Chapter 14
Survival Analysis

Chapter Table of Contents

Chapter 14
Survival Analysis

Introduction

Survival data often consists of a response variable that measures the duration of time until a specified event occurs and a set of independent variables thought to be associated with the event-time variable. Component lifetimes in industrial reliability, durations of jobs, and survival times in a clinical trial are examples of event times. The purpose of survival analysis is to model the underlying distribution of event times and to assess the dependence of the event time on other explanatory variables. In many situations, the event time is not observed due to withdrawal or termination of the study; this phenomenon is known as *censoring*. Survival analysis methods correctly use both the censored and uncensored observations.
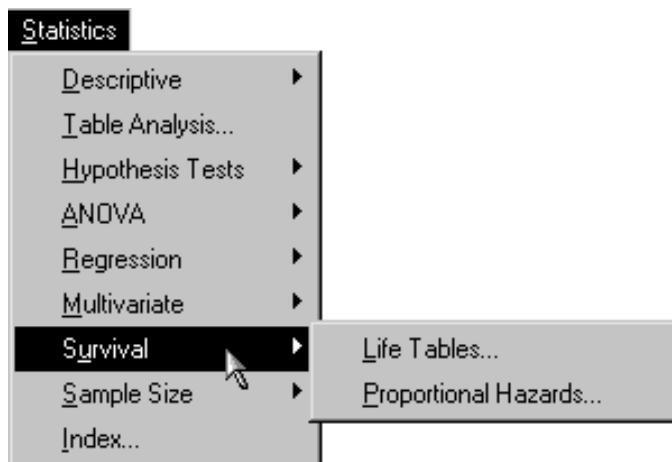
Figure 14.1. Survival Analysis Menu

Usually, a first step in the analysis of survival data is the estimation of the distribution of the 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 time *t* is the probability that an experimental unit from the population will have a lifetime exceeding *t*. The product limit and actuarial methods are popular techniques for estimating survival distributions.

Proportional hazards regression is a useful technique for assessing the relationship between survival times and a set of explanatory variables. The proportional hazards model of Cox (1972) assumes a parametric form for the effects of explanatory variables on survival times and allows an unspecified form for the underlying survivor function. The proportional hazards model is also known as Cox regression.

Survival Analysis Task Features

The Life Tables task provides both the actuarial (also known as lifetable) method and product-limit method (also known as the Kaplan-Meier method). You can define strata and test the homogeneity of survival functions across strata with rank tests and a likelihood ratio test based on an underlying exponential distribution. In addition, you can test the association between covariates and the lifetime variable with the log-rank test and the Wilcoxon test. Plots provided are the survival function, $-\log$ (survival function), $\log(-\log(\text{survival function}))$, hazard function, and probability density function.

The Proportional Hazards task performs Cox regression. You can choose from five different model selection techniques, select from four different methods for handling tied event times, and produce a survivor function plot with confidence intervals.

The examples in this chapter demonstrate how you can use the Survival tasks in the Analyst Application to analyze survival data.

Life Tables

The data set analyzed in this task contains the survival times of rats in a small randomized trial. Forty rats were exposed to a carcinogen and assigned to one of two treatment groups. The survival time is the time from randomization to death. The event of interest is death from cancer induced by the carcinogen, and interest lies in whether the survival distributions differ between the two treatments. Four rats died of other causes, and their survival times are regarded as censored observations. The data set Exposed contains four variables: Days, Status, Treatmnt, and Gender. The Days variable contains survival times in days from randomization to death, and the Status variable has the value 0 for censored observations and 1 for uncensored observations. The Treatmnt variable has the value 1 if the rat was administered the first treatment or 2 if the rat was administered the second treatment, and the Gender variable has the value F if the rat is female and M if the rat is male.

Open the Exposed Data Set

The data are provided in the Analyst Sample Library. To open the Exposed data set, follow these steps:

1. Select **Tools** → **Sample Data** . . .
2. Select Exposed.
3. Click **OK** to create the sample data set in your Sasuser directory.
4. Select **File** → **Open By SAS Name** . . .
5. Select Sasuser from the list of **Libraries**.
6. Select Exposed from the list of members.
7. Click **OK** to bring the Exposed data set into the data table.

Request the Life Tables Analysis

To specify the Life Tables task, follow these steps.

1. Select **Statistics** → **Survival** → **Life Tables** . . .

2. Select Days as the time variable.

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. The analysis methodology must correctly use the censored observations as well as the noncensored observations. In this analysis, the values of Days are considered censored if the value of Status is 0; otherwise, they are considered event times.

3. Select Status as the censoring variable.

4. Specify **0** as the censoring value by directly typing **0** in the **Censoring values:** field or by clicking on the down arrow under **Censoring values:** and selecting **0** from the list. You can remove censoring values by deleting the values in the field.

5. Select Treatmnt as the strata variable.

Figure 14.2 displays the dialog with Days specified as the time variable, Status specified as the censoring variable, 0 selected as the censoring value, and Treatmnt specified as the strata variable.

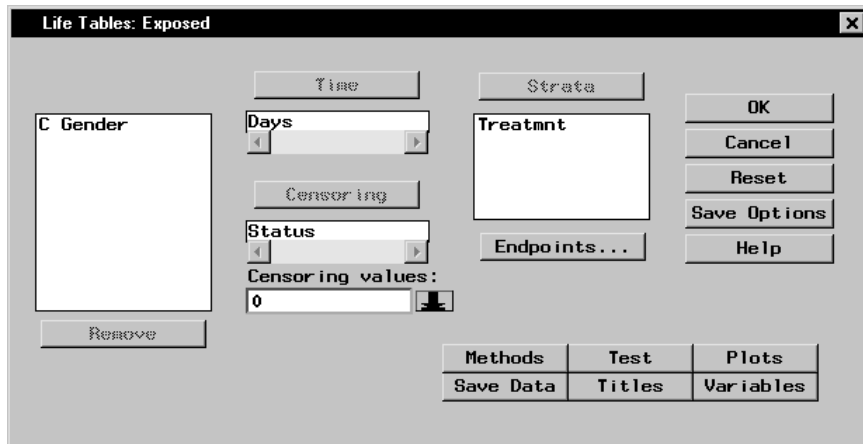**Figure 14.2.** Life Tables Dialog

Request A Survivor Function Plot

To produce a plot of the survivor function, follow these steps:

1. Click **Plots** to open the Plots dialog.

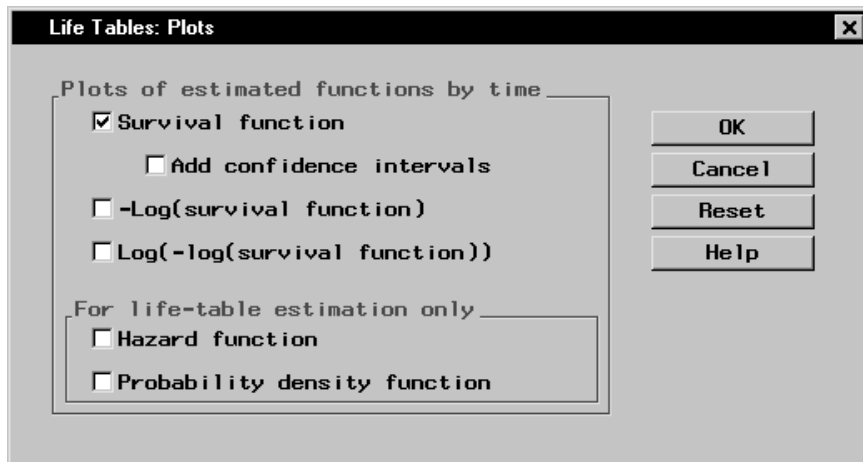2. Select **Survival function**.

3. Click **OK**.

Figure 14.3. Life Tables: Plots Dialog

When you have completed your selections, click **OK** in the main dialog to produce the analysis.

Review the Results

The results are presented in the project tree under the **Life Tables** folder, as displayed in Figure 14.4. The three nodes represent the life tables output, the survivor distribution function plot, and the SAS programming statements (labeled **Code**) that generated the output.
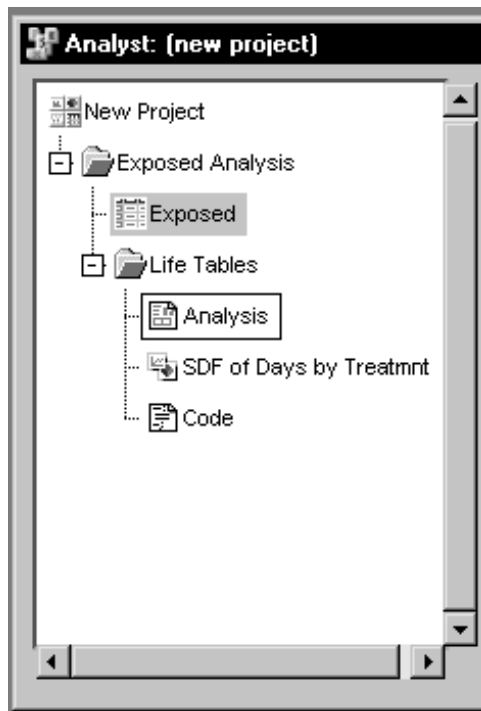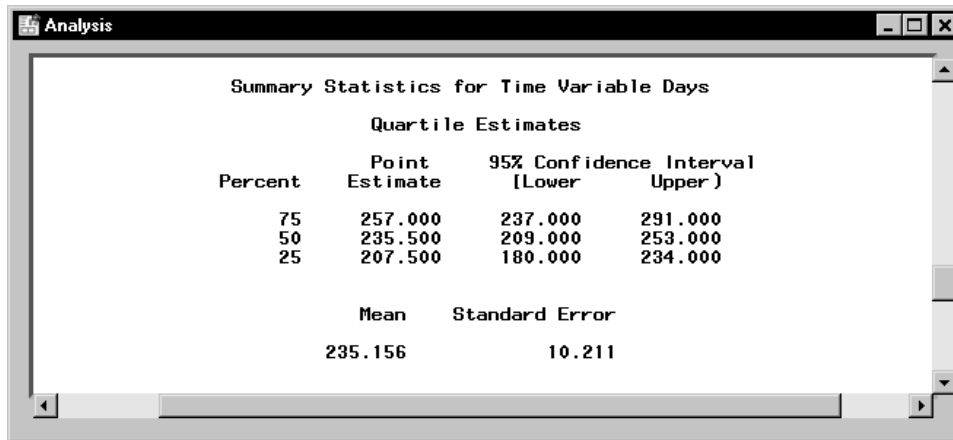

Figure 14.4. Life Tables: Project Tree

You can double-click on any node in the project tree to view the contents in a separate window.

```
  Analysis                                                      _ □ ✕

              Summary Statistics for Time Variable Days

                         Quartile Estimates

                        Point        95% Confidence Interval
             Percent    Estimate      [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 14.5. Life Tables: Results

Figure 14.5 displays summary statistics for the survival times for rats administered treatment 2. Of greatest interest is the 50th percentile, which is the median survival time. Here, rats administered treatment 2 have a median survival time of 235.5 days with a 95-percent confidence interval of 209 to 253. The mean survival time is 235.156 with a standard error of 10.211.
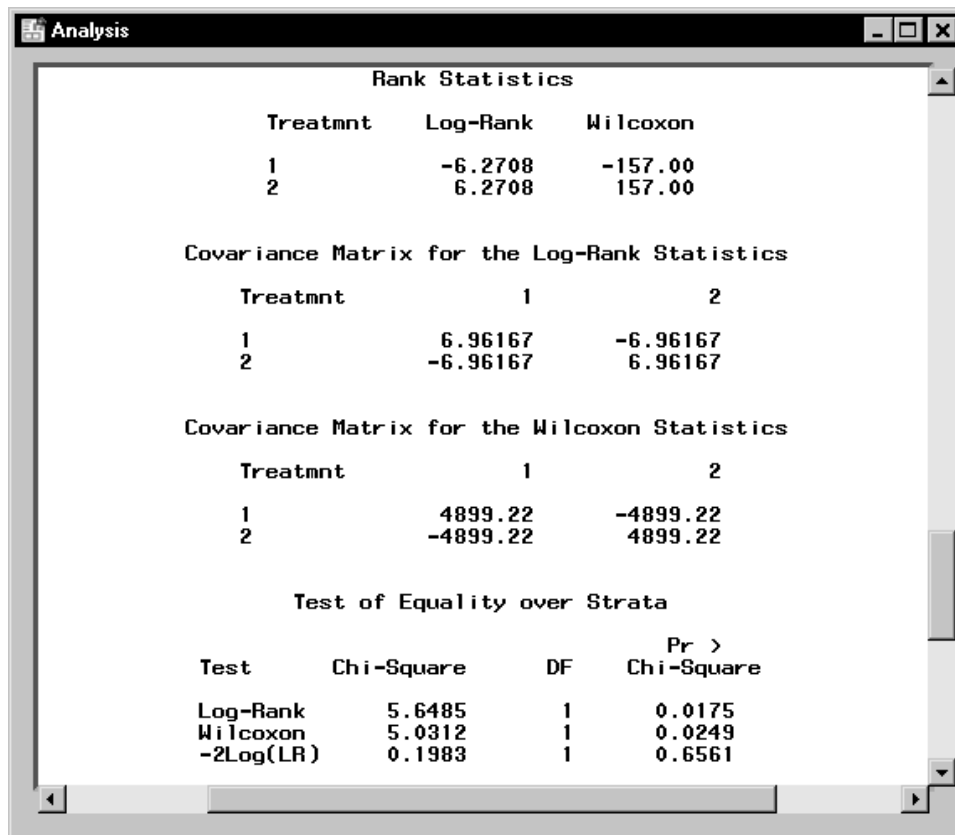
```
                          Rank Statistics

                Treatmnt      Log-Rank      Wilcoxon

                    1          -6.2708       -157.00
                    2           6.2708        157.00


         Covariance Matrix for the Log-Rank Statistics

             Treatmnt             1              2

                 1            6.96167        -6.96167
                 2           -6.96167         6.96167


         Covariance Matrix for the Wilcoxon Statistics

             Treatmnt             1              2

                 1            4899.22        -4899.22
                 2           -4899.22         4899.22


              Test of Equality over Strata

                                            Pr >
         Test        Chi-Square     DF    Chi-Square

         Log-Rank      5.6485        1      0.0175
         Wilcoxon      5.0312        1      0.0249
         -2Log(LR)     0.1983        1      0.6561
```

Figure 14.6. Life Tables: Test for Equality over Strata

The "Test for Equality over Strata" table contains rank and likelihood-based statistics for testing homogeneity of survivor functions across strata. 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), where rats in the first treatment group live significantly longer than those in the second treatment group. The log-rank test, which places more weight on larger survival times, has a lower p-value than the Wilcoxon test, which places more weight on early survival times.
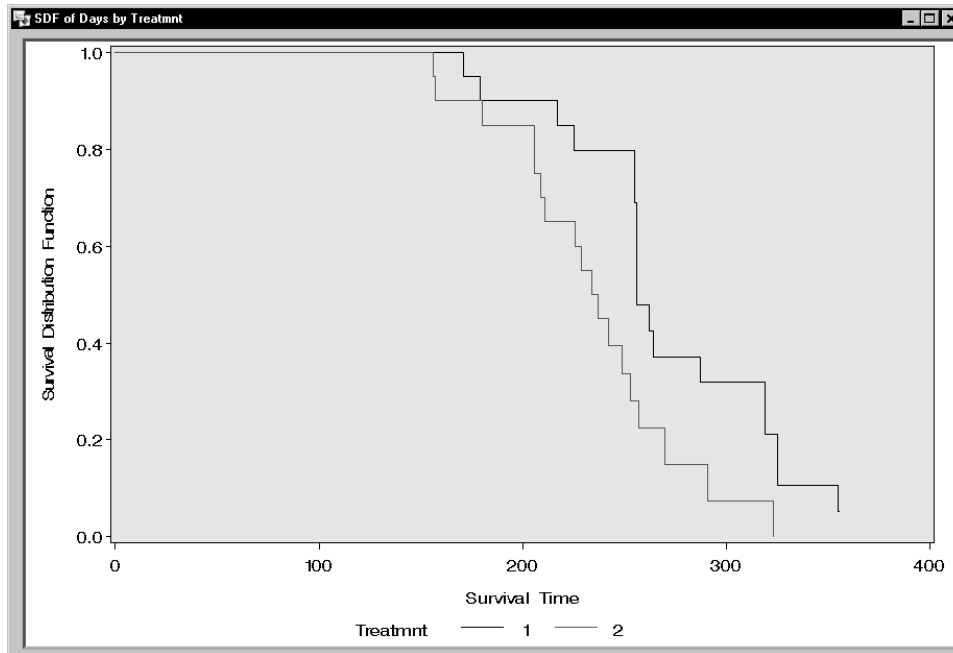
Figure 14.7.   Life Tables: Survivor Distribution Plot

Figure 14.7 displays the survivor function against time for each of
the two treatments. The gap between the two curves distinguishes
between the survival distributions, where the curve for treatment 1
decreases after the curve for treatment 2. The difference in displayed
survival curves reinforces the conclusions that the rats in the first
treatment group live longer than rats in the second group.

Proportional Hazards

The example in this section contains information on a different study
that explores survival times of rats exposed to a carcinogen. Two
groups of rats received different pretreatment regimes and were ex-
posed to a carcinogen. Investigators recorded the survival times of
the rats from exposure to death from vaginal cancer. Interest lies
in whether the survival curves differ between the two groups. The

data set Rats contains the variables Days, Status, and Group. The variable Days is the survival time in days. Status is the censoring variable and has the value 0 if the observation is censored and 1 if the observation is not censored. The Group variable indicates the pretreatment group, which takes the value 0 for the first treatment and 1 for the second treatment.

Open the Rats Data Set

The data are provided in the Analyst Sample Library. To open the Rats data set, follow these steps:

1. Select **Tools → Sample Data** . . .

2. Select Rats.

3. Click **OK** to create the sample data set in your Sasuser directory.

4. Select **File → Open By SAS Name** . . .

5. Select Sasuser from the list of **Libraries**.

6. Select Rats from the list of members.

7. Click **OK** to bring the Rats data set into the data table.

To request proportional hazards regression, follow these steps:

1. Select **Statistics → Survival → Proportional Hazards** . . .

2. Select Days as the time variable.

The values of Days are considered censored if the value of Status is 0; otherwise, they are considered event times.

3. Select Status as the censoring variable.

4. Specify **0** as the censoring value by directly typing **0** in the **Censoring values:** field or by clicking the down arrow under **Censoring values:** and selecting **0** from the list.

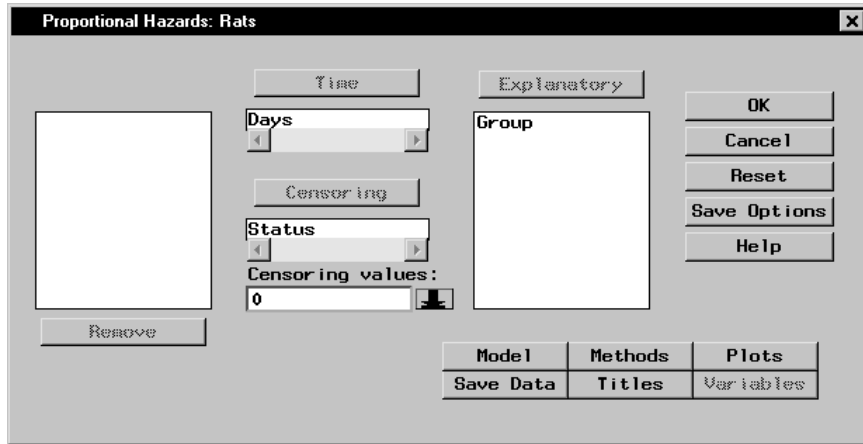5. Select Group as the explanatory variable.

Figure 14.8. Proportional Hazards Dialog

Click **OK** in the **Proportional Hazards** main dialog to produce the results for the proportional hazards task.

Review the Results

The results are presented in the project tree under the **Proportional Hazards** folder. Double-click on the icon labeled **Analysis** to display the corresponding information in an independent window.


```
              Testing Global Null Hypothesis: BETA=0

        Test              Chi-Square       DF      Pr > ChiSq

        Likelihood Ratio    2.8784          1        0.0898
        Score               3.0001          1        0.0833
        Wald                2.9254          1        0.0872

            Analysis of Maximum Likelihood Estimates

                 Parameter     Standard                            Hazard
Variable   DF     Estimate       Error    Chi-Square   Pr > ChiSq   Ratio

Group       1     -0.59590      0.34840     2.9254       0.0872      0.551
```

Figure 14.9. Proportional Hazards: Results

Figure 14.9 displays likelihood statistics and the analysis of parameter estimates. Since Group takes only two values, the null hypothesis for no difference between two groups is identical to the null hypothesis that the regression coefficient for Group is 0. All three tests in the "Testing Global Null Hypothesis: BETA=0" table suggest that the two pretreatment groups may not be the same. In this model, the hazards ratio (or risk ratio) for Group, defined as the exponentiation of the regression coefficient for Group, is the ratio of hazard functions between the two groups. The estimate is 0.551, implying that the hazard function for group 1 is smaller than the hazard function for group 0. In other words, rats in group 1 lived longer than those in group 0.

In this example, the comparison of two survival curves is put in the form of a proportional hazards model. This approach is essentially the same as the log-rank (Mantel-Haenszel) test. In fact, if there are no ties in the survival times, the likelihood score test in the Cox regression analysis is identical to the log-rank test. The advantage of the Cox regression approach is the ability to adjust for the other variables by including them in the model. For example, including a variable that contains the initial body weights of the rats could expand the present model.

References

Allison, P. D. (1995), *Survival Analysis Using the SAS System: A Practical Guide.* Cary, NC: SAS Institute Inc.

Cox, D. R. (1972), "Regression Models and Life-Tables (with Discussion)," *Journal of the Royal Statistical Society, Series B*, 34, 187–200.

SAS Institute Inc. (1999), *SAS/STAT User's Guide, Version 7-1*, Cary, NC: SAS Institute Inc.

The Analyst Application, First Edition