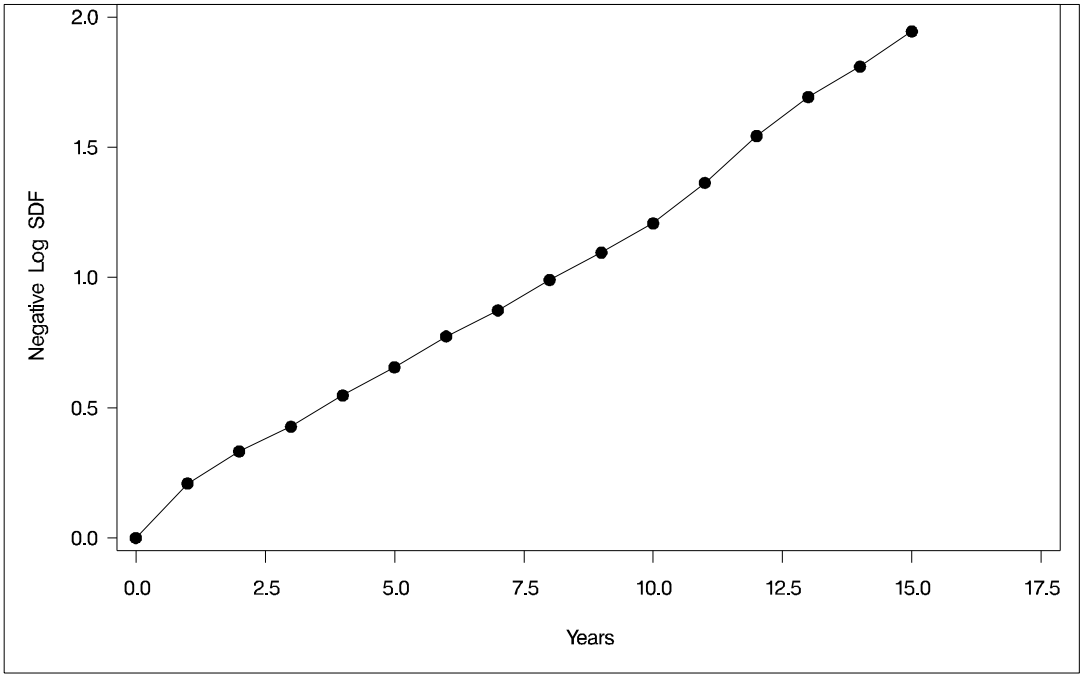
NOTE: There were 2 observations with missing values, negative time values or frequency values less than 1.

Output 37.2.2 shows the number of event and censored observations. The percentage of the patients that have withdrawn from the study is 32.8%.

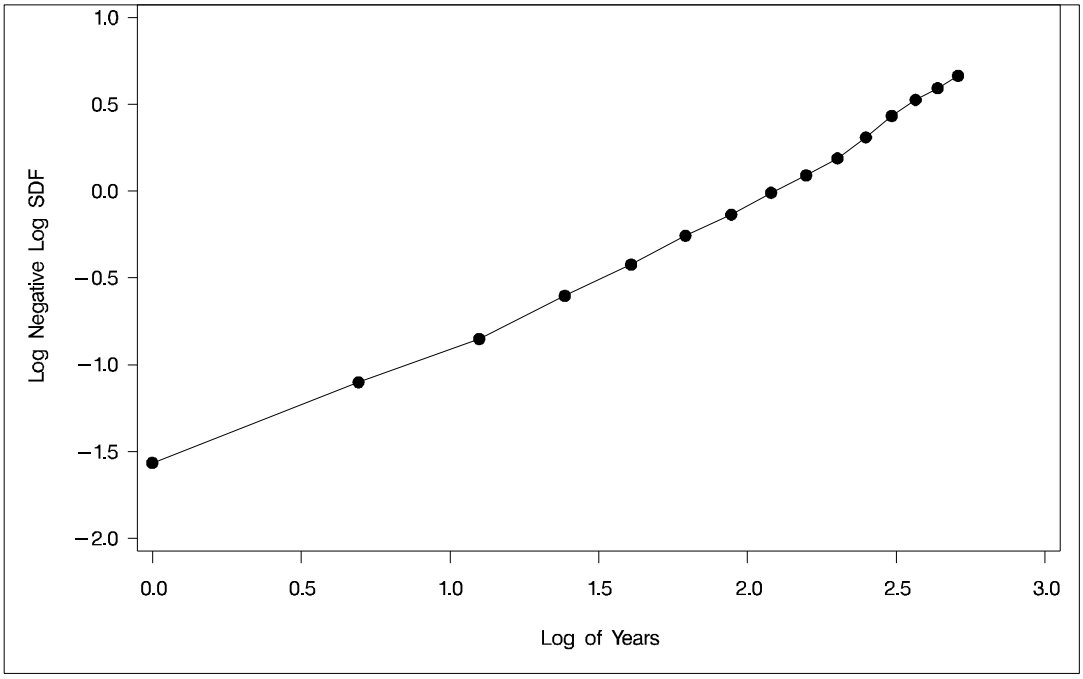
Output 37.2.3. Life Table Survivor Function Estimate



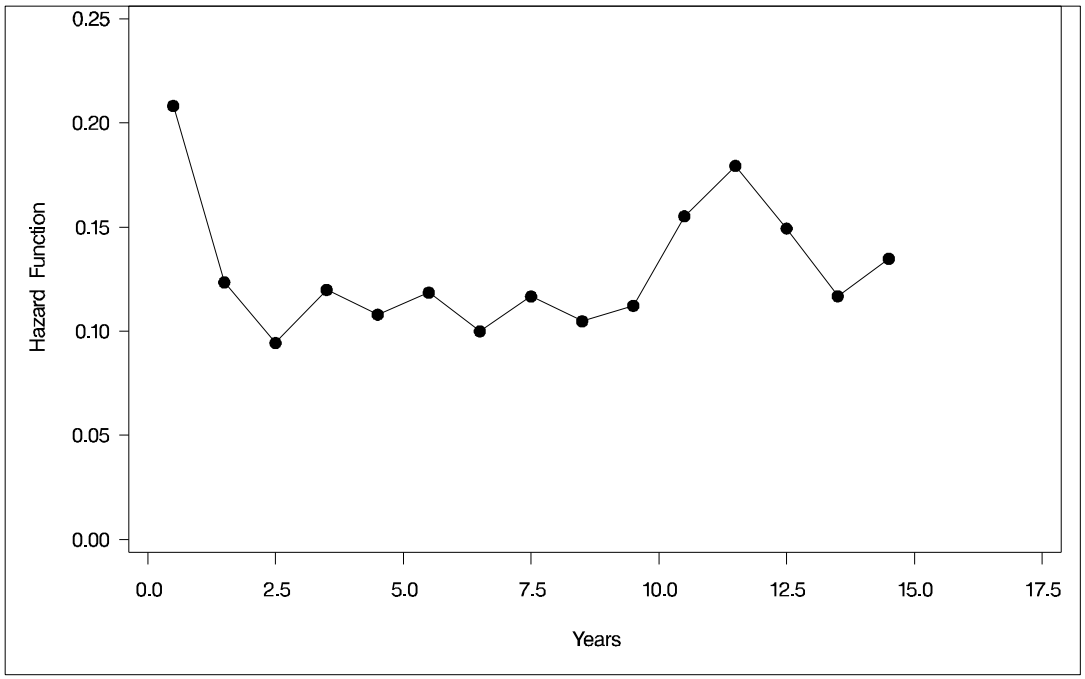
Output 37.2.4. Log of Survivor Function Estimate



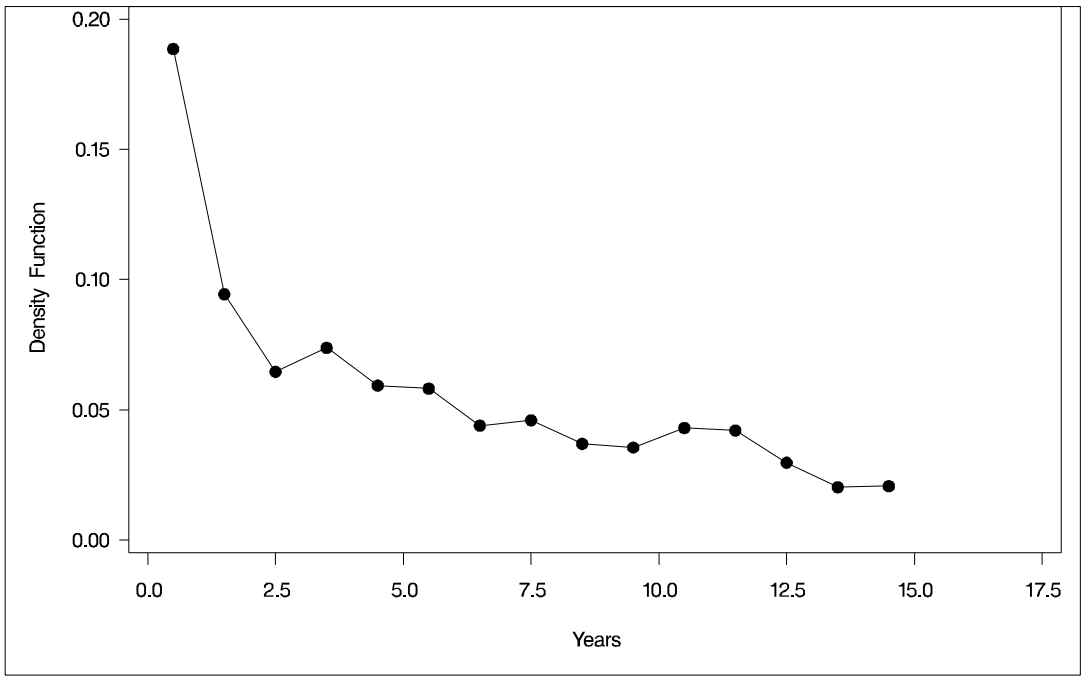
Output 37.2.5. Log of Negative Log of Survivor Function Estimate



Output 37.2.6. Hazard Function Estimate



Output 37.2.7. Density Function Estimate



Output 37.2.3 displays the graph of the life table survivor function estimates versus years after diagnosis. The median survival time, read from the survivor function curve, is 5.33 years, and the 25th and 75th percentiles are 1.04 and 11.13 years, respectively.

As discussed in Lee (1992), the graph of the estimated hazard function (Output 37.2.6) shows that the death rate is highest in the first year of diagnosis. From the end of the first year to the end of the tenth year, the death rate remains relatively constant, fluctuating between 0.09 and 0.12. The death rate is generally higher after the tenth year. This could indicate that a patient who has survived the first year has a better chance than a patient who has just been diagnosed. The profile of the median residual lifetimes also supports this interpretation.

An exponential model may be appropriate for the survival of these male patients with angina pectoris since the curve of the log of the survivor function estimate versus years of diagnosis (Output 37.2.4) approximates a straight line through the origin. Visually, the density estimate (Output 37.2.7) resembles that of an exponential distribution.

References

Brookmeyer, R. and Crowley, J. (1982), "A Confidence Interval for the Median Survival Time," *Biometrics*, 38, 29–41.

Collett, D. (1994), *Modeling Survival Data In Medical Research*, London: Chapman and Hall.

Cox, D.R. and Oakes, D. (1984), *Analysis of Survival Data*, London: Chapman and Hall.

Elandt-Johnson, R.C. and Johnson, N.L. (1980), *Survival Models and Data Analysis*, New York: John Wiley & Sons.

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

Lawless, J.E. (1982), *Statistical Models and Methods for Lifetime Data*, New York: John Wiley & Sons.

Lee, E.T. (1992), *Statistical Methods for Survival Data Analysis*, Second Edition, New York: John Wiley & Sons.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/STAT® User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999.

SAS/STAT® User's Guide, Version 8

ISBN 1–58025–494–2

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999