# SAS/STAT® 9.2 User's Guide

# The PHREG Procedure

## (Book Excerpt)

SAS® Documentation

This document is an individual chapter from *SAS/STAT® 9.2 User's Guide*.

The correct bibliographic citation for the complete manual is as follows: SAS Institute Inc. 2008. *SAS/STAT® 9.2 User's Guide*. Cary, NC: SAS Institute Inc.



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

1st electronic book, March 2008
2nd electronic book, February 2009

# Chapter 64

# The PHREG Procedure

**Contents**

# Overview: PHREG Procedure

The analysis of survival data requires special techniques because the data are almost always incomplete, and familiar parametric assumptions might be unjustifiable. Investigators follow subjects until they reach a prespecified endpoint (for example, death). However, subjects sometimes withdraw from a study, or the study is completed before the endpoint is reached. In these cases, the survival times (also known as failure times) are *censored*; subjects survived to a certain time beyond which their status is unknown. The uncensored survival times are sometimes referred to as *event* times. Methods of survival analysis must account for both censored and uncensored data.

Many types of models have been used for survival data. Two of the more popular types of models are the accelerated failure time model (Kalbfleisch and Prentice 1980) and the Cox proportional hazards model (Cox 1972). Each has its own assumptions about the underlying distribution of the survival times. Two closely related functions often used to describe the distribution of survival times are the survivor function and the hazard function (see the section "Failure Time Distribution" on page 4573 for definitions). The accelerated failure time model assumes a parametric form for the effects of the explanatory variables and usually assumes a parametric form for the underlying survivor function. Cox's proportional hazards model also assumes a parametric form for the effects of the explanatory variables, but it allows an unspecified form for the underlying survivor function.

The PHREG procedure performs regression analysis of survival data based on the Cox proportional hazards model. Cox's semiparametric model is widely used in the analysis of survival data to explain the effect of explanatory variables on hazard rates.

The survival time of each member of a population is assumed to follow its own hazard function, $\lambda_i(t)$, expressed as

$$\lambda_i(t) = \lambda(t; \mathbf{Z}_i) = \lambda_0(t) \exp(\mathbf{Z}_i' \boldsymbol{\beta})$$

where $\lambda_0(t)$ is an arbitrary and unspecified baseline hazard function, $\mathbf{Z}_i$ is the vector of explanatory variables for the $i$th individual, and $\boldsymbol{\beta}$ is the vector of unknown regression parameters associated with the explanatory variables. The vector $\boldsymbol{\beta}$ is assumed to be the same for all individuals. The survivor function can be expressed as

$$S(t; \mathbf{Z}_i) = [S_0(t)]^{\exp(\mathbf{Z}_i' \boldsymbol{\beta})}$$

where $S_0(t) = \exp(-\int_0^t \lambda_0(u) du)$ is the baseline survivor function. To estimate $\boldsymbol{\beta}$, Cox (1972, 1975) introduced the partial likelihood function, which eliminates the unknown baseline hazard $\lambda_0(t)$ and accounts for censored survival times.

The partial likelihood of Cox also allows time-dependent explanatory variables. An explanatory variable is time-dependent if its value for any given individual can change over time. Time-dependent variables have many useful applications in survival analysis. You can use a time-dependent variable to model the effect of subjects changing treatment groups. Or you can include time-dependent variables such as blood pressure or blood chemistry measures that vary with time during the course of a study. You can also use time-dependent variables to test the validity of the proportional hazards model.

An alternative way to fit models with time-dependent explanatory variables is to use the counting process style of input. The counting process formulation enables PROC PHREG to fit a superset of the Cox model, known as the multiplicative hazards model. This extension also includes recurrent events data and left truncation of failure times. The theory of these models is based on the counting process pioneered by Andersen and Gill (1982), and the model is often referred to as the Andersen-Gill model.

Multivariate failure time data arise when each study subject can potentially experience several events (for instance, multiple infections after surgery) or when there exists some natural or artificial clustering of subjects (for instance, a litter of mice) that induces dependence among the failure times of the same cluster. Data in the former situation are referred to as multiple events data, which include recurrent events data as a special case; data in the latter situation are referred to as clustered data. You can use PROC PHREG to carry out various methods of analyzing these data.

The population under study can consist of a number of subpopulations, each of which has its own baseline hazard function. PROC PHREG performs a stratified analysis to adjust for such subpopulation differences. Under the stratified model, the hazard function for the $j$th individual in the $i$th stratum is expressed as

$$\lambda_{ij}(t) = \lambda_{i0}(t) \exp(\mathbf{Z}'_{ij}\boldsymbol{\beta})$$

where $\lambda_{i0}(t)$ is the baseline hazard function for the $i$th stratum, and $\mathbf{Z}_{ij}$ is the vector of explanatory variables for the individual. The regression coefficients are assumed to be the same for all individuals across all strata.

Ties in the failure times can arise when the time scale is genuinely discrete or when survival times generated from the continuous-time model are grouped into coarser units. The PHREG procedure includes four methods of handling ties. The *discrete* logistic model is available for discrete time-scale data. The other three methods apply to continuous time-scale data. The *exact* method computes the exact conditional probability under the model that the set of observed tied event times occurs before all the censored times with the same value or before larger values. *Breslow* and *Efron* methods provide approximations to the exact method.

Variable selection is a typical exploratory exercise in multiple regression when the investigator is interested in identifying important prognostic factors from a large number of candidate variables. The PHREG procedure provides four selection methods: forward selection, backward elimination, stepwise selection, and best subset selection. The best subset selection method is based on the likelihood score statistic. This method identifies a specified number of best models containing one, two, or three variables and so on, up to the single model containing all of the explanatory variables.

The PHREG procedure also enables you to include an offset variable in the model; to weight the observations in the input data; to test linear hypotheses about the regression parameters; to perform conditional logistic regression analysis for matched case-control studies; to output survivor function estimates, residuals, and regression diagnostics; and to estimate the survivor function for a new set of covariates.

PROC PHREG can also be used to fit the multinomial logit choice model to discrete choice data. See http://support.sas.com/resources/papers/tnote/tnote_marketresearch.html for more information about discrete choice modeling and the multinomial logit model. Look for the “Discrete Choice” report.

The PHREG procedure now uses ODS Graphics to create graphs as part of its output. For example, the ASSESS statement uses a graphical method that uses ODS Graphics to check the adequacy of the model. See Chapter 21, “Statistical Graphics Using ODS,” for general information about ODS Graphics.

There have been a number of enhancements to PROC PHREG with this release. The most noticeable additions are the CLASS statement for specifying categorical variables; the CONTRAST statement for estimating and testing linear contrasts; the BAYES statement for performing a Bayesian analysis; and the HAZARDRATIO statement for estimating customized hazard ratios.

The CLASS statement enables you to specify categorical variables (also known as factors or CLASS variables) to be used in the analysis. Model effects, including covariates, main effects (CLASS variables), crossed effects (interactions), and nested effects, can be specified in the same way as in the GLM procedure. The CLASS statement supports less-than-full-rank parameterization as in the GLM procedure as well as various full-rank parameterization methods such as reference coding, effect coding, and orthogonal polynomial coding. For some of the full-rank coding schemes, you can designate a specific value (category or level) of the CLASS variable as the reference level. The CLASS statement also enables you to specify the ordering of the categories of CLASS variables, to reverse the ordering of the categories, and to treat categories with missing values as valid categories.

With the new way of specifying model effects, the CONTRAST statement enables you to test customized hypotheses concerning the regression parameters. Each CONTRAST statement also provides estimation of individual rows of contrasts, which is particularly useful in comparing the hazards between the categories of a CLASS variable.

The BAYES statement invokes a Bayesian analysis of the Cox model or the piecewise constant baseline hazard model (also known as the piecewise exponential model ). In essence, the Bayesian paradigm treats parameters as random variables, and inference (measurement of uncertainty) about parameters is based on the posterior distribution of the parameters. A posterior distribution is a weighted likelihood function of the data with a prior distribution that uses the Bayes theorem. Without any past experience or knowledge of what prior distribution to use, you can always start with a noninformative prior. Knowledge of the prior is accumulated over time, and the Bayesian approach can be viewed as a process of learning from experience. A closed form of the posterior distribution is hard to come by, and a Markov chain Monte Carlo method is used to simulate samples from the posterior distribution. See Chapter 7, “Introduction to Bayesian Analysis Procedures,” for an introduction to the basic concepts in Bayesian statistics. You can also refer to the section “Bayesian Analysis: Advantages and Disadvantages” on page 149 for a discussion of the advantages and disadvantages of Bayesian analysis. For the Cox model, the partial likelihood is used as the likelihood, which is justified by Sinha, Ibrahim, and Chen (2003). PROC PHREG generates a chain of posterior distribution samples by the Gibbs sampler, using the adaptive rejection sampling algorithm (Gilks and Wild 1992; Gilks, Best, and Tan 1995) to sample each parameter value from its full conditional distribution. Summary statistics (mean, standard deviation, percentiles, HPD intervals, and equal-tail credible intervals) and convergence diagnostics (autocorrelations; Gelman-Rubin, Geweke, Raftery-Lewis, and Heidelberger-Welch tests; and the effective sample size) are computed for each parameter, as well as the covariance and correlation matrices of the posterior samples. Trace plots, posterior density plots, and autocorrelation function plots are also provided using ODS Graphics.

The HAZARDRATIO statement identifies the variable for which hazard ratios are to be evaluated. For a continuous variable, the hazard ratio compares the hazards for a given change in the variable. For a CLASS variable, a hazard ratio compares the hazards of two levels of the variable. The HAZARDRATIO statement enables you to obtain hazard ratios even in the presence of interactions and nested effects.

Other enhancements include plotting the baseline functions through ODS Graphics, computing profile-likelihood-based confidence limits for hazard ratios, and allowing the bias-reducing penalized likelihood optimization as discussed by Firth (1993) and Heinze and Schemper (2001).

For both the BASELINE and OUTPUT statements, the default method of estimating a survivor function is now based on the Breslow (1972) estimator—that is, METHOD=CH. There are a few other changes in the BASELINE statement with this release. The option NOMEAN in the past releases has become obsolete—that is, requested statistics at the sample average values of the covariates are no longer computed and added to the OUT= data set. However, if the COVARIATES= data set is not specified, the requested statistics are computed and output for the covariate set that consists of the reference levels for the CLASS variables and sample averages for the continuous variable. In addition to the requested statistics, the OUT= data set also contains all variables in the COVARIATES= data set.

The remaining sections of this chapter contain information about how to use PROC PHREG, information about the underlying statistical methodology, and some sample applications of the procedure. The section "Getting Started: PHREG Procedure" on page 4520 introduces PROC PHREG with two examples. The section "Syntax: PHREG Procedure" on page 4529 describes the syntax of the procedure. The section "Details: PHREG Procedure" on page 4573 summarizes the statistical techniques employed in PROC PHREG. The section "Examples: PHREG Procedure" on page 4638 includes eight additional examples of useful applications. Experienced SAS/STAT software users might decide to proceed to the "Syntax" section, while other users might choose to read both the "Getting Started" and "Examples" sections before proceeding to "Syntax" and "Details."

## Getting Started: PHREG Procedure

This section uses the two-sample vaginal cancer mortality data from Kalbfleisch and Prentice (1980, p. 2) in two examples to illustrate some of the basic features of PROC PHREG. The first example carries out a classical Cox regression analysis and the second example performs a Bayesian analysis of the Cox model.

Two groups of rats received different pretreatment regimes and then were exposed to a carcinogen. Investigators recorded the survival times of the rats from exposure to mortality from vaginal cancer. Four rats died of other causes, so their survival times are censored. Interest lies in whether the survival curves differ between the two groups.

The following DATA step creates the data set Rats, which contains the variable Days (the survival time in days), the variable Status (the censoring indicator variable: 0 if censored and 1 if not censored), and the variable Group (the pretreatment group indicator).

```
data Rats;
   label Days  ='Days from Exposure to Death';
   input Days Status Group @@;
   datalines;
143 1 0   164 1 0   188 1 0   188 1 0
190 1 0   192 1 0   206 1 0   209 1 0
213 1 0   216 1 0   220 1 0   227 1 0
230 1 0   234 1 0   246 1 0   265 1 0
304 1 0   216 0 0   244 0 0   142 1 1
156 1 1   163 1 1   198 1 1   205 1 1
232 1 1   232 1 1   233 1 1   233 1 1
233 1 1   233 1 1   239 1 1   240 1 1
261 1 1   280 1 1   280 1 1   296 1 1
296 1 1   323 1 1   204 0 1   344 0 1
;
run;
```

By using ODS Graphics, PROC PHREG allows you to plot the survival curve for Group=0 and the survival curve for Group=1, but first you must save these two covariate values in a SAS data set as in the following DATA step:

```
data Regimes;
  Group=0;
  output;
  Group=1;
  output;
run;
```

## Classical Method of Maximum Likelihood

PROC PHREG fits the Cox model by maximizing the partial likelihood and computes the baseline survivor function by using the Breslow (1972) estimate. The following statements produce Figure 64.1 and Figure 64.2:

```
ods graphics on;
proc phreg data=Rats plot(overlay)=survival;
   model Days*Status(0)=Group;
   baseline covariates=regimes out=_null_;
run;
ods graphics off;
```

In the MODEL statement, the response variable, Days, is crossed with the censoring variable, Status, with the value that indicates censoring is enclosed in parentheses. The values of Days are considered censored if the value of Status is 0; otherwise, they are considered event times. Graphics results are enabled through ODS Graphics with the specification of the **ods graphics on** statement. The survival curves for the two observations in the data set Regime, specified in the COVARIATES= option in the BASELINE statement, are requested through the PLOTS= option with the OVERLAY option for overlaying both survival curves in the same plot.

Figure 64.2 shows a typical printed output of a classical analysis. Since Group takes only two values, the null hypothesis for no difference between the 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 (see the section "Testing the Global Null Hypothesis" on page 4591) suggest that the survival curves for the two pretreatment groups might not be the same. In this model, the hazard ratio (or risk ratio) for Group, defined as the exponentiation of the regression coefficient for Group, is the ratio of the hazard functions between the two groups. The estimate is 0.551, implying that the hazard function for Group=1 is smaller than that for Group=0. In other words, rats in Group=1 lived longer than those in Group=0. This conclusion is also revealed in the plot of the survivor functions in Figure 64.2.

**Figure 64.1** Comparison of Two Survival Curves

```
                          The PHREG Procedure

                           Model Information

       Data Set                   WORK.RATS
       Dependent Variable         Days            Days from Exposure to Death
       Censoring Variable         Status
       Censoring Value(s)         0
       Ties Handling              BRESLOW

                 Number of Observations Read           40
                 Number of Observations Used           40

            Summary of the Number of Event and Censored Values

                                                     Percent
                  Total        Event     Censored     Censored

                     40           36            4        10.00

                           Convergence Status

              Convergence criterion (GCONV=1E-8) satisfied.

                         Model Fit Statistics

                                 Without              With
                 Criterion      Covariates       Covariates

                 -2 LOG L          204.317          201.438
                 AIC               204.317          203.438
                 SBC               204.317          205.022
```

**Figure 64.1** *continued*

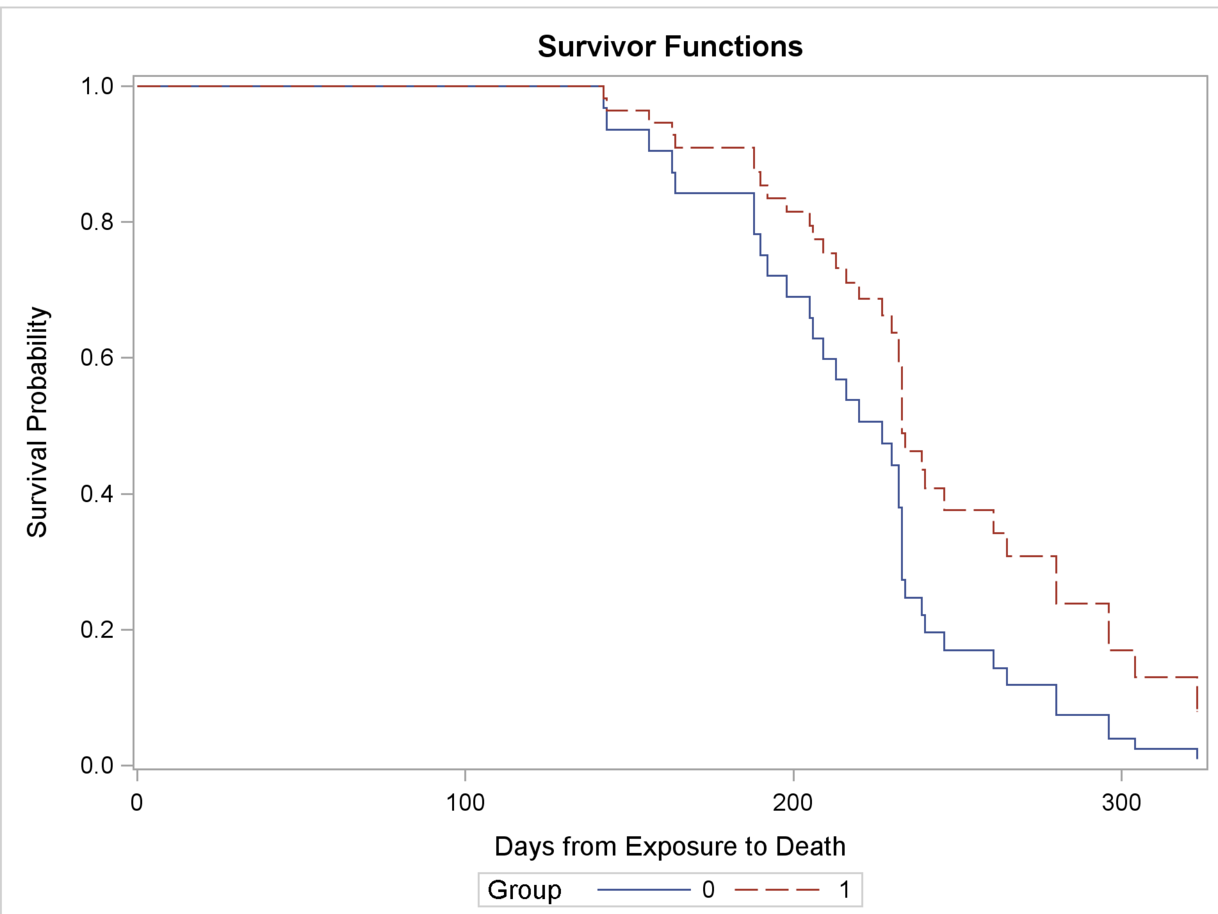
Testing Global Null Hypothesis: BETA=0

| 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 | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| Group | 1 | -0.59590 | 0.34840 | 2.9254 | 0.0872 | 0.551 |

**Figure 64.2** Survivorship for the Two Pretreatment Regimes




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, the present model could be expanded by including a variable that contains the initial body weights of the rats.

Next, consider a simple test of the validity of the proportional hazards assumption. The proportional hazards model for comparing the two pretreatment groups is given by the following:

$$\lambda(t) = \begin{cases} \lambda_0(t) & \text{if GROUP} = 0 \\ \lambda_0(t)e^{\beta_1} & \text{if GROUP} = 1 \end{cases}$$

The ratio of hazards is $e^{\beta_1}$, which does not depend on time. If the hazard ratio changes with time, the proportional hazards model assumption is invalid. Simple forms of departure from the proportional hazards model can be investigated with the following time-dependent explanatory variable $x = x(t)$:

$$x(t) = \begin{cases} 0 & \text{if GROUP} = 0 \\ \log(t) - 5.4 & \text{if GROUP} = 1 \end{cases}$$

Here, $\log(t)$ is used instead of $t$ to avoid numerical instability in the computation. The constant, 5.4, is the average of the logs of the survival times and is included to improve interpretability. The hazard ratio in the two groups then becomes $e^{\beta_1 - 5.4\beta_2} t^{\beta_2}$, where $\beta_2$ is the regression parameter for the time-dependent variable $x$. The term $e^{\beta_1}$ represents the hazard ratio at the geometric mean of the survival times. A nonzero value of $\beta_2$ would imply an increasing ($\beta_2 > 0$) or decreasing ($\beta_2 < 0$) trend in the hazard ratio with time.

The following statements implement this simple test of the proportional hazards assumption. The MODEL statement includes the time-dependent explanatory variable X, which is defined subsequently by the programming statement. At each event time, subjects in the risk set (those alive just before the event time) have their X values changed accordingly.

```
proc phreg data=Rats;
   model Days*Status(0)=Group X;
   X=Group*(log(Days) - 5.4);
run;
```

The analysis of the parameter estimates is displayed in Figure 64.3. The Wald chi-square statistic for testing the null hypothesis that $\beta_2 = 0$ is 0.0158. The statistic is not statistically significant when compared to a chi-square distribution with one degree of freedom ($p = 0.8999$). Thus, you can conclude that there is no evidence of an increasing or decreasing trend over time in the hazard ratio.

**Figure 64.3** A Simple Test of Trend in the Hazard Ratio

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| Group | 1 | -0.59976 | 0.34837 | 2.9639 | 0.0851 | 0.549 |
| X | 1 | -0.22952 | 1.82489 | 0.0158 | 0.8999 | 0.795 |

## Bayesian Analysis

PROC PHREG uses the partial likelihood of the Cox model as the likelihood and generates a chain of posterior distribution samples by the Gibbs Sampler. Summary statistics, convergence diagnostics, and diagnostic plots are provided for each parameter. The following statements generate Figure 64.4–Figure 64.10:

```
ods graphics on;
proc phreg data=Rats;
   model Days*Status(0)=Group;
   bayes seed=1 outpost=Post;
run;
ods graphics off;
```

The BAYES statement invokes the Bayesian analysis. The SEED= option is specified to maintain reproducibility; the OUTPOST= option saves the posterior distribution samples in a SAS data set for postprocessiong; no other options are specified in the BAYES statement. By default, a uniform prior distribution is assumed on the regression coefficient Group. The uniform prior is a flat prior on the real line with a distribution that reflects ignorance of the location of the parameter, placing equal probability on all possible values the regression coefficient can take. Using the uniform prior in the following example, you would expect the Bayesian estimates to resemble the classical results of maximizing the likelihood. If you can elicit an informative prior on the regression coefficients, you should use the COEFFPRIOR= option to specify it.

You should make sure that the posterior distribution samples have achieved convergence before using them for Bayesian inference. PROC PHREG produces three convergence diagnostics by default. If you enable ODS Graphics before calling PROC PHREG as in the preceding program, diagnostics plots are also displayed.

The results of this analysis are shown in the following figures.

The "Model Information" table in Figure 64.4 summarizes information about the model you fit and the size of the simulation.

**Figure 64.4** Model Information

The PHREG Procedure

Bayesian Analysis

Model Information

| | | |
|---|---|---|
| Data Set | WORK.RATS | |
| Dependent Variable | Days | Days from Exposure to Death |
| Censoring Variable | Status | |
| Censoring Value(s) | 0 | |
| Model | Cox | |
| Ties Handling | BRESLOW | |
| Burn-In Size | 2000 | |
| MC Sample Size | 10000 | |
| Thinning | 1 | |

PROC PHREG first fits the Cox model by maximizing the partial likelihood. The only parameter in the model is the regression coefficient of Group. The maximum likelihood estimate (MLE) of the parameter and its 95% confidence interval are shown in Figure 64.5.

**Figure 64.5** Classical Parameter Estimates

Maximum Likelihood Estimates

| Parameter | DF | Estimate | Standard Error | 95% Confidence Limits | |
|---|---|---|---|---|---|
| Group | 1 | -0.5959 | 0.3484 | -1.2788 | 0.0870 |

Since no prior is specified for the regression coefficient, the default uniform prior is used. This information is displayed in the "Uniform Prior for Regression Coefficients" table in Figure 64.6.

**Figure 64.6** Coefficient Prior

Uniform Prior for Regression Coefficients

| Parameter | Prior |
|---|---|
| Group | Constant |

The "Fit Statistics" table in Figure 64.7 lists information about the fitted model. The table displays the DIC (deviance information criterion) and pD (effective number of parameters). See the section "Fit Statistics" on page 4622 for details.

**Figure 64.7** Fit Statistics

Fit Statistics

| | |
|---|---|
| DIC (smaller is better) | 203.444 |
| pD (Effective Number of Parameters) | 1.003 |

Summary statistics of the posterior samples are displayed in the "Posterior Summaries" table and "Posterior Intervals" table as shown in Figure 64.8. Note that the mean and standard deviation of the posterior samples are comparable to the MLE and its standard error, respectively, due to the use of the uniform prior.

**Figure 64.8** Summary Statistics

The PHREG Procedure

Bayesian Analysis

Posterior Summaries

| Parameter | N | Mean | Standard Deviation | Percentiles 25% | 50% | 75% |
|---|---|---|---|---|---|---|
| Group | 10000 | -0.5998 | 0.3511 | -0.8326 | -0.5957 | -0.3670 |

Posterior Intervals

| Parameter | Alpha | Equal-Tail Interval | | HPD Interval | |
|---|---|---|---|---|---|
| Group | 0.050 | -1.3042 | 0.0721 | -1.2984 | 0.0756 |

PROC PHREG provides diagnostics to assess the convergence of the generated Markov chain. Figure 64.9 shows three of these diagnostics: the lag1, lag5, lag10, and lag50 autocorrelations; the Geweke diagnostic; and the effective sample size. There is no indication that the Markov chain has not reached convergence. Refer to the section "Statistical Diagnostic Tests" on page 160 for information about interpreting these diagnostics.

**Figure 64.9** Convergence Diagnostics

The PHREG Procedure

Bayesian Analysis

Posterior Autocorrelations

| Parameter | Lag 1 | Lag 5 | Lag 10 | Lag 50 |
|---|---|---|---|---|
| Group | -0.0079 | 0.0091 | -0.0161 | 0.0101 |

Geweke Diagnostics

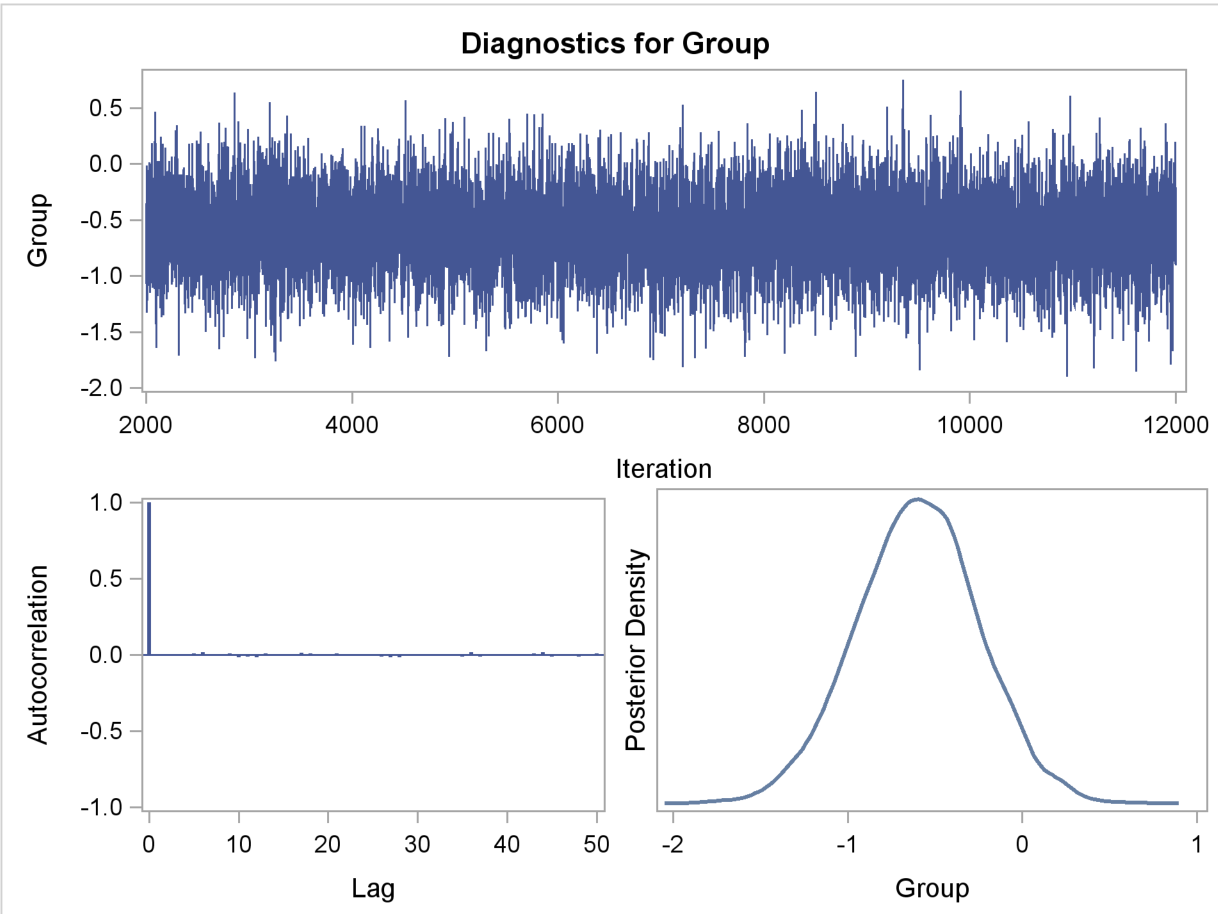
| Parameter | z | Pr > \|z\| |
|---|---|---|
| Group | 0.0149 | 0.9881 |

Effective Sample Sizes

| Parameter | ESS | Correlation Time | Efficiency |
|---|---|---|---|
| Group | 10000.0 | 1.0000 | 1.0000 |

You can also assess the convergence of the generated Markov chain by examining the trace plot, the autocorrelation function plot, and the posterior density plot. Figure 64.10 displays a panel of these three plots for the parameter Group. This graphical display is automatically produced when ODS Graphics is enabled. Note that the trace of the samples centers on –0.6 with only small fluctuations, the autocorrelations are quite small, and the posterior density appears bell-shaped—all exemplifying the behavior of a converged Markov chain.

**Figure 64.10** Diagnostic Plots




The proportional hazards model for comparing the two pretreatment groups is

$$\lambda(t) = \begin{cases} \lambda_0(t) & \text{if Group=0} \\ \lambda_0(t)e^{\beta} & \text{if Group=1} \end{cases}$$

The probability that the hazard of Group=0 is greater than that of Group=1 is

$$\Pr(\lambda_0(t) > \lambda_0(t)e^{\beta}) = \Pr(\beta < 0)$$

This probability can be enumerated from the posterior distribution samples by computing the fraction of samples with a coefficient less than 0. The following DATA step and PROC MEANS perform this calculation:

```
data New;
   set Post;
   Indicator=(Group < 0);
   label Indicator='Group < 0';
   run;
proc means data=New(keep=Indicator) n mean;
   run;
```

**Figure 64.11** Prob(Hazard(Group=0) > Hazard(Group=1))

```
                    The MEANS Procedure

         Analysis Variable : Indicator Group < 0

                   N            Mean
              ---------------------
               10000       0.9581000
              ---------------------
```

The PROC MEANS results are displayed in Output 64.11. There is a 95.8% chance that the hazard rate of Group=0 is greater than that of Group=1. The result is consistent with the fact that the average survival time of Group=0 is less than that of Group=1.

## Syntax: PHREG Procedure

The following statements are available in PROC PHREG. Items within < > are optional.

**PROC PHREG** < *options* > ;
  **ASSESS** *keyword* < */options* > ;
  **BASELINE** < *OUT=SAS-data-set* > < *COVARIATES=SAS-data-set* > < *keyword=name* . . . *keyword=name* > < */options* > ;
  **BAYES** < *options* > ;
  **BY** *variables* ;
  **CLASS** *variable* < *(options)* > < . . . *variable* < *(options)* > > < */options* > ;
  **CONTRAST** < *'label'* > *effect values* < *,. . . , effect values* > < */options* > ;
  **FREQ** *variable* ;
  **HAZARDRATIO** < *'label'* > *variable* < */options* > ;
  **ID** *variables* ;
  **MODEL** *response* < **censor(list)* > = *variables* < */options* > ;
  **OUTPUT** < *OUT=SAS-data-set* > < *keyword=name . . . keyword=name* > < */options* > ;
  *programming statements* ;
  **STRATA** *variable* < *(list)* > < . . . *variable* < *(list)* > > < */option* > ;
  **< label: >TEST** *equation1* < *,. . . , equationk* > < */options* > ;
  **WEIGHT** *variable* < */option* > ;

The PROC PHREG and MODEL statements are required. The CLASS statement, if present, must precede the MODEL statement, and the ASSESS or CONTRAST statement, if present, must come after the MODEL statement. The BAYES statement, that invokes a Bayesian analysis, is not compatible with the ASSESS, CONTRAST, ID, OUTPUT, and TEST statements, as well as a number of options in the PROC PHREG and MODEL statements. See the section "Specifics for Bayesian Analysis" on page 4614 for details.

The rest of this section provides detailed syntax information for each statement, beginning with the PROC PHREG statement. The remaining statements are covered in alphabetical order.

## PROC PHREG Statement

**PROC PHREG** < *options* > **;**

You can specify the following options in the PROC PHREG statement.

**ALPHA=***number*

specifies the level of significance $\alpha$ for $100(1-\alpha)\%$ confidence intervals. The value *number* must be between 0 and 1; the default value is 0.05, which results in 95% intervals. This value is used as the default confidence level for limits computed by the BASELINE, BAYES, CONTRAST, HAZARDRATIO, and MODEL statements. You can override this default by specifying the ALPHA= option in the separate statements.

**COVOUT**

adds the estimated covariance matrix of the parameter estimates to the OUTEST= data set. The COVOUT option has no effect unless the OUTEST= option is specified.

**COVM**

requests that the model-based covariance matrix (which is the inverse of the observed information matrix) be used in the analysis if the COVS option is also specified. The COVM option has no effect if the COVS option is not specified.

**COVSANDWICH < (AGGREGATE) >**

**COVS < (AGGREGATE) >**

requests the robust sandwich estimate of Lin and Wei (1989) for the covariance matrix. When this option is specified, this robust sandwich estimate is used in the Wald tests for testing the global null hypothesis, null hypotheses of individual parameters, and the hypotheses in the CONTRAST and TEST statements. In addition, a modified score test is computed in the testing of the global null hypothesis, and the parameter estimates table has an additional StdErrRatio column, which contains the ratios of the robust estimate of the standard error relative to the corresponding model-based estimate. Optionally, you can specify the keyword AGGREGATE enclosed in parentheses after the COVSANDWICH (or COVS) option, which requests a summing up of the score residuals for each distinct ID pattern in the computation of the robust sandwich covariance estimate. This AGGREGATE option has no effects if the ID statement is not specified.

**DATA=***SAS-data-set*

names the SAS data set containing the data to be analyzed. If you omit the DATA= option, the procedure uses the most recently created SAS data set.

**INEST=***SAS-data-set*

names the SAS data set that contains initial estimates for all the parameters in the model. BY-group processing is allowed in setting up the INEST= data set. See the section "INEST= Input Data Set" on page 4625 for more information.

**MULTIPASS**

requests that, for each Newton-Raphson iteration, PROC PHREG recompile the risk sets corresponding to the event times for the (start,stop) style of response and recomputes the values of the time-dependent variables defined by the programming statements for each observation in the risk sets. If the MULTIPASS option is not specified, PROC PHREG computes all risk sets and all the variable values and saves them in a utility file. The MULTIPASS option decreases required disk space at the expense of increased execution time; however, for very large data, it might actually save time since it is time-consuming to write and read large utility files. This option has an effect only when the (start,stop) style of response is used or when there are time-dependent explanatory variables.

**NAMELEN=***n*

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

**NOPRINT**

suppresses all displayed output. Note that this option temporarily disables the Output Delivery System (ODS); see Chapter 20, "Using the Output Delivery System," for more information.

**NOSUMMARY**

suppresses the summary display of the event and censored observation frequencies.

**OUTEST=***SAS-data-set*

creates an output SAS data set that contains estimates of the regression coefficients. The data set also contains the convergence status and the log likelihood. If you use the COVOUT option, the data set also contains the estimated covariance matrix of the parameter estimators. See the section "OUTEST= Output Data Set" on page 4624 for more information.

**PLOTS**< **(***global-plot-options***)** > **=** *plot-request*

**PLOTS**< **(***global-plot-options***)** > **= (***plot-request* < . . . < *plot-request* > >**)**

controls the baseline functions plots produced through ODS Graphics. Each observation in the COVARIATES= data set in the BASELINE statement represents a set of covariates for which a curve is produced for each plot request and for each stratum. You can use the ROWID= option in the BASELINE statement to specify a variable in the COVARIATES= data set for identifying the curves produced for the covariate sets. If the ROWID= option is not specified, the curves produced are identified by the covariate values if there is only a single covariate or by the observation numbers of the COVARIATES= data set if the model has two or more covariates. If the COVARIATES= data set is not specified, a reference set of covariates consisting of the reference levels for the CLASS variables and the average values for the continuous variables is used. For plotting more than one curve, you can use the

OVERLAY= option to group the curves in separate plots. When you specify one plot request, you can omit the parentheses around the plot request. Here are some examples:

```
plots=survival
plots=(survival cumhaz)
```

You must enable ODS Graphics before requesting plots, for example, like this:

```
ods graphics on;

proc phreg plots(cl)=survival;
  model Time*Status(0)=X1-X5;
  baseline covariates=One;
run;

ods graphics off;
```

The *global plot options* include the following:

**CL<=EQTAIL | HPD>**

displays the pointwise interval limits for the specified curves. For the classical analysis, CL displays the confidence limits. For the Bayesian analysis, CL=EQTAIL displays the equal-tail credible limits and CL=HPD displays the HPD limits. Specifying just CL in a Bayesian analysis defaults to CL=HPD.

**OVERLAY** < =*overlay-option* >

specifies how the curves for the various strata and covariate sets are overlaid. If the STRATA statement is not specified, specifying OVERLAY without any option will overlay the curves for all the covariate sets. The available *overlay options* are as follows:

**BYGROUP**

**GROUP**

overlays, for each stratum, all curves for the covariate sets that have the same GROUP= value in the COVARIATES= data set in the same plot.

**INDIVIDUAL**

**IND**

displays, for each stratum, a separate plot for each covariate set.

**BYROW**

**ROW**

displays, for each covariate set, a separate plot containing the curves for all the strata.

**BYSTRATUM**

**STRATUM**

displays, for each stratum, a separate plot containing the curves for all sets of covariates.

The default is OVERLAY=BYGROUP if the GROUP= option is specified in the BASELINE statement or if the COVARIATES= data set contains the _GROUP_ variable; otherwise the default is OVERLAY=INDIVIDUAL.

**TIMERANGE=(**< *min* > < *,max* >**)**

**TIMERANGE=**< *min* > < *,max* >

**RANGE=(**< *min* > < *,max* >**)**

**RANGE=**< *min* > < *,max* >

specifies the range of values on the time axis to clip the display. The *min* and *max* values are the lower and upper bounds of the range. By default, *min* is 0 and *max* is the largest event time.

The *plot requests* include the following:

**CUMHAZ**

plots the estimated cumulative hazard function for each set of covariates in the COVARIATES= data set in the BASELINE statement. If the COVARIATES= data set is not specified, the estimated cumulative hazard function is plotted for the reference set of covariates consisting of reference levels for the CLASS variables and average values for the continuous variables.

**MCF**

plots the estimated mean cumulative function for each set of covariates in the COVARIATES= data set in the BASELINE statement. If the COVARIATES= data set is not specified, the estimated mean cumulative function is plotted for the reference set of covariates consisting of reference levels for the CLASS variables and average values for the continuous variables.

**NONE**

suppresses all the plots in the procedure. Specifying this option is equivalent to disabling ODS Graphics for the entire procedure.

**SURVIVAL**

plots the estimated survivor function for each set of covariates in the COVARIATES= data set in the BASELINE statement. If COVARIATES= data set is not specified, the estimated survivor function is plotted for the reference set of covariates consisting of reference levels for the CLASS variables and average values for the continuous variables.

**SIMPLE**

displays simple descriptive statistics (mean, standard deviation, minimum, and maximum) for each explanatory variable in the MODEL statement.

## ASSESS Statement

**ASSESS** < **VAR=***(list)* > < **PH** > < */options* > **;**

The ASSESS statement performs the graphical and numerical methods of Lin, Wei, and Ying (1993) for checking the adequacy of the Cox regression model. The methods are derived from cumulative sums of martingale residuals over follow-up times or covariate values. You can assess the functional form of a covariate or you can check the proportional hazards assumption for each covariate in the Cox model. PROC PHREG uses ODS Graphics for the graphical displays. You must specify at least one of the following options to create an analysis.

**VAR=(***variable-list***)**

specifies the *list* of explanatory *variables* for which their functional forms are assessed. For each variable on the list, the observed cumulative martingale residuals are plotted against the values of the explanatory variable along with 20 (or $n$ if NPATHS=$n$ is specified) simulated residual patterns.

**PROPORTIONALHAZARDS**

**PH**

requests the checking of the proportional hazards assumption. For each explanatory variable in the model, the observed score process component is plotted against the follow-up time along with 20 (or $n$ if NPATHS=$n$ is specified) simulated patterns.

The following options can be specified after a slash (/):

**NPATHS=*n***

specifies the number of simulated residual patterns to be displayed in a cumulative martingale residual plot or a score process plot. The default is $n$=20.

**CRPANEL**

requests that a plot with four panels, each containing the observed cumulative martingale residuals and two simulated residual patterns, be created.

**RESAMPLE** < **=*n*** >

requests that the Kolmogorov-type supremum test be computed on 1,000 simulated patterns or on $n$ simulated patterns if $n$ is specified.

**SEED=*n***

specifies an integer seed for the random number generator used in creating simulated realizations for plots and for the Kolmogorov-type supremum tests. Specifying a seed enables you to reproduce identical graphs and $p$-values for the model assessments from the same PHREG specification. If the SEED= option is not specified, or if you specify a nonpositive seed, a random seed is derived from the time of day.

## BASELINE Statement

**BASELINE** < **OUT=***SAS-data-set*> < **COVARIATES=***SAS-data-set*> < **TIMELIST=***list*> < *keyword=name . . . keyword=name*> < */options*> **;**

The BASELINE statement creates a new SAS data set that contains the baseline function estimates at the event times of each stratum for every set of covariates (**x**) given in the COVARIATES= data set. If the COVARIATES= data set is not specified, a reference set of covariates consisting of the reference levels for the CLASS variables and the average values for the continuous variables is used. No BASELINE data set is created if the model contains a time-dependent variable defined by means of programming statement.

The following options are available in the BASELINE statement.

**OUT=***SAS-data-set*

names the output BASELINE data set. If you omit the OUT= option, the data set is created and given a default name by using the DATA*n* convention. See the section "OUT= Output Data Set in the BASELINE Statement" on page 4626 for more information.

**COVARIATES=***SAS-data-set*

names the SAS data set that contains the sets of explanatory variable values for which the quantities of interest are estimated. All variables in the COVARIATES= data set are copied to the OUT= data set. Thus, any variable in the COVARIATES= data set can be used to identify the covariate sets in the OUT= data set.

**TIMELIST=***list*

specifies a list of time points at which the survival function estimates, cumulative function estimates, or MCF estimates are computed. The following specifications are equivalent:

```
timelist=5,20 to 50 by 10
timelist= 5 20 30 40 50
```

If the TIMELIST= option is not specified, the default is to carry out the prediction at all event times and at time 0. This option can be used only for the Bayesian analysis.

*keyword=name*

specifies the statistics to be included in the OUT= data set and assigns names to the variables that contain these statistics. Specify a *keyword* for each desired statistic, an equal sign, and the name of the variable for the statistic. Not all *keywords* listed in Table 64.1 (and discussed in the text that follows) are appropriate for both the classical analysis and the Bayesian analysis; and the table summaries the choices for each analysis.

**Table 64.1** Summary of the Keyword Choices

| Keyword | Classical | Bayesian |
|---|---|---|
| **Survivor Function** | | |
| SURVIVAL | x | x |
| STDERR | x | x |
| LOWER | x | x |

**Table 64.1** *continued*

| Options | Classical | Bayesian |
|---|---|---|
| UPPER | x | x |
| LOWERHPD | | x |
| UPPERHPD | | x |
| | | |
| **Cumulative Hazard Function** | | |
| CUMHAZ | x | x |
| STDCUMHAZ | x | x |
| LOWERCUMHAZ | x | x |
| UPPERCUMHAZ | x | x |
| LOWERHPDCUMHAZ | | x |
| UPPERHPDCUMHAZ | | x |
| | | |
| **Cumulative Mean Function** | | |
| CMF | x | |
| STDCMF | x | |
| LOWERCMF | x | |
| UPPERCMF | x | |
| | | |
| **Others** | | |
| XBETA | x | x |
| STDXBETA | x | x |
| LOGSURV | x | |
| LOGLOGS | x | |

The available *keywords* are as follows.

**CMF**

**MCF**

specifies the cumulative mean function estimate for recurrent events data. Specifying CMF=_ALL_ is equivalent to specifying CMF=CMF, STDCMF=StdErrCMF, LOWERCMF=LowerCMF, and UPPERCMF=UpperCMF. Nelson (2002) refers to the mean function estimate as MCF (mean cumulative function).

**CUMHAZ**

specifies the cumulative hazard function estimate. Specifying CUMHAZ=_ALL_ is equivalent to specifying CUMHAZ=CumHaz, STDCUMHAZ=StdErrCumHaz, LOWERCUMHAZ=LowerCumHaz, and UPPERCUMHAZ=UpperCumHaz. For a Bayesian analysis, CUMHAZ=_ALL_ also includes LOWERHPDCUMHAZ=LowerHPDCumHaz and UpperHPDCUMHAZ=UpperHPDCumHaz.

**LOGLOGS**

specifies the log of the negative log of SURVIVAL.

**LOGSURV**

specifies the log of SURVIVAL.

**LOWER**

**L**

specifies the lower pointwise confidence limit for the survivor function. For a Bayesian analysis, this is the lower limit of the equal-tail credible interval for the survivor function. The confidence level is determined by the ALPHA= option.

**LOWERCMF**

**LOWERMCF**

specifies the lower pointwise confidence limit for the cumulative mean function. The confidence level is determined by the ALPHA= option.

**LOWERHPD**

specifies the lower limit of the HPD interval for the survivor function. The confidence level is determined by the ALPHA= option.

**LOWERHPDCUMHAZ**

specifies the lower limit of the HPD interval for the cumulative hazard function. The confidence level is determined by the ALPHA= option.

**LOWERCUMHAZ**

specifies the lower pointwise confidence limit for the cumulative hazard function. For a Bayesian analysis, this is the lower limit of the equal-tail credible interval for the cumulative hazard function. The confidence level is determined by the ALPHA= option.

**STDERR**

specifies the standard error of the survivor function estimator. For a Bayesian analysis, this is the standard deviation of the posterior distribution of the survivor function.

**STDCMF**

**STDMCF**

specifies the estimated standard error of the cumulative mean function estimator.

**STDCUMHAZ**

specifies the estimated standard error of the cumulative hazard function estimator. For a Bayesian analysis, this is the standard deviation of the posterior distribution of the cumulative hazard function.

**STDXBETA**

specifies the estimated standard error of the linear predictor estimator. For a Bayesian analysis, this is the standard deviation of the posterior distribution of the linear predictor.

**SURVIVAL**

specifies the survivor function ($S(t) = [S_0(t)]^{\exp(\boldsymbol{\beta}'\mathbf{x})}$) estimate. Specifying SURVIVAL=_ALL_ is equivalent to specifying SURVIVAL=Survival, STDERR=StdErrSurvival, LOWER=LowerSurvival, and UPPER=UpperSurvival; and for a Bayesian analyis, SURVIVAL=_ALL_ also specifies LOWERHPD=LowerHPDSurvival and UPPERHPD=UpperHPDSurvival.

**UPPER**

**U**

specifies the upper pointwise confidence limit for the survivor function. For a Bayesian analysis, this is the upper limit of the equal-tail credible interval for the survivor function. The confidence level is determined by the ALPHA= option.

**UPPERCMF**

**UPPERMCF**

specifies the upper pointwise confidence limit for the cumulative mean function. The confidence level is determined by the ALPHA= option.

**UPPERCUMHAZ**

specifies the upper pointwise confidence limit for the cumulative hazard function. For a Bayesian analysis, this is the upper limit of the equal-tail credible interval for the cumulative hazard function. The confidence level is determined by the ALPHA= option.

**UPPERHPD**

specifies the upper limit of the equal-tail credible interval for the survivor function. The confidence level is determined by the ALPHA= option.

**UPPERHPDCUMHAZ**

specifies the upper limit of the equal-tail credible interval for the cumulative hazard function. The confidence level is determined by the ALPHA= option.

**XBETA**

specifies the estimate of the linear predictor $\mathbf{x}'\boldsymbol{\beta}$.

The following options can appear in the BASELINE statement after a slash (/). The METHOD= and CLTYPE= options apply only to the estimate of the survivor function in the classical analysis. For the Bayesian analysis, the survivor function is estimated by the Breslow (1972) method.

**ALPHA=*value***

specifies the significance level of the confidence interval for the survivor function. The value must be between 0 and 1. The default is the value of the ALPHA= option in the PROC PHREG statement, or 0.05 if that option is not specified.

**CLTYPE=*method***

specifies the method used to compute the confidence limits for $S(t,\mathbf{z})$, the survivor function for a subject with a fixed covariate vector $\mathbf{z}$ at event time $t$. The CLTYPE= option can take the following values:

**LOG**

specifies that the confidence limits for $\log(S(t,\mathbf{z}))$ be computed using the normal theory approximation. The confidence limits for $S(t,\mathbf{z})$ are obtained by back-transforming the confidence limits for $\log(S(t,\mathbf{z}))$. The default is CLTYPE=LOG.

**LOGLOG**

specifies that the confidence limits for the $\log(-\log(S(t,\mathbf{z})))$ be computed using normal theory approximation. The confidence limits for $S(t,\mathbf{z})$ are obtained by back-transforming the confidence limits for $\log(-\log(S(t,\mathbf{z})))$.

**NORMAL**

specifies that the confidence limits for $S(t, \mathbf{z})$ be computed directly using normal theory approximation.

**GROUP=***variable*

names a numeric variable in the COVARIATES= data set to group the baseline function curves for the observations into separate plots. This option has no effect if the PLOTS= option in the PROC PHREG statement is not specified. Curves for the covariate sets with the same value of the GROUP= variable are overlaid in the same plot.

**METHOD=***method*

specifies the method used to compute the survivor function estimates. The two available methods are as follows:

**BRESLOW**

**CH**

**EMP**

**NELSON**

specifies that the Breslow (1972) method be used to compute the survivor function—that is, that the survivor function be estimated by exponentiating the negative empirical cumulative hazard function.

**PL**

specifies that the product-limit estimate of the survivor function be computed.

The default is METHOD=BRESLOW.

**ROWID=***variable*

**ID=***variable*

**ROW=***variable*

names a variable in the COVARIATES= data set for identifying the baseline function curves in the plots. This option has no effect if the PLOTS= option in the PROC PHREG statement is not specified. Values of this variable are used to label the curves for the corresponding rows in the COVARIATES= data set. You can specify ROWID=_OBS_ to use the observation numbers in the COVARIATES= data set for identification.

For recurrent events data, both CMF= and CUMHAZ= statistics are the Nelson estimators, but their standard error are not the same. Confidence limits for the cumulative mean function and cumulative hazard function are based on the log transform.

## BAYES Statement

**BAYES** < *options* > **;**

The BAYES statement requests a Bayesian analysis of the regression model by using Gibbs sampling. The Bayesian posterior samples (also known as the chain) for the regression parameters can be output to a SAS data set. Table 64.2 summarizes the options available in the BAYES statement.

**Table 64.2** BAYES Statement Options

| Option | Description |
|---|---|
| **Monte Carlo Options** | |
| INITIAL= | specifies initial values of the chain |
| NBI= | specifies the number of burn-in iterations |
| NMC= | specifies the number of iterations after burn-in |
| SEED= | specifies the random number generator seed |
| THINNING= | controls the thinning of the Markov chain |
| **Model and Prior Options** | |
| COEFFPRIOR= | specifies the prior of the regression coefficients |
| PIECEWISE= | specifies details of the piecewise exponential model |
| **Summaries and Diagnostics of the Posterior Samples** | |
| DIAGNOSTICS= | displays convergence diagnostics |
| PLOTS= | displays diagnostic plots |
| STATISTICS= | displays summary statistics |
| **Posterior Samples** | |
| OUTPOST= | names a SAS data set for the posterior samples |

The following list describes these options and their suboptions.

**COEFFPRIOR=UNIFORM | NORMAL < (*normal-option*) >**

**CPRIOR=UNIFORM | NORMAL < (*option*) >**

**COEFF=UNIFORM | NORMAL < (*option*) >**

specifies the prior distribution for the regression coefficients. The default is COEFFPRIOR=UNIFORM. The available prior distributions are as follows:

**UNIFORM**

specifies a flat prior—that is, the prior that is proportional to a constant ($p(\beta_1, \ldots, \beta_k) \propto 1$ for all $-\infty < \beta_i < \infty$).

**NORMAL< (*normal-option*) >**

specifies a normal distribution. The *normal-options* include the following:

**INPUT=*SAS-data-set***

specifies a SAS data set containing the mean and covariance information of the normal prior. The data set must contain the _TYPE_ variable to identify the observation type, and it must contain a variable to represent each regression coefficient. If the data set also contains the _NAME_ variable, values of this variable are used to identify the covariances for the _TYPE_='COV' observations; otherwise, the _TYPE_='COV' observations are assumed to be in the same order as the explanatory variables in the MODEL statement. PROC PHREG reads the mean vector from the observation with _TYPE_='MEAN' and the covariance matrix from observations with _TYPE_='COV'. For an independent normal prior, the variances can be specified with _TYPE_='VAR'; alternatively, the precisions (inverse of the variances) can be specified with _TYPE_='PRECISION'.

**RELVAR** **<** **=***c* **>**

specifies a normal prior $N(\mathbf{0}, c\mathbf{J})$, where $\mathbf{J}$ is a diagonal matrix with diagonal elements equal to the variances of the corresponding ML estimator. By default, *c*=$10^6$.

**VAR=***c*

specifies the normal prior $N(\mathbf{0}, c\mathbf{I})$, where $\mathbf{I}$ is the identity matrix.

If you do not specify a *normal-option*, the normal prior $N(\mathbf{0}, 10^6\mathbf{I})$, where $\mathbf{I}$ is the identity matrix, is used. See the section "Normal Prior" on page 4619 for details.

**DIAGNOSTICS=ALL | NONE |** *keyword* **|** *(keyword-list)*

**DIAG=ALL | NONE |** *keyword* **|** *(keyword-list)*

controls the number of diagnostics produced. You can request all the diagnostics in the following list by specifying DIAGNOSTICS=ALL. If you do not want any of these diagnostics, you specify DIAGNOSTICS=NONE. If you want some but not all of the diagnostics, or if you want to change certain settings of these diagnostics, you specify a subset of the following keywords. The default is DIAGNOSTICS=(AUTOCORR GEWEKE ESS).

**AUTOCORR** **<** **(LAGS=** *numeric-list***)** **>**

computes the autocorrelations of lags given by LAGS= list for each parameter. Elements in the list are truncated to integers and repeated values are removed. If the LAGS= option is not specified, autocorrections of lags 1, 5, 10, and 50 are computed for each variable. See the section "Autocorrelations" on page 169 for details.

**ESS**

computes the effective sample size of Kass et al. (1998), the correlation time, and the efficiency of the chain for each parameter. See the section "Effective Sample Size" on page 169 for details.

**MCSE**

**MCERROR**

computes the Monte Carlo standard error for each parameter. The Monte Caro standard error, which measures the simulation accuracy, is the standard error of the posterior mean estimate and is calculated as the posterior standard deviation divided by the square root of the effective sample size. See the section "Standard Error of the Mean Estimate" on page 170 for details.

**HEIDELBERGER** **<** *(heidel-options)* **>**

computes the Heidelberger and Welch tests for each parameter. See the section "Heidelberger and Welch Diagnostics" on page 165 for details. The tests consist of a stationary test and a halfwidth test. The former tests the null hypothesis that the sample values form a stationary process. If the stationarity test is passed, a halfwidth test is then carried out. Optionally, you can specify one or more of the following *heidel-options*:

**SALPHA=***value*

specifies the $\alpha$ level ($0 < \alpha < 1$) for the stationarity test. The default is the value of the ALPHA= option in the PROC PHREG statement, or 0.05 if that option is not specified.

**HALPHA=***value*

specifies the $\alpha$ level ($0 < \alpha < 1$) for the halfwidth test. The default is the value of the ALPHA= option in the PROC PHREG statement, or 0.05 if that option is not specified.

**EPS=***value*

specifies a small positive number $\epsilon$ such that if the halfwidth is less than $\epsilon$ times the sample mean of the retaining samples, the halfwidth test is passed.

**GELMAN** < **=(***gelman-options***)** >

computes the Gelman and Rubin convergence diagnostics. See the section “Gelman and Rubin Diagnostics” on page 161 for details. You can specify one or more of the following *gelman-options*:

**NCHAIN=***number*

**N=***number*

specifies the number of parallel chains used to compute the diagnostic and has to be 2 or larger. The default is NCHAIN=3. The NCHAIN= option is ignored when the INITIAL= option is specified in the BAYES statement, and in such a case, the number of parallel chains is determined by the number of valid observations in the INITIAL= data set.

**ALPHA=***value*

specifies the significance level for the upper bound. The default is the value of the ALPHA= option in the PROC PHREG statement, or 0.05 if that option is not specified (resulting in a 97.5% bound).

.

**GEWEKE** < **=***geweke-options* >

computes the Geweke diagnostics. See the section “Geweke Diagnostics” on page 163 for details. The diagnostic is essentially a two-sample $t$-test between the first $f_1$ portion and the last $f_2$ portion of the chain. The default is $f_1$=0.1 and $f_2$=0.5, but you can choose other fractions by using the following *geweke-options*:

**FRAC1=***value*

specifies the early $f_1$ fraction of the Markov chain.

**FRAC2=***value*

specifies the latter $f_2$ fraction of the Markov chain.

**RAFTERY** < **(***raftery-options***)** >

computes the Raftery and Lewis diagnostics. See the section “Raftery and Lewis Diagnostics” on page 166 for details. The diagnostic evaluates the accuracy of the estimated quantile ($\hat{\theta}_Q$ for a given $Q \in (0,1)$) of a chain. $\hat{\theta}_Q$ can achieve any degree of accuracy when the chain is allowed to run for a long time. A stopping criterion is when the estimated probability $\hat{P}_Q = \Pr(\theta \le \hat{\theta}_Q)$ reaches within $\pm R$ of the value $Q$ with probability $S$; that is, $\Pr(Q - R \le \hat{P}_Q \le Q + R) = S$. The following *raftery-options* enable you to specify $Q$, $R$, $S$, and a precision level $\epsilon$ for a stationary test.

**QUANTILE=***value*

**Q=***value*

specifies the order (a value between 0 and 1) of the quantile of interest. The default is 0.025.

**ACCURACY=***value*

**R=***value*

specifies a small positive number as the margin of error for measuring the accuracy of estimation of the quantile. The default is 0.005.

**PROBABILITY=***value*

**P=***value*

specifies the probability of attaining the accuracy of the estimation of the quantile. The default is 0.95.

**EPSILON=***value*

**EPS=***value*

specifies the tolerance level (a small positive number) for the test. The default is 0.001.

**INITIAL=***SAS-data-set*

specifies the SAS data set that contains the initial values of the Markov chains. The INITIAL= data set must contain a variable for each parameter in the model. You can specify multiple rows as the initial values of the parallel chains for the Gelman-Rubin statistics, but posterior summary statistics, diagnostics, and plots are computed only for the first chain.

**NBI=***number*

specifies the number of burn-in iterations before the chains are saved. The default is 2000.

**NMC=***number*

specifies the number of iterations after the burn-in. The default is 10000.

**OUTPOST=***SAS-data-set*

**OUT=***SAS-data-set*

names the SAS data set that contains the posterior samples. See the section “OUTPOST= Output Data Set in the BAYES Statement” on page 4626 for more information. Alternatively, you can output the posterior samples into a data set, as shown in the following example in which the data set is named PostSamp.

```
ODS OUTPUT PosteriorSample = PostSamp;
```

**PIECEWISE < =***keyword* **< (< NINTERVAL=***number* **> < INTERVAL=(***numeric-list***) > < PRIOR=***option* **>) > >**

specifies that the piecewise constant baseline hazard model be used in the Bayesian analysis. You can specify one of the following two *keywords*:

**HAZARD**

models the baseline hazard parameters in the original scale. The hazard parameters are named Lambda1, Lambda2, . . ., and so on.

**LOGHAZARD**

models the baseline hazard parameters in the log scale. The log-hazard parameters are named Alpha1, Alpha2, . . ., and so on.

Specifying PIECEWISE by itself is the same as specifying PIECEWISE=LOGHAZARD.

You can choose one of the following two options to specify the partition of the time axis into intervals of constant baseline hazards:

**NINTERVAL=*number***

**N=*number***

specifies the number of intervals with constant baseline hazard rates. PROC PHREG partitions the time axis into the given number of intervals with approximately equal number of events in each interval.

**INTERVAL=(*numeric-list*)**

specifies the list of numbers that partition the time axis into disjoint intervals with constant baseline hazard in each interval. For example, INTERVAL=(100, 150, 200, 250, 300) specifies a model with a constant hazard in the intervals [0,100), [100,150), [150,200), [200,250), [250,300), and [300,$\infty$). Each interval must contain at least one event; otherwise, the posterior distribution can be improper, and inferences cannot be derived from an improper posterior distribution.

If neither NINTERVAL= nor INTERVAL= is specified, the default is NINTERVAL=8.

To specify the prior for the baseline hazards $(\lambda_1, \ldots, \lambda_J)$ in the original scale, you specify the following:

**PRIOR = IMPROPER | UNIFORM | GAMMA< (*gamma-option*) > | ARGAMMA< (*argamma-option*) >**

The default is PRIOR=IMPROPER. The available prior options include the following:

**IMPROPER**

specifies the noninformative and improper prior $p(\lambda_1, \ldots, \lambda_J) \propto \prod_i \lambda_i{}^{-1}$ for all $\lambda_i > 0$.

**UNIFORM**

specifies a uniform prior on the real line; that is, $p(\lambda_i) \propto 1$ for all $\lambda_i > 0$.

**GAMMA < (*gamma-option*) >**

specifies an independent gamma prior $G(a, b)$ with density $f(t) = \frac{b(bt)^{a-1}e^{-bt}}{\Gamma(a)}$, which can be followed by one of the following *gamma-options* enclosed in parentheses. The hyperparameters $a$ and $b$ are the shape and inverse-scale parameters of the gamma distribution, respectively. See the section "Independent Gamma Prior" on page 4618 for details. The default is $G(10^{-4}, 10^{-4})$ for each $\lambda_j$, setting the prior mean to 1 with variance $10^4$. This prior is proper and reasonably noninformative.

**INPUT=*SAS-data-set***

specifies a data set containing the hyperparameters of the independent gamma prior. The data set must contain the _TYPE_ variable to identify the observation type, and it must contain the variables named Lambda1, Lambda2, . . . , and so forth, to represent the hazard parameters. The observation with _TYPE_='SHAPE' identifies the shape parameters, and the observation with _TYPE_='ISCALE' identifies the inverse-scale parameters.

**RELSHAPE< =*c*>**

specifies independent $G(c\hat{\lambda}_j, c)$ distribution, where $\hat{\lambda}_j$'s are the MLEs of the hazard rates. This prior has mean $\hat{\lambda}_j$ and variance $\frac{\hat{\lambda}_j}{c}$. By default, *c*=$10^{-4}$.

**SHAPE=*a* and SCALE=*b***

together specify the Gamma$(a, b)$ prior.

**SHAPE=*c***

**ISCALE=*c***

specifies the $G(c, c)$ prior.

**ARGAMMA < (*argamma-option*) >**

specifies an autoregressive gamma prior of order 1, which can be followed by one of the following *argamma-options*. See the section "AR1 Prior" on page 4618 for details.

**INPUT=*SAS-data-set***

specifies a data set containing the hyperparameters of the correlated gamma prior. The data set must contain the _TYPE_ variable to identify the observation type, and it must contain the variables named Lambda1, Lambda2, . . . , and so forth, to represent the hazard parameters. The observation with _TYPE_='SHAPE' identifies the shape parameters, and the observation with _TYPE_='ISCALE' identifies the *relative* inverse-scale parameters; that is, if $a_j$ and $b_j$ are, respectively, the SHAPE and ISCALE values for $\lambda_j, 1 \leq j \leq J$, then $\lambda_1 \sim G(a_1, b_1)$, and $\lambda_j \sim G(a_j, b_j/\lambda_{j-1})$ for $2 \leq j \leq J$.

**SHAPE=*a* and SCALE=*b***

together specify that $\lambda_1 \sim G(a, b)$ and $\lambda_j \sim G(a, b/\lambda_{j-1})$ for $2 \leq j \leq J$.

**SHAPE=*c***

**ISCALE=*c***

specifies that $\lambda_1 \sim G(c, c)$ and $\lambda_j \sim G(c, c/\lambda_{j-1})$ for $2 \leq j \leq J$.

To specify the prior for $\alpha_1, \ldots, \alpha_J$, the hazard parameters in the log scale, you specifying the following:

**PRIOR=UNIFORM | NORMAL< (*normal-option*) >**

The default is PRIOR=UNIFORM. The available prior options are as follows:

**UNIFORM**

specifies the uniform prior on the real line; that is, $\alpha_i \propto 1$ for all $-\infty < \alpha_i < \infty$.

**NORMAL< (*normal-option*) >**

specifies a normal prior distribution on the log-hazard parameters. The *normal options* include the following. If you do not specify an option, the normal prior $N(\mathbf{0}, 10^6\mathbf{I})$, where $\mathbf{I}$ is the identity matrix, is used.

**INPUT=*SAS-data-set***

specifies a SAS data set containing the mean and covariance information of the normal prior. The data set must contain the _TYPE_ variable to identify the observation type, and it must contain variables named Alpha1, Alpha2, ..., and so forth, to represent the log-hazard parameters. If the data set also contains the _NAME_ variable, the value of this variable will be used to identify the covariances for the _TYPE_='COV' observations; otherwise, the _TYPE_='COV' observations are assumed to be in the same order as the explanatory variables in the MODEL statement. PROC PHREG reads the mean vector from the observation with _TYPE_='MEAN' and the covariance matrix from observations with _TYPE_='COV'. See the section "Normal Prior" on page 4619 for details. For an independent normal prior, the variances can be specified with _TYPE_='VAR'; alternatively, the precisions (inverse of the variances) can be specified with _TYPE_='PRECISION'.

If you have a joint normal prior for the log-hazard parameters and the regression coefficients, specify the same data set containing the mean and covariance information of the multivariate normal distribution in both the COEFFPRIOR=NORMAL(INPUT=) and the PIECEWISE=LOGHAZARD(PRIOR=NORMAL(INPUT=)) options. See the section "Joint Multivariate Normal Prior for Log-Hazards and Regression Coefficients" on page 4619 for details.

**RELVAR < =*c* >**

specifies the normal prior $N(\mathbf{0}, c\mathbf{J})$, where $\mathbf{J}$ is a diagonal matrix with diagonal elements equal to the variances of the corresponding ML estimator. By default, $c$=$10^6$.

**VAR=*c***

specifies the normal prior $N(\mathbf{0}, c\mathbf{I})$, where $\mathbf{I}$ is the identity matrix.

**PLOTS < (*global-plot-options*) > = *plot-request***

**PLOTS < (*global-plot-options*) > = (*plot-requests*)**

controls the diagnostic plots produced through ODS Graphics. Three types of plots can be requested: trace plots, autocorrelation function plots, and kernel density plots. By default, the plots are displayed in panels unless the global plot option UNPACK is specified. If you specify more than one type of plots, the plots are displayed by parameters unless the global plot option GROUPBY=TYPE is specified. When you specify only one plot request, you can omit the parentheses around the plot request. For example:

```
plots=none
plots(unpack)=trace
plots=(trace autocorr)
```

You must enable ODS Graphics before requesting plots, for example, like this:

```
ods graphics on;

proc phreg;
   model y=x;
   bayes plots=trace;
   run;
end;

ods graphics off;
```

If you have enabled ODS Graphics but do not specify the PLOTS= option in the BAYES statement, then PROC PHREG produces, for each parameter, a panel containing the trace plot, the autocorrelation function plot, and the density plot. This is equivalent to specifying plots=(trace autocorr density).

The global plot options include the following:

**FRINGE**
creates a fringe plot on the X axis of the density plot.

**GROUPBY = PARAMETER | TYPE**
specifies how the plots are to be grouped when there is more than one type of plots. The choices are as follows:

**TYPE**
specifies that the plots be grouped by type.

**PARAMETER**
specifies that the plots be grouped by parameter.

GROUPBY=PARAMETER is the default.

**SMOOTH**
displays a fitted penalized B-spline curve each trace plot.

**UNPACKPANEL**
**UNPACK**
specifies that all paneled plots be unpacked, meaning that each plot in a panel is displayed separately.

The plot requests include the following:

**ALL**
specifies all types of plots. PLOTS=ALL is equivalent to specifying PLOTS=(TRACE AUTOCORR DENSITY).

**AUTOCORR**
displays the autocorrelation function plots for the parameters.

**DENSITY**
displays the kernel density plots for the parameters.

**NONE**
suppresses all diagnostic plots.

**TRACE**
displays the trace plots for the parameters. See the section "Visual Analysis via Trace Plots" on page 156 for details.

Consider a model with four parameters, X1–X4. Displays for various specification are depicted as follows.

1. PLOTS=(TRACE AUTOCORR) displays the trace and autocorrelation plots for each parameter side by side with two parameters per panel:

| | | |
|---|---|---|
| Display 1 | Trace(X1) | Autocorr(X1) |
| | Trace(X2) | Autocorr(X2) |
| Display 2 | Trace(X3) | Autocorr(X3) |
| | Trace(X4) | Autocorr(X4) |

2. PLOTS(GROUPBY=TYPE)=(TRACE AUTOCORR) displays all the paneled trace plots, followed by panels of autocorrelation plots:

| | | |
|---|---|---|
| Display 1 | Trace(X1) | |
| | Trace(X2) | |
| Display 2 | Trace(X3) | |
| | Trace(X4) | |
| Display 3 | Autocorr(X1) | Autocorr(X2) |
| | Autocorr(X3) | Autocorr(X4) |

3. PLOTS(UNPACK)=(TRACE AUTOCORR) displays a separate trace plot and a separate correlation plot, parameter by parameter:

| | |
|---|---|
| Display 1 | Trace(X1) |
| Display 2 | Autocorr(X1) |
| Display 3 | Trace(X2) |
| Display 4 | Autocorr(X2) |
| Display 5 | Trace(X3) |
| Display 6 | Autocorr(X3) |
| Display 7 | Trace(X4) |
| Display 8 | Autocorr(X4) |

4. PLOTS(UNPACK GROUPBY=TYPE) = (TRACE AUTOCORR) displays all the separate trace plots followed by the separate autocorrelation plots:

| | |
|---|---|
| Display 1 | Trace(X1) |
| Display 2 | Trace(X2) |
| Display 3 | Trace(X3) |
| Display 4 | Trace(X4) |
| Display 5 | Autocorr(X1) |
| Display 6 | Autocorr(X2) |
| Display 7 | Autocorr(X3) |
| Display 8 | Autocorr(X4) |

**SEED=***number*

specifies an integer seed ranging from 1 to $2^{31}$–1 for the random number generator in the simulation. Specifying a seed enables you to reproduce identical Markov chains for the same specification. If the SEED= option is not specified, or if you specify a nonpositive seed, a random seed is derived from the time of day.

**STATISTICS < (***global-options***) > = ALL | NONE |** *keyword* **|** *(keyword-list)*

**STATS < (***global-statoptions***) > = ALL | NONE |** *keyword* **|** *(keyword-list)*

controls the number of posterior statistics produced. Specifying STATISTICS=ALL is equivalent to specifying STATISTICS=(SUMMARY INTERVAL COV CORR). If you do not want any posterior statistics, you specify STATISTICS=NONE. The default is STATISTICS=(SUMMARY INTERVAL). See the section "Summary Statistics" on page 170 for details. The *global-options* include the following:

**ALPHA=***numeric-list*

controls the probabilities of the credible intervals. The ALPHA= values must be between 0 and 1. Each ALPHA= value produces a pair of 100(1–ALPHA)% equal-tail and HPD intervals for each parameters. The default is the value of the ALPHA= option in the PROC PHREG statement, or 0.05 if that option is not specified (yielding the 95% credible intervals for each parameter).

**PERCENT=***numeric-list*

requests the percentile points of the posterior samples. The PERCENT= values must be between 0 and 100. The default is PERCENT= 25, 50, 75, which yield the 25th, 50th, and 75th percentile points for each parameter.

The list of *keywords* includes the following:

**CORR**

produces the posterior correlation matrix.

**COV**

produces the posterior covariance matrix.

**SUMMARY**

produces the means, standard deviations, and percentile points for the posterior samples. The default is to produce the 25th, 50th, and 75th percentile points, but you can use the global PERCENT= option to request specific percentile points.

**INTERVAL**

produces equal-tail credible intervals and HPD intervals. The defult is to produce the 95% equal-tail credible intervals and 95% HPD intervals, but you can use the global ALPHA= option to request intervals of any probabilities.

**THINNING=***number*

**THIN=***number*

controls the thinning of the Markov chain. Only one in every $k$ samples is used when THINNING=$k$, and if NBI=$n_0$ and NMC=$n$, the number of samples kept is

$$\left[\frac{n_0+n}{k}\right]-\left[\frac{n_0}{k}\right]$$

where $[a]$ represents the integer part of the number $a$. The default is THINNING=1.

## BY Statement

**BY** *variables* **;**

You can specify a BY statement with PROC PHREG to obtain separate analyses on observations in groups defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. The *variables* are one or more variables in the input data set.

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

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

For more information about the BY statement, see *SAS Language Reference: Concepts*. For more information about the DATASETS procedure, see the *Base SAS Procedures Guide*.

## CLASS Statement

**CLASS** *variable* < *(options)* > < *variable* < *(options)* > . . . > < */ options* > **;**

The CLASS statement names the categorical variables to be used in the analysis. The CLASS statement must precede the MODEL statement. You can specify various *options* for each variable by enclosing them in parentheses after the variable name. You can also specify global *options* for the CLASS statement by placing them after a slash (/). Global *options* are applied to all the variables specified in the CLASS statement. If you specify more than one CLASS statement, the global *options* specified in any one CLASS statement apply to all CLASS statements. However, individual CLASS variable *options* override the global *options*.

**CPREFIX=** *n*

specifies that, at most, the first *n* characters of a CLASS variable name be used in creating names for the corresponding dummy variables. The default is $32 - \min(32, \max(2, f))$, where $f$ is the formatted length of the CLASS variable.

**DESCENDING**

**DESC**

reverses the sorting order of the categorical variable.

**LPREFIX=** *n*

specifies that, at most, the first *n* characters of a CLASS variable label be used in creating labels for the corresponding dummy variables.

**MISSING**

allows a missing value (for example,'.' for a numeric variable and blanks for a character variable) as a valid value for the CLASS variable.

**ORDER=DATA | FORMATTED | FREQ | INTERNAL**

specifies the sorting order for the categories of cateogrical variables. This ordering determines which parameters in the model correspond to each level in the data, so the ORDER= option can be useful when you use the CONTRAST statement. When the default ORDER=FORMATTED is in effect for numeric variables for which you have supplied no explicit format, the levels are ordered by their internal values. The following table shows how PROC PHREG interprets values of the ORDER= option.

| **Value of ORDER=** | **Levels Sorted By** |
|---|---|
| DATA | order of appearance in the input data set |
| FORMATTED | external formatted value, except for numeric variables with no explicit format, which are sorted by their unformatted (internal) value |
| FREQ | descending frequency count; levels with the most observations come first in the order |
| INTERNAL | unformatted value |

By default, ORDER=FORMATTED. For FORMATTED and INTERNAL, the sort order is machine dependent. For more information about sorting order, see the chapter on the SORT procedure in the *Base SAS Procedures Guide* and the discussion of BY-group processing in *SAS Language Reference: Concepts*.

**PARAM=***keyword*

specifies the parameterization method for the categorical variable or variables. Design matrix columns are created from CLASS variables according to the following coding schemes. The default is PARAM=REF. If PARAM=ORTHPOLY or PARAM=POLY, and the CLASS levels are numeric, then the ORDER= option in the CLASS statement is ignored, and the internal, unformatted values are used. See the section "CLASS Variable Parameterization" on page 4574 for further details.

| | |
|---|---|
| EFFECT | specifies effect coding. |
| GLM | specifies less-than-full-rank, reference-cell coding; this option can be used only as a global option. |
| ORDINAL | specifies the cumulative parameterization for an ordinal CLASS variable. |
| POLYNOMIAL<br>POLY | specifies polynomial coding. |
| REFERENCE<br>REF | specifies reference cell coding. |
| ORTHEFFECT | orthogonalizes PARAM=EFFECT. |
| ORTHORDINAL | orthogonalizes PARAM=ORDINAL. |
| ORTHPOLY | orthogonalizes PARAM=POLYNOMIAL. |
| ORTHREF | orthogonalizes PARAM=REFERENCE. |

The EFFECT, POLYNOMIAL, REFERENCE, and ORDINAL coding schemes and their orthogonal parameterizations are full-rank parameterization. The REF= option in the CLASS statement determines the reference level for the EFFECT, REFERENCE, and their orthogonal parameterizations.

Parameter names for a CLASS predictor variable are constructed by concatenating the CLASS variable name with the CLASS levels. However, for the POLYNOMIAL and orthogonal parameterizations, parameter names are formed by concatenating the CLASS variable name and keywords that reflect the parameterization.

**REF=***'level' | keyword*

specifies the reference level for PARAM=EFFECT or PARAM=REF. For an individual variable, you can specify a specific *level* of the variable in the REF= option. For a global or individual variable REF= *option*, you can use one of the following *keywords*. The default is REF=LAST.

| | |
|---|---|
| FIRST | designates the first ordered level as reference. |
| LAST | designates the last ordered level as reference. |

**TRUNCATE**

specifies that class levels should be determined using no more than the first 16 characters of the formatted values of CLASS variables. This is a global option, not an individual CLASS variable option.

## CONTRAST Statement

**CONTRAST** *'label' row-description* < *,. . . row-description* >< */options* > **;**

where a *row-description* is: *effect values* <,. . . *effect values*>

The CONTRAST statement provides a mechanism for obtaining customized hypothesis tests. It is similar to the CONTRAST statement in PROC GLM and PROC CATMOD, depending on the coding schemes used with any categorical variables involved.

The CONTRAST statement enables you to specify a matrix, $\mathbf{L}$, for testing the hypothesis $\mathbf{L}\boldsymbol{\beta} = \mathbf{0}$. You must be familiar with the details of the model parameterization that PROC PHREG uses (for more information, see the PARAM= option in the section "CLASS Statement" on page 4551). Optionally, the CONTRAST statement enables you to estimate each row, $\boldsymbol{l}_i'\boldsymbol{\beta}$, of $\mathbf{L}\boldsymbol{\beta}$ and test the hypothesis $\boldsymbol{l}_i'\boldsymbol{\beta} = 0$. Computed statistics are based on the asymptotic chi-square distribution of the Wald statistic.

There is no limit to the number of CONTRAST statements that you can specify, but they must appear after the MODEL statement.

The following parameters are specified in the CONTRAST statement:

*label* identifies the contrast on the output. A label is required for every contrast specified, and it must be enclosed in quotes.

*effect* identifies an effect that appears in the MODEL statement. You do not need to include all effects that are included in the MODEL statement.

*values* are constants that are elements of the $\mathbf{L}$ matrix associated with the effect. To correctly specify your contrast, it is crucial to know the ordering of parameters within each effect and the variable levels associated with any parameter. The "Class Level Information" table shows the ordering of levels within variables. The E option, described later in this section, enables you to verify the proper correspondence of *values* to parameters.

The rows of $\mathbf{L}$ are specified in order and are separated by commas. Multiple degree-of-freedom hypotheses can be tested by specifying multiple *row-descriptions*. For any of the full-rank parameterizations, if an effect is not specified in the CONTRAST statement, all of its coefficients in the $\mathbf{L}$ matrix are set to 0. If too many values are specified for an effect, the extra ones are ignored. If too few values are specified, the remaining ones are set to 0.

When you use effect coding (by specifying PARAM=EFFECT in the CLASS statement), all parameters are directly estimable (involve no other parameters). For example, suppose an effect coded CLASS variable `A` has four levels. Then there are three parameters ($\alpha_1, \alpha_2, \alpha_3$) representing the first three levels, and the fourth parameter is represented by

$$-\alpha_1 - \alpha_2 - \alpha_3$$

To test the first versus the fourth level of A, you would test

$$\alpha_1 = -\alpha_1 - \alpha_2 - \alpha_3$$

or, equivalently,

$$2\alpha_1 + \alpha_2 + \alpha_3 = 0$$

which, in the form $\mathbf{L}\boldsymbol{\beta} = 0$, is

$$\begin{bmatrix} 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = 0$$

Therefore, you would use the following CONTRAST statement:

```
contrast '1 vs. 4' A 2 1 1;
```

To contrast the third level with the average of the first two levels, you would test

$$\frac{\alpha_1 + \alpha_2}{2} = \alpha_3$$

or, equivalently,

$$\alpha_1 + \alpha_2 - 2\alpha_3 = 0$$

Therefore, you would use the following CONTRAST statement:

```
contrast '1&2 vs. 3' A 1 1 -2;
```

Other CONTRAST statements involving classification variables with PARAM=EFFECT are constructed similarly. For example:

```
contrast '1 vs. 2    '  A  1 -1  0;
contrast '1&2 vs. 4  '  A  3  3  2;
contrast '1&2 vs. 3&4'  A  2  2  0;
contrast 'Main Effect'  A  1  0  0,
                        A  0  1  0,
                        A  0  0  1;
```

When you use the less-than-full-rank parameterization (by specifying PARAM=GLM in the CLASS statement), each row is checked for estimability. If PROC PHREG finds a contrast to be nonestimable, it displays missing values in corresponding rows in the results. PROC PHREG handles missing level combinations of categorical variables in the same manner as PROC GLM. Parameters corresponding to missing level combinations are not included in the model. This convention can affect the way in which you specify the **L** matrix in your CONTRAST statement. If the elements of **L** are not specified for an effect that contains a specified effect, then the elements of the specified effect are distributed over the levels of the higher-order effect just as the GLM procedure does for its CONTRAST and ESTIMATE statements. For example, suppose that the model contains effects A and B and their interaction A*B. If you specify a CONTRAST statement involving A alone, the **L** matrix contains nonzero terms for both A and A*B, since A*B contains A.

The Cox model contains no explicit intercept parameter, so it is not valid to specify one in the CONTRAST statement. As a consequence, you can test or estimate only homogeneous linear combinations (those with zero-intercept coefficients, such as contrasts that represent group differences) for the GLM parameterization.

The degrees of freedom are the number of linearly independent constraints implied by the CONTRAST statement—that is, the rank of $\mathbf{L}$.

You can specify the following options after a slash (/).

**ALPHA=** *p*

specifies the level of significance $p$ for the $100(1-p)\%$ confidence interval for each contrast when the ESTIMATE option is specified. The value $p$ must be between 0 and 1. By default, $p$ is equal to the value of the ALPHA= option in the PROC PHREG statement, or 0.05 if that option is not specified.

**E**

requests that the $\mathbf{L}$ matrix be displayed.

**ESTIMATE=***keyword*

requests that each individual contrast (that is, each row, $l_i'\boldsymbol{\beta}$, of $\mathbf{L}\boldsymbol{\beta}$) or exponentiated contrast ($e^{l_i'\boldsymbol{\beta}}$) be estimated and tested. PROC PHREG displays the point estimate, its standard error, a Wald confidence interval, and a Wald chi-square test for each contrast. The significance level of the confidence interval is controlled by the ALPHA= option. You can estimate the contrast or the exponentiated contrast ($e^{l_i'\boldsymbol{\beta}}$), or both, by specifying one of the following *keywords*:

| | |
|---|---|
| PARM | specifies that the contrast itself be estimated. |
| EXP | specifies that the exponentiated contrast be estimated. |
| BOTH | specifies that both the contrast and the exponentiated contrast be estimated. |

**SINGULAR=***number*

tunes the estimability check. This option is ignored when the full-rank parameterization is used. If $\boldsymbol{v}$ is a vector, define ABS($\boldsymbol{v}$) to be the largest absolute value of the elements of $\boldsymbol{v}$. For a row vector $\boldsymbol{l}'$ of the contrast matrix $\mathbf{L}$, define $c$ to be equal to ABS($\boldsymbol{l}$) if ABS($\boldsymbol{l}$) is greater than 0; otherwise, $c$ equals 1. If ABS($\boldsymbol{l}' - \boldsymbol{l}'\mathbf{T}$) is greater than $c * number$, then $\boldsymbol{l}$ is declared nonestimable. The $\mathbf{T}$ matrix is the Hermite form matrix $\mathbf{I}_0^-\mathbf{I}_0$, where $\mathbf{I}_0^-$ represents a generalized inverse of the information matrix $\mathbf{I}_0$ of the null model. The value for $number$ must be between 0 and 1; the default value is 1E–4.

## FREQ Statement

**FREQ** *variable* < */option* > **;**

The *variable* in the FREQ statement identifies the variable (in the input data set) containing the frequency of occurrence of each observation. PROC PHREG treats each observation as if it appears $n$ times, where $n$ is the value of the FREQ variable for the observation. If not an integer, the frequency value is truncated to an integer. If the frequency value is missing, the observation is not used in the estimation of the regression parameters.

The following option can be specified in the FREQ statement after a slash (/):

**NOTRUNCATE**

**NOTRUNC**

specifies that frequency values are not truncated to integers.

## HAZARDRATIO Statement

**HAZARDRATIO** < *'label'*> *variable* < / *options*> **;**

The HAZARDRATIO statement enables you to request hazard ratios for any variable in the model at customized settings. For example, if the model contains the interaction of a CLASS variable A and a continuous variable X, the following specification displays a table of hazard ratios comparing the hazards of each pair of levels of A at X=3:

```
hazardratio A / at (X=3) diff=ALL;
```

The HAZARDRATIO statement identifies the variable whose hazard ratios are to be evaluated. If the variable is a continuous variable, the hazard ratio compares the hazards for a given change (by default, a increase of 1 unit) in the variable. For a CLASS variable, a hazard ratio compares the hazards of two levels of the variable. More than one HAZARDRATIO statement can be specified, and an optional label (specified as a quoted string) helps identify the output.

Options for the HAZARDRATIO statement are as follows.

**ALPHA=***number*

specifies the alpha level of the interval estimates for the hazard ratios. The value must be between 0 and 1. The default is the value of the ALPHA= option in the PROC PHREG statement, or 0.05 if that option is not specified.

**AT (***variable***=ALL | REF |** *list* **< . . .** *variable***=ALL | REF|** *list* **> )**

specifies the variables that interact with the variable of interest and the corresponding values of the interacting variables. If the interacting variable is continuous and a numeric list is specified after the equal sign, hazard ratios are computed for each value in the list. If the interacting variable is a CLASS variable, you can specify, after the equal sign, a list of quoted strings corresponding to various levels of the CLASS variable, or you can specify the keyword ALL or REF. Hazard ratios are computed at each value of the list if the list is specified, or at each level of the interacting variable if ALL is specified, or at the reference level of the interacting variable if REF is specified.

If this option is not specified, PROC PHREG finds all the variables that interact with the variable of interest. If an interacting *variable* is a CLASS variable, *variable*= ALL is the default; if the interacting *variable* is continuous, *variable*=$m$ is the default, where $m$ is the average of all the sampled values of the continuous *variable*.

Suppose the model contains two interactions: an interaction A*B of CLASS variables A and B, and another interaction A*X of A with a continuous variable X. If 3.5 is the average of the sampled values of X, the following two HAZARDRATIO statements are equivalent:

```
hazardratio A;
hazardratio A / at (B=ALL X=3.5);
```

**CL=WALD | PL |BOTH**

specifies whether to create the Wald or profile-likelihood confidence limits, or both for the classical analyis. By default, Wald confidence limits are produced. This option is not applicable to a Bayesian analysis.

**DIFF=ALL | REF**

specifies which differences to consider for the level comparisons of a CLASS variable. The default is DIFF=ALL. This option is ignored in the estimation of hazard ratios for a continuous variable. DIFF=ALL requests all differences, and DIFF=REF requests comparisons between the reference level and all other levels of the CLASS variable.

**E**

displays the vector $\mathbf{h}$ of linear coefficients such that $\mathbf{h}'\boldsymbol{\beta}$ is the log-hazard ratio, with $\boldsymbol{\beta}$ being the vector of regression coefficients.

**PLCONV=*value***

controls the convergence criterion for the profile-likelihood confidence limits. The quantity *value* must be a positive number, with a default value of 1E–4. The PLCONV= option has no effect if profile-likelihood confidence intervals (CL=PL) are not requested.

**PLMAXIT=*n***

specifies the maximum number of iterations to achieve the convergence of the profile-likelihood confidence limits. By default, PLMAXITER=25. If convergence is not attained in *n* iterations, the corresponding profile-likelihood confidence limit for the hazard ratio is set to missing. The PLMAXITER= option has no effect if profile-likelihood confidence intervals (CL=PL) are not requested.

**PLSINGULAR=*value***

specifies the tolerance for testing the singularity of the Hessian matrix in the computation of the profile-likelihood confidence limits. The test requires that a pivot for sweeping this matrix be at least this number times a norm of the matrix. Values of the PLSINGULAR= option must be numeric. By default, *value* is the machine epsilon times 1E7, which is approximately 1E–9. The PLSINGULAR= option has no effect if profile-likelihood confidence intervals (CL=PL) are not requested.

**UNITS=*value***

specifies the units of change in the continuous explanatory variable for which the customized hazard ratio is estimated. The default is UNITS=1. This option is ignored in the computation of the hazard ratios for a CLASS variable.

## ID Statement

**ID** *variables* **;**

The ID statement specifies additional variables for identifying observations in the input data. These variables are placed in the OUT= data set created by the OUTPUT statement. In the computation of the robust sandwich variance estimate, you can aggregate over distinct values of these ID variables.

Only variables in the input data set can be included in the ID statement.

## MODEL Statement

**MODEL** *response < *censor ( list ) > = effects < /options >* **;**

**MODEL** *(t1, t2) < *censor(list) > = effects < /options >* **;**

The MODEL statement identifies the variables to be used as the failure time variables, the optional censoring variable, and the explanatory effects, including covariates, main effects, interactions, nested effects; see the section "Specification of Effects" on page 2486 of Chapter 39, "The GLM Procedure," for more information. A note of caution: specifying the effect `T*A` in the MODEL statement, where `T` is the time variable and `A` is a CLASS variable, does not make the effect time-dependent. See the section "Clarification of the Time and CLASS Variables Usage" on page 4576 for more information.

Two forms of MODEL syntax can be specified; the first form allows one response variable, while the second form allows two variables for the counting process style of input (see the section "Counting Process Style of Input" on page 4582 for more information).

In the first MODEL statement, preceding the equal sign, is the name of the failure time variable. This can optionally be followed by an asterisk, the name of the censoring variable, and a list of censoring values (separated by blanks or commas if there is more than one) enclosed in parentheses. If the censoring variable takes on one of these values, the corresponding failure time is considered to be censored. The variables following the equal sign are the explanatory variables (sometimes called independent variables or covariates) for the model.

Instead of a single failure time variable, the second MODEL statement identifies a pair of failure time variables. Their names are enclosed in parentheses, and they signify the endpoints of a semi-closed interval $(t1, t2]$ during which the subject is at risk. If the censoring variable takes on one of the censoring values, the time $t2$ is considered to be censored.

The censoring variable and the explanatory variables must be numeric. The failure time variables must contain nonnegative values. Any observation with a negative failure time is excluded from the analysis, as is any observation with a missing value for any of the variables listed in the MODEL statement.

Table 64.3 summarizes the options available in the MODEL statement, which can be specified after a slash (/). Four convergence criteria are allowed for the maximum likelihood optimization:

ABSFCONV=, FCONV=, GCONV=, and XCONV=. If you specify more than one convergence criterion, the optimization is terminated as soon as one of the criteria is satisfied. If none of the criteria is specified, the default is GCONV=1E−8.

**Table 64.3** MODEL Statement Options

| Option | Description |
|---|---|
| **Model Specification Options** | |
| NOFIT | suppresses model fitting |
| OFFSET= | specifies offset variable |
| SELECTION= | specifies effect selection method |
| **Effect Selection Options** | |
| BEST= | controls the number of models displayed for SCORE selection |
| DETAILS | requests detailed results at each step |
| HIERARCHY= | specifies whether and how hierarchy is maintained and whether a single effect or multiple effects are allowed to enter or leave the model per step |
| INCLUDE= | specifies number of effects included in every model |
| MAXSTEP= | specifies maximum number of steps for STEPWISE selection |
| SEQUENTIAL | adds or deletes effects in sequential order |
| SLENTRY= | specifies significance level for entering effects |
| SLSTAY= | specifies significance level for removing effects |
| START= | specifies number of variables in first model |
| STOP= | specifies number of variables in final model |
| STOPRES | adds or deletes variables by residual chi-square criterion |
| **Maximum Likelihood Optimization Options** | |
| ABSFCONV= | specifies absolute function convergence criterion |
| FCONV= | specifies relative function convergence criterion |
| FIRTH | specifies Firth's penalized likelihood method |
| GCONV= | specifies relative gradient convergence criterion |
| XCONV= | specifies relative parameter convergence criterion |
| MAXITER= | specifies maximum number of iterations |
| RIDGEINIT= | specifies the initial ridging value |
| RIDGING= | specifies the technique to improve the log likelihood function when its value is worse than that of the previous step |
| SINGULAR= | specifies tolerance for testing singularity |
| **Confidence Interval Options** | |
| ALPHA= | specifies $\alpha$ for the $100(1-\alpha)\%$ confidence intervals |
| PLCONV= | specifies profile-likelihood convergence criterion |
| RISKLIMITS= | computes confidence intervals for hazard ratios |
| **Display Options** | |
| CORRB | displays correlation matrix |
| COVB | displays covariance matrix |
| ITPRINT | displays iteration history |

**Table 64.3** *continued*

| Option | Description |
|---|---|
| NODUMMYPRINT | suppresses "Class Level Information" table |
| **Miscellaneous Options** | |
| ENTRYTIME= | specifies the delayed entry time variable |
| TIES= | specifies the method of handling ties in failure times |

**ALPHA=***value*

sets the significance level used for the confidence limits for the hazard ratios. The value must be between 0 and 1. The default is the value of the ALPHA= option in the PROC PHREG statement, or 0.05 if that option is not specified. This option has no effect unless the RISKLIMITS option is specified.

**ABSFCONV=***value*

**CONVERGELIKE=***value*

specifies the absolute function convergence criterion. Termination requires a small change in the objective function (log partial likelihood function) in subsequent iterations,

$$|l_k - l_{k-1}| < value$$

where $l_k$ is the value of the objective function at iteration $k$.

**BEST=***n*

is used exclusively with the SCORE model selection method. The BEST=$n$ option specifies that $n$ models with the highest-score chi-square statistics are to be displayed for each model size. If the option is omitted and there are no more than 10 explanatory variables, then all possible models are listed for each model size. If the option is omitted and there are more than 10 explanatory variables, then the number of models selected for each model size is, at most, equal to the number of explanatory variables listed in the MODEL statement.

See Example 64.2 for an illustration of the SCORE selection method and the BEST= option.

**CORRB**

displays the estimated correlation matrix of the parameter estimates.

**COVB**

displays the estimated covariance matrix of the parameter estimates.

**DETAILS**

produces a detailed display at each step of the model-building process. It produces an "Analysis of Variables Not in the Model" table before displaying the variable selected for entry for FORWARD or STEPWISE selection. For each model fitted, it produces the "Analysis of Maximum Likelihood Estimates" table.

See Example 64.1 for a discussion of these tables.

**ENTRYTIME=***variable*

**ENTRY=***variable*

specifies the name of the variable that represents the left truncation time. This option has no effect when the counting process style of input is specified. See the section "Left Truncation of Failure Times" on page 4583 for more information.

**FCONV=***value*

specifies the relative function convergence criterion. Termination requires a small relative change in the objective function (log partial likelihood function) in subsequent iterations,

$$\frac{|l_k - l_{k-1}|}{|l_{k-1}| + 1\mathrm{E}-6} < value$$

where $l_k$ is the value of the objective function at iteration $k$.

**FIRTH**

performs Firth's penalized maximum likelihood estimation to reduce bias in the parameter estimates (Heinze and Schemper 2001; Firth 1993). This method is useful when the likelihood is monotone—that is, the likelihood converges to finite value while at least one estimate diverges to infinity.

**GCONV=***value*

specifies the relative gradient convergence criterion. Termination requires that the normalized prediction function reduction is small,

$$\frac{\mathbf{g}_k \mathbf{H}_k^{-1} \mathbf{g}_k}{|l_k| + 1\mathrm{E}-6} < value$$

where $l_k$ is the log partial likelihood, $\mathbf{g}_k$ is the gradient vector (first partial derivatives of the log partial likelihood), and $\mathbf{H}_k$ is the negative Hessian matrix (second partial derivatives of the log partial likelihood), all at iteration $k$.

**HIERARCHY=***keyword*

**HIER=***keyword*

specifies whether and how the model hierarchy requirement is applied and whether a single effect or multiple effects are allowed to enter or leave the model in one step. You can specify that only CLASS variable effects, or both CLASS and continuous variable effects, be subject to the hierarchy requirement. The HIERARCHY= option is ignored unless you also specify the FORWARD, BACKWARD, or STEPWISE selection method.

Model hierarchy refers to the requirement that, for any term to be in the model, all effects contained in the term must be present in the model. For example, in order for the interaction A*B to enter the model, the main effects A and B must be in the model. Likewise, neither effect A nor B can leave the model while the interaction A*B is in the model.

The keywords you can specify in the HIERARCHY= option are described as follows:

**NONE**

indicates that the model hierarchy is not maintained. Any single effect can enter or leave the model at any given step of the selection process.

**SINGLE**

indicates that only one effect can enter or leave the model at one time, subject to the model hierarchy requirement. For example, suppose that you specify the main effects A and B and the interaction of A*B in the model. In the first step of the selection process, either A or B can enter the model. In the second step, the other main effect can enter the model. The interaction effect can enter the model only when both main effects have already been entered. Also, before A or B can be removed from the model, the A*B interaction must first be removed. All effects (CLASS and continuous variables) are subject to the hierarchy requirement.

**SINGLECLASS**

is the same as HIERARCHY=SINGLE except that only CLASS effects are subject to the hierarchy requirement.

**MULTIPLE**

indicates that more than one effect can enter or leave the model at one time, subject to the model hierarchy requirement. In a forward selection step, a single main effect can enter the model, or an interaction can enter the model together with all the effects that are contained in the interaction. In a backward elimination step, an interaction itself, or the interaction together with all the effects that the interaction contains, can be removed. All effects (CLASS and continuous variable) are subject to the hierarchy requirement.

**MULTIPLECLASS**

is the same as HIERARCHY=MULTIPLE except that only CLASS effects are subject to the hierarchy requirement.

The default value is HIERARCHY=SINGLE, which means that model hierarchy is to be maintained for all effects (that is, both CLASS and continuous variable effects) and that only a single effect can enter or leave the model at each step.

**INCLUDE=*n***

includes the first *n* effects in the MODEL statement in every model. By default, INCLUDE=0. The INCLUDE= option has no effect when SELECTION=NONE.

**ITPRINT**

displays the iteration history, including the last evaluation of the gradient vector.

**MAXITER=*n***

specifies the maximum number of iterations allowed. The default value for *n* is 25. If convergence is not attained in *n* iterations, the displayed output and all data sets created by PROC PHREG contain results that are based on the last maximum likelihood iteration.

**MAXSTEP=*n***

specifies the maximum number of times the explanatory variables can move in and out of the model before the STEPWISE model-building process ends. The default value for *n* is twice the number of explanatory variables in the MODEL statement. The option has no effect for other model selection methods.

**NODUMMYPRINT**

**NODESIGNPRINT**

**NODP**

suppresses the “Class Level Information” table, which shows how the design matrix columns for the CLASS variables are coded.

**NOFIT**

performs the global score test, which tests the joint significance of all the explanatory variables in the MODEL statement. No parameters are estimated. If the NOFIT option is specified along with other MODEL statement options, NOFIT takes precedence, and all other options are ignored except the TIES= option.

**OFFSET=***name*

specifies the name of an offset variable, which is an explanatory variable with a regression coefficient fixed as one. This option can be used to incorporate risk weights for the likelihood function.

**PLCONV=***value*

controls the convergence criterion for confidence intervals based on the profile-likelihood function. The quantity *value* must be a positive number, with a default value of 1E−4. The PLCONV= option has no effect if profile-likelihood based confidence intervals are not requested.

**RIDGING=***keyword*

specifies the technique to improve the log likelihood when its value is worse than that of the previous step. The available *keywords* are as follows:

**ABSOLUTE**

specifies that the diagonal elements of the negative (expected) Hessian be inflated by adding the ridge value.

**RELATIVE**

specifies that the diagonal elements be inflated by the factor equal to 1 plus the ridge value.

**NONE**

specifies the crude line-search method of taking half a step be used instead of ridging.

The default is RIDGING=RELATIVE.

**RIDGEINIT=***value*

specifies the initial ridge value. The maximum ridge value is 2000 times the maximum of 1 and the initial ridge value. The initial ridge value is raised to 1E–4 if it is less than 1E–4. By default, RIDGEINIT=1E–4. This option has no effect for RIDGING=ABSOLUTE.

**RISKLIMITS**< **=***keyword* >

**RL**< **=***keyword* >

produces confidence intervals for hazard ratios of main effects not involved in interactions or nestings. Computation of these confidence intervals is based on the profile likelihood or based on individual Wald tests. The confidence coefficient can be specified with the ALPHA= option. You can specify one of the following keywords:

**PL**

requests profile-likelihood confidence limits.

**WALD**

requests confidence limits based on the Wald tests.

**BOTH**

request both profile-likelihood and Wald confidence limits.

Classification main effects that use parameterizations other than REF, EFFECT, or GLM are ignored. If you need to compute hazard ratios for an effect involved in interactions or nestings, or using some other parameterization, then you should specify a HAZARDRATIO statement for that effect.

**SELECTION=*method***

specifies the method used to select the model. The *method*s available are as follows:

**BACKWARD**

**B**

requests backward elimination.

**FORWARD**

**F**

requests forward selection.

**NONE**

**N**

fits the complete model specified in the MODEL statement. This is the default value.

**SCORE**

requests best subset selection. It identifies a specified number of models with the highest-score chi-square statistic for all possible model sizes ranging from one explanatory variable to the total number of explanatory variables listed in the MODEL statement. This option is not allowed if an explanatory effect in the MODEL statement contains a CLASS variable.

**STEPWISE**

**S**

requests stepwise selection.

For more information, see the section "Effect Selection Methods" on page 4611.

**SEQUENTIAL**

forces variables to be added to the model in the order specified in the MODEL statement or to be eliminated from the model in the reverse order of that specified in the MODEL statement.

**SINGULAR=***value*

specifies the singularity criterion for determining linear dependencies in the set of explanatory variables. The default value is $10^{-12}$.

**SLENTRY=***value*

**SLE=***value*

specifies the significance level (a value between 0 and 1) for entering an explanatory variable into the model in the FORWARD or STEPWISE method. For all variables not in the model, the one with the smallest $p$-value is entered if the $p$-value is less than or equal to the specified significance level. The default value is 0.05.

**SLSTAY=***value*

**SLS=***value*

specifies the significance level (a value between 0 and 1) for removing an explanatory variable from the model in the BACKWARD or STEPWISE method. For all variables in the model, the one with the largest $p$-value is removed if the $p$-value exceeds the specified significance level. The default value is 0.05.

**START=***n*

begins the FORWARD, BACKWARD, or STEPWISE selection process with the first $n$ effects listed in the MODEL statement. The value of $n$ ranges from 0 to $s$, where $s$ is the total number of effects in the MODEL statement. The default value of $n$ is $s$ for the BACKWARD method and 0 for the FORWARD and STEPWISE methods. Note that START=$n$ specifies only that the first $n$ effects appear in the first model, while INCLUDE=$n$ requires that the first $n$ effects be included in every model. For the SCORE method, START=$n$ specifies that the smallest models contain $n$ effects, where $n$ ranges from 1 to $s$; the default value is 1. The START= option has no effect when SELECTION=NONE.

**STOP=***n*

specifies the maximum (FORWARD method) or minimum (BACKWARD method) number of effects to be included in the final model. The effect selection process is stopped when $n$ effects are found. The value of $n$ ranges from 0 to $s$, where $s$ is the total number of effects in the MODEL statement. The default value of $n$ is $s$ for the FORWARD method and 0 for the BACKWARD method. For the SCORE method, STOP=$n$ specifies that the smallest models contain $n$ effects, where $n$ ranges from 1 to $s$; the default value of $n$ is $s$. The STOP= option has no effect when SELECTION=NONE or STEPWISE.

**STOPRES**

**SR**

specifies that the addition and deletion of variables be based on the result of the likelihood score test for testing the joint significance of variables not in the model. This score chi-square statistic is referred to as the residual chi-square. In the FORWARD method, the STOPRES option enters the explanatory variables into the model one at a time until the residual chi-square becomes insignificant (that is, until the $p$-value of the residual chi-square exceeds the SLENTRY= value). In the BACKWARD method, the STOPRES option removes variables from the model one at a time until the residual chi-square becomes significant (that is, until the $p$-value of the residual chi-square becomes less than the SLSTAY= value). The STOPRES option has no effect for the STEPWISE method.

**TIES=***method*

specifies how to handle ties in the failure time. The available *methods* are as follows:

**BRESLOW**

uses the approximate likelihood of Breslow (1974). This is the default value.

**DISCRETE**

replaces the proportional hazards model by the discrete logistic model

$$\frac{\lambda(t;\mathbf{z})}{1-\lambda(t;\mathbf{z})} = \frac{\lambda_0(t)}{1-\lambda_0(t)} \exp(\mathbf{z}'\boldsymbol{\beta})$$

where $\lambda_0(t)$ and $h(t;\mathbf{z})$ are discrete hazard functions.

**EFRON**

uses the approximate likelihood of Efron (1977).

**EXACT**

computes the exact conditional probability under the proportional hazards assumption that all tied event times occur before censored times of the same value or before larger values. This is equivalent to summing all terms of the marginal likelihood for $\boldsymbol{\beta}$ that are consistent with the observed data (Kalbfleisch and Prentice 1980; DeLong, Guirguis, and So 1994).

The EXACT method can take a considerable amount of computer resources. If ties are not extensive, the EFRON and BRESLOW methods provide satisfactory approximations to the EXACT method for the continuous time-scale model. In general, Efron's approximation gives results that are much closer to the EXACT method results than Breslow's approximation does. If the time scale is genuinely discrete, you should use the DISCRETE method. The DISCRETE method is also required in the analysis of case-control studies when there is more than one case in a matched set. If there are no ties, all four methods result in the same likelihood and yield identical estimates. The default, TIES=BRESLOW, is the most efficient method when there are no ties.

**XCONV=***value*

**CONVEREPARM=***value*

specifies the relative parameter convergence criterion. Termination requires a small relative parameter change in subsequent iterations,

$$\max_i |\delta_k^{(i)}| < value$$

where

$$\delta_k^{(i)} = \begin{cases} \theta_k^{(i)} - \theta_{k-1}^{(i)} & |\theta_{k-1}^{(i)}| < .01 \\ \frac{\theta_k^{(i)} - \theta_{k-1}^{(i)}}{\theta_{k-1}^{(i)}} & \text{otherwise} \end{cases}$$

where $\theta_k^{(i)}$ is the estimate of the $i$th parameter at iteration $k$.

## OUTPUT Statement

**OUTPUT** < *OUT=SAS-data-set* > < *keyword=name . . . keyword=name* > < */options* > **;**

The OUTPUT statement creates a new SAS data set containing statistics calculated for each observation. These can include the estimated linear predictor ($\mathbf{z}'_j\hat{\boldsymbol{\beta}}$) and its standard error, survival distribution estimates, residuals, and influence statistics. In addition, this data set includes the time variable, the explanatory variables listed in the MODEL statement, the censoring variable (if specified), and the BY, STRATA, FREQ, and ID variables (if specified).

For observations with missing values in the time variable or any explanatory variables, the output statistics are set to missing. However, for observations with missing values only in the censoring variable or the FREQ variable, survival estimates are still computed. Therefore, by adding observations with missing values in the FREQ variable or the censoring variable, you can compute the survivor function estimates for new observations or for settings of explanatory variables not present in the data without affecting the model fit.

No OUTPUT data set is created if the model contains a time-dependent variable defined by means of programming statements.

The following list explains specifications in the OUTPUT statement.

**OUT=***SAS-data-set*

names the output data set. If you omit the OUT= option, the OUTPUT data set is created and given a default name by using the DATA*n* convention. See the section “OUT= Output Data Set in the OUTPUT Statement” on page 4625 for more information.

*keyword=name*

specifies the statistics included in the OUTPUT data set and names the new variables that contain the statistics. Specify a keyword for each desired statistic (see the following list of keywords), an equal sign, and either a variable or a list of variables to contain the statistic. The keywords that accept a list of variables are DFBETA, RESSCH, RESSCO, and WTRESSCH. For these keywords, you can specify as many names in *name* as the number of explanatory variables specified in the MODEL statement. If you specify $k$ names and $k$ is less than the total number of explanatory variables, only the changes for the first $k$ parameter estimates are output. The keywords and the corresponding statistics are as follows:

**ATRISK**

**NUM_LEFT**

specifies the number of subjects at risk at the observation time $\tau_j$ (or at the right endpoint of the at-risk interval when a counting process MODEL specification is used).

**DFBETA**

specifies the approximate changes in the parameter estimates ($\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(j)}$) when the $j$th observation is omitted. These variables are a weighted transform of the score residual variables and are useful in assessing local influence and in computing robust variance estimates.

**LD**

specifies the approximate likelihood displacement when the observation is left out. This diagnostic can be used to assess the impact of each observation on the overall fit of the model.

**LMAX**

specifies the relative influence of observations on the overall fit of the model. This diagnostic is useful in assessing the sensitivity of the fit of the model to each observation.

**LOGLOGS**

specifies the log of the negative log of SURVIVAL.

**LOGSURV**

specifies the log of SURVIVAL.

**RESDEV**

specifies the deviance residual $\hat{D}_j$. This is a transform of the martingale residual to achieve a more symmetric distribution.

**RESMART**

specifies the martingale residual $\hat{M}_j$. The residual at the observation time $\tau_j$ can be interpreted as the difference over $[0, \tau_j]$ in the observed number of events minus the expected number of events given by the model.

**RESSCH**

specifies the Schoenfeld residuals. These residuals are useful in assessing the proportional hazards assumption.

**RESSCO**

specifies the score residuals. These residuals are a decomposition of the first partial derivative of the log likelihood. They can be used to assess the leverage exerted by each subject in the parameter estimation. They are also useful in constructing robust sandwich variance estimators.

**STDXBETA**

specifies the standard error of the estimated linear predictor, $\sqrt{\mathbf{z}'_j \hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})\mathbf{z}_j}$.

**SURVIVAL**

specifies the survivor function estimate $\hat{S}_j = [\hat{S}_0(\tau_j)]^{\exp(\mathbf{z}'_j \hat{\boldsymbol{\beta}})}$, where $\tau_j$ is the observation time.

**WTRESSCH**

specifies the weighted Schoenfeld residuals. These residuals are useful in investigating the nature of nonproportionality if the proportional hazard assumption does not hold.

**XBETA**

specifies the estimate of the linear predictor, $\mathbf{z}'_j \hat{\boldsymbol{\beta}}$.

The following options can appear in the OUTPUT statement after a slash (/) as follows:

**ORDER=***sort_order*
specifies the order of the observations in the OUTPUT data set. Available values for *sort_order* are as follows:

| | |
|---|---|
| DATA | requests that the output observations be sorted the same as the input data set. |
| SORTED | requests that the output observations be sorted by strata and descending order of the time variable within each stratum. |

The default is ORDER=DATA.

**METHOD=***method*
specifies the method used to compute the survivor function estimates. The two available methods are as follows:

**BRESLOW**
**CH**
**EMP**
specifies that the empirical cumulative hazard function estimate of the survivor function be computed; that is, the survivor function is estimated by exponentiating the negative empirical cumulative hazard function.

**PL**
specifies that the product-limit estimate of the survivor function be computed.

The default is METHOD=BRESLOW.

## Programming Statements

Programming statements are used to create or modify the values of the explanatory variables in the MODEL statement. They are especially useful in fitting models with time-dependent explanatory variables. Programming statements can also be used to create explanatory variables that are not time dependent. For example, you can create indicator variables from a categorical variable and incorporate them into the model. PROC PHREG programming statements cannot be used to create or modify the values of the response variable, the censoring variable, the frequency variable, or the strata variables.

The following DATA step statements are available in PROC PHREG:

```
ABORT
ARRAY
assignment statements
CALL
DO
iterative DO
DO UNTIL
```

```
DO WHILE
END
GOTO
IF-THEN/ELSE
LINK-RETURN
PUT
SELECT
SUM statement
```

By default, the PUT statement in PROC PHREG writes results to the Output window instead of the Log window. If you want the results of the PUT statements to go to the Log window, add the following statement before the PUT statements:

```
FILE LOG;
```

DATA step functions are also available. Use these programming statements the same way you use them in the DATA step. For detailed information, refer to *SAS Language Reference: Dictionary*.

Consider the following example of using programming statements in PROC PHREG. Suppose blood pressure is measured at multiple times during the course of a study investigating the effect of blood pressure on some survival time. By treating the blood pressure as a time-dependent explanatory variable, you are able to use the value of the most recent blood pressure at each specific point of time in the modeling process rather than using the initial blood pressure or the final blood pressure. The values of the following variables are recorded for each patient, if they are available. Otherwise, the variables contain missing values.

| | |
|---|---|
| Time | survival time |
| Censor | censoring indicator (with 0 as the censoring value) |
| BP0 | blood pressure on entry to the study |
| T1 | time 1 |
| BP1 | blood pressure at T1 |
| T2 | time 2 |
| BP2 | blood pressure at T2 |

The following programming statements create a variable BP. At each time T, the value of BP is the blood pressure reading for that time, if available. Otherwise, it is the last blood pressure reading.

```
proc phreg;
   model Time*Censor(0)=BP;
   BP = BP0;
   if Time>=T1 and T1^=. then BP=BP1;
   if Time>=T2 and T2^=. then BP=BP2;
run;
```

For other illustrations of using programming statements, see the section “Classical Method of Maximum Likelihood” on page 4521 and Example 64.6.

## STRATA Statement

**STRATA** *variable* < *( list )* > < *. . . variable* < *( list )* > > < */option* > **;**

The proportional hazards assumption might not be realistic for all data. If so, it might still be reasonable to perform a stratified analysis. The STRATA statement names the variables that determine the stratification. Strata are formed according to the nonmissing values of the STRATA variables unless the MISSING option is specified. In the STRATA statement, *variable* is a variable with values that are used to determine the strata levels, and *list* is an optional list of values for a numeric variable. Multiple variables can appear in the STRATA statement.

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 variable. If the variable is numeric and is followed by a list, then the levels for that variable correspond to the intervals defined by the list. The corresponding strata are formed by the combination of levels and unique values. The list can include numeric values separated by commas or blanks, *value* to *value* by *value* range specifications, or combinations of these.

For example, the specification

```
strata age (5, 10 to 40 by 10) sex ;
```

indicates that the levels for age are to be less than 5, 5 to 10, 10 to 20, 20 to 30, 30 to 40, and greater than 40. (Note that observations with exactly the cutpoint value fall into the interval preceding the cutpoint.) Thus, with the sex variable, this STRATA statement specifies 12 strata altogether.

The following option can be specified in the STRATA statement after a slash (/):

**MISSING**
allows missing values ('.' for numeric variables and blanks for character variables) as valid STRATA variable values. Otherwise, observations with missing STRATA variable values are deleted from the analysis.

## TEST Statement

< *label:* > **TEST** *equation1* < *,. . . , equationk* > < */options* > **;**

The TEST statement tests linear hypotheses about the regression coefficients. PROC PHREG performs a Wald test for the joint hypothesis specified in a single TEST statement. Each equation specifies a linear hypothesis; multiple equations (rows of the joint hypothesis) are separated by commas. The *label*, which must be a valid SAS name, is used to identify the resulting output and should always be included. You can submit multiple TEST statements.

The form of an equation is as follows:

*term* < ± *term . . .* > < = < ± *term* < ± *term . . .* > > >

where *term* is a variable or a constant or a constant times a variable. The variable is any explanatory variable in the MODEL statement. When no equal sign appears, the expression is set to 0. The following program illustrates possible uses of the TEST statement:

```
proc phreg;
   model time= A1 A2 A3 A4;
   Test1: test A1, A2;
   Test2: test A1=0,A2=0;
   Test3: test A1=A2=A3;
   Test4: test A1=A2,A2=A3;
run;
```

Note that the first and second TEST statements are equivalent, as are the third and fourth TEST statements.

The following options can be specified in the TEST statement after a slash (/):

**AVERAGE**

enables you to assess the average effect of the variables in the given TEST statement. An overall estimate of the treatment effect is computed as a weighted average of the treatment coefficients as illustrated in the following statement:

```
TREATMENT: test trt1, trt2, trt3, trt4 / average;
```

Let $\beta_1, \beta_2, \beta_3$, and $\beta_4$ be corresponding parameters for trt1, trt2, trt3, and trt4, respectively. Let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)'$ be the estimated coefficient vector and let $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})$ be the corresponding variance estimate. Assuming $\beta_1 = \beta_2 = \beta_3 = \beta_4$, let $\bar{\beta}$ be the average treatment effect. The effect is estimated by $\mathbf{c}'\hat{\boldsymbol{\beta}}$, where $\mathbf{c} = [\mathbf{1}_4'\hat{\mathbf{V}}^{-1}(\hat{\boldsymbol{\beta}})\mathbf{1}_4]^{-1}\hat{\mathbf{V}}^{-1}(\hat{\boldsymbol{\beta}})\mathbf{1}_4$ and $\mathbf{1}_4 = (1,1,1,1)'$. A test of the null hypothesis $H_0 : \bar{\beta} = 0$ is also included, which is more sensitive than the multivariate test for testing the null hypothesis $H_0 : \beta 1 = \beta 2 = \beta 3 = \beta 4 = 0$.

**E**

specifies that the linear coefficients and constants be printed. When the AVERAGE option is specified along with the E option, the optimal weights of the average effect are also printed in the same tables as the coefficients.

**PRINT**

displays intermediate calculations. This includes $\mathbf{L}\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})\mathbf{L}'$ bordered by $(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{c})$, and $[\mathbf{L}\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})\mathbf{L}']^{-1}$ bordered by $[\mathbf{L}\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})\mathbf{L}']^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{c})$, where $\mathbf{L}$ is a matrix of linear coefficients and $\mathbf{c}$ is a vector of constants.

See the section "Testing Linear Hypotheses about Regression Coefficients" on page 4594 for details.

## WEIGHT Statement

**WEIGHT** *variable* < */option* > **;**

The *variable* in the WEIGHT statement identifies the variable in the input data set that contains the case weights. When the WEIGHT statement appears, each observation in the input data set is weighted by the value of the WEIGHT variable. The WEIGHT values can be nonintegral and are not truncated. Observations with negative, zero, or missing values for the WEIGHT variable are not used in the model fitting. When the WEIGHT statement is not specified, each observation is assigned a weight of 1. The WEIGHT statement is available for TIES=BRESLOW and TIES=EFRON only.

The following option can be specified in the WEIGHT statement after a slash (/):

**NORMALIZE**

**NORM**

causes the weights specified by the WEIGHT *variable* to be normalized so that they add up the actual sample size. With this option, the estimated covariance matrix of the parameter estimators is invariant to the scale of the WEIGHT variable.

# Details: PHREG Procedure

## Failure Time Distribution

Let $T$ be a nonnegative random variable representing the failure time of an individual from a homogeneous population. The survival distribution function (also known as the survivor function) of $T$ is written as

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

A mathematically equivalent way of specifying the distribution of $T$ is through its hazard function. The hazard function $\lambda(t)$ specifies the instantaneous failure rate at $t$. If $T$ is a continuous random variable, $\lambda(t)$ is expressed as

$$\lambda(t) = \lim_{\Delta t \rightarrow 0^+} \frac{\Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}$$

where $f(t)$ is the probability density function of $T$. If $T$ is discrete with masses at $x_1 < x_2 < \ldots$, then $\lambda(t)$ is given by

$$\lambda(t) = \sum_j \lambda_j \delta(t - x_j)$$

where

$$\delta(u) = \begin{cases} 0 & \text{if } u < 0 \\ 1 & \text{otherwise} \end{cases}$$

$$\lambda_j = \Pr(T = x_j \mid T \geq x_j) = \frac{\Pr(T = x_j)}{S(x_j)}$$

for $j = 1, 2, \ldots$

## CLASS Variable Parameterization

Consider a model with one CLASS variable A with four levels, 1, 2, 5, and 7. Details of the possible choices for the PARAM= option follow.

**EFFECT**

Three columns are created to indicate group membership of the nonreference levels. For the reference level, all three dummy variables have a value of $-1$. For instance, if the reference level is 7 (REF=7), the design matrix columns for A are as follows:

| | Design Matrix | | |
|---|---|---|---|
| **A** | **A1** | **A2** | **A5** |
| 1 | 1 | 0 | 0 |
| 2 | 0 | 1 | 0 |
| 5 | 0 | 0 | 1 |
| 7 | $-1$ | $-1$ | $-1$ |

Parameter estimates of CLASS main effects, using the effect coding scheme, estimate the difference in the effect of each nonreference level compared to the average effect over all four levels.

**GLM**

As in PROC GLM, four columns are created to indicate group membership. The design matrix columns for A are as follows:

| | Design Matrix | | | |
|---|---|---|---|---|
| **A** | **A1** | **A2** | **A5** | **A7** |
| 1 | 1 | 0 | 0 | 0 |
| 2 | 0 | 1 | 0 | 0 |
| 5 | 0 | 0 | 1 | 0 |
| 7 | 0 | 0 | 0 | 1 |

Parameter estimates of CLASS main effects, using the GLM coding scheme, estimate the difference in the effects of each level compared to the last level.

**ORDINAL**

Three columns are created to indicate group membership of the higher levels of the effect. For the first level of the effect (which for A is 1), all three dummy variables have a value of 0. The design matrix columns for A are as follows:

| | Design Matrix | | |
|---|---|---|---|
| **A** | **A2** | **A5** | **A7** |
| 1 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 |
| 5 | 1 | 1 | 0 |
| 7 | 1 | 1 | 1 |

In the ORDINAL coding scheme, the first level of the effect is a control or baseline level. Parameter estimates of CLASS main effects estimate the differences between effects of successive levels. When the parameters have the same sign, the effect is monotonic across the levels.

**POLYNOMIAL**

**POLY**

Three columns are created. The first represents the linear term ($x$), the second represents the quadratic term ($x^2$), and the third represents the cubic term ($x^3$), where $x$ is the level value. If the CLASS levels are not numeric, they are translated into 1, 2, 3, ... according to their sorting order. The design matrix columns for A are as follows:

| | Design Matrix | | |
|---|---|---|---|
| **A** | **APOLY1** | **APOLY2** | **APOLY3** |
| 1 | 1 | 1 | 1 |
| 2 | 2 | 4 | 8 |
| 5 | 5 | 25 | 125 |
| 7 | 7 | 49 | 343 |

**REFERENCE**

**REF**

Three columns are created to indicate group membership of the nonreference levels. For the reference level, all three dummy variables have a value of 0. For instance, if the reference level is 7 (REF=7), the design matrix columns for A are as follows:

| | Design Matrix | | |
|---|---|---|---|
| **A** | **A1** | **A2** | **A5** |
| 1 | 1 | 0 | 0 |
| 2 | 0 | 1 | 0 |
| 5 | 0 | 0 | 1 |
| 7 | 0 | 0 | 0 |

Parameter estimates of CLASS main effects, using the reference coding scheme, estimate the difference in the effect of each nonreference level compared to the effect of the reference level.

**ORTHEFFECT**

The columns are obtained by applying the Gram-Schmidt orthogonalization to the columns for PARAM=EFFECT. The design matrix columns for A are as follows:

| | Design Matrix | | |
|---|---|---|---|
| **A** | **AOEFF1** | **AOEFF2** | **AOEFF3** |
| 1 | 1.41421 | −0.81650 | −0.57735 |
| 2 | 0.00000 | 1.63299 | −0.57735 |
| 5 | 0.00000 | 0.00000 | 1.73205 |
| 7 | −1.41421 | −0.81649 | −0.57735 |

**ORTHORDINAL**

The columns are obtained by applying the Gram-Schmidt orthogonalization to the columns for PARAM=ORDINAL. The design matrix columns for A are as follows:

| | Design Matrix | | |
|---|---|---|---|
| **A** | **AOORD1** | **AOORD2** | **AOORD3** |
| 1 | −1.73205 | 0.00000 | 0.00000 |
| 2 | 0.57735 | −1.63299 | 0.00000 |
| 5 | 0.57735 | 0.81650 | −1.41421 |
| 7 | 0.57735 | 0.81650 | 1.41421 |

**ORTHPOLY**

The columns are obtained by applying the Gram-Schmidt orthogonalization to the columns for PARAM=POLY. The design matrix columns for A are as follows:

| | Design Matrix | | |
|---|---|---|---|
| **A** | **AOPOLY1** | **AOPOLY2** | **AOPOLY5** |
| 1 | −1.153 | 0.907 | −0.921 |
| 2 | −0.734 | −0.540 | 1.473 |
| 5 | 0.524 | −1.370 | −0.921 |
| 7 | 1.363 | 1.004 | 0.368 |

**ORTHREF**

The columns are obtained by applying the Gram-Schmidt orthogonalization to the columns for PARAM=REFERENCE. The design matrix columns for A are as follows:

| | Design Matrix | | |
|---|---|---|---|
| **A** | **AOREF1** | **AOREF2** | **AOREF3** |
| 1 | 1.73205 | 0.00000 | 0.00000 |
| 2 | −0.57735 | 1.63299 | 0.00000 |
| 5 | −0.57735 | −0.81650 | 1.41421 |
| 7 | −0.57735 | −0.81650 | −1.41421 |

## Clarification of the Time and CLASS Variables Usage

The following DATA step creates an artificial data set, Foo, to be used in this section. There are four variables in Foo: the variable T contains the failure times; the variable Status is the censoring indicator variable with the value 1 for an uncensored failure time and the value 0 for a censored time; the variable A is a categorical variable with values 1, 2, and 3 representing three different categories; and the variable MirrorT is an exact copy of T.

```
Data Foo;
   input T Status A @@;
   MirrorT = T;
   datalines;
 23         1       1    7          0       1
 23         1       1   10          1       1
```

```
  20        0        1   13        0        1
  24        1        1   10        1        1
  18        1        2    6        1        2
  18        0        2    6        1        2
  13        0        2   13        1        2
   9        0        2   15        1        2
   8        1        3    6        1        3
  12        0        3    4        1        3
  11        1        3    8        1        1
   6        1        3    7        1        3
   7        1        3   12        1        3
   9        1        2   15        1        2
   3        1        2   14        0        3
   6        1        1   13        1        2
   ;
```

## Time Variable on the Right Side of the MODEL Statement

When the time variable is explicitly used in an explanatory effect in the MODEL statement, the effect is NOT time dependent. In the following specification, T is the time variable, but T does not play the role of the time variable in the explanatory effect T*A:

```
proc phreg data=Foo;
   class A;
   model T*Status(0)=T*A;
   run;
```

The parameter estimates of this model are shown in Figure 64.12.

**Figure 64.12** T*A Effect

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

| Parameter | | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio | Label |
|---|---|---|---|---|---|---|---|---|
| T*A | 1 | 1 | -0.16549 | 0.05042 | 10.7734 | 0.0010 | . | A 1 * T |
| T*A | 2 | 1 | -0.11852 | 0.04181 | 8.0344 | 0.0046 | . | A 2 * T |

To verify that the effect T*A in the MODEL statement is not time dependent, T is replaced by MirrorT, which is an exact copy of T, as in the following statements:

```
proc phreg data=Foo;
   class A;
   model T*Status(0)=A*MirrorT;
   run;
```

The results of fitting this model (Figure 64.13) are identical to those of the previous model (Figure 64.12), except for the parameter names and labels. The effect A*MirrorT is not time dependent, so neither is A*T.

**Figure 64.13** T*A Effect

```
                         The PHREG Procedure

                Analysis of Maximum Likelihood Estimates

                      Parameter      Standard
Parameter       DF     Estimate         Error    Chi-Square    Pr > ChiSq

MirrorT*A 1      1     -0.16549       0.05042       10.7734        0.0010
MirrorT*A 2      1     -0.11852       0.04181        8.0344        0.0046

                 Analysis of Maximum Likelihood Estimates

                                 Hazard
                 Parameter        Ratio    Label

                 MirrorT*A 1          .    A 1 * MirrorT
                 MirrorT*A 2          .    A 2 * MirrorT
```

## CLASS Variables and Programming Statements

In PROC PHREG, the levelization of CLASS variables is determined by the CLASS statement and the input data, and is not affected by user-supplied programming statements. Consider the following statements, which produce the results in Figure 64.14. Variable A is declared as a CLASS variable in the CLASS statement. By default, the reference parameterization is used with A=3 as the reference level. Two regression coefficients are estimated for the two dummy variables of A.

```
proc phreg data=Foo;
   class A;
   model T*Status(0)=A;
   run;
```

Figure 64.14 shows the dummy variables of A and the regression coefficients estimates.

**Figure 64.14** Design Variable and Regression Coefficient Estimates

```
                         The PHREG Procedure

                       Class Level Information

                                          Design
                      Class     Value     Variables

                      A         1           1      0
                                2           0      1
                                3           0      0
```

**Figure 64.14** *continued*

Analysis of Maximum Likelihood Estimates

| Parameter | | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio | Label |
|---|---|---|---|---|---|---|---|---|
| A | 1 | 1 | -1.40925 | 0.64802 | 4.7293 | 0.0297 | 0.244 | A 1 |
| A | 2 | 1 | -0.65705 | 0.51764 | 1.6112 | 0.2043 | 0.518 | A 2 |

Now consider the programming statement that attempts to change the value of the CLASS variable A as in the following specification:

```
proc phreg data=Foo;
   class A;
   model T*Status(0)=A;
   if A=3 then A=2;
   run;
```

Results of this analysis are shown in Figure 64.15 and are identical to those in Figure 64.14. The `if A=3 then A=2` programming statement has no effects on the design variables for A, which have already been determined.

**Figure 64.15** Design Variable and Regression Coefficient Estimates

The PHREG Procedure

Class Level Information

| Class | Value | Design Variables | |
|---|---|---|---|
| A | 1 | 1 | 0 |
| | 2 | 0 | 1 |
| | 3 | 0 | 0 |

Analysis of Maximum Likelihood Estimates

| Parameter | | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio | Label |
|---|---|---|---|---|---|---|---|---|
| A | 1 | 1 | -1.40925 | 0.64802 | 4.7293 | 0.0297 | 0.244 | A 1 |
| A | 2 | 1 | -0.65705 | 0.51764 | 1.6112 | 0.2043 | 0.518 | A 2 |

Additionally any variable used in a programming statement that has already been declared in the CLASS statement is NOT treated as a collection of the corresponding design variables. Consider the following statements:

```
proc phreg data=Foo;
   class A;
   model T*Status(0)=A X;
   X=T*A;
   run;
```

The CLASS variable A generates two design variables as explanatory variables. The variable X created by the **X=T*A** programming statement is a single time-dependent covariate whose values are evaluated using the exact values of A given in the data.

**Figure 64.16** Single Time-Dependent Variable X*A

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

| Parameter | | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio | Label |
|---|---|---|---|---|---|---|---|---|
| A | 1 | 1 | 0.15798 | 1.69338 | 0.0087 | 0.9257 | 1.171 | A 1 |
| A | 2 | 1 | 0.00898 | 0.87573 | 0.0001 | 0.9918 | 1.009 | A 2 |
| X | | 1 | 0.09268 | 0.09535 | 0.9448 | 0.3311 | 1.097 | |

To generalize the simple test of proportional hazard assumption for the design variables of A (as in the section the "Classical Method of Maximum Likelihood" on page 4521), you specify the following statements, which are not the same as in the preceding program or the specification in the section "Time Variable on the Right Side of the MODEL Statement" on page 4577:

```
proc phreg data=Foo;
   class A;
   model T*Status(0)=A X1 X2;
   X1= T*(A=1);
   X2= T*(A=2);
   run;
```

Results of this test are shown in Figure 64.17.

**Figure 64.17** Simple Test of Proportional Hazards Assumption

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

| Parameter | | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio | Label |
|---|---|---|---|---|---|---|---|---|
| A | 1 | 1 | -0.00766 | 1.69435 | 0.0000 | 0.9964 | 0.992 | A 1 |
| A | 2 | 1 | -0.88132 | 1.64298 | 0.2877 | 0.5917 | 0.414 | A 2 |
| X1 | | 1 | -0.15522 | 0.20174 | 0.5920 | 0.4417 | 0.856 | |
| X2 | | 1 | 0.01155 | 0.18858 | 0.0037 | 0.9512 | 1.012 | |

## Partial Likelihood Function for the Cox Model

Let $\mathbf{Z}_l(t)$ denote the vector explanatory variables for the $l$th individual at time $t$. Let $t_1 < t_2 < \ldots < t_k$ denote the $k$ distinct, ordered event times. Let $d_i$ denote the multiplicity of failures at $t_i$; that is, $d_i$ is the size of the set $\mathcal{D}_i$ of individuals that fail at $t_i$. Let $w_l$ be the weight associated with the $l$th individual. Using this notation, the likelihood functions used in PROC PHREG to estimate $\boldsymbol{\beta}$ are described in the following sections.

### Continuous Time Scale

Let $\mathcal{R}_i$ denote the risk set just before the $i$th ordered event time $t_i$. Let $\mathcal{R}_i^*$ denote the set of individuals whose event or censored times exceed $t_i$ or whose censored times are equal to $t_i$.

#### *Exact Likelihood*

$$L_1(\boldsymbol{\beta}) = \prod_{i=1}^{k} \left\{ \int_0^\infty \prod_{j \in \mathcal{D}_i} \left[ 1 - \exp\left( -\frac{\mathrm{e}^{\boldsymbol{\beta}'\mathbf{Z}_j(t_i)}}{\sum_{l \in \mathcal{R}_i^*} \mathrm{e}^{\boldsymbol{\beta}'\mathbf{Z}_l(t_i)}} t \right) \right] \exp(-t)dt \right\}$$

#### *Breslow Likelihood*

$$L_2(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\mathrm{e}^{\boldsymbol{\beta}' \sum_{j \in \mathcal{D}_i} \mathbf{Z}_j(t_i)}}{\left[ \sum_{l \in \mathcal{R}_i} \mathrm{e}^{\boldsymbol{\beta}'\mathbf{Z}_l(t_i)} \right]^{d_i}}$$

Incorporating weights, the Breslow likelihood becomes

$$L_2(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\mathrm{e}^{\boldsymbol{\beta}' \sum_{j \in \mathcal{D}_i} w_j \mathbf{Z}_j(t_i)}}{\left[ \sum_{l \in \mathcal{R}_i} w_l \mathrm{e}^{\boldsymbol{\beta}'\mathbf{Z}_l(t_i)} \right]^{\sum_{j \in \mathcal{D}_i} w_i}}$$

#### *Efron Likelihood*

$$L_3(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\mathrm{e}^{\boldsymbol{\beta}' \sum_{j \in \mathcal{D}_i} \mathbf{Z}_j(t_i)}}{\prod_{j=1}^{d_i} \left( \sum_{l \in \mathcal{R}_i} \mathrm{e}^{\boldsymbol{\beta}'\mathbf{Z}_l(t_i)} - \frac{j-1}{d_i} \sum_{l \in \mathcal{D}_i} \mathrm{e}^{\boldsymbol{\beta}'\mathbf{Z}_l(t_i)} \right)}$$

Incorporating weights, the Efron likelihood becomes

$$L_3(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\mathrm{e}^{\boldsymbol{\beta}' \sum_{j \in \mathcal{D}_i} w_j \mathbf{Z}_j(t_i)}}{\left[ \prod_{j=1}^{d_i} \left( \sum_{l \in \mathcal{R}_i} w_l \mathrm{e}^{\boldsymbol{\beta}' \mathbf{Z}_l(t_i)} - \frac{j-1}{d_i} \sum_{l \in \mathcal{D}_i} w_l \mathrm{e}^{\boldsymbol{\beta}' \mathbf{Z}_l(t_i)} \right) \right]^{\frac{1}{d_i} \sum_{j \in \mathcal{D}_i} w_j}}$$

## Discrete Time Scale

Let $\mathcal{Q}_i$ denote the set of all subsets of $d_i$ individuals from the risk set $\mathcal{R}_i$. For each $\mathbf{q} \in \mathcal{Q}_i$, $\mathbf{q}$ is a $d_i$-tuple $(q_1, q_2, \ldots, q_{d_i})$ of individuals who might have failed at $t_i$.

### *Discrete Logistic Likelihood*

$$L_4(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\mathrm{e}^{\boldsymbol{\beta}' \sum_{j \in \mathcal{D}_i} \mathbf{Z}_j(t_i)}}{\sum_{\boldsymbol{q} \in \mathcal{Q}_i} \mathrm{e}^{\boldsymbol{\beta}' \sum_{l=1}^{d_j} \mathbf{Z}_{q_l}(t_i)}}$$

The computation of $L_4(\boldsymbol{\beta})$ and its derivatives is based on an adaptation of the recurrence algorithm of Gail, Lubin, and Rubinstein (1981) to the logarithmic scale. When there are no ties on the event times (that is, $d_i \equiv 1$), all four likelihood functions $L_1(\boldsymbol{\beta})$, $L_2(\boldsymbol{\beta})$, $L_3(\boldsymbol{\beta})$, and $L_4(\boldsymbol{\beta})$ reduce to the same expression. In a stratified analysis, the partial likelihood is the product of the partial likelihood functions for the individual strata.

# Counting Process Style of Input

In the counting process formulation, data for each subject are identified by a triple $\{N, Y, \mathbf{Z}\}$ of counting, at-risk, and covariate processes. Here, $N(t)$ indicates the number of events that the subject experiences over the time interval $(0, t]$; $Y(t)$ indicates whether the subject is at risk at time $t$ (one if at risk and zero otherwise); and $\mathbf{Z}(t)$ is a vector of explanatory variables for the subject at time $t$. The sample path of $N$ is a step function with jumps of size +1 at the event times, and $N(0) = 0$. Unless $\mathbf{Z}(t)$ changes continuously with time, the data for each subject can be represented by multiple observations, each identifying a semiclosed time interval $(t1, t2]$, the values of the explanatory variables over that interval, and the event status at $t2$. The subject remains at risk during the interval $(t1, t2]$, and an event might occur at $t2$. Values of the explanatory variables for the subject remain unchanged in the interval. This style of data input was originated by Therneau (1994).

For example, a patient has a tumor recurrence at weeks 3, 10, and 15 and is followed up to week 23. The explanatory variables are Trt (treatment), Z1 (initial tumor number), and Z2 (initial tumor size), and, for this patient, the values of Trt, Z1, and Z2 are (1,1,3). The data for this patient are represented by the following four observations:

| T1 | T2 | Event | Trt | Z1 | Z2 |
|---|---|---|---|---|---|
| 0 | 3 | 1 | 1 | 1 | 3 |
| 3 | 10 | 1 | 1 | 1 | 3 |
| 10 | 15 | 1 | 1 | 1 | 3 |
| 15 | 23 | 0 | 1 | 1 | 3 |

Here (T1,T2] contains the at-risk intervals. The variable Event is a censoring variable indicating whether a recurrence has occurred at T2; a value of 1 indicates a tumor recurrence, and a value of 0 indicates nonrecurrence. The PHREG procedure fits the multiplicative hazards model, which is specified as follows:

```
proc phreg;
   model (T1,T2) * Event(0) = Trt Z1 Z2;
run;
```

Another useful application of the counting process formulation is delayed entry of subjects into the risk set. For example, in studying the mortality of workers exposed to a carcinogen, the survival time is chosen to be the worker's age at death by malignant neoplasm. Any worker joining the workplace at a later age than a given event failure time is not included in the corresponding risk set. The variables of a worker consist of Entry (age at which the worker entered the workplace), Age (age at death or age censored), Status (an indicator of whether the observation time is censored, with the value 0 identifying a censored time), and X1 and X2 (explanatory variables thought to be related to survival). The specification for such an application is as follows:

```
proc phreg;
   model (Entry, Age) * Status(0) = X1 X2;
run;
```

Alternatively, you can use a time-dependent variable to control the risk set, as illustrated in the following specification:

```
proc phreg;
   model Age * Status(0) = X1 X2;
   if Age < Entry then X1= .;
run;
```

Here, X1 becomes a time-dependent variable. At a given death time $t$, the value of X1 is reevaluated for each subject with Age $\geq t$; subjects with Entry $> t$ are given a missing value in X1 and are subsequently removed from the risk set. Computationally, this approach is not as efficient as the one that uses the counting process formulation.

## Left Truncation of Failure Times

Left truncation arises when individuals come under observation only some known time after the natural time origin of the phenomenon under study. The risk set just prior to an event time does not include individuals whose left truncation times exceed the given event time. Thus, any contribution to the likelihood must be conditional on the truncation limit having been exceeded.

An alternative way to specify left truncation in PROC PHREG is through the counting process style of input. The following specifications are equivalent:

```
proc phreg data=one;
   model t2*dead(0)=x1-x10/entry=t1;
   title 'The ENTRY= option is Specified';
run;

proc phreg data=one;
   model (t1,t2)*dead(0)=x1-x10;
   title 'Counting Process Style of Input';
run;
```

## The Multiplicative Hazards Model

Consider a set of $n$ subjects such that the counting process $N_i \equiv \{N_i(t), t \geq 0\}$ for the $i$th subject represents the number of observed events experienced over time $t$. The sample paths of the process $N_i$ are step functions with jumps of size $+1$, with $N_i(0) = 0$. Let $\boldsymbol{\beta}$ denote the vector of unknown regression coefficients. The multiplicative hazards function $\Lambda(t, \mathbf{Z}_i(t))$ for $N_i$ is given by

$$Y_i(t)d\Lambda(t, \mathbf{Z}_i(t)) = Y_i(t)\exp(\boldsymbol{\beta}'\mathbf{Z}_i(t))d\Lambda_0(t)$$

where

- $Y_i(t)$ indicates whether the $i$th subject is at risk at time $t$ (specifically, $Y_i(t) = 1$ if at risk and $Y_i(t) = 0$ otherwise)
- $\mathbf{Z}_i(t)$ is the vector of explanatory variables for the $i$th subject at time $t$
- $\Lambda_0(t)$ is an unspecified baseline hazard function

Refer to Fleming and Harrington (1991) and Andersen et al. (1992). The Cox model is a special case of this multiplicative hazards model, where $Y_i(t) = 1$ until the first event or censoring, and $Y_i(t) = 0$ thereafter.

The partial likelihood for $n$ independent triplets $(N_i, Y_i, \mathbf{Z}_i), i = 1, \ldots, n$, has the form

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \prod_{t \geq 0} \left\{ \frac{Y_i(t)\exp(\boldsymbol{\beta}'\mathbf{Z}_i(t))}{\sum_{j=1}^{n} Y_j(t)\exp(\boldsymbol{\beta}'\mathbf{Z}_j(t))} \right\}^{\Delta N_i(t)}$$

where $\Delta N_i(t) = 1$ if $N_i(t) - N_i(t-) = 1$, and $\Delta N_i(t) = 0$ otherwise.

## Hazard Ratios

Consider a dichotomous risk factor variable X that takes the value 1 if the risk factor is present and 0 if the risk factor is absent. The log-hazard function is given by

$$\log[\lambda(t|X)] = \log[\lambda_0(t)] + \beta_1 X$$

where $\lambda_0(t)$ is the baseline hazard function.

The hazard ratio $\psi$ is defined as the ratio of the hazard for those with the risk factor ($X = 1$) to the hazard without the risk factor ($X = 0$). The log of the hazard ratio is given by

$$\log(\psi) \equiv \log[\psi(X = 1, X = 0)] = \log[\lambda(t|X = 1)] - \log[\lambda(t|X = 0)] = \beta_1$$

In general, the hazard ratio can be computed by exponentiating the difference of the log-hazard between any two population profiles. This is the approach taken by the HAZARDRATIO statement, so the computations are available regardless of parameterization, interactions, and nestings. However, as shown in the preceding equation for $\log(\psi)$, hazard ratios of main effects can be computed as functions of the parameter estimates, and the remainder of this section is concerned with this methodology.

The parameter, $\beta_1$, associated with X represents the change in the log-hazard from $X = 0$ to $X = 1$. So the hazard ratio is obtained by simply exponentiating the value of the parameter associated with the risk factor. The hazard ratio indicates how the hazard change as you change $X$ from 0 to 1. For instance, $\psi = 2$ means that the hazard when $X = 1$ is twice the hazard when $X = 0$.

Suppose the values of the dichotomous risk factor are coded as constants $a$ and $b$ instead of 0 and 1. The hazard when $X = a$ becomes $\lambda(t)\exp(a\beta_1)$, and the hazard when $X = b$ becomes $\lambda(t)\exp(b\beta_1)$. The hazard ratio corresponding to an increase in $X$ from $a$ to $b$ is

$$\psi = \exp[(b - a)\beta_1] = [\exp(\beta_1)]^{b-a} \equiv [\exp(\beta_1)]^c$$

Note that for any $a$ and $b$ such that $c = b - a = 1, \psi = \exp(\beta_1)$. So the hazard ratio can be interpreted as the change in the hazard for any increase of one unit in the corresponding risk factor. However, the change in hazard for some amount other than one unit is often of greater interest. For example, a change of one pound in body weight might be too small to be considered important, while a change of 10 pounds might be more meaningful. The hazard ratio for a change in $X$ from $a$ to $b$ is estimated by raising the hazard ratio estimate for a unit change in $X$ to the power of $c = b - a$ as shown previously.

For a polytomous risk factor, the computation of hazard ratios depends on how the risk factor is parameterized. For illustration, suppose that Cell is a risk factor with four categories: Adeno, Large, Small, and Squamous.

For the effect parameterization scheme (PARAM=EFFECT) with Squamous as the reference group, the design variables for Cell are as follows:

| | Design Variables | | |
|---|---|---|---|
| **Cell** | $X_1$ | $X_2$ | $X_3$ |
| Adeno | 1 | 0 | 0 |
| Large | 0 | 1 | 0 |
| Small | 0 | 0 | 1 |
| Squamous | −1 | −1 | −1 |

The log-hazard for Adeno is

$$\begin{aligned}\log[\lambda(t|\text{Adeno})] &= \log[\lambda_0(t)] + \beta_1(X_1 = 1) + \beta_2(X_2 = 0) + \beta_3(X_3 = 0) \\ &= \lambda_0(t) + \beta_1\end{aligned}$$

The log-hazard for Squamous is

$$\begin{aligned}\log[\lambda(t|\text{Squamous})] &= \log[\lambda_0(t)] + \beta_1(X_1 = -1) + \beta_2(X_2 = -1) + \beta_3(X_3 = -1)) \\ &= \log[\lambda_0(t)] - \beta_1 - \beta_2 - \beta_3\end{aligned}$$

Therefore, the log-hazard ratio of Adeno versus Squamous

$$\begin{aligned}\log[\psi(\text{Adeno}, \text{Squamous})] &= \log[\lambda(t|\text{Adeno})] - \log[\lambda(t|\text{Squamous})] \\ &= 2\beta_1 + \beta_2 + \beta_3\end{aligned}$$

For the reference cell parameterization scheme (PARAM=REF) with Squamous as the reference cell, the design variables for Cell are as follows:

| | Design Variables | | |
|---|---|---|---|
| **Cell** | $X_1$ | $X_2$ | $X_3$ |
| Adeno | 1 | 0 | 0 |
| Large | 0 | 1 | 0 |
| Small | 0 | 0 | 1 |
| Squamous | 0 | 0 | 0 |

The log-hazard ratio of Adeno versus Squamous is given by

$$\begin{aligned}&\log(\psi(\text{Adeno}, \text{Squamous})) \\ &\quad= \log[\lambda(t|\text{Adeno})] - \log[\lambda(t|\text{Squamous})] \\ &\quad= (\log[\lambda_0(t)] + \beta_1(X_1 = 1) + \beta_2(X_2 = 0) + \beta_3(X_3 = 0)) - \\ &\quad\phantom{=}\ (\log[\lambda_0(t)] + \beta_1(X_1 = 0) + \beta_2(X_2 = 0) + \beta_3(X_3 = 0)) \\ &\quad= \beta_1\end{aligned}$$

For the GLM parameterization scheme (PARAM=GLM), the design variables are as follows:

| | Design Variables | | | |
|---|---|---|---|---|
| **Cell** | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
| Adeno | 1 | 0 | 0 | 0 |
| Large | 0 | 1 | 0 | 0 |
| Small | 0 | 0 | 1 | 0 |
| Squamous | 0 | 0 | 0 | 1 |

The log-hazard ratio of Adeno versus Squamous is

$$
\begin{aligned}
&\log(\psi(\text{Adeno}, \text{Squamous})) \\
&\quad = \log[\lambda(t|\text{Adeno})] - \log[\lambda(t|\text{Squamous})] \\
&\quad = \log[\lambda_0(t)] + \beta_1(X_1 = 1) + \beta_2(X_2 = 0) + \beta_3(X_3 = 0) + \beta_4(X_4 = 0)) - \\
&\qquad (\log(\lambda_0(t)) + \beta_1(X_1 = 0) + \beta_2(X_2 = 0) + \beta_3(X_3 = 0) + \beta_4(X_4 = 1)) \\
&\quad = \beta_1 - \beta_4
\end{aligned}
$$

Consider Cell as the only risk factor in the Cox regression in Example 64.3. The computation of hazard ratio of Adeno versus Squamous for various parameterization schemes is tabulated in Table 64.4.

**Table 64.4** Hazard Ratio Comparing Adeno to Squamous

| | Parameter Estimates | | | | |
|---|---|---|---|---|---|
| **PARAM=** | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\beta}_4$ | **Hazard Ratio Estimates** |
| EFFECT | 0.5772 | –0.2115 | 0.2454 | | $\exp(2 \times 0.5772 - 0.2115 + 0.2454) = 3.281$ |
| REF | 1.8830 | 0.3996 | 0.8565 | | $\exp(1.8830) = 3.281$ |
| GLM | 1.8830 | 0.3996 | 0.8565 | 0.0000 | $\exp(1.8830) = 3.281$ |

The fact that the log-hazard ratio ($\log(\psi)$) is a linear function of the parameters enables the HAZARDRATIO statement to compute the hazard ratio of the main effect even in the presence of interactions and nest effects. The section “Hazard Ratios” on page 4584 details the estimation of the hazard ratios in a classical analysis.

To customize hazard ratios for specific units of change for a continuous risk factor, you can use the UNITS= option in a HAZARDRATIO statement to specify a list of relevant units for each explanatory variable in the model. Estimates of these customized hazard ratios are given in a separate table. Let $(L_j, U_j)$ be a confidence interval for $\log(\psi)$. The corresponding lower and upper confidence limits for the customized hazard ratio $\exp(c\beta_j)$ are $\exp(cL_j)$ and $\exp(cU_j)$, respectively (for $c > 0$), or $\exp(cU_j)$ and $\exp(cL_j)$, respectively (for $c < 0$).

## Specifics for Classical Analysis

### Proportional Rates/Means Models for Recurrent Events

Let $N(t)$ be the number of events experienced by a subject over the time interval $(0, t]$. Let $dN(t)$ be the increment of the counting process $N$ over $[t, t + dt)$. The rate function is given by

$$
d\mu_{\mathbf{Z}}(t) = E[dN(t)|Z(t)] = e^{\boldsymbol{\beta}'\mathbf{Z}(t)} d\mu_0(t)
$$

where $\mu_0(.)$ is an unknown continuous function. If the $\mathbf{Z}$ are time independent, the rate model is reduced to the mean model

$$\mu_Z(t) = \mathrm{e}^{\boldsymbol{\beta}' Z} \mu_0(t)$$

The partial likelihood for $n$ independent triplets $(N_i, Y_i, \mathbf{Z}_i), i = 1, \ldots, n$, of counting, at-risk, and covariate processes is the same as that of the multiplicative hazards model. However, a robust sandwich estimate is used for the covariance matrix of the parameter estimator instead of the model-based estimate.

Let $T_{ki}$ be the $k$th event time of the $i$th subject. Let $C_i$ be the censoring time of the $i$th subject. The at-risk indicator and the failure indicator are, respectively,

$$Y_i(t) = I(C_i \geq t) \text{and} \Delta_{ki} = I(T_{ki} \leq C_i)$$

Denote

$$S^{(0)}(\boldsymbol{\beta}, t) = \sum_{i=1}^{n} Y_i(t) \mathrm{e}^{\boldsymbol{\beta}' \mathbf{Z}_i(t)} \text{and} \bar{\mathbf{Z}}(\boldsymbol{\beta}, t) = \frac{\sum_{i=1}^{n} Y_i(t) \mathrm{e}^{\boldsymbol{\beta}' \mathbf{Z}_i(t)} \mathbf{Z}_i(t)}{S^{(0)}(\boldsymbol{\beta}, t)}$$

Let $\hat{\boldsymbol{\beta}}$ be the maximum likelihood estimate of $\boldsymbol{\beta}$, and let $\mathcal{I}(\hat{\boldsymbol{\beta}})$ be the observed information matrix. The robust sandwich covariance matrix estimate is given by

$$\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}}) \sum_{i=1}^{n} \Big[ W_i(\hat{\boldsymbol{\beta}}) W_i'(\hat{\boldsymbol{\beta}}) \Big] \mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})$$

where

$$\begin{aligned} W_i(\boldsymbol{\beta}) \quad = \quad & \sum_k \Delta_{ki} \Big\{ Z_i(T_{ki}) - \bar{\mathbf{Z}}(\boldsymbol{\beta}, T_{ki}) \Big\} - \\ & \sum_{i=1}^{n} \sum_l \frac{\Delta_{lj} Y_i(T_{lj}) \mathrm{e}^{\boldsymbol{\beta}' \mathbf{Z}_i(T_{lj})}}{S^0(\boldsymbol{\beta}, T_{lj})} \Big\{ Z_i(T_{lj}) - \bar{\mathbf{Z}}(\boldsymbol{\beta}, T_{lj}) \Big\} \end{aligned}$$

For a given realization of the covariates $\boldsymbol{\xi}$, the Nelson estimator is used to predict the mean function

$$\hat{\mu}_{\boldsymbol{\xi}}(t) = \mathrm{e}^{\hat{\boldsymbol{\beta}}' \boldsymbol{\xi}} \sum_{i=1}^{n} \sum_k \frac{I(T_{ki} \leq t) \Delta_{ki}}{S^{(0)}(\hat{\boldsymbol{\beta}}, T_{ki})}$$

with standard error estimate given by

$$\hat{\sigma}^2(\hat{\mu}_{\boldsymbol{\xi}}(t)) = \sum_{i=1}^{n} \Big( \frac{1}{n} \hat{\Psi}_i(t, \boldsymbol{\xi}) \Big)^2$$

where

$$\begin{aligned} \frac{1}{n} \hat{\Psi}_i(\boldsymbol{\xi}, t) \quad = \quad & \mathrm{e}^{\hat{\boldsymbol{\beta}}' \boldsymbol{\xi}} \Bigg\{ \sum_k \frac{I(T_{ki} \leq t) \Delta_{ik}}{S^{(0)}(\hat{\boldsymbol{\beta}}, T_{ki})} - \sum_{j=1}^{n} \sum_k \frac{Y_i(T_{kj}) \mathrm{e}^{\hat{\boldsymbol{\beta}}' \mathbf{Z}_i(T_{kj})} I(T_{kj} \leq t) \Delta_{kj}}{[S^{(0)}(\hat{\boldsymbol{\beta}}, T_{kj})]^2} - \\ & \Bigg[ \sum_{i=1}^{n} \sum_k \frac{I(T_{ki} \leq t) \Delta_{ik} [\bar{\mathbf{Z}}(\hat{\boldsymbol{\beta}}, T_{ki}) - \boldsymbol{\xi}]}{S^{(0)}(\hat{\boldsymbol{\beta}}, T_{ki})} \Bigg] \\ & \times \mathcal{I}^{-1}(\hat{\boldsymbol{\beta}}) \int_0^{\tau} [\mathbf{Z}_i(u) - \bar{\mathbf{Z}}(\hat{\boldsymbol{\beta}}, u)] d\hat{M}_i(u) \Bigg\} \end{aligned}$$

Since the cumulative mean function is always nonnegative, the log transform is used to compute confidence intervals. The $100(1-\alpha)\%$ pointwise confidence limits for $\mu_{\boldsymbol{\xi}}(t)$ are

$$\hat{\mu}_{\boldsymbol{\xi}}(t)\mathrm{e}^{\pm z_{\alpha/2}\frac{\hat{\sigma}(\hat{\mu}_{\boldsymbol{\xi}}(t))}{\hat{\mu}_{\boldsymbol{\xi}}(t)}}$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

## Newton-Raphson Method

Let $L(\boldsymbol{\beta})$ be one of the likelihood functions described in the previous subsections. Let $l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$. Finding $\boldsymbol{\beta}$ such that $L(\boldsymbol{\beta})$ is maximized is equivalent to finding the solution $\hat{\boldsymbol{\beta}}$ to the likelihood equations

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0$$

With $\hat{\boldsymbol{\beta}}^0 = \mathbf{0}$ as the initial solution, the iterative scheme is expressed as

$$\hat{\boldsymbol{\beta}}^{j+1} = \hat{\boldsymbol{\beta}}^{j} - \left[\frac{\partial^2 l(\hat{\boldsymbol{\beta}}^j)}{\partial \boldsymbol{\beta}^2}\right]^{-1} \frac{\partial l(\hat{\boldsymbol{\beta}}^j)}{\partial \boldsymbol{\beta}}$$

The term after the minus sign is the Newton-Raphson step. If the likelihood function evaluated at $\hat{\boldsymbol{\beta}}^{j+1}$ is less than that evaluated at $\hat{\boldsymbol{\beta}}^{j}$, then $\hat{\boldsymbol{\beta}}^{j+1}$ is recomputed using half the step size. The iterative scheme continues until convergence is obtained—that is, until $\hat{\boldsymbol{\beta}}_{j+1}$ is sufficiently close to $\hat{\boldsymbol{\beta}}_{j}$. Then the maximum likelihood estimate of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{j+1}$.

The model-based variance estimate of $\hat{\boldsymbol{\beta}}$ is obtained by inverting the information matrix $\mathcal{I}(\hat{\boldsymbol{\beta}})$

$$\hat{\mathbf{V}}_m(\hat{\boldsymbol{\beta}}) = \mathcal{I}^{-1}(\hat{\boldsymbol{\beta}}) = -\left[\frac{\partial^2 l(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^2}\right]^{-1}$$

## Firth’s Modification for Maximum Likelihood Estimation

In fitting a Cox model, the phenomenon of monotone likelihood is observed if the likelihood converges to a finite value while at least one parameter diverges (Heinze and Schemper 2001).

Let $\mathbf{x}_l(t)$ denote the vector explanatory variables for the $l$th individual at time $t$. Let $t_1 < t_2 < \ldots < t_m$ denote the $k$ distinct, ordered event times. Let $d_j$ denote the multiplicity of failures at $t_j$; that is, $d_j$ is the size of the set $\mathcal{D}_j$ of individuals that fail at $t_j$. Let $\mathcal{R}_j$ denote the risk set just before $t_j$. Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ be the vector of regression parameters. The Breslow log partial likelihood is given by

$$l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta}) = \sum_{j=1}^{m} \left\{ \boldsymbol{\beta}' \sum_{l \in \mathcal{D}_j} \mathbf{x}_l(t_j) - d_j \log \sum_{h \in \mathcal{R}j} \mathrm{e}^{\boldsymbol{\beta}' \mathbf{x}_h(t_j)} \right\}$$

Denote

$$\mathbf{S}_j^{(a)}(\boldsymbol{\beta}) = \sum_{h \in \mathcal{R}j} \mathrm{e}^{\boldsymbol{\beta}'\mathbf{x}_h(t_j)}[\mathbf{x}_h(t_j)]^{\otimes a} \qquad a = 0, 1, 2$$

Then the score function is given by

$$\begin{aligned}
\mathbf{U}(\boldsymbol{\beta}) &\equiv (U(\beta_1), \ldots, U(\beta_k))' \\
&= \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\
&= \sum_{j=1}^{m} \left\{ \sum_{l \in \mathcal{D}j} \mathbf{x}_l(t_j) - d_j \frac{\mathbf{S}_j^{(1)}(\boldsymbol{\beta})}{S_j^{0}(\boldsymbol{\beta})} \right\}
\end{aligned}$$

and the Fisher information matrix is given by

$$\begin{aligned}
\mathcal{I}(\boldsymbol{\beta}) &= -\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \\
&= \sum_{j=1}^{m} d_j \left\{ \frac{\mathbf{S}_j^{(2)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})} - \left[\frac{\mathbf{S}_j^{(1)}(\boldsymbol{\beta})}{\mathbf{S}_j^{(0)}(\boldsymbol{\beta})}\right]\left[\frac{\mathbf{S}_j^{(1)}(\boldsymbol{\beta})}{\mathbf{S}_j^{(0)}(\boldsymbol{\beta})}\right]' \right\}
\end{aligned}$$

Heinze (1999); Heinze and Schemper (2001) applied the idea of Firth (1993) by maximizing the penalized partial likelihood

$$l^*(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) + 0.5 \log(|\mathcal{I}(\boldsymbol{\beta})|)$$

The score function $\mathbf{U}(\boldsymbol{\beta})$ is replaced by the modified score function by $\mathbf{U}^*(\boldsymbol{\beta}) \equiv (U^*(\beta_1), \ldots, U^*(\beta_k))'$, where

$$U^*(\beta_r) = U(\beta_r) + 0.5\mathrm{tr}\left\{\mathcal{I}^{-1}(\boldsymbol{\beta})\frac{\partial \mathcal{I}(\boldsymbol{\beta})}{\partial \beta_r}\right\} \qquad r = 1, \ldots, k$$

The Firth estimate is obtained iteratively as

$$\boldsymbol{\beta}^{(s+1)} = \boldsymbol{\beta}^{(s)} + \mathcal{I}^{-1}(\boldsymbol{\beta}^{(s)})\mathbf{U}^*(\boldsymbol{\beta}^{(s)})$$

The covariance matrix $\hat{\mathbf{V}}$ is computed as $\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}}$ is the maximum penalized partial likelihood estimate.

### Explicit formulae for $\frac{\partial \mathcal{I}(\boldsymbol{\beta})}{\partial \beta_r}, r = 1, \ldots, k$

Denote

$$\begin{aligned}
\mathbf{x}_h(t) &= (x_{h1}(t), \ldots, x_{hk}(t))' \\
\mathbf{Q}_{jr}^{(a)}(\boldsymbol{\beta}) &= \sum_{h \in \mathcal{R}j} \mathrm{e}^{\boldsymbol{\beta}'\mathbf{x}_h(t_j)} x_{hr}(t_j)[\mathbf{x}_h(t_j)]^{\otimes a} \qquad a = 0, 1, 2; r = 1, \ldots, k
\end{aligned}$$

Then

$$
\begin{aligned}
\frac{\partial \mathcal{I}(\boldsymbol{\beta})}{\partial \beta_r} &= \sum_{j=1}^{m} d_j \Biggl\{ \left[ \frac{\mathbf{Q}_{jr}^{(2)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})} - \frac{\mathbf{Q}_{jr}^{(0)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})} \frac{\mathbf{S}_j^{(2)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})} \right] - \\
& \left[ \frac{\mathbf{Q}_{jr}^{(1)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})} - \frac{\mathbf{Q}_{jr}^{(0)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})} \frac{\mathbf{S}_j^{(1)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})} \right] \left[ \frac{\mathbf{S}_j^{(1)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})} \right]' - \\
& \left[ \frac{\mathbf{S}_j^{(1)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})} \right] \left[ \frac{\mathbf{Q}_{jr}^{(1)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})} - \frac{\mathbf{Q}_{jr}^{(0)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})} \frac{\mathbf{S}_j^{(1)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})} \right]' \Biggr\} \qquad r = 1, \ldots, k
\end{aligned}
$$

## Robust Sandwich Variance Estimate

For the $i$th subject, $i = 1, \ldots, n$, let $X_i$, $w_i$, and $\mathbf{Z}_i(t)$ be the observed time, weight, and the covariate vector at time $t$, respectively. Let $\Delta_i$ be the event indicator and let $Y_i(t) = I(X_i \geq t)$. Let

$$
S^{(r)}(\boldsymbol{\beta}, t) = \sum_{j=1}^{n} w_j Y_j(t) e^{\boldsymbol{\beta}' \mathbf{Z}_j(t)} \mathbf{Z}_j^{\bigotimes r}(t), r = 0, 1
$$

Let $\bar{\mathbf{Z}}(\boldsymbol{\beta}, t) = \frac{S^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)}$. The score residual for the $i$th subject is

$$
\mathbf{L}_i(\boldsymbol{\beta}) = \Delta_i \left\{ \mathbf{Z}_i(X_i) - \bar{\mathbf{Z}}(\beta, X_i) \right\} - \sum_{j=1}^{n} \Delta_j \frac{w_j Y_i(X_j) e^{\boldsymbol{\beta}' \mathbf{Z}_i(X_j)}}{S^{(0)}(\boldsymbol{\beta}, X_j)} \left\{ \mathbf{Z}_i(X_j) - \bar{\mathbf{Z}}(\boldsymbol{\beta}, X_j) \right\}
$$

For TIES=EFRON, the computation of the score residuals is modified to comply with the Efron partial likelihood. See the section "Residuals" on page 4605 for more information.

The robust sandwich variance estimate of $\hat{\boldsymbol{\beta}}$ derived by Binder (1992), who incorporated weights into the analysis, is

$$
\hat{\mathbf{V}}_s(\hat{\boldsymbol{\beta}}) = \mathcal{I}^{-1}(\hat{\boldsymbol{\beta}}) \left[ \sum_{j=1}^{n} (w_j \mathbf{L}_j(\hat{\boldsymbol{\beta}}))^{\bigotimes 2} \right] \mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})
$$

where $\mathcal{I}(\hat{\boldsymbol{\beta}})$ is the observed information matrix, and $\mathbf{a}^{\bigotimes 2} = \mathbf{a}\mathbf{a}'$. Note that when $w_i \equiv 1$,

$$
\hat{\mathbf{V}}_s(\hat{\boldsymbol{\beta}}) = \mathbf{D}'\mathbf{D}
$$

where $\mathbf{D}$ is the matrix of DFBETA residuals. This robust variance estimate was proposed by Lin and Wei (1989) and Reid and Crèpeau (1985).

## Testing the Global Null Hypothesis

The following statistics can be used to test the global null hypothesis $H_0$: $\boldsymbol{\beta}$=$\mathbf{0}$. Under mild assumptions, each statistic has an asymptotic chi-square distribution with $p$ degrees of freedom given the null hypothesis. The value $p$ is the dimension of $\boldsymbol{\beta}$. For clustered data, the likelihood ratio test, the score test, and the Wald test assume independence of observations within a cluster, while the robust Wald test and the robust score test do not need such an assumption.

### Likelihood Ratio Test

$$\chi^2_{LR} = 2\left[l(\hat{\boldsymbol{\beta}}) - l(\mathbf{0})\right]$$

### Score Test

$$\chi^2_S = \left[\frac{\partial l(\mathbf{0})}{\partial \boldsymbol{\beta}}\right]' \left[-\frac{\partial^2 l(\mathbf{0})}{\partial \boldsymbol{\beta}^2}\right]^{-1} \left[\frac{\partial l(\mathbf{0})}{\partial \boldsymbol{\beta}}\right]$$

### Wald's Test

$$\chi^2_W = \hat{\boldsymbol{\beta}}' \left[-\frac{\partial^2 l(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^2}\right] \hat{\boldsymbol{\beta}}$$

### Robust Score Test

$$\chi^2_{RS} = \left[\sum_i \mathbf{L}_i^0\right]' \left[\sum_i \mathbf{L}_i^0 \mathbf{L}_i^{0\prime}\right]^{-1} \left[\sum_i \mathbf{L}_i^0\right]$$

where $\mathbf{L}_i^0$ is the score residual of the $i$th subject at $\boldsymbol{\beta}$=$\mathbf{0}$; that is, $\mathbf{L}_i^0$=$\mathbf{L}_i(\mathbf{0}, \infty)$, where the score process $\mathbf{L}_i(\boldsymbol{\beta}, t)$ is defined in the section "Residuals" on page 4605.

### Robust Wald's Test

$$\chi^2_{RW} = \hat{\boldsymbol{\beta}}'[\hat{\mathbf{V}}_s(\hat{\boldsymbol{\beta}})]^{-1}\hat{\boldsymbol{\beta}}$$

where $\hat{\mathbf{V}}_s(\hat{\boldsymbol{\beta}})$ is the sandwich variance estimate (see the section "Robust Sandwich Variance Estimate" on page 4591 for details).

## Confidence Limits for a Hazard Ratio

Let $\mathbf{e}_j$ be the $j$th unit vector—that is, the $j$th entry of the vector is 1 and all other entries are 0. The hazards ratio for the explanatory variable with regression coefficient $\beta_j = \mathbf{e}'_j\boldsymbol{\beta}$ is defined as $\exp(\beta_j)$. In general, a log-hazard ratio can be written as $\mathbf{h}'\boldsymbol{\beta}$, a linear combination of the regression coefficients, and the hazard ratio $\exp(\mathbf{h}'\boldsymbol{\beta})$ is obtained by replacing $\mathbf{e}_j$ with $\mathbf{h}$.

### Point Estimate

The hazards ratio $\exp \mathbf{e}'_j \boldsymbol{\beta}$ is estimated by $\exp(\mathbf{e}'_j \hat{\beta})$, where $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimate of the $\boldsymbol{\beta}$.

### Wald's Confidence Limits

The $100(1-\alpha)\%$ confidence limits for the hazard ratio are calculated as

$$\exp\left(\mathbf{e}'_j \hat{\beta} \pm z_{\alpha/2}\sqrt{\mathbf{e}'_j \hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})\mathbf{e}_j}\right)$$

where $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})$ is estimated covariance matrix, and $z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentile point of the standard normal distribution.

### Profile-Likelihood Confidence Limits

The construction of the profile-likelihood-based confidence interval is derived from the asymptotic $\chi^2$ distribution of the generalized likelihood ratio test of Venzon and Moolgavkar (1988). Suppose that the parameter vector is $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ and you want to compute a confidence interval for $\beta_j$. The profile-likelihood function for $\beta_j = \gamma$ is defined as

$$l_j^*(\gamma) = \max_{\boldsymbol{\beta} \in \mathcal{B}_j(\gamma)} l(\boldsymbol{\beta})$$

where $\mathcal{B}_j(\gamma)$ is the set of all $\boldsymbol{\beta}$ with the $j$th element fixed at $\gamma$, and $l(\boldsymbol{\beta})$ is the log-likelihood function for $\boldsymbol{\beta}$. If $l_{\max} = l(\hat{\boldsymbol{\beta}})$ is the log likelihood evaluated at the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$, then $2(l_{\max} - l_j^*(\beta_j))$ has a limiting chi-square distribution with one degree of freedom if $\beta_j$ is the true parameter value. Let $l_0 = l_{\max} - 0.5\chi_1^2(1-\alpha)$, where $\chi_1^2(1-\alpha)$ is the $100(1-\alpha)$ percentile of the chi-square distribution with one degree of freedom. A $100(1-\alpha)\%$ confidence interval for $\beta_j$ is

$$\{\gamma : l_j^*(\gamma) \geq l_0\}$$

The endpoints of the confidence interval are found by solving numerically for values of $\beta_j$ that satisfy equality in the preceding relation. To obtain an iterative algorithm for computing the confidence limits, the log-likelihood function in a neighborhood of $\boldsymbol{\beta}$ is approximated by the quadratic function

$$\tilde{l}(\boldsymbol{\beta} + \boldsymbol{\delta}) = l(\boldsymbol{\beta}) + \boldsymbol{\delta}'\mathbf{g} + \frac{1}{2}\boldsymbol{\delta}'\mathbf{V}\boldsymbol{\delta}$$

where $\mathbf{g} = \mathbf{g}(\boldsymbol{\beta})$ is the gradient vector and $\mathbf{V} = \mathbf{V}(\boldsymbol{\beta})$ is the Hessian matrix. The increment $\boldsymbol{\delta}$ for the next iteration is obtained by solving the likelihood equations

$$\frac{d}{d\boldsymbol{\delta}}\{\tilde{l}(\boldsymbol{\beta} + \boldsymbol{\delta}) + \lambda(\mathbf{e}'_j\boldsymbol{\delta} - \gamma)\} = \mathbf{0}$$

where $\lambda$ is the Lagrange multiplier, $\mathbf{e}_j$ is the $j$th unit vector, and $\gamma$ is an unknown constant. The solution is

$$\boldsymbol{\delta} = -\mathbf{V}^{-1}(\mathbf{g} + \lambda \mathbf{e}_j)$$

By substituting this $\boldsymbol{\delta}$ into the equation $\tilde{l}(\boldsymbol{\beta} + \boldsymbol{\delta}) = l_0$, you can estimate $\lambda$ as

$$\lambda = \pm \left( \frac{2(l_0 - l(\boldsymbol{\beta}) + \frac{1}{2}\mathbf{g}'\mathbf{V}^{-1}\mathbf{g})}{\mathbf{e}_j'\mathbf{V}^{-1}\mathbf{e}_j} \right)^{\frac{1}{2}}$$

The upper confidence limit for $\beta_j$ is computed by starting at the maximum likelihood estimate of $\boldsymbol{\beta}$ and iterating with positive values of $\lambda$ until convergence is attained. The process is repeated for the lower confidence limit, using negative values of $\lambda$.

Convergence is controlled by value $\epsilon$ specified with the PLCONV= option in the MODEL statement (the default value of $\epsilon$ is 1E–4). Convergence is declared on the current iteration if the following two conditions are satisfied:

$$|l(\boldsymbol{\beta}) - l_0| \le \epsilon$$

and

$$(\mathbf{g} + \lambda \mathbf{e}_j)'\mathbf{V}^{-1}(\mathbf{g} + \lambda \mathbf{e}_j) \le \epsilon$$

The profile-likelihood confidence limits for the hazard ratio $\exp(\mathbf{e}_j'\boldsymbol{\beta})$ are obtained by exponentiating these confidence limits.

## Testing Linear Hypotheses about Regression Coefficients

Linear hypotheses for $\boldsymbol{\beta}$ are expressed in matrix form as

$$H_0\colon \mathbf{L}\boldsymbol{\beta} = \mathbf{c}$$

where $\mathbf{L}$ is a matrix of coefficients for the linear hypotheses, and $\mathbf{c}$ is a vector of constants. The Wald chi-square statistic for testing $H_0$ is computed as

$$\chi^2_W = \left(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{c}\right)' \left[\mathbf{L}\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})\mathbf{L}'\right]^{-1} \left(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{c}\right)$$

where $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})$ is the estimated covariance matrix. Under $H_0$, $\chi^2_W$ has an asymptotic chi-square distribution with $r$ degrees of freedom, where $r$ is the rank of $\mathbf{L}$.

### *Optimal Weights for the AVERAGE option in the TEST Statement*

Let $\boldsymbol{\beta}_0 = (\beta_{i_1}, \ldots, \beta_{i_s})'$, where $\{\beta_{i_1}, \ldots, \beta_{i_s}\}$ is a subset of $s$ regression coefficients. For any vector $\mathbf{e} = (e_1, \ldots, e_s)'$ of length $s$,

$$\mathbf{e}'\hat{\boldsymbol{\beta}}_0 \sim N(\mathbf{e}'\boldsymbol{\beta}_0, \mathbf{e}'\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_0)\mathbf{e})$$

To find $\mathbf{e}$ such that $\mathbf{e}'\hat{\boldsymbol{\beta}}_0$ has the minimum variance, it is necessary to minimize $\mathbf{e}'\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_0)\mathbf{e}$ subject to $\sum_{i=1}^{k} e_i = 1$. Let $\mathbf{1}_s$ be a vector of 1's of length $s$. The expression to be minimized is

$$\mathbf{e}'\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_0)\mathbf{e} - \lambda(\mathbf{e}'\mathbf{1}_s - 1)$$

where $\lambda$ is the Lagrange multipler. Differentiating with respect to $\mathbf{e}$ and $\lambda$, respectively, yields

$$\begin{aligned} \hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_0)\mathbf{e} - \lambda\mathbf{1}_s &= \mathbf{0} \\ \mathbf{e}'\mathbf{1}_s - 1 &= 0 \end{aligned}$$

Solving these equations gives

$$\mathbf{e} = [\mathbf{1}_s'\hat{\mathbf{V}}^{-1}(\hat{\boldsymbol{\beta}}_0)\mathbf{1}_s]^{-1}\hat{\mathbf{V}}^{-1}(\hat{\boldsymbol{\beta}}_0)\mathbf{1}_s$$

This provides a one degree-of-freedom test for testing the null hypothesis $H_0 : \beta_{i_1} = \ldots = \beta_{i_s} = 0$ with normal test statistic

$$Z = \frac{\mathbf{e}'\hat{\boldsymbol{\beta}}_0}{\sqrt{\mathbf{e}'\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_0)\mathbf{e}}}$$

This test is more sensitive than the multivariate test specified by the TEST statement

```
Multivariate: test X1, ..., Xs;
```

where X1, ..., X$s$ are the variables with regression coefficients $\beta_{i_1}, \ldots, \beta_{i_s}$, respectively.

## Analysis of Multivariate Failure Time Data

Multivariate failure time data arise when each study subject can potentially experience several events (for instance, multiple infections after surgery) or when there exists some natural or artificial clustering of subjects (for instance, a litter of mice) that induces dependence among the failure times of the same cluster. Data in the former situation are referred to as multiple events data, and data in the latter situation are referred to as clustered data. The multiple events data can be further classified into ordered and unordered data. For ordered data, there is a natural ordering of the multiple failures within a subject, which includes recurrent events data as a special case. For unordered data, the multiple event times result from several concurrent failure processes.

Multiple events data can be analyzed by the Wei, Lin, and Weissfeld (1989), or WLW, method based on the marginal Cox models. For the special case of recurrent events data, you can fit the intensity model (Andersen and Gill 1982), the proportional rates/means model (Pepe and Cai 1993; Lawless and Nadeau 1995; Lin et al. 2000), or the stratified models for total time and gap time proposed by Prentice, Williams, and Peterson (1981), or PWP. For clustered data, you can carry out the analysis of Lee, Wei, and Amato (1992) based on the marginal Cox model. To use PROC PHREG to perform these analyses correctly and effectively, you have to array your data in a specific way to produce the correct risk sets.

All examples described in this section can be found in the program *phrmult.sas* in the SAS/STAT sample library. Furthermore, the "Examples" section in this chapter contains two examples to illustrate the methods of analyzing recurrent events data and clustered data.

### Marginal Cox Models for Multiple Events Data

Suppose there are $n$ subjects and each subject can experience up to $K$ potential events. Let $\mathbf{Z}_{ki}(.)$ be the covariate process associated with the $k$th event for the $i$th subject. The marginal Cox models are given by

$$\lambda_k(t;\mathbf{Z}_{ki}) = \lambda_{k0} e^{\boldsymbol{\beta}_k' \mathbf{Z}_{ki}(t)}, k = 1, \ldots, K; i = 1, \ldots, n$$

where $\lambda_{k0}(t)$ is the (event-specific) baseline hazard function for the $k$th event and $\boldsymbol{\beta}_k$ is the (event-specific) column vector of regression coefficients for the $k$th event. WLW estimates $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$ by the maximum partial likelihood estimates $\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_K$, respectively, and uses a robust sandwich covariance matrix estimate for $(\hat{\boldsymbol{\beta}}_1', \ldots, \hat{\boldsymbol{\beta}}_K')'$ to account for the dependence of the multiple failure times.

By using a properly prepared input data set, you can estimate the regression parameters for all the marginal Cox models and compute the robust sandwich covariance estimates in one PROC PHREG invocation. For convenience of discussion, suppose each subject can potentially experience $K$=3 events and there are two explanatory variables Z1 and Z2. The event-specific parameters to be estimated are $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{21})'$ for the first marginal model, $\boldsymbol{\beta}_2 = (\beta_{12}, \beta_{22})'$ for the second marginal model, and $\boldsymbol{\beta}_3 = (\beta_{13}, \beta_{23})'$ for the third marginal model. Inference of these parameters is based on the robust sandwich covariance matrix estimate of the parameter estimators. It is necessary that each row of the input data set represent the data for a potential event of a subject. The input data set should contain the following:

- an ID variable for identifying the subject so that all observations of the same subject have the same ID value
- an Enum variable to index the multiple events. For example, Enum=1 for the first event, Enum=2 for the second event, and so on.
- a Time variable to represent the observed time from some time origin for the event. For recurrence events data, it is the time from the study entry to each recurrence.
- a Status variable to indicate whether the Time value is a censored or uncensored time. For example, Status=1 indicates an uncensored time and Status=0 indicates a censored time.
- independent variables (Z1 and Z2)

The WLW analysis can be carried out by specifying the following:

```
proc phreg covs(aggregate);
   model Time*Status(0)=Z11 Z12 Z13 Z21 Z22 Z23;
   strata Enum;
   id ID;
   Z11= Z1 * (Enum=1);
   Z12= Z1 * (Enum=2);
   Z13= Z1 * (Enum=3);
   Z21= Z2 * (Enum=1);
   Z22= Z2 * (Enum=2);
   Z23= Z2 * (Enum=3);
   run;
```

The variable Enum is specified in the STRATA statement so that there is one marginal Cox model for each distinct value of Enum. The variables Z11, Z12, Z13, Z21, Z22, and Z23 in the MODEL statement are event-specific variables derived from the independent variables Z1 and Z2 by the given programming statements. In particular, the variables Z11, Z12, and Z13 are event-specific variables for the explanatory variable Z1; the variables Z21, Z22, and Z23 are event-specific variables for the explanatory variable Z2. For $j = 1, 2$, and $k = 1, 2, 3$, variable Zjk contains the same values as the explanatory variable Zj for the rows that correspond to $k$th marginal model and the value 0 for all other rows; as such, $\beta_{jk}$ is the regression coefficient for Zjk. You can avoid using the programming statements in PROC PHREG if you create these event-specific variables in the input data set by using the same programming statements in a DATA step.

The option COVS(AGGREGATE) is specified in the PROC statement to obtain the robust sandwich estimate of the covariance matrix, and the score residuals used in computing the middle part of the sandwich estimate are aggregated over identical ID values. You can also include TEST statements in the PROC PHREG code to test various linear hypotheses of the regression parameters based on the robust sandwich covariance matrix estimate.

Consider the AIDS study data in Wei, Lin, and Weissfeld (1989) from a randomized clinical trial to assess the antiretroviral capacity of ribavirin over time in AIDS patients. Blood samples were collected at weeks 4, 8, and 12 from each patient in three treatment groups (placebo, low dose of ribavirin, and high dose). For each serum sample, the failure time is the number of days before virus positivity was detected. If the sample was contaminated or it took a longer period of time than was achievable in the laboratory, the sample was censored. For example:

- Patient #1 in the placebo group has uncensored times 9, 6, and 7 days (that is, it took 9 days to detect viral positivity in the first blood sample, 6 days for the second blood sample, and 7 days for the third blood sample).
- Patient #14 in the low-dose group of ribavirin has uncensored times of 16 and 17 days for the first and second sample, respectively, and a censored time of 21 days for the third blood sample.
- Patient #28 in the high-dose group has an uncensored time of 21 days for the first sample, no measurement for the second blood sample, and a censored time of 25 days for the third sample.

For a full-rank parameterization, two design variables are sufficient to represent three treatment groups. Based on the reference coding with placebo as the reference, the values of the two dummy explanatory variables Z1 and Z2 representing the treatments are as follows:

| Treatment Group | Z1 | Z2 |
|---|---|---|
| Placebo | 0 | 0 |
| Low dose ribavirin | 1 | 0 |
| High dose ribavirin | 0 | 1 |

The bulk of the task in using PROC PHREG to perform the WLW analysis lies in the preparation of the input data set. As discussed earlier, the input data set should contain the ID, Enum, Time, and Status variables, and event-specific independent variables Z11, Z12, Z13, Z21, Z22, and Z23. Data for the three patients described earlier are arrayed as follows:

| ID | Time | Status | Enum | Z1 | Z2 |
|---|---|---|---|---|---|
| 1 | 9 | 1 | 1 | 0 | 0 |
| 1 | 6 | 1 | 2 | 0 | 0 |
| 1 | 7 | 1 | 3 | 0 | 0 |
| | | | | | |
| 14 | 16 | 1 | 1 | 1 | 0 |
| 14 | 17 | 1 | 2 | 1 | 0 |
| 14 | 21 | 0 | 3 | 1 | 0 |
| | | | | | |
| 28 | 21 | 1 | 1 | 0 | 1 |
| 28 | 25 | 0 | 3 | 0 | 1 |

The first three rows are data for Patient #1 with event times at 9, 6, and 7 days, one row for each event. The next three rows are data for Patient #14, who has an uncensored time of 16 days for the first serum sample, an uncensored time of 17 days for the second sample, and a censored time of 21 days for the third sample. The last two rows are data for Patient #28 of the high-dose group (Z1=0 and Z2=1). Since the patient did not have a second serum sample, there are only two rows of data.

To perform the WLW analysis, you specify the following statements:

```
proc phreg covs(aggregate);
   model Time*Status(0)=Z11 Z12 Z13 Z21 Z22 Z23;
   strata Enum;
   id ID;
   Z11= Z1 * (Enum=1);
   Z12= Z1 * (Enum=2);
   Z13= Z1 * (Enum=3);
   Z21= Z2 * (Enum=1);
   Z22= Z2 * (Enum=2);
   Z23= Z2 * (Enum=3);
   EqualLowDose: test Z11=Z12, Z12=Z23;
   AverageLow: test Z11,Z12,Z13 / e average;
run;
```

Two linear hypotheses are tested using the TEST statements. The specification

```
   EqualLowDose: test Z11=Z12, Z12=Z13;
```

tests the null hypothesis $\beta_{11} = \beta_{12} = \beta_{13}$ of identical low-dose effects across three marginal models. The specification

```
 AverageLow: test Z11,Z12,Z13 / e average;
```

tests the null hypothesis of no low-dose effects (that is, $\beta_{11} = \beta_{12} = \beta_{13} = 0$). The AVERAGE option computes the optimal weights for estimating the average low-dose effect $\beta_1^* = \beta_{11} = \beta_{12} = \beta_{13}$ and performs a 1 DF test for testing the null hypothesis that $\beta_1^* = 0$. The E option displays the coefficients for the linear hypotheses, including the optimal weights.

### Marginal Cox Models for Clustered Data

Suppose there are $n$ clusters with $K_i$ members in the $i$th cluster, $i = 1, \ldots, n$. Let $\mathbf{Z}_{ki}(.)$ be the covariate process associated with the $k$th member of the $i$th cluster. The marginal Cox model is given by

$$\lambda(t;\mathbf{Z}_{ki}) = \lambda_0(t)\mathrm{e}^{\boldsymbol{\beta}'\mathbf{Z}_{ki}(t)} k = 1,\ldots,K_i; i = 1,\ldots,n$$

where $\lambda_0(t)$ is an arbitrary baseline hazard function and $\boldsymbol{\beta}$ is the vector of regression coefficients. Lee, Wei, and Amato (1992) estimate $\boldsymbol{\beta}$ by the maximum partial likelihood estimate $\hat{\boldsymbol{\beta}}$ under the independent working assumption, and use a robust sandwich covariance estimate to account for the intracluster dependence.

To use PROC PHREG to analyze the clustered data, each member of a cluster is represented by an observation in the input data set. The input data set to PROC PHREG should contain the following:

- an ID variable to identify the cluster so that members of the same cluster have the same ID value
- a Time variable to represent the observed survival time of a member of a cluster
- a Status variable to indicate whether the Time value is an uncensored or censored time. For example, Status=1 indicates an uncensored time and Status=0 indicates a censored time.
- the explanatory variables thought to be related to the failure time

Consider a tumor study in which one of three female rats of the same litter was randomly subjected to a drug treatment. The failure time is the time from randomization to the detection of tumor. If a rat died before the tumor was detected, the failure time was censored. For example:

- In litter #1, the drug-treated rat has an uncensored time of 101 days, one untreated rat has a censored time of 49 days, and the other untreated rat has a failure time of 104 days.
- In litter #3, the drug-treated rat has a censored time of 104 days, one untreated rat has a censored time of 102 days, and the other untreated rat has a censored time of 104 days.

In this example, a litter is a cluster and the rats of the same litter are members of the cluster. Let Trt be a 0-1 variable representing the treatment a rat received, with value 1 for drug treatment and 0 otherwise. Data for the two litters of rats described earlier contribute six observations to the input data set:

| Litter | Time | Status | Trt |
|---|---|---|---|
| 1 | 101 | 1 | 1 |
| 1 | 49 | 0 | 0 |
| 1 | 104 | 1 | 0 |
| | | | |
| 3 | 104 | 0 | 1 |
| 3 | 102 | 0 | 0 |
| 3 | 104 | 0 | 0 |

The analysis of Lee, Wei, and Amato (1992) can be performed by PROC PHREG as follows:

```
proc phreg covs(aggregate);
   model Time*Status(0)=Treatment;
   id Litter;
run;
```

### Intensity and Rate/Mean Models for Recurrent Events Data

Suppose each subject experiences recurrences of the same phenomenon. Let $N(t)$ be the number of events a subject experiences over the interval [0,$t$] and let $\mathbf{Z}(.)$ be the covariate process of the subject.

The intensity model (Andersen and Gill 1982) is given by

$$\lambda_{\mathbf{Z}}(t)dt = E\{dN(t)|\mathcal{F}_{t-}\} = \lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{Z}(t)}dt$$

where $\mathcal{F}_t$ represents all the information of the processes $N$ and $\mathbf{Z}$ up to time $t$, $\lambda_0(t)$ is an arbitrary baseline intensity function, and $\boldsymbol{\beta}$ is the vector of regression coefficients. This model consists of two components: (1) all the influence of the prior events on future recurrences, if there is any, is mediated through the time-dependent covariates, and (2) the covariates have multiplicative effects on the instantaneous rate of the counting process. If the covariates are time invariant, the risk of recurrences is unaffected by the past events.

The proportional rates and means models (Pepe and Cai 1993; Lawless and Nadeau 1995; Lin et al. 2000) assume that the covariates have multiplicative effects on the mean and rate functions of the counting process. The rate function is given by

$$d\mu_{\mathbf{Z}}(t) = E\{dN(t)|\mathbf{Z}(t)\} = e^{\boldsymbol{\beta}'\mathbf{Z}(t)}d\mu_0(t)$$

where $\mu_0(t)$ is an unknown continuous function and $\boldsymbol{\beta}$ is the vector of regression parameters. If $\mathbf{Z}$ is time invariant, the mean function is given by

$$\mu_{\mathbf{Z}}(t) = E\{N(t)|\mathbf{Z}\} = e^{\boldsymbol{\beta}'\mathbf{Z}}\mu_0(t)$$

For both the intensity and the proportional rates/means models, estimates of the regression coefficients are obtained by solving the partial likelihood score function. However, the covariance matrix estimate for the intensity model is computed as the inverse of the observed information matrix, while that for the proportional rates/means model is given by a sandwich estimate. For a given pattern of fixed covariates, the Nelson estimate for the cumulative intensity function is the same for the cumulative mean function, but their standard errors are not the same.

To fit the intensity or rate/mean model by using PROC PHREG, the counting process style of input is needed. A subject with $K$ events contributes $K$+1 observations to the input data set. The $k$th observation of the subject identifies the time interval from the $(k-1)$th event or time 0 (if $k = 1$) to the $k$th event, $k = 1, \ldots, K$. The $(K+1)$th observation represents the time interval from the $K$th event to time of censorship. The input data set should contain the following variables:

- a TStart variable to represent the $(k-1)$th recurrence time or the value 0 if $k = 1$
- a TStop variable to represent the $k$th recurrence time or the follow-up time if $k = K + 1$
- a Status variable indicating whether the TStop time is a recurrence time or a censored time; for example, Status=1 for a recurrence time and Status=0 for censored time
- explanatory variables thought to be related to the recurrence times

If the rate/mean model is used, the input data should also contain an ID variable for identifying the subjects.

Consider the chronic granulomatous disease (CGD) data listed in Fleming and Harrington (1991). The disease is a rare disorder characterized by recurrent pyrogenic infections. The study is a placebo-controlled randomized clinical trial conducted by the International CGD Cooperative Study to assess the effect of gamma interferon to reduce the rate of infection. For each study patient the times of recurrent infections along with a number of prognostic factors were collected. For example:

- Patient #17404, age 38, in the gamma interferon group had a follow-up time of 293 without any infection.
- Patient #204001, age 12, in the placebo group had an infection at 219 days, a recurrent infection at 373 days, and was followed up to 414 days.

Let Trt be the variable representing the treatment status with value 1 for gamma interferon and value 2 for placebo. Let Age be a covariate representing the age of the CGD patient. Data for the two CGD patients described earlier are given in the following table.

| ID | TStart | TStop | Status | Trt | Age |
|---|---|---|---|---|---|
| 174054 | 0 | 293 | 0 | 1 | 38 |
| | | | | | |
| 204001 | 0 | 219 | 1 | 2 | 12 |
| 204001 | 219 | 373 | 1 | 2 | 12 |
| 204001 | 373 | 414 | 0 | 2 | 12 |

Since Patient #174054 had no infection through the end of the follow-up period (293 days), there is only one observation representing the period from time 0 to the end of the follow-up. Data for Patient #204001 are broken into three observations, since there are two infections. The first observation represents the period from time 0 to the first infection, the second observation represents the period from the first infection to the second infection, and the third time period represents the period from the second infection to the end of the follow-up.

The following specification fits the intensity model:

```
proc phreg;
   model (TStart,TStop)*Status(0)=Trt Age;
   run;
```

You can predict the cumulative intensity function for a given pattern of fixed covariates by specifying the CUMHAZ= option in the BASELINE statement. Suppose you are interested in two fixed patterns, one for patients of age 30 in the gamma interferon group and the other for patients of age 1 in the placebo group. You first create the SAS data set as follows:

```
data Pattern;
   Trt=1; Age=30;
   output;
   Trt=2; Age=1;
   output;
   run;
```

You then include the following BASELINE statement in the PROC PHREG specification. The CUMHAZ=_all_ option produces the cumulative hazard function estimates, the standard error estimates, and the lower and upper pointwise confidence limits.

```
baseline covariates=Pattern out=out1 cumhaz=_all_;
```

The following specification of PROC PHREG fits the mean model and predicts the cumulative mean function for the two patterns of covariates in the Pattern data set:

```
proc phreg covs(aggregate);
   model (Tstart,Tstop)*Status(0)=Trt Age;
   baseline covariates=Pattern out=out2 cmf=_all_;
   id ID;
```

The COV(AGGREGATE) option, along with the ID statement, computes the robust sandwich covariance matrix estimate. The CMF=_ALL_ option adds the cumulative mean function estimates, the standard error estimates, and the lower and upper pointwise confidence limits to the OUT=Out2 data set.

### PWP Models for Recurrent Events Data

Let $N(t)$ be the number of events a subject experiences by time $t$. Let $\mathbf{Z}(t)$ be the covariate vectors of the subject at time $t$. For a subject who has $K$ events before censorship takes place, let $t_0 = 0$, let $t_k$ be the $k$th recurrence time, $k = 1, \ldots, K$, and let $t_{K+1}$ be the censored time. Prentice, Williams, and Peterson (1981) consider two time scales, a total time from the beginning of the study and a gap time from immediately preceding failure. The PWP models are stratified Cox-type models that allow the shape of the hazard function to depend on the number of preceding events and possibly on other characteristics of $\{N(t)$ and $\mathbf{Z}(t)\}$. The total time and gap time models are given, respectively, as follows:

$$\lambda(t|\mathcal{F}_{t-}) = \lambda_{0k}(t)e^{\boldsymbol{\beta}_k'\mathbf{Z}(t)}, \qquad t_{k-1} < t \le t_k$$

$$\lambda(t|\mathcal{F}_{t-}) = \lambda_{0k}(t - t_{k-1})e^{\boldsymbol{\beta}_k'\mathbf{Z}(t)}, \qquad t_{k-1} < t \le t_k$$

where $\lambda_{0k}$ is an arbitrary baseline intensity functions, and $\boldsymbol{\beta}_k$ is a vector of stratum-specific regression coefficients. Here, a subject moves to the $k$th stratum immediately after his $(k-1)$th recurrence time and remains there until the $k$th recurrence occurs or until censorship takes place. For instance, a subject who experiences only one event moves from the first stratum to the second stratum after the event occurs and remains in the second stratum until the end of the follow-up.

You can use PROC PHREG to carry out the analyses of the PWP models, but you have to prepare the input data set to provide the correct risk sets. The input data set for analyzing the total time is the same as the AG model with an additional variable to represent the stratum that the subject is in. A subject with $K$ events contributes $K$+1 observations to the input data set, one for each stratum that the subject moves to. The input data should contain the following variables:

- a TStart variable to represent the $(k-1)$th recurrence time or the value 0 if $k = 1$
- a TStop variable to represent the $k$th recurrence time or the time of censorship if $k = K + 1$
- a Status variable with value 1 if the Time value is a recurrence time and value 0 if the Time value is a censored time

- an Enum variable representing the index of the stratum that the subject is in. For a subject who has only one event at $t_1$ and is followed to time $t_c$, Enum=1 for the first observation (where Time=$t_1$ and Status=1) and Enum=2 for the second observation (where Time=$t_c$ and Status=0).
- explanatory variables thought to be related to the recurrence times

To analyze gap times, the input data set should also include a GapTime variable that is equal to (TStop − TStart).

Consider the data of two subjects in CGD data described in the previous section:

- Patients #174054, age 38, in the gamma interferon group had a follow-up time of 293 without any infection.
- Patient #204001, age 12, in the placebo group had an infection at 219 days, a recurrent infection at 373 days, and a follow-up time of 414 days.

To illustrate, suppose all subjects have at most two observed events. The data for the two subjects in the input data set are as follows:

| ID | TStart | TStop | Gaptime | Status | Enum | Trt | Age |
|---|---|---|---|---|---|---|---|
| 174054 | 0 | 293 | 293 | 0 | 1 | 1 | 38 |
| | | | | | | | |
| 204001 | 0 | 219 | 219 | 1 | 1 | 2 | 12 |
| 204001 | 219 | 373 | 154 | 1 | 2 | 2 | 12 |
| 204001 | 373 | 414 | 41 | 0 | 3 | 2 | 12 |

Subject #174054 contributes only one observation to the input data, since there is no observed event. Subject #204001 contributes three observations, since there are two observed events.

To fit the total time model of PWP with stratum-specific slopes, either you can create the stratum-specific explanatory variables (Trt1, Trt2, and Trt3 for Trt, and Age1, Age2, and Age3 for Age) in a DATA step, or you can specify them in PROC PHREG by using programming statements as follows:

```
proc phreg;
   model (TStart,TStop)*Status(0)=Trt1 Trt2 Trt3 Age1 Age2 Age3;
   strata Enum;
   Trt1= Trt * (Enum=1);
   Trt2= Trt * (Enum=2);
   Trt3= Trt * (Enum=3);
   Age1= Age * (Enum=1);
   Age2= Age * (Enum=2);
   Age3= Age * (Enum=3);
   run;
```

To fit the total time model of PWP with the common regression coefficients, you specify the following:

```
proc phreg;
   model (TStart,TStop)*Status(0)=Trt Age;
   strata Enum;
   run;
```

To fit the gap time model of PWP with stratum-specific regression coefficients, you specify the following:

```
proc phreg;
   model Gaptime*Status(0)=Trt1 Trt2 Trt3 Age1 Age2 Age3;
   strata Enum;
   Trt1= Trt * (Enum=1);
   Trt2= Trt * (Enum=2);
   Trt3= Trt * (Enum=3);
   Age1= Age * (Enum=1);
   Age2= Age * (Enum=2);
   Age3= Age * (Enum=3);
run;
```

To fit the gap time model of PWP with common regression coefficients, you specify the following:

```
proc phreg;
   model Gaptime*Status(0)=Trt Age;
   strata Enum;
   run;
```

## Model Fit Statistics

Suppose the model contains $p$ regression parameters. Let $\Delta_j$ and $f_j$ be the event indicator and the frequency, respectively, of the $j$th observation. The three criteria displayed by the PHREG procedure are calculated as follows:

- $-2$ Log Likelihood:

  $$-2 \text{ Log L} = -2 \log(L_n(\hat{\boldsymbol{\beta}}))$$

  where $L_n(.)$ is a partial likelihood function for the corresponding TIES= option as described in the section "Partial Likelihood Function for the Cox Model" on page 4581, and $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimate of the regression parameter vector.

- Akaike Information Criterion:

  $$\text{AIC} = -2 \text{ Log L} + 2p$$

- Schwarz Bayesian (Information) Criterion:

  $$\text{SBC} = -2 \text{ Log L} + p \log(\sum_j f_j \Delta_j)$$

The $-2$ Log Likelihood statistic has a chi-square distribution under the null hypothesis (that all the explanatory effects in the model are zero) and the procedure produces a $p$-value for this statistic.

The AIC and SBC statistics give two different ways of adjusting the $-2$ Log Likelihood statistic for the number of terms in the model and the number of observations used. These statistics should be used when comparing different models for the same data (for example, when you use the METHOD=STEPWISE option in the MODEL statement); lower values of the statistic indicate a more desirable model.

## Residuals

This section describes the computation of residuals (RESMART=, RESDEV=, RESSCH=, and RESSCO=) in the OUTPUT statement.

First, consider TIES=BRESLOW. Let

$$
\begin{aligned}
S^{(0)}(\boldsymbol{\beta}, t) &= \sum_i Y_i(t) \mathrm{e}^{\boldsymbol{\beta}' \mathbf{Z}_i(t)} \\
S^{(1)}(\boldsymbol{\beta}, t) &= \sum_i Y_i(t) \mathrm{e}^{\boldsymbol{\beta}' \mathbf{Z}_i(t)} \mathbf{Z}_i(t) \\
\bar{\mathbf{Z}}(\boldsymbol{\beta}, t) &= \frac{S^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} \\
d\Lambda_0(\boldsymbol{\beta}, t) &= \sum_i \frac{dN_i(t)}{S^{(0)}(\boldsymbol{\beta}, t)} \\
dM_i(\boldsymbol{\beta}, t) &= dN_i(t) - Y_i(t) \mathrm{e}^{\boldsymbol{\beta}' \mathbf{Z}_i(t)} d\Lambda_0(\boldsymbol{\beta}, t)
\end{aligned}
$$

The martingale residual at $t$ is defined as

$$
\hat{M}_i(t) = \int_0^t dM_i(\hat{\boldsymbol{\beta}}, s) = N_i(t) - \int_0^t Y_i(s) \mathrm{e}^{\hat{\boldsymbol{\beta}}' \mathbf{Z}_i(s)} d\Lambda_0(\hat{\boldsymbol{\beta}}, s)
$$

Here $\hat{M}_i(t)$ estimates the difference over $(0, t]$ between the observed number of events for the $i$th subject and a conditional expected number of events. The quantity $\hat{M}_i \equiv \hat{M}_i(\infty)$ is referred to as the martingale residual for the $i$th subject. When the counting process MODEL specification is used, the RESMART= variable contains the component $(\hat{M}_i(t_2) - \hat{M}_i(t_1))$ instead of the martingale residual at $t_2$. The martingale residual for a subject can be obtained by summing up these component residuals within the subject. For the Cox model with no time-dependent explanatory variables, the martingale residual for the $i$th subject with observation time $t_i$ and event status $\Delta_i$ is

$$
\hat{M}_i = \Delta_i - \mathrm{e}^{\hat{\boldsymbol{\beta}}' \mathbf{Z}_i} \int_0^{t_i} d\Lambda_0(\hat{\boldsymbol{\beta}}, s)
$$

The deviance residuals $D_i$ are a transform of the martingale residuals:

$$
D_i = \mathrm{sign}(\hat{M}_i) \sqrt{2 \left[ -\hat{M}_i - N_i(\infty) \log \left( \frac{N_i(\infty) - \hat{M}_i}{N_i(\infty)} \right) \right]}
$$

The square root shrinks large negative martingale residuals, while the logarithmic transformation expands martingale residuals that are close to unity. As such, the deviance residuals are more symmetrically distributed around zero than the martingale residuals. For the Cox model, the deviance residual reduces to the form

$$D_i = \mathrm{sign}(\hat{M}_i)\sqrt{2[-\hat{M}_i - \Delta_i \log(\Delta_i - \hat{M}_i)]}$$

When the counting process MODEL specification is used, values of the RESDEV= variable are set to missing because the deviance residuals can be calculated only on a per-subject basis.

The Schoenfeld (1982) residual vector is calculated on a per-event-time basis. At the $j$th event time $t_{i_j}$ of the $i$th subject, the Schoenfeld residual

$$\hat{\mathbf{U}}_i(t_{i_j}) = \mathbf{Z}_i(t_{i_j}) - \bar{\mathbf{Z}}(\hat{\boldsymbol{\beta}}, t_{i_j})$$

is the difference between the $i$th subject covariate vector at $t_{i_j}$ and the average of the covariate vectors over the risk set at $t_{i_j}$. Under the proportional hazards assumption, the Schoenfeld residuals have the sample path of a random walk; therefore, they are useful in assessing time trend or lack of proportionality. Harrell (1886) proposed a $z$-transform of the Pearson correlation between these residuals and the rank order of the failure time as a test statistic for nonproportional hazards. Therneau, Grambsch, and Fleming (1990) considered a Kolmogorov-type test based on the cumulative sum of the residuals.

The score process for the $i$th subject at time $t$ is

$$\mathbf{L}_i(\boldsymbol{\beta}, t) = \int_0^t [\mathbf{Z}_i(s) - \bar{\mathbf{Z}}(\boldsymbol{\beta}, s)] dM_i(\boldsymbol{\beta}, s)$$

The vector $\hat{\mathbf{L}}_i \equiv \mathbf{L}_i(\hat{\boldsymbol{\beta}}, \infty)$ is the score residual for the $i$th subject. When the counting process MODEL specification is used, the RESSCO= variables contain the components of $(\mathbf{L}_i(\hat{\boldsymbol{\beta}}, t_2) - \mathbf{L}_i(\hat{\boldsymbol{\beta}}, t_1))$ instead of the score process at $t_2$. The score residual for a subject can be obtained by summing up these component residuals within the subject.

The score residuals are a decomposition of the first partial derivative of the log likelihood. They are useful in assessing the influence of each subject on individual parameter estimates. They also play an important role in the computation of the robust sandwich variance estimators of Lin and Wei (1989) and Wei, Lin, and Weissfeld (1989).

For TIES=EFRON, the preceding computation is modified to comply with the Efron partial likelihood. Consider an uncensored time t. For a given time $t$, let $\Delta_i(t)$=1 if the $t$ is an event time of the $i$th subject and 0 otherwise. Let $d(t) = \sum_i \Delta_i(t)$, which is the number of subjects that have an event at $t$. For $1 \le k \le d(t)$, let

$$
\begin{aligned}
S^{(0)}(\boldsymbol{\beta}, k, t) &= \sum_i Y_i(t) \left\{ 1 - \frac{k-1}{d(t)} \Delta_i(t) \right\} \mathrm{e}^{\boldsymbol{\beta}' \mathbf{Z}_i(t)} \\
S^{(1)}(\boldsymbol{\beta}, k, t) &= \sum_i Y_i(t) \left\{ 1 - \frac{k-1}{d(t)} \Delta_i(t) \right\} \mathrm{e}^{\boldsymbol{\beta}' \mathbf{Z}_i(t)} \mathbf{Z}_i(t) \\
\bar{\mathbf{Z}}(\boldsymbol{\beta}, k, t) &= \frac{S^{(1)}(\boldsymbol{\beta}, k, t)}{S^{(0)}(\boldsymbol{\beta}, k, t)} \\
d\Lambda_0(\boldsymbol{\beta}, k, t) &= \sum_i \frac{dN_i(t)}{S^{(0)}(\boldsymbol{\beta}, k, t)} \\
dM_i(\boldsymbol{\beta}, k, t) &= dN_i(t) - Y_i(t) \left( 1 - \Delta_i(t) \frac{k-1}{d(t)} \right) \mathrm{e}^{\boldsymbol{\beta}' \mathbf{Z}_i(t)} d\Lambda_0(\boldsymbol{\beta}, k, t)
\end{aligned}
$$

The martingale residual at $t$ for the $i$th subject is defined as

$$
\hat{M}_i(t) = \int_0^t \frac{1}{d(s)} \sum_{k=1}^{d(s)} dM_i(\hat{\boldsymbol{\beta}}, k, s) = N_i(t) - \int_0^t \frac{1}{d(s)} \sum_{k=1}^{d(s)} Y_i(s) \left( 1 - \Delta_i(s) \frac{k-1}{d(s)} \right) \mathrm{e}^{\hat{\boldsymbol{\beta}}' \mathbf{Z}_i(s)} d\Lambda_0(\hat{\boldsymbol{\beta}}, k, s)
$$

Deviance residuals are computed by using the same transform on the corresponding martingle residuals as in TIES=BRESLOW.

The Schoenfeld residual vector for the $i$th subject at event time $t_{i_j}$ is

$$
\hat{\mathbf{U}}_i(t_{i_j}) = \mathbf{Z}_i(t_{i_j}) - \frac{1}{d(t_{i_j})} \sum_{k=1}^{d(t_{i_j})} \bar{\mathbf{Z}}(\hat{\boldsymbol{\beta}}, k, t_{i_j})
$$

The score process for the $i$th subject at time $t$ is given by

$$
\mathbf{L}_i(\boldsymbol{\beta}, t) = \int_0^t \frac{1}{d(s)} \sum_{k=1}^{d(s)} \left( \mathbf{Z}_i(s) - \bar{\mathbf{Z}}(\boldsymbol{\beta}, k, s) \right) dM_i(\boldsymbol{\beta}, k, s)
$$

For TIES=DISCRETE or TIES=EXACT, it is difficult to come up with modifications that are consistent with the correponding partial likelihood. Residuals for these TIES= methods are computed by using the same formulae as in TIES=BRESLOW.

## Diagnostics Based on Weighted Residuals

The vector of weighted Schoenfeld residuals, $\mathbf{r}_i$, is computed as

$$
\mathbf{r}_i = n_e \mathcal{I}^{-1}(\hat{\boldsymbol{\beta}}) \hat{\mathbf{U}}_i(t_i)
$$

where $n_e$ is the total number of events and $\hat{\mathbf{U}}_i(t_i)$ is the vector of Schoenfeld residuals at the event time $t_i$. The components of $\mathbf{r}_i$ are output to the WTRESSCH= variables.

The weighted Schoenfeld residuals are useful in assessing the proportional hazards assumption. The idea is that most of the common alternatives to the proportional hazards can be cast in terms of a time-varying coefficient model

$$\lambda(t, \mathbf{Z}) = \lambda_0(t) \exp(\beta_1(t) Z_1 + \beta_2(t) Z_2 + \ldots)$$

where $\lambda(t, \mathbf{Z})$ and $\lambda_0(t)$ are hazard rates. Let $\hat{\beta}_j$ and $r_{ij}$ be the $j$th component of $\hat{\boldsymbol{\beta}}$ and $\mathbf{r}_i$, respectively. Grambsch and Therneau (1994) suggest using a smoothed plot of $(\hat{\beta}_j + r_{ij})$ versus $t_i$ to discover the functional form of the time-varying coefficient $\beta_j(t)$. A zero slope indicates that the coefficient is not varying with time.

The weighted score residuals are used more often than their unscaled counterparts in assessing local influence. Let $\hat{\boldsymbol{\beta}}_{(i)}$ be the estimate of $\boldsymbol{\beta}$ when the $i$th subject is left out, and let $\delta\hat{\boldsymbol{\beta}}_i = \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}$. The $j$th component of $\delta\hat{\boldsymbol{\beta}}_i$ can be used to assess any untoward effect of the $i$th subject on $\hat{\beta}_j$. The exact computation of $\delta\hat{\boldsymbol{\beta}}_i$ involves refitting the model each time a subject is omitted. Cain and Lange (1984) derived the following approximation of $\boldsymbol{\Delta}_i$ as weighted score residuals:

$$\delta\hat{\boldsymbol{\beta}}_i = \mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})\hat{\mathbf{L}}_i$$

Here, $\hat{\mathbf{L}}_i$ is the vector of the score residuals for the $i$th subject. Values of $\delta\hat{\boldsymbol{\beta}}_i$ are output to the DFBETA= variables. Again, when the counting process MODEL specification is used, the DFBETA= variables contain the component $\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})(\mathbf{L}_i(\hat{\boldsymbol{\beta}}, t_2) - \mathbf{L}_i(\hat{\boldsymbol{\beta}}, t_1))$, where the score process $\mathbf{L}_i(\boldsymbol{\beta}, t)$ is defined in the section "Residuals" on page 4605. The vector $\delta\hat{\boldsymbol{\beta}}_i$ for the $i$th subject can be obtained by summing these components within the subject.

Note that these DFBETA statistics are a transform of the score residuals. In computing the robust sandwich variance estimators of Lin and Wei (1989) and Wei, Lin, and Weissfeld (1989), it is more convenient to use the DFBETA statistics than the score residuals (see Example 64.10).

## Influence of Observations on Overall Fit of the Model

The LD statistic approximates the likelihood displacement, which is the amount by which minus twice the log likelihood ($-2 \log L(\hat{\boldsymbol{\beta}})$), under a fitted model, changes when each subject in turn is left out. When the $i$th subject is omitted, the likelihood displacement is

$$2 \log L(\hat{\boldsymbol{\beta}}) - 2 \log L(\hat{\boldsymbol{\beta}}_{(i)})$$

where $\hat{\boldsymbol{\beta}}_{(i)}$ is the vector of parameter estimates obtained by fitting the model without the $i$th subject. Instead of refitting the model without the $i$th subject, Pettitt and Bin Daud (1989) propose that the likelihood displacement for the $i$th subject be approximated by

$$LD_i = \hat{\mathbf{L}}_i' \mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})\hat{\mathbf{L}}_i$$

wher $\hat{\mathbf{L}}_i$ is the score residual vector of the $i$th subject. This approximation is output to the LD= variable.

The LMAX statistic is another global influence statistic. This statistic is based on the symmetric matrix

$$\mathbf{B} = \mathbf{L}\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})\mathbf{L}'$$

where $\mathbf{L}$ is the matrix with rows that are the score residual vectors $\hat{\mathbf{L}}_i$. The elements of the eigenvector associated with the largest eigenvalue of the matrix $\mathbf{B}$, standardized to unit length, give a measure of the sensitivity of the fit of the model to each observation in the data. The influence of the $i$th subject on the global fit of the model is proportional to the magnitude of $\zeta_i$, where $\zeta_i$ is the $i$th element of the vector $\boldsymbol{\zeta}$ that satisfies

$$\mathbf{B}\boldsymbol{\zeta} = \lambda_{\max}\boldsymbol{\zeta} \text{ and } \boldsymbol{\zeta}'\boldsymbol{\zeta} = 1$$

with $\lambda_{\max}$ being the largest eigenvalue of $\mathbf{B}$. The sign of $\zeta_i$ is irrelevant, and its absolute value is output to the LMAX= variable.

When the counting process MODEL specification is used, the LD= and LMAX= variables are set to missing, because these two global influence statistics can be calculated on a per-subject basis only.

## Survivor Function Estimation for the Cox Model

Two estimators of the survivor function are available: one is the product-limit estimator (Kalbfleisch and Prentice 1980, pp. 84–86) and the other is the Breslow (1972) estimator based on the empirical cumulative hazard function.

### *Product-Limit Estimates*

Let $\mathcal{C}_i$ denote the set of individuals censored in the half-open interval $[t_i, t_{i+1})$, where $t_0 = 0$ and $t_{k+1} = \infty$. Let $\gamma_l$ denote the censoring times in $[t_i, t_{i+1})$; $l$ ranges over $\mathcal{C}_i$ .

The likelihood function for all individuals is given by

$$\mathcal{L} = \prod_{i=0}^{k} \left\{ \prod_{l \in \mathcal{D}_i} \left( [S_0(t_i)]^{\exp(\mathbf{z}_l'\boldsymbol{\beta})} - [S_0(t_i + 0)]^{\exp(\mathbf{z}_l'\boldsymbol{\beta})} \right) \prod_{l \in \mathcal{C}_i} [S_0(\gamma_l + 0)]^{\exp(\mathbf{z}_l'\boldsymbol{\beta})} \right\}$$

where $\mathcal{D}_0$ is empty. The likelihood $\mathcal{L}$ is maximized by taking $S_0(t) = S_0(t_i + 0)$ for $t_i < t \le t_{i+1}$ and allowing the probability mass to fall only on the observed event times $t_1, \ldots, t_k$. By considering a discrete model with hazard contribution $1-\alpha_i$ at $t_i$, you take $S_0(t_i) = S_0(t_{i-1}+0) = \prod_{j=0}^{i-1} \alpha_j$, where $\alpha_0 = 1$. Substitution into the likelihood function produces

$$\mathcal{L} = \prod_{i=0}^{k} \left\{ \prod_{j \in \mathcal{D}_i} \left( 1 - \alpha_i^{\exp(\mathbf{z}_j'\boldsymbol{\beta})} \right) \prod_{l \in \mathcal{R}_i - \mathcal{D}_i} \alpha_i^{\exp(\mathbf{z}_l'\boldsymbol{\beta})} \right\}$$

If you replace $\boldsymbol{\beta}$ with $\hat{\boldsymbol{\beta}}$ estimated from the partial likelihood function and then maximize with respect to $\alpha_1, \ldots, \alpha_k$, the maximum likelihood estimate $\hat{\alpha}_i$ of $\alpha_i$ becomes a solution of

$$\sum_{j \in \mathcal{D}_i} \frac{\exp(\mathbf{z}_j'\hat{\boldsymbol{\beta}})}{1 - \hat{\alpha}_i^{\exp(\mathbf{z}_j'\hat{\boldsymbol{\beta}})}} = \sum_{l \in \mathcal{R}_i} \exp(\mathbf{z}_l'\hat{\boldsymbol{\beta}})$$

When only a single failure occurs at $t_i$, $\hat{\alpha}_i$ can be found explicitly. Otherwise, an iterative solution is obtained by the Newton method.

The estimated baseline cumulative hazard function is

$$\hat{H}_0(t) = -\log(\hat{S}_0(t))$$

where $\hat{S}_0(t)$ is the estimated baseline survivor function given by

$$\hat{S}_0(t) = \hat{S}_0(t_{i-1}+0) = \prod_{j=0}^{i-1} \hat{\alpha}_j, t_{i-1} < t \le t_i$$

For details, refer to Kalbfleisch and Prentice (1980). For a given realization of the explanatory variables $\boldsymbol{\xi}$, the product-limit estimate of the survival function at $\mathbf{Z} = \boldsymbol{\xi}$ is

$$\hat{S}(t, \boldsymbol{\xi}) = [\hat{S}_0(t)]^{\exp(\boldsymbol{\beta}'\boldsymbol{\xi})}$$

### Empirical Cumulative Hazards Function Estimates

Let $\boldsymbol{\xi}$ be a given realization of the explanatory variables. The empirical cumulative hazard function estimate at $\mathbf{Z} = \boldsymbol{\xi}$ is

$$\hat{\Lambda}(t, \boldsymbol{\xi}) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\sum_{j=1}^n Y_j(s) \exp(\hat{\boldsymbol{\beta}}'(\mathbf{z}_j - \boldsymbol{\xi}))}$$

The variance estimator of $\hat{\Lambda}(t, \boldsymbol{\xi})$ is given by the following (Tsiatis 1981):

$$\begin{aligned} & v\hat{a}r\{n^{\frac{1}{2}}(\hat{\Lambda}(t, \boldsymbol{\xi}) - \Lambda(t, \boldsymbol{\xi}))\} \\ = \ & n\left\{ \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{[\sum_{j=1}^n Y_j(s) \exp(\hat{\boldsymbol{\beta}}'(\mathbf{z}_j - \boldsymbol{\xi}))]^2} + \mathbf{H}'(t, \boldsymbol{\xi})\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})\mathbf{H}(t, \boldsymbol{\xi}) \right\} \end{aligned}$$

where $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})$ is the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$ and

$$\mathbf{H}(t, \boldsymbol{\xi}) = \sum_{i=1}^{n} \int_0^t \frac{\sum_{l=1}^n Y_l(s)(\mathbf{Z}_l - \boldsymbol{\xi}) \exp(\hat{\boldsymbol{\beta}}'(\mathbf{z}_l - \boldsymbol{\xi}))}{[\sum_{j=1}^n Y_j(s) \exp(\hat{\boldsymbol{\beta}}'(\mathbf{z}_j - \boldsymbol{\xi}))]^2} dN_i(s)$$

For the marginal model, the variance estimator computation follows Spiekerman and Lin (1998).

The empirical cumulative hazard function (CH) estimate of the survivor function for $\mathbf{Z} = \boldsymbol{\xi}$ is

$$\tilde{S}(t, \boldsymbol{\xi}) = \exp(-\hat{\Lambda}(t, \boldsymbol{\xi}))$$

### Confidence Intervals for the Survivor Function

Let $\hat{S}(t, \boldsymbol{\xi})$ and $\tilde{S}(t, \boldsymbol{\xi})$ correspond to the product-limit (PL) and empirical cumulative hazard function (CH) estimates of the survivor function for $\mathbf{Z} = \boldsymbol{\xi}$, respectively. Both the standard error of $\log(\hat{S}(t, \boldsymbol{\xi}))$ and the standard error of $\log(\tilde{S}(t, \boldsymbol{\xi}))$ are approximated by $\tilde{\sigma}_0(t, \boldsymbol{\xi})$, which is the square root of the variance estimate of $\hat{\Lambda}(t, \boldsymbol{\xi})$; refer to Kalbfleisch and Prentice (1980, p. 116). By the delta method, the standard errors of $\hat{S}(t, \boldsymbol{\xi})$ and $\tilde{S}(t, \boldsymbol{\xi})$ are given by

$$\hat{\sigma}_1(t, \boldsymbol{\xi}) \doteq \hat{S}(t, \boldsymbol{\xi})\tilde{\sigma}_0(t, \boldsymbol{\xi}) \quad \text{and} \quad \tilde{\sigma}_1(t, \boldsymbol{\xi}) \doteq \tilde{S}(t, \boldsymbol{\xi})\tilde{\sigma}_0(t, \boldsymbol{\xi})$$

respectively. The standard errors of log[–log($\hat{S}(t, \boldsymbol{\xi})$)] and log[–log($\tilde{S}(t, \boldsymbol{\xi})$)] are given by

$$\hat{\sigma}_2(t, \boldsymbol{\xi}) \doteq \frac{-\tilde{\sigma}_0(t, \boldsymbol{\xi})}{\log(\hat{S}(t, \boldsymbol{\xi}))} \quad \text{and} \quad \tilde{\sigma}_2(t, \boldsymbol{\xi}) \doteq \frac{\tilde{\sigma}_0(t, \boldsymbol{\xi})}{\hat{\Lambda}(t, \boldsymbol{\xi})}$$

respectively.

Let $z_{\alpha/2}$ be the upper $100(1-\frac{\alpha}{2})$ percentile point of the standard normal distribution. A $100(1-\alpha)\%$ confidence interval for the survivor function $S(t, \boldsymbol{\xi})$ is given in the following table.

| CLTYPE | Method | Confidence Limits |
|---|---|---|
| LOG | PL | $\exp[\log(\hat{S}(t, \boldsymbol{\xi})) \pm z_{\frac{\alpha}{2}}\tilde{\sigma}_0(t, \boldsymbol{\xi})]$ |
| LOG | CH | $\exp[\log(\tilde{S}(t, \boldsymbol{\xi})) \pm z_{\frac{\alpha}{2}}\tilde{\sigma}_0(t, \boldsymbol{\xi})]$ |
| LOGLOG | PL | $\exp\{-\exp[\log(-\log(\hat{S}(t, \boldsymbol{\xi}))) \pm z_{\frac{\alpha}{2}}\hat{\sigma}_2(t, \boldsymbol{\xi})]\}$ |
| LOGLOG | CH | $\exp\{-\exp[\log(-\log(\tilde{S}(t, \boldsymbol{\xi}))) \pm z_{\frac{\alpha}{2}}\tilde{\sigma}_2(t, \boldsymbol{\xi})]\}$ |
| NORMAL | PL | $\hat{S}(t, \boldsymbol{\xi}) \pm z_{\frac{\alpha}{2}}\hat{\sigma}_1(t, \boldsymbol{\xi})$ |
| NORMAL | CH | $\tilde{S}(t, \boldsymbol{\xi}) \pm z_{\frac{\alpha}{2}}\tilde{\sigma}_1(t, \boldsymbol{\xi})$ |

## Effect Selection Methods

Five variable selection methods are available. The simplest method (and the default) is SELECTION=NONE, for which PROC PHREG fits the complete model as specified in the MODEL statement. The other four methods are FORWARD for forward selection, BACKWARD for backward elimination, STEPWISE for stepwise selection, and SCORE for best subsets selection. These methods are specified with the SELECTION= option in the MODEL statement.

When SELECTION=FORWARD, PROC PHREG first estimates parameters for variables forced into the model. These variables are the first $n$ effects in the MODEL statement, where $n$ is the number specified by the START= or INCLUDE= option in the MODEL statement ($n$ is zero by default). Next, the procedure computes the adjusted chi-square statistics for each variable not in the model and examines the largest of these statistics. If it is significant at the SLSENTRY= level, the corresponding variable is added to the model. Once a variable is entered in the model, it is never removed from the model. The process is repeated until none of the remaining variables meet the specified level for entry or until the STOP= value is reached.

When SELECTION=BACKWARD, parameters for the complete model as specified in the MODEL statement are estimated unless the START= option is specified. In that case, only the parameters for the first $n$ effects in the MODEL statement are estimated, where $n$ is the number specified by the START= option. Results of the Wald test for individual parameters are examined. The least significant variable that does not meet the SLSSTAY= level for staying in the model is removed. Once a variable is removed from the model, it remains excluded. The process is repeated until no other variable in the model meets the specified level for removal or until the STOP= value is reached.

The SELECTION=STEPWISE option is similar to the SELECTION=FORWARD option except that variables already in the model do not necessarily remain. Variables are entered into and removed from the model in such a way that each forward selection step can be followed by one or more backward elimination steps. The stepwise selection process terminates if no further variable can be added to the model or if the variable just entered into the model is the only variable removed in the subsequent backward elimination.

For SELECTION=SCORE, PROC PHREG uses the branch-and-bound algorithm of Furnival and Wilson (1974) to find a specified number of models with the highest likelihood score (chi-square) statistic for all possible model sizes, from 1, 2, or 3 variables, and so on, up to the single model containing all of the explanatory variables. The number of models displayed for each model size is controlled by the BEST= option. You can use the START= option to impose a minimum model size, and you can use the STOP= option to impose a maximum model size. For instance, with BEST=3, START=2, and STOP=5, the SCORE selection method displays the best three models (that is, the three models with the highest score chi-squares) containing 2, 3, 4, and 5 variables. One of the limitations of the branch-and-bound algorithm is that it works only when each explanatory effect contains exactly one parameter—the SELECTION=SCORE option is not allowed when an explanatory effect in the MODEL statement contains a CLASS variable.

The SEQUENTIAL and STOPRES options can alter the default criteria for adding variables to or removing variables from the model when they are used with the FORWARD, BACKWARD, or STEPWISE selection method.

## Assessment of the Proportional Hazards Model

The proportional hazards model specifies that the hazard function for the failure time $T$ associated with a $p \times 1$ column covariate vector $\mathbf{Z}$ takes the form

$$\lambda(t;\mathbf{Z}) = \lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{Z}}$$

where $\lambda_0(.)$ is an unspecified baseline hazard function and $\boldsymbol{\beta}$ is a $p \times 1$ column vector of regression parameters. Lin, Wei, and Ying (1993) present graphical and numerical methods for model assessment based on the cumulative sums of martingale residuals and their transforms over certain coordinates (such as covariate values or follow-up times). The distributions of these stochastic processes under the assumed model can be approximated by the distributions of certain zero-mean Gaussian processes whose realizations can be generated by simulation. Each observed residual pattern can then be compared, both graphically and numerically, with a number of realizations from the null distribution. Such comparisons enable you to assess objectively whether the observed residual pattern reflects anything beyond random fluctuation. These procedures are useful in determining appropriate functional forms of covariates and assessing the proportional hazards assumption. You use the ASSESS statement to carry out these model-checking procedures.

For a sample of $n$ subjects, let $(X_i, \Delta_i, \mathbf{Z}_i)$ be the data of the $i$th subject; that is, $X_i$ represents the observed failure time, $\Delta_i$ has a value of 1 if $X_i$ is an uncensored time and 0 otherwise, and $\mathbf{Z}_i = (Z_{1i}, \ldots, Z_{pi})'$ is a $p$-vector of covariates. Let $N_i(t) = \Delta_i I(X_i \le t)$ and $Y_i(t) = I(X_i \ge t)$. Let

$$S^{(0)}(\boldsymbol{\beta}, t) = \sum_{i=1}^{n} Y_i(t)e^{\boldsymbol{\beta}'\mathbf{Z}_i} \quad \text{and} \quad \bar{\mathbf{Z}}(\boldsymbol{\beta}, t) = \frac{\sum_{i=1}^{n} Y_i(t)e^{\boldsymbol{\beta}'\mathbf{Z}_i}\mathbf{Z}_i}{S^{(0)}(\boldsymbol{\beta}, t)}$$

Let $\hat{\boldsymbol\beta}$ be the maximum partial likelihood estimate of $\boldsymbol\beta$, and let $\mathcal{I}(\hat{\boldsymbol\beta})$ be the observed information matrix.

The martingale residuals are defined as

$$\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(u)\mathrm{e}^{\hat{\boldsymbol\beta}'\mathbf{Z}_i} d\hat{\Lambda}_0(u), i = 1, \ldots, n$$

where $\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{S^{(0)}(\hat{\boldsymbol\beta},u)}$.

The empirical score process $\mathbf{U}(\hat{\boldsymbol\beta}, t) = (U_1(\hat{\boldsymbol\beta}, t), \ldots, U_p(\hat{\boldsymbol\beta}, t))'$ is a transform of the martingale residuals:

$$\mathbf{U}(\hat{\boldsymbol\beta}, t) = \sum_{i=1}^n \mathbf{Z}_i \hat{M}_i(t)$$

### Checking the Functional Form of a Covariate

To check the functional form of the $j$th covariate, consider the partial-sum process of $\hat{M}_i = \hat{M}_i(\infty)$:

$$W_j(z) = \sum_{i=1}^n I(Z_{ji} \le z)\hat{M}_i$$

Under that null hypothesis that the model holds, $W_j(z)$ can be approximated by the zero-mean Gaussian process

$$\begin{aligned}\hat{W}_j(z) \quad = \quad & \sum_{l=1}^n \Delta_l \left\{ I(Z_{jl} \le z) - \frac{\sum_{i=1}^n Y_i(X_l)\mathrm{e}^{\boldsymbol\beta'\mathbf{Z}_i} I(Z_{ij} \le z)}{S^{(0)}(\hat{\boldsymbol\beta}, X_l)} \right\} G_l - \\ & \sum_{k=1}^n \int_0^\infty Y_k(s)\mathrm{e}^{\hat{\boldsymbol\beta}'\mathbf{Z}_k} I(Z_{jk} \le z)[\mathbf{Z}_k - \bar{\mathbf{Z}}(\hat{\boldsymbol\beta}, s)]' d\hat{\Lambda}_0(s) \\ & \times\mathcal{I}^{-1}(\hat{\boldsymbol\beta}) \sum_{l=1}^n \Delta_l[\mathbf{Z}_l - \bar{\mathbf{Z}}(\hat{\boldsymbol\beta}, X_l)]G_l\end{aligned}$$

where $(G_1, \ldots, G_n)$ are independent standard normal variables that are independent of $(X_i, \Delta_i, \mathbf{Z}_i)$, $i = 1, \ldots, n$.

You can assess the functional form of the $j$th covariate by plotting a small number of realizations (the default is 20) of $\hat{W}_j(z)$ on the same graph as the observed $W_j(z)$ and visually comparing them to see how typical the observed pattern of $W_j(z)$ is of the null distribution samples. You can supplement the graphical inspection method with a Kolmogorov-type supremum test. Let $s_j$ be the observed value of $S_j = \sup_z |W_j(z)|$ and let $\hat{S}_j = \sup_z |\hat{W}_j(z)|$. The $p$-value $\Pr(S_j \ge s_j)$ is approximated by $\Pr(\hat{S}_j \ge s_j)$, which in turn is approximated by generating a large number of realizations (1000 is the default) of $\hat{W}_j(.)$.

### Checking the Proportional Hazards Assumption

Consider the standardized empirical score process for the $j$th component of $\mathbf{Z}$

$$U_j^*(t) = [\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})_{jj}]^{\frac{1}{2}} U_j(\hat{\boldsymbol{\beta}}, t),$$

Under the null hypothesis that the model holds, $U_j^*(t)$ can be approximated by

$$\begin{aligned}
\hat{U}_j^*(t) \quad = \quad & [\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})_{jj}]^{\frac{1}{2}} \Bigg\{ \sum_{l=1}^{n} I(X_l \le t)\Delta_l [Z_{jl} - \bar{Z}_j(\hat{\boldsymbol{\beta}}, t)] G_l - \\
& \sum_{k=1}^{n} \int_0^t Y_k(s) e^{\hat{\boldsymbol{\beta}}'\mathbf{Z}_k} Z_{jk} [\mathbf{Z}_k - \bar{\mathbf{Z}}(\hat{\boldsymbol{\beta}}, s)]' d\hat{\Lambda}_0(s) \\
& \times \mathcal{I}^{-1}(\hat{\boldsymbol{\beta}}) \sum_{l=1}^{n} \Delta_l [\mathbf{Z}_l - \bar{\mathbf{Z}}(\hat{\boldsymbol{\beta}}, X_l)] G_l \Bigg\}
\end{aligned}$$

where $\bar{Z}_j(\hat{\boldsymbol{\beta}}, t)$ is the $j$th component of $\bar{\mathbf{Z}}(\hat{\boldsymbol{\beta}}, t)$, and $(G_1, \ldots, G_n)$ are independent standard normal variables that are independent of $(X_i, \Delta_i, \mathbf{Z}_i, (i = 1, \ldots, n)$.

You can assess the proportional hazards assumption for the $j$th covariate by plotting a few realizations of $\hat{U}_j^*(t)$ on the same graph as the observed $U_j^*(t)$ and visually comparing them to see how typical the observed pattern of $U_j^*(t)$ is of the null distribution samples. Again you can supplement the graphical inspection method with a Kolmogorov-type supremum test. Let $s_j^*$ be the observed value of $S_j^* = \sup_t |U_j^*(t)|$ and let $\hat{S}_j^* = \sup_t |\hat{U}_j^*(t)|$. The $p$-value $\Pr[S_j^* \ge s_j^*]$ is approximated by $\Pr[\hat{S}_j^* \ge s_j^*]$, which in turn is approximated by generating a large number of realizations (1000 is the default) of $\hat{U}_j^*(.)$.

## Specifics for Bayesian Analysis

To request a Bayesian analysis, you specify the new BAYES statement in addition to the PROC PHREG statement and the MODEL statement. You include a CLASS statement if you have effects that involve categorical variables. The FREQ or WEIGHT statement can be included if you have a frequency or weight variable, respectively, in the input data. The STRATA statement can be used to carry out a stratified analysis for the Cox model, but it is not allowed in the piecewise constant baseline hazard model. Programming statements can be used to create time-dependent covariates for the Cox model, but they are not allowed in the piecewise constant baseline hazard model. However, you can use the counting process style of input to accommodate time-dependent covariates that are not continuously changing with time for the piecewise constant baseline hazard model and the Cox model as well. The HAZARDRATIO statement enables you to request a hazard ratio analysis based on the posterior samples. The ASSESS, CONTRAST, ID, OUTPUT, and TEST statements, if specified, are ignored. Also ignored are the COVM and COVS options in the PROC PHREG statement and the following options in the MODEL statement: BEST=, CORRB, COVB, DETAILS, HIERARCHY=, INCLUDE=, MAXSTEP=, NOFIT, PLCONV=, SELECTION=, SEQUENTIAL, SLENTRY=, and SLSTAY=.

## Piecewise Constant Baseline Hazard Model

### Single Failure Time Variable

Let $\{(t_i, \mathbf{x}_i, \delta_i), i = 1, 2, \ldots, n\}$ be the observed data. Let $a_0 = 0 < a_1 < \ldots < a_{J-1} < a_J = \infty$ be a partition of the time axis.

### Hazards in Original Scale

The hazard function for subject $i$ is

$$h(t|\mathbf{x}_i; \boldsymbol{\theta}) = h_0(t) \exp(\boldsymbol{\beta}' \mathbf{x}_i)$$

where

$$h_0(t) = \lambda_j a_{j-1} \le t < a_j (j = 1, \ldots, J)$$

The baseline cumulative hazard function is

$$H_0(t) = \sum_{j=1}^{J} \lambda_j \Delta_j(t)$$

where

$$\Delta_j(t) = \begin{cases} 0 & t < a_{j-1} \\ t - a_{j-1} & a_{j-1} \le t < a_j \\ a_j - a_{j-1} & t \ge a_j \end{cases}$$

The log likelihood is given by

$$\begin{aligned} l(\boldsymbol{\lambda}, \boldsymbol{\beta}) &= \sum_{i=1}^{n} \delta_i \left[ \sum_{j=1}^{J} I(a_{j-1} \le t_i < a_j) \log \lambda_j + \boldsymbol{\beta}' \mathbf{x}_i \right] - \sum_{i=1}^{n} \left[ \sum_{j=1}^{J} \Delta_j(t_i) \lambda_j \right] \exp(\boldsymbol{\beta}' \mathbf{x}_i) \\ &= \sum_{j=1}^{J} d_j \log \lambda_j + \sum_{i=1}^{n} \delta_i \boldsymbol{\beta}' \mathbf{x}_i - \sum_{j=1}^{J} \lambda_j \left[ \sum_{i=1}^{n} \Delta_j(t_i) \exp(\boldsymbol{\beta}' \mathbf{x}_i) \right] \end{aligned}$$

where $d_j = \sum_{i=1}^{n} \delta_i I(a_{j-1} \le t_i < a_j)$.

Note that for $1 \le j \le J$, the full conditional for $\lambda_j$ is log-concave only when $d_j > 0$, but the full conditionals for the $\beta$'s are always log-concave.

For a given $\boldsymbol{\beta}$, $\frac{\partial l}{\partial \boldsymbol{\lambda}} = 0$ gives

$$\tilde{\lambda_j}(\boldsymbol{\beta}) = \frac{d_j}{\sum_{i=1}^{n} \Delta_j(t_i) \exp(\boldsymbol{\beta}' \mathbf{x}_i)} (j = 1, \ldots, J)$$

Substituting these values into $l(\boldsymbol{\lambda}, \boldsymbol{\beta})$ gives the profile log likelihood for $\boldsymbol{\beta}$

$$l_p(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \boldsymbol{\beta}' \mathbf{x}_i - \sum_{j=1}^{J} d_j \log \left[ \sum_{l=1}^{n} \Delta_j(t_l) \exp(\boldsymbol{\beta}' \mathbf{x}_l) \right] + c$$

where $c = \sum_j (d_j \log d_j - d_j)$. Since the constant $c$ does not depend on $\boldsymbol{\beta}$, it can be discarded from $l_p(\boldsymbol{\beta})$ in the optimization.

The MLE $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained by maximizing

$$l_p(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \boldsymbol{\beta}'\mathbf{x}_i - \sum_{j=1}^{J} d_j \log\Bigl[\sum_{l=1}^{n} \Delta_j(t_l) \exp(\boldsymbol{\beta}'\mathbf{x}_l)\Bigr]$$

with respect to $\boldsymbol{\beta}$, and the MLE $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ is given by

$$\hat{\boldsymbol{\lambda}} = \tilde{\boldsymbol{\lambda}}(\hat{\boldsymbol{\beta}})$$

Let

$$\begin{aligned}
\mathbf{S}_j^{(r)}(\boldsymbol{\beta}) &= \sum_{l=1}^{n} \Delta_j(t_l) \mathrm{e}^{\boldsymbol{\beta}'\mathbf{x}_l} \mathbf{x}_l^{\otimes r} r = 0,1,2 (j = 1, \ldots, J) \\
\mathbf{E}_j(\boldsymbol{\beta}) &= \frac{\mathbf{S}_j^{(1)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})}
\end{aligned}$$

The partial derivatives of $l_p(\boldsymbol{\beta})$ are

$$\begin{aligned}
\frac{\partial l_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{n} \delta_i \mathbf{x}_i - \sum_{j=1}^{J} d_j \mathbf{E}_j(\boldsymbol{\beta}) \\
-\frac{\partial^2 l_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} &= \sum_{j=1}^{J} d_j \left\{ \frac{\mathbf{S}_j^{(2)}(\boldsymbol{\beta})}{S_j^{(0)}(\boldsymbol{\beta})} - \Bigl[\mathbf{E}_j(\boldsymbol{\beta})\Bigr]\Bigl[\mathbf{E}_j(\boldsymbol{\beta})\Bigr]' \right\}
\end{aligned}$$

The asymptotic covariance matrix for $(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\beta}})$ is obtained as the inverse of the information matrix given by

$$\begin{aligned}
-\frac{\partial^2 l(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\lambda}^2} &= \mathcal{D}\left(\frac{d_1}{\hat{\lambda}_1^2}, \ldots, \frac{d_J}{\hat{\lambda}_J^2}\right) \\
-\frac{\partial^2 l(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^2} &= \sum_{j=1}^{J} \hat{\lambda}_j \mathbf{S}_j^{(2)}(\hat{\boldsymbol{\beta}}) \\
-\frac{\partial^2 l(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\beta}} &= (\mathbf{S}_1^{(1)}(\hat{\boldsymbol{\beta}}), \ldots, \mathbf{S}_J^{(1)}(\hat{\boldsymbol{\beta}}))
\end{aligned}$$

See Example 6.5.1 in Lawless (2003) for details.

### Hazards in Log Scale

By letting

$$\alpha_j = \log(\lambda_j) j = 1, \ldots, J$$

you can build a prior correlation among the $\lambda_j$'s by using a correlated prior $\boldsymbol{\alpha} \sim N(\boldsymbol{\alpha}_0, \Sigma_\alpha)$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)'$.

The log likelihood is given by

$$l(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{j=1}^{J} d_j \alpha_j + \sum_{i=1}^{n} \delta_i \boldsymbol{\beta}' \mathbf{x}_i - \sum_{j=1}^{J} \mathrm{e}^{\alpha_j} S_j^{(0)}(\boldsymbol{\beta})$$

Then the MLE of $\lambda_j$ is given by

$$\mathrm{e}^{\hat{\alpha}_j} = \hat{\lambda}_j = \frac{d_j}{S_j^0(\hat{\boldsymbol{\beta}})}$$

Note that the full conditionals for $\alpha$'s and $\beta$'s are always log-concave.

The asymptotic covariance matrix for $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ is obtained as the inverse of the information matrix formed by

$$\begin{aligned}
-\frac{\partial^2 l(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\alpha}^2} &= \mathcal{D}\Big(\mathrm{e}^{\hat{\alpha}_j} S_j^0(\hat{\boldsymbol{\beta}}), \ldots, \mathrm{e}^{\hat{\alpha}_J} S_J^0(\hat{\boldsymbol{\beta}})\Big) \\
-\frac{\partial^2 l(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^2} &= \sum_{j=1}^{J} \mathrm{e}^{\hat{\alpha}_j} \mathbf{S}_j^{(2)}(\hat{\boldsymbol{\beta}}) \\
-\frac{\partial^2 l(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\beta}} &= (\mathrm{e}^{\hat{\alpha}_j} \mathbf{S}_1^{(1)}(\hat{\boldsymbol{\beta}}), \ldots, \mathrm{e}^{\hat{\alpha}_j} \mathbf{S}_J^{(1)}(\hat{\boldsymbol{\beta}}))
\end{aligned}$$

### Counting Process Style of Input

Let $\{((s_j, t_i], \mathbf{x}_i, \delta_i), i = 1, 2, \ldots, n\}$ be the observed data. Let $a_0 = 0 < a_1 < \ldots < a_k$ be a partition of the time axis, where $a_k > t_i$ for all $i = 1, 2, \ldots, n$.

Replacing $\Delta_j(t_i)$ with

$$\Delta_j((s_i, t_i]) = \begin{cases} 0 & t_i < a_{j-1} \vee s_i > a_j \\ t_i - \max(s_i, a_{j-1}) & a_{j-1} \le t_i < a_j \\ a_j - \max(s_i, a_{j-1}) & t_i \ge a_j \end{cases}$$

the formulation for the single failure time variable applies.

## Priors for Model Parameters

For a Cox model, the model parameters are the regression coefficients. For a piecewise exponential model, the model parameters consist of the regression coefficients and the hazards or log-hazards. The priors for the hazards and the priors for the regression coefficients are assumed to be independent, while you can have a joint multivariate normal prior for the log-hazards and the regression coefficients.

### Hazard Parameters

Let $\lambda_1, \ldots, \lambda_J$ be the constant baseline hazards.

**Improper Prior** The joint prior density is given by

$$p(\lambda_1, \ldots, \lambda_J) = \prod_{j=1}^{J} \frac{1}{\lambda_j} . \forall \lambda_j > 0$$

This prior is improper (nonintegrable), but the posterior distribution is proper as long as there is at least one event time in each of the constant hazard intervals.

**Uniform Prior** The joint prior density is given by

$$p(\lambda_1, \ldots, \lambda_J) \propto 1, \forall \lambda_j > 0$$

This prior is improper (nonintegrable), but the posteriors are proper as long as there is at least one event time in each of the constant hazard intervals.

**Gamma Prior** The gamma distribution $G(a, b)$ has a pdf

$$f_{a,b}(t) = \frac{b(bt)^{a-1} e^{-bt}}{\Gamma(a)}, t > 0$$

where $a$ is the shape parameter and $b^{-1}$ is the scale parameter. The mean is $\frac{a}{b}$ and the variance is $\frac{a}{b^2}$.

**Independent Gamma Prior** Suppose for $j = 1, \ldots, J$, $\lambda_j$ has an independent $G(a_j, b_j)$ prior. The joint prior density is given by

$$p(\lambda_1, \ldots, \lambda_J) \propto \prod_{j=1}^{J} \left\{ \lambda_j^{a_j - 1} e^{-b_j \lambda_j} \right\}, \forall \lambda_j > 0$$

**AR1 Prior** $\lambda_1, \ldots, \lambda_J$ are correlated as follows:

$$\begin{array}{lcl}
\lambda_1 & \sim & G(a_1, b_1) \\
\lambda_2 & \sim & G\left(a_2, \frac{b_2}{\lambda_1}\right) \\
\ldots & & \ldots \\
\lambda_J & \sim & G\left(a_J, \frac{b_J}{\lambda_{J-1}}\right)
\end{array}$$

The joint prior density is given by

$$p(\lambda_1, \ldots, \lambda_J) \propto \lambda_1^{a_1 - 1} e^{-b_1 \lambda_1} \prod_{j=2}^{J} \left( \frac{b_j}{\lambda_{j-1}} \right)^{a_j} \lambda_j^{a_j - 1} e^{-\frac{b_j}{\lambda_{j-1}} \lambda_j}$$

#### *Log-Hazard Parameters*

Write $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)' \equiv (\log\lambda_1, \ldots, \log\lambda_J)'$.

**Uniform Prior** The joint prior density is given by

$$p(\alpha_1 \ldots \alpha_J) \propto 1, \forall -\infty < \alpha_i < \infty$$

Note that the uniform prior for the log-hazards is the same as the improper prior for the hazards.

**Normal Prior** Assume $\boldsymbol{\alpha}$ has a multivariate normal prior with mean vector $\boldsymbol{\alpha}_0$ and covariance matrix $\boldsymbol{\Psi}_0$. The joint prior density is given by

$$p(\boldsymbol{\alpha}) \propto \mathrm{e}^{-\frac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)'\boldsymbol{\Psi}_0^{-1}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)}$$

#### *Regression Coefficients*

Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ be the vector of regression coefficients.

**Uniform Prior** The joint prior density is given by

$$p(\beta_1, \ldots, \beta_k) \propto 1, \forall -\infty < \beta_i < \infty$$

This prior is improper, but the posterior distributions for $\boldsymbol{\beta}$ are proper.

**Normal Prior** Assume $\boldsymbol{\beta}$ has a multivariate normal prior with mean vector $\boldsymbol{\beta}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$. The joint prior density is given by

$$p(\boldsymbol{\beta}) \propto \mathrm{e}^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}$$

**Joint Multivariate Normal Prior for Log-Hazards and Regression Coefficients** Assume $(\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ has a multivariate normal prior with mean vector $(\alpha_0', \boldsymbol{\beta}_0')'$ and covariance matrix $\boldsymbol{\Phi}_0$. The joint prior density is given by

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}) \propto \mathrm{e}^{-\frac{1}{2}[(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)',(\boldsymbol{\beta}-\boldsymbol{\beta}_0)']\boldsymbol{\Phi}_0^{-1}[(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)',(\boldsymbol{\beta}-\boldsymbol{\beta}_0)']'}$$

### Posterior Distribution

Denote the observed data by $D$.

#### *Cox Model*

$$\pi(\boldsymbol{\beta}|D) \propto L_P(D|\boldsymbol{\beta})p(\boldsymbol{\beta})$$

where $L_P(D|\boldsymbol{\beta})$ is the partial likelihood function with regression coefficients $\boldsymbol{\beta}$ as parameters.

### Piecewise Exponential Model

#### Hazard Parameters

$$\pi(\boldsymbol{\lambda},\boldsymbol{\beta}|D) \propto L_H(D|\boldsymbol{\lambda},\boldsymbol{\beta})p(\boldsymbol{\lambda})p(\boldsymbol{\beta})$$

where $L_H(D|\boldsymbol{\lambda},\boldsymbol{\beta})$ is the likelihood function with hazards $\boldsymbol{\lambda}$ and regression coefficients $\boldsymbol{\beta}$ as parameters.

#### Log-Hazard Parameters

$$\pi(\boldsymbol{\alpha},\boldsymbol{\beta}|D) \propto \begin{cases} L_{LH}(D|\boldsymbol{\alpha},\boldsymbol{\beta})p(\boldsymbol{\alpha},\boldsymbol{\beta}) & \text{if} \quad (\boldsymbol{\alpha}',\boldsymbol{\beta}')' \sim \text{MVN} \\ L_{LH}(D|\boldsymbol{\alpha},\boldsymbol{\beta})p(\boldsymbol{\alpha})p(\boldsymbol{\beta}) & \text{otherwise} \end{cases}$$

where $L_{LH}(D|\boldsymbol{\alpha},\boldsymbol{\beta})$ is the likelihood function with log-hazards $\boldsymbol{\alpha}$ and regression coefficients $\boldsymbol{\beta}$ as parameters.

## Sampling from the Posterior Distribution

PROC PHREG uses a Gibbs sampler to generate the posterior samples. See the section "Gibbs Sampler" on page 154 for a general discussion.

Let $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_k)'$ be the parameter vector. For the Cox model, the $\theta_i$'s are the regression coefficients $\beta_i$'s, and for the piecewise constant baseline hazard model, the $\theta_i$'s consist of the baseline hazards $\lambda_i$'s (or log baseline hazards $\alpha_i$'s) and the regression coefficients $\beta_j$'s. Let $L(D|\boldsymbol{\theta})$ be the likelihood function, where $D$ is the observed data. Note that for the Cox model, the likelihood contains the infinite-dimensional baseline hazard function and the Gamma process is perhaps the most commonly used prior process (Ibrahim, Chen, and Sinha 2001); however, Sinha, Ibrahim, and Chen (2003) justify using the partial likelihood as the likelihood function for the Bayesian analysis. Let $p(\boldsymbol{\theta})$ be the prior distribution. The full conditional distribution of $\theta_i$ is proportional to the joint distribution; that is,

$$\pi(\theta_i|\theta_j, i \neq j, D) \propto L(D|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

For instance, the one-dimensional conditional distribution of $\theta_1$ given $\theta_j = \theta_j^*, 2 \le j \le k$, is computed as

$$\pi(\theta_1|\theta_j = \theta_j^*, 2 \le j \le k, D) = L(D|\boldsymbol{\theta} = (\theta_1,\theta_2^*,\ldots,\theta_k^*)')p(\boldsymbol{\theta} = (\theta_1,\theta_2^*,\ldots,\theta_k^*)')$$

Suppose you have a set of arbitrary starting values $\{\theta_1^{(0)},\ldots,\theta_k^{(0)}\}$. Using the ARMS (adaptive rejection Metropolis sampling) algorithm of Gilks, Best, and Tan (1995), you do the following:

$$\text{draw } \theta_1^{(1)} \text{ from } \pi(\theta_1|\theta_2^{(0)},\ldots,\theta_k^{(0)},D)$$

$$\text{draw } \theta_2^{(1)} \text{ from } \pi(\theta_2|\theta_1^{(1)},\theta_3^{(0)},\ldots,\theta_k^{(0)},D)$$

$$\ldots\ldots$$

$$\text{draw } \theta_k^{(1)} \text{ from } \pi(\theta_k|\theta_1^{(1)},\ldots,\theta_{k-1}^{(1)},D)$$

This completes one iteration of the Gibbs sampler. After one iteration, you have $\{\theta_1^{(1)}, \ldots, \theta_k^{(1)}\}$. After $n$ iterations, you have $\{\theta_1^{(n)}, \ldots, \theta_k^{(n)}\}$.

You can output these posterior samples into a SAS data set through ODS. The following SAS statement outputs the posterior samples into the SAS data set Post:

```
ods output PosteriorSample=Post;
```

The data set also includes the variable LogPost, representing the log of the posterior log likelihood.

## Starting Values of the Markov Chains

When the BAYES statement is specified, PROC PHREG generates one Markov chain containing the approximate posterior samples of the model parameters. Additional chains are produced when the Gelman-Rubin diagnostics are requested. Starting values (or initial values) can be specified in the INITIAL= data set in the BAYES statement. If INITIAL= option is not specified, PROC PHREG picks its own initial values for the chains. If the prior distribution of the parameter $\omega$ is proper, the starting values for $\omega$ are based on the estimated mean ($\hat{\omega}$) and standard deviation ($\hat{s}(\hat{\omega})$) of the posterior distribution given the MLE. If the prior distribution of $\omega$ is improper, the starting values for $\omega$ are based on the MLE and its standard error estimate; that is, $\hat{\omega}$ is the MLE of $\omega$ and $\hat{s}(\hat{\omega})$ is the standard error of maximum likelihood estimator.

Denote $[x]$ as the integral value of $x$.

### *Constant Baseline Hazards $\lambda_i$'s*

For the first chain that the summary statistics and diagnostics are based on, the initial values are

$$\lambda_i^{(0)} = \hat{\lambda}_i$$

For subsequent chains, the starting values are picked in two different ways according to the total number of chains specified. If the total number of chains specified is less than or equal to 10, initial values of the $r$th chain ($2 \leq r \leq 10$) are given by

$$\lambda_i^{(0)} = \hat{\lambda}_i \mathrm{e}^{\pm\left([\frac{r}{2}]+2\right)\hat{s}(\hat{\lambda}_i)}$$

with the plus sign for odd $r$ and minus sign for even $r$. If the total number of chains is greater than 10, initial values are picked at random over a wide range of values. Let $u_i$ be a uniform random number between 0 and 1; the initial value for $\lambda_i$ is given by

$$\lambda_i^{(0)} = \hat{\lambda}_i \mathrm{e}^{16(u_i-0.5)\hat{s}(\hat{\lambda}_i)}$$

### *Regression Coefficients and Log-Hazard Parameters $\theta_i$'s*

The $\theta_i$'s are the regression coefficients $\beta_i$'s, and in the piecewise exponential model, include the log-hazard parameters $\alpha_i$'s. For the first chain that the summary statistics and regression diagnostics are based on, the initial values are

$$\theta_i^{(0)} = \hat{\theta}_i$$

If the number of chains requested is less than or equal to 10, initial values for the $r$th chain ($2 \leq r \leq 10$) are given by

$$\theta_i^{(0)} = \hat{\theta}_i \pm \left(2 + \left[\frac{r}{2}\right]\right)\hat{s}(\hat{\theta}_i)$$

with the plus sign for odd $r$ and minus sign for even $r$. When there are more than 10 chains, the initial value for the $\theta_i$ is picked at random over the range $(\hat{\theta}_i - 8\hat{s}(\hat{\theta}_i), \hat{\theta}_i + 8\hat{s}(\hat{\theta}_i))$; that is,

$$\theta_i^{(0)} = \hat{\theta}_i + 16(u_i - 0.5)\hat{s}(\hat{\theta}_i)$$

where $u_i$ is a uniform random number between 0 and 1.

## Fit Statistics

Denote the observed data by $D$. Let $\boldsymbol{\theta}$ be the vector of parameters of length $k$. Let $L(D|\boldsymbol{\theta})$ be the likelihood. The deviance information criterion (DIC) proposed in Spiegelhalter et al. (2002) is a Bayesian model assessment tool. Let $\mathrm{Dev}(\boldsymbol{\theta}) = -2\log L(D|\boldsymbol{\theta})$. Let $\overline{\mathrm{Dev}(\theta)}$ and $\bar{\boldsymbol{\theta}}$ be the corresponding posterior means of $\mathrm{Dev}(\boldsymbol{\theta})$ and $\boldsymbol{\theta}$, respectively. The deviance information criterion is computed as

$$\mathrm{DIC} = 2\overline{\mathrm{Dev}(\boldsymbol{\theta})} - \mathrm{Dev}(\bar{\boldsymbol{\theta}})$$

Also computed is

$$pD = \overline{\mathrm{Dev}(\boldsymbol{\theta})} - \mathrm{Dev}(\bar{\boldsymbol{\theta}})$$

where $pD$ is interpreted as the effective number of parameters.

Note that $\mathrm{Dev}(\boldsymbol{\theta})$ defined here does not have the standardizing term as in the section "Deviance Information Criterion (DIC)" on page 172. Nevertheless, DIC calculated here is still useful for variable selection.

## Posterior Distribution for Quantities of Interest

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ be the parameter vector. For the Cox model, the $\theta_i$'s are the regression coefficients $\beta_i$'s, and for the piecewise constant baseline hazard model, the $\theta_i$'s consist of the baseline hazards $\lambda_i$'s (or log baseline hazards $\alpha_i$'s) and the regression coefficients $\beta_j$'s. Let $\mathcal{S} = \{\boldsymbol{\theta}^{(r)}, r = 1, \ldots, N\}$ be the chain representing the posterior distribution for $\boldsymbol{\theta}$.

Consider a quantity of interest $\tau$ that can be expressed as a function $f(\boldsymbol{\theta})$ of the parameter vector $\boldsymbol{\theta}$. You can construct the posterior distribution of $\tau$ by evaluating the function $f(\boldsymbol{\theta}^{(r)})$ for each $\boldsymbol{\theta}^{(r)}$ in $\mathcal{S}$. The posterior chain for $\tau$ is $\{f(\boldsymbol{\theta}^{(r)}), r = 1, \ldots, N\}$. Summary statistics such as mean, standard deviation, percentiles, and credible intervals are used to describe the posterior distribution of $\tau$.

### *Hazard Ratio*

As shown in the section "Hazard Ratios" on page 4584, a log-hazard ratio is a linear combination of the regression coefficients. Let **h** be the vector of linear coefficients. The posterior sample for this hazard ratio is the set $\{\exp(\mathrm{h}'\boldsymbol{\beta}^{(r)}), r = 1, \ldots, N\}$.

### Survival Distribution

Let x be a covariate vector of interest.

**Cox Model** Let $\{(t_i, \mathbf{z}_i, \delta_i), i = 1, 2, \ldots, n\}$ be the observed data. Define

$$Y_i(t) = \begin{cases} 1 & t < t_i \\ 0 & \text{otherwise} \end{cases}$$

Consider the $r$th draw $\boldsymbol{\beta}^{(r)}$ of $\mathcal{S}$. The baseline cumulative hazard function at time $t$ is given by

$$H_0(t|\boldsymbol{\beta}^{(r)}) = \sum_{i:t_i \le t} \frac{\delta_i}{\sum_{l=1}^{n} Y_l(t_i) \exp(\mathbf{z}_l' \boldsymbol{\beta}^{(r)})}$$

For the given covariate vector x, the cumulative hazard function at time $t$ is

$$H(t; \mathbf{x}|\boldsymbol{\beta}^{(r)}) = H_0(t|\boldsymbol{\beta}^{(r)}) \exp(\mathbf{x}' \boldsymbol{\beta}^{(r)})$$

and the survival function at time $t$ is

$$S(t; \mathbf{x}|\boldsymbol{\beta}^{(r)}) = \exp[-H^{(r)}(t; \mathbf{x}|\boldsymbol{\beta}^{(r)})]$$

**Piecewise Exponential Model** Let $0 = a_0 < a_1 < \ldots < a_J < \infty$ be a partition of the time axis. Consider the $r$th draw $\boldsymbol{\theta}^{(r)}$ in $\mathcal{S}$, where $\boldsymbol{\theta}^{(r)}$ consists of $\boldsymbol{\lambda}^{(r)} = (\lambda_1^{(r)}, \ldots, \lambda_J^{(r)})'$ and $\boldsymbol{\beta}^{(r)}$. The baseline cumulative hazard function at time $t$ is

$$H_0(t|\boldsymbol{\lambda}^{(r)}) = \sum_{j=1}^{J} \lambda_j^{(r)} \Delta_j(t)$$

where

$$\Delta_j(t) = \begin{cases} 0 & t < a_{j-1} \\ t - a_{j-1} & a_{j-1} \le t < a_j \\ a_j - a_{j-1} & t \ge a_j \end{cases}$$

For the given covariate vector $\mathbf{x}$, the cumulative hazard function at time $t$ is

$$H(t; \mathbf{x}|\boldsymbol{\lambda}^{(r)}, \boldsymbol{\beta}^{(r)}) = H_0(t|\boldsymbol{\lambda}^{(r)}) \exp(\mathbf{x}' \boldsymbol{\beta}^{(r)})$$

and the survival function at time $t$ is

$$S(t; \mathbf{x}|\boldsymbol{\lambda}^{(r)}, \boldsymbol{\beta}^{(r)}) = \exp[-H(t; \mathbf{x}|\boldsymbol{\lambda}^{(r)}, \boldsymbol{\beta}^{(r)})]$$

## Computational Resources

Let $n$ be the number of observations in a BY group. Let $p$ be the number of explanatory variables. The minimum working space (in bytes) needed to process the BY group is

$$\max\{12n, 24p^2 + 160p\}$$

Extra memory is needed for certain TIES= options. Let $k$ be the maximum multiplicity of tied times. The TIES=DISCRETE option requires extra memory (in bytes) of

$$4k(p^2 + 4p)$$

The TIES=EXACT option requires extra memory (in bytes) of

$$24(k^2 + 5k)$$

If sufficient space is available, the input data are also kept in memory. Otherwise, the input data are reread from the utility file for each evaluation of the likelihood function and its derivatives, with the resulting execution time substantially increased.

## Input and Output Data Sets

### OUTEST= Output Data Set

The OUTEST= data set contains one observation for each BY group containing the maximum likelihood estimates of the regression coefficients. If you also use the COVOUT option in the PROC PHREG statement, there are additional observations containing the rows of the estimated covariance matrix. If you specify SELECTION=FORWARD, BACKWARD, or STEPWISE, only the estimates of the parameters and covariance matrix for the final model are output to the OUTEST= data set.

#### *Variables in the OUTEST= Data Set*

The OUTEST= data set contains the following variables:

- any BY variables specified
- _TIES_, a character variable of length 8 with four possible values: BRESLOW, DISCRETE, EFRON, and EXACT. These are the four values of the TIES= option in the MODEL statement.

- _TYPE_, a character variable of length 8 with two possible values: PARMS for parameter estimates or COV for covariance estimates. If both the COVM and COVS options are specified in the PROC LIFETEST statement along with the COVOUT option, _TYPE_='COVM' for the model-based covariance estimates and _TYPE_='COVS' for the robust sandwich covariance estimates.
- _STATUS_, a character variable indicating whether the estimates have converged
- _NAME_, a character variable containing the name of the TIME variable for the row of parameter estimates and the name of each explanatory variable to label the rows of covariance estimates
- one variable for each regression coefficient and one variable for the offset variable if the OFFSET= option is specified. If an explanatory variable is not included in the final model in a variable selection process, the corresponding parameter estimates and covariances are set to missing.
- _LNLIKE_, a numeric variable containing the last computed value of the log likelihood

### *Parameter Names in the OUTEST= Data Set*

For continuous explanatory variables, the names of the parameters are the same as the corresponding variables. For CLASS variables, the parameter names are obtained by concatenating the corresponding CLASS variable name with the CLASS category; see the PARAM= option in the CLASS statement and the section "CLASS Variable Parameterization" on page 4574 for more details. For interaction and nested effects, the parameter names are created by concatenating the names of each component effect.

## INEST= Input Data Set

You can specify starting values for the maximum likelihood iterative algorithm in the INEST= data set. The INEST= data set has the same structure as the OUTEST= data set but is not required to have all the variables or observations that appear in the OUTEST= data set.

The INEST= data set must contain variables that represent the regression coefficients of the model. If BY processing is used, the INEST= data set should also include the BY variables, and there must be one observation for each BY group. If the INEST= data set also contains the _TYPE_ variable, only observations with _TYPE_ value 'PARMS' are used as starting values.

## OUT= Output Data Set in the OUTPUT Statement

The OUT= data set in the OUTPUT statement contains all the variables in the input data set, along with statistics you request by specifying *keyword=name* options. The new variables contain a variety of diagnostics that are calculated for each observation in the input data set.

### OUT= Output Data Set in the BASELINE Statement

The OUT= data set in the BASELINE statement contains all the variables in the COVARIATES= data set, along with statistics you request by specifying *keyword=name* options. For unstratified input data, there are $1{+}n$ observations in the OUT= data set for each observation in the COVARIATES= data set, where $n$ is the number of distinct event times in the input data. For input data that are stratified into $k$ strata, with $n_i$ distinct events in the $i$th stratum, $i = 1, \ldots, k$, there are $1{+}n_i$ observations for the $i$th stratum in the OUT= data set for each observation in the COVARIATES= data set.

### OUTPOST= Output Data Set in the BAYES Statement

The OUTPOST= data set contains the generated posterior samples. There are $2{+}n$ variables, where $n$ is the number of model parameters. The variable Iteration represents the iteration number and the variable LogPost contains the log posterior likelihood values. The other $n$ variables represent the draws of the Markov chain for the model parameters.

## Displayed Output

If you use the NOPRINT option in the PROC PHREG statement, the procedure does not display any output. Otherwise, PROC PHREG displays results of the analysis in a collection of tables. The tables are listed separately for the maximum likelihood analysis and for the Bayesian analysis.

### Maximum Likelihood Analysis Displayed Output

#### *Model Information*

The “Model Information” table displays the two-level name of the input data set, the name and label of the failure time variable, the name and label of the censoring variable and the values indicating censored times, the model (either the Cox model or the piecewise constant baseline hazard model), the name and label of the OFFSET variable, the name and label of the FREQ variable, the name and label of the WEIGHT variable, and the method of handling ties in the failure time for the Cox model. For ODS purposes, the name of the “Model Information” table is “ModelInfo.”

#### *Number of Observations*

The “Number of Observations” table displays the number of observations read and used in the analysis. For ODS purposes, the name of the “Number of Observations” is “NObs.”

### Summary of the Number of Event and Censored Values

The “Summary of the Number of Event and Censored Values” table displays, for each stratum, the breakdown of the number of events and censored values. For ODS purposes, the name of the “Summary of the Number of Event and Censored Values” table is “CensoredSummary.”

### Class Level Information

The “Class Level Information” table is displayed when there are CLASS variables in the model. The table lists the categories of every CLASS variable that is used in the model and the corresponding design variable values. For ODS purposes, the name of the “Class Level Information” table is “ClassLevelInfo.”

### Descriptive Statistics for Continuous Explanatory Variables

The “Simple Statistics for Continuous Explanatory Variables” table is displayed when you specify the SIMPLE option in the PROC PHREG statement. The table contains, for each stratum, the mean, standard deviation, and minimum and maximum for each continuous explanatory variable in the MODEL statement. For ODS purposes, the name of the “Descriptive Statistics for Continuous Explanatory Variables” table is “SimpleStatistics.”

### Frequency Distribution of CLASS Variables

The “Frequency Distribution of CLASS Variables” table is displayed if you specify the SIMPLE option in the PROC PHREG statement and there are CLASS variables in the model. The table lists the frequency of the levels of the CLASS variables. For ODS purposes, the name of the “Frequency Distribution of CLASS Variables” table is “ClassLevelFreq.”

### Maximum Likelihood Iteration History

The “Maximum Likelihood Iteration History” table is displayed if you specify the ITPRINT option in the MODEL statement. The table contains the iteration number, ridge value or step size, log likelihood, and parameter estimates at each iteration. For ODS purposes, the name of the “Maximum Likelihood Iteration History” table is “IterHistory.”

### Gradient of Last Iteration

The “Gradient of Last Iteration” table is displayed if you specify the ITPRINT option in the MODEL statement. For ODS purposes, the name of the “Gradient of Last Iteration” table is “LastGradient.”

### Convergence Status

The “Convergence Status” table displays the convergence status of the Newton-Raphson maximization. For ODS purposes, the name of the “Convergence Status” table is “ConvergenceStatus.”

### Model Fit Statistics

The "Model Fit Statistics" table displays the values of $-2$ log likelihood for the null model and the fitted model, the AIC, and SBC. For ODS purposes, the name of the "Model Fit Statistics" table is "FitStatistics."

### Testing Global Null Hypothesis: BETA=0

The "Testing Global Null Hypothesis: BETA=0" table displays results of the likelihood ratio test, the score test, and the Wald test for testing the hypothesis that all parameters are zero. For ODS purpose, the name of the "Testing Global Null Hypothesis: BETA=0" table is "GlobalTests."

### Type 3 Tests

The "Type 3 Tests" table is displayed if the model contains CLASS variables. The table displays type 3 Wald chi-square statistic, the degrees of freedom, and the *p*-value for each effect in the model. For ODS purposes, the name of "Type 3 Tests" is "Type3."

### Analysis of Maximum Likelihood Estimates

The "Analysis of Maximum Likelihood Estimates" table displays the maximum likelihood estimate of the parameter; the estimated standard error, computed as the square root of the corresponding diagonal element of the estimated covariance matrix; the ratio of the robust standard error estimate to the model-based standard error estimate if you specify the COVS option in the PROC statement; the Wald Chi-Square statistic, computed as the square of the parameter estimate divided by its standard error estimate; the degrees of freedom of the Wald chi-square statistic, which has a value of 1 unless the corresponding parameter is redundant or infinite, in which case the value is 0; the *p*-value of the Wald chi-square statistic with respect to a chi-square distribution with one degree of freedom; the hazard ratio estimate; and the confidence limits for the hazard ratio if you specified the RISKLIMITS option in the MODEL statement. For ODS purposes, the name of the "Analysis of Maximum Likelihood Estimates" table is "ParameterEstimates."

### Regression Models Selected by Score Criterion

The "Regression Models Selected by Score Criterion" table is displayed if you specify SELECTION=SCORE in the MODEL statement. The table contains the number of explanatory variables in each model, the score chi-square statistic, and the names of the variables included in the model. For ODS purposes, the name of the "Regression Models Selected by Score Criterion" table is "BestSubsets."

### Analysis of Effects Eligible for Entry

The "Analysis of Effects Eligible for Entry" table is displayed if you use the FORWARD or STEPWISE selection method and you specify the DETAILS option in the MODEL statement. The table contains the score chi-square statistic for testing the significance of each variable not in the model (after adjusting for the variables already in the model), and the *p*-value of the chi-square statistic with respect to a chi-square distribution with one degree of freedom. This table is produced before a variable is selected for entry in a forward selection step. For ODS purposes, the name of the "Analysis of Effects Eligible for Entry" table is "EffectsToEntry."

### Analysis of Effects Eligible for Removal

The "Analysis of Effects Eligible for Removal" table is displayed if you use the BACKWARD or STEPWISE selection method and you specify the DETAILS option in the MODEL statement. The table contains the Wald chi-square statistic for testing the significance of each candidate effect for removal, the degrees of freedom of the Wald chi-square, and the corresponding *p*-value. This table is produced before an effect is selected for removal. For ODS purposes, the name of the "Analysis of Effects Eligible for Removal" table is "EffectsToRemoval."

### Summary of Backward Elimination

The "Summary of Backward Elimination" table is displayed if you specify the SELECTION=BACKWARD option in the MODEL statement. The table contains the step number, the effects removed at each step, the corresponding chi-square statistic, the degrees of freedom, and the *p*-value. For ODS purpose, the name of the "Summary of Backward Elimination" table is "ModelBuildingSummary."

### Summary of Forward Selection

The "Summary of Forward Selection" table is displayed if you specify the SELECTION=FORWARD option in the MODEL statement. The table contains the step number, the effects entered at each step, the corresponding chi-square statistic, the degrees of freedom, and the *p*-value. For ODS purpose, the name of the "Summary of Forward Selection" table is "ModelBuildingSummary."

### Summary of Stepwise Selection

The "Summary of Stepwise Selection" table is displayed if you specify SELECTION=STEPWISE is specified in the MODEL statement. The table contains the step number, the effects entered or removed at each step, the corresponding chi-square statistic, the degrees of freedom, and the corresponding *p*-value. For ODS purpose, the name of the "Summary of Stepwise Selection" table is "ModelBuildingSummary."

### Covariance Matrix

The "Covariance Matrix" table is displayed if you specify the COVB option in the MODEL statement. The table contains the estimated covariance matrix for the parameter estimates. For ODS purposes, the name of the "Covariance Matrix" table is "CovB."

### *Correlation Matrix*

The "Correlation Matrix" table is displayed if you specify the COVB option in the MODEL statement. The table contains the estimated correlation matrix for the parameter estimates. For ODS purposes, the name of the "Correlation Matrix" table is "CorrB."

### *Hazard Ratios for label*

The "Hazard Ratios for *label*" table is displayed if you specify the HAZARDRATIO statement. The table displays the estimate and confidence limits for each hazard ratio. For ODS purposes, the name of the "Hazard Ratios for *label*" table is "HazardRatios."

### *Linear Coefficients for label*

The "Linear Coefficients *label*" table is displayed if you specify the E option in the TEST statement with *label* being the TEST statement label. The table contains the coefficients and constants of the linear hypothesis. For ODS purposes, the name of the "Linear Coefficients for *label*" table is "TestCoeff."

### *L[cov(b)]L' and Lb-c*

The "L[cov(b)]L' and Lb-c" table is displayed if you specified the PRINT option in a TEST statement with *label* being the TEST statement label. The table displays the intermediate calculations of the Wald test. For ODS purposes, the name of the "L[cov(b)]L' and Lb-c" table is "TestPrint1."

### *Ginv(L[cov(b)]L') and Ginv(L[cov(b)]L')(Lb-c)*

The "Ginv(L[cov(b)]L') and Ginv(L[cov(b)]L')(Lb-c)" table is displayed if you specified the PRINT option in a TEST statement with *label* being the TEST statement label. The table displays the intermediate calculations of the Wald test. For ODS purposes, the name of the "Ginv(L[cov(b)]L') and Ginv(L[cov(b)]L')(Lb-c)" table is "TestPrint2."

### label *Test Results*

The "*label* Test Results" table is displayed if you specify a TEST statement with *label* being the TEST statement label. The table contains the Wald chi-square statistic, the degrees of freedom, and the $p$-value. For ODS purposes, the name of "*label* Test Results" table is "TestStmts."

### *Average Effect for label*

The "Average Effect for *label*" table is displayed if the AVERAGE option is specified in a TEST statement with *label* being the TEST statement label. The table contains the weighted average of the parameter estimates for the variables in the TEST statement, the estimated standard error, the z-score, and the $p$-value. For ODS purposes, the name of the "Average Effect for *label*" is "TestAverage."

### Reference Set of Covariates for Plotting

The “Reference Set of Covariates for Plotting” table is displayed if the PLOTS= option is requested without specifying the COVARIATES= data set in the BASELINE statement. The tables contains the values of the covariates for the reference set, where the reference levels are used for the CLASS variables and the sample averages for the continuous variables.

## Bayesian Analysis Displayed Output

### Model Information

The “Model Information” table displays the two-level name of the input data set, the name and label of the failure time variable, the name and label of the censoring variable and the values indicating censored times, the model (either the Cox model or the piecewise constant baseline hazard model), the name and label of the OFFSET variable, the name and label of the FREQ variable, the name and label of the WEIGHT variable, the method of handling ties in the failure time, the number of burn-in iterations, the number of iterations after the burn-in, and the number of thinning iterations. For ODS purposes, the name of the “Model Information” table is “ModelInfo.”

### Number of Observations

The “Number of Observations” table displays the number of observations read and used in the analysis. For ODS purposes, the name of the “Number of Observations” is “NObs.”

### Summary of the Number of Event and Censored Values

The “Summary of the Number of Event and Censored Values” table displays, for each stratum, the breakdown of the number of events and censored values. This table is not produced if the NONSUMMARY option is specified in the PROC PHREG statement. For ODS purposes, the name of the “Summary of the Number of Event and Censored Values” table is “CensoredSummary.”

### Descriptive Statistics for Continuous Explanatory Variables

The “Simple Statistics for Continuous Explanatory Variables” table is displayed when you specify the SIMPLE option in the PROC PHREG statement. The table contains, for each stratum, the mean, standard deviation, and minimum and maximum for each continuous explanatory variable in the MODEL statement. For ODS purposes, the name of the “Descriptive Statistics for Continuous Explanatory Variables” table is “SimpleStatistics.”

### Class Level Information

The “Class Level Information” table is displayed if there are CLASS variables in the model. The table lists the categories of every CLASS variable in the model and the corresponding design variable values. For ODS purposes, the name of the “Class Level Information” table is “ClassLevelInfo.”

### Frequency Distribution of CLASS Variables

The “Frequency Distribution of CLASS Variables” table is displayed if you specify the SIMPLE option in the PROC PHREG statement and there are CLASS variables in the model. The table lists the frequency of the levels of the CLASS variables. For ODS purposes, the name of the “Frequency Distribution of CLASS Variables” table is “ClassLevelFreq.”

### Regression Parameter Information

The “Regression Parameter Information” table displays the names of the parameters and the corresponding level information of effects containing the CLASS variables. For ODS purposes, the name of the “Regression Parameter Information” table is “ParmInfo.”

### Constant Baseline Hazard Time Intervals

The “Constant Baseline Hazard Time Intervals” table displays the intervals of constant baseline hazard and the corresponding numbers of failure times and event times. This table is produced only if you specify the PIECEWISE option in the BAYES statement. For ODS purposes, the name of the “Constant Baseline Hazard Time Intervals” table is “Interval.”

### Maximum Likelihood Estimates

The “Maximum Likelihood Estimates” table displays, for each parameter, the maximum likelihood estimate, the estimated standard error, and the 95% confidence limits. For ODS purposes, the name of the “Maximum Likelihood Estimates” table is “ParameterEstimates.”

### Hazard Prior

The “Hazard Prior” table is displayed if you specify the PIECEWISE=HAZARD option in the BAYES statement. It describes the prior distribution of the hazard parameters. For ODS purposes, the name of the “Hazard Prior” table is “HazardPrior.”

### Log-Hazard Prior

The “Log-Hazard Prior” table is displayed if you specify the PIECEWISE=LOGHAZARD option in the BAYES statement. It describes the prior distribution of the log-hazard parameters. For ODS purposes, the name of the “Log-Hazard Prior” table is “HazardPrior.”

### Coefficient Prior

The “Coefficient Prior” table displays the prior distribution of the regression coefficients. For ODS purposes, the name of the “Coefficient Prior” table is “CoeffPrior.”

### Initial Values

The “Initial Values” table is displayed if you specify the INITIAL option in the BAYES statement. The table contains the initial values of the parameters for the Gibbs sampling. For ODS purposes, the name of the “Initial Values” table is “InitialValues.”

### Fit Statistics

The “Fit Statistics” table displays the DIC and pD statistics for each parameter. For ODS purposes, the name of the “Fit Statistics” table is “FitStatistics.”

### Posterior Summaries

The “Posterior Summaries” table displays the size of the posterior sample, the mean, the standard error, and the percentiles for each model parameter. For ODS purposes, the name of the “Posterior Summaries” table is “PostSummaries.”

### Posterior Intervals

The “Posterior Intervals” table displays the equal-tail interval and the HPD interval for each model parameter. For ODS purposes, the name of the “Posterior Intervals” table is “PostIntervals.”

### Posterior Covariance Matrix

The “Posterior Covariance Matrix” table is produced if you include COV in the SUMMARY= option in the BAYES statement. This tables displays the sample covariance of the posterior samples. For ODS purposes, the name of the “Posterior Covariance Matrix” table is “Cov.”

### Posterior Correlation Matrix

The “Posterior Correlation Matrix” table is displayed if you include CORR in the SUMMARY= option in the BAYES statement. The table contains the sample correlation of the posterior samples. For ODS purposes, the name of the “Posterior Correlation Matrix” table is “Corr.”

### Posterior Autocorrelations

The “Posterior Autocorrelations” table displays the lag 1, lag 5, lag 10, and lag 50 autocorrelations for each parameter. For ODS purposes, the name of the “Posterior Autocorrelations” table is “AutoCorr.”

### Gelman-Rubin Diagnostics

The “Gelman-Rubin Diagnostics” table is produced if you include GELMAN in the DIAGNOSTIC= option in the BAYES statement. This table displays the estimate of the potential scale reduction factor and its 97.5% upper confidence limit for each parameter. For ODS purposes, the name of the “Gelman-Rubin Diagnostics” table is “Gelman.”

### Geweke Diagnostics

The "Geweke Diagnostics" table displays the Geweke statistic and its $p$-value for each parameter. For ODS purposes, the name of the "Geweke Diagnostics" table is "Geweke."

### Raftery-Lewis Diagnostics

The "Raftery-Lewis Diagnostics" tables is produced if you include RAFTERY in the DIAGNOSTIC= option in the BAYES statement. This table displays the Raftery and Lewis diagnostics for each variable. For ODS purposes, the name of the "Raftery-Diagnostics" table is "Raftery."

### Heidelberger-Welch Diagnostics

The "Heidelberger-Welch Diagnostics" table is displayed if you include HEIDELBERGER in the DIAGNOSTIC= option in the BAYES statement. This table describes the results of a stationary test and a halfwidth test for each parameter. For ODS purposes, the name of the "Heidelberger-Welch Diagnostics" table is "Heidelberger."

### Effective Sample Sizes

The "Effective Sample Sizes" table displays, for each parameter, the effective sample size, the correlation time, and the efficiency. For ODS purposes, the name of the "Effective Sample Sizes" table is "ESS."

### Hazard Ratios for label

The "Hazard Ratios for *label*" table is displayed if you specify the HAZARDRATIO statement. The table displays the posterior summary for each hazard ratio. The summary includes the mean, standard error, quartiles, and equal-tailed and HPD intervals. For ODS purposes, the name of the "Hazard Ratios for *label*" table is "HazardRatios."

### Reference Set of Covariates for Plotting

The "Reference Set of Covariates for Plotting" table is displayed if the PLOTS= option is requested without specifying the COVARIATES= data set in the BASELINE statement. The table contains the values of the covariates for the reference set, where the reference levels are used for the CLASS variables and the sample averages for the continuous variables.

## ODS Table Names

PROC PHREG 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 separately in Table 64.5 for the maximum likelihood analysis and in

Table 64.6 for the Bayesian analysis. For more information about ODS, see Chapter 20, "Using the Output Delivery System."

**Table 64.5** ODS Tables for a Maximum Likelihood Analysis Produced by PROC PHREG

| ODS Table Name | Description | Statement / Option |
|---|---|---|
| BestSubsets | Best subset selection | MODEL / SELECTION=SCORE |
| CensoredSummary | Summary of event and censored observations | default |
| ClassLevelFreq | Frequency distribution of CLASS variables | CLASS, PROC / SIMPLE |
| ClassLevelInfo | CLASS variable levels and design variables | CLASS |
| ContrastCoeff | L matrix for contrasts | CONTRAST / E |
| ContrastEstimate | Individual contrast estimates | CONTRAST / ESTIMATE= |
| ContrastTest | Wald test for contrasts | CONTRAST |
| ConvergenceStatus | Convergence status | default |
| CorrB | Estimated correlation matrix of parameter estimators | MODEL / CORRB |
| CovB | Estimated covariance matrix of parameter estimators | MODEL / COVB |
| EffectsToEnter | Analysis of effects for entry | MODEL / SELECTION=F\|S |
| EffectsToremove | Analysis of effects for removal | MODEL / SELECTION=B\|S |
| FitStatistics | Model fit statistics | default |
| FunctionalFormSupTest | Supremum test for functional form | ASSESS / VAR= |
| GlobalScore | Global chi-square test | MODEL / NOFIT |
| GlobalTests | Tests of the global null hypothesis | default |
| HazardRatios | Hazard ratios and confidence limits | HAZARDRATIO |
| IterHistory | Iteration history | MODEL /ITPRINT |
| LastGradient | Last evaluation of gradient | MODEL / ITPRINT |
| ModelBuildingSummary | Summary of model building | MODEL / SELECTION=B\|F\|S |
| ModelInfo | Model information | default |
| NObs | Number of observations | default |
| ParameterEstimates | Maximum likelihood estimates of model parameters | default |
| ProportionalHazardsSupTest | Supremum test for proportional hazards assumption | ASSESS / PH |
| ResidualChiSq | Residual chi-square | MODEL / SELECTION=F\|B |
| ReferenceSet | Reference set of covariates for plotting | PROC / PLOTS= |
| SimpleStatistics | Summary statistics of input continuous explanatory variables | PROC / SIMPLE |
| TestAverage | Average effect for test | TEST / AVERAGE |
| TestCoeff | Coefficients for linear hypotheses | TEST / E |
| TestPrint1 | **L**[cov(**b**)]**L**' and **Lb-c** | TEST / PRINT |
| TestPrint2 | Ginv(**L**[cov(**b**)]**L**') and Ginv(**L**[cov(**b**)]**L**')(**Lb-c**) | TEST / PRINT |

**Table 64.5** *continued*

| ODS Table Name | Description | Statement / Option |
|---|---|---|
| TestStmts | Linear hypotheses testing results | TEST |
| Type3 | Type 3 chi-square tests | CLASS |

**Table 64.6** ODS Table for a Bayesian Analysis Produced by PROC PHREG

| ODS Table Name | Description | Statement / Option |
|---|---|---|
| AutoCorr | Autocorrelations of the posterior samples | BAYES |
| CensoredSummary | Numbers of the event and censored observations | PROC |
| ClassLevelFreq | Frequency distribution of CLASS variables | CLASS, PROC / SIMPLE |
| ClassLevelInfo | CLASS variable levels and design variables | CLASS |
| CoeffPrior | Prior distribution of the regression coefficients | BAYES |
| Corr | Posterior correlation matrix | BAYES / SUMMARY=CORR |
| Cov | Posterior covariance Matrix | BAYES / SUMMARY=COV |
| ESS | Effective sample sizes | BAYES / DIAGNOSTICS=ESS |
| FitStatistics | Fit statistics | BAYES |
| Gelman | Gelman-Rubin convergence diagnostics | BAYES / DIAGNOSTICS=GELMAN |
| Geweke | Geweke convergence diagnostics | BAYES |
| HazardPrior | Prior distribution of the baseline hazards | BAYES / PIECEWISE |
| HazardRatios | Posterior summary statistics for hazard ratios | HAZARDRATIO |
| Heidelberger | Heidelberger-Welch convergence diagnostics | BAYES / DIAGNOSTICS=HEIDELBERGER |
| InitialValues | Initial values of the Markov chains | BAYES |
| ModelInfo | Model information | default |
| NObs | Number of observations | default |
| MCError | Monte Carlo standard errors | BAYES / DIAGNOSTICS=MCERROR |
| ParameterEstimates | Maximum likelihood estimates of model parameters | default |
| ParmInfo | Names of regression coefficients | CLASS,BAYES |
| Partition | Partition of constant baseline hazard intervals | BAYES / PIECEWISE |
| PostIntervals | equal-tail and high probability density intervals of the posterior samples | BAYES |
| PosteriorSample | Posterior samples | BAYES / (for ODS output data set only) |

**Table 64.6** *continued*

| ODS Table Name | Description | Statement / Option |
|---|---|---|
| PostSummaries | Summary statistics of the posterior samples | BAYES |
| Raftery | Raftery-Lewis convergence diagnostics | BAYES / DIAGNOSTICS=RAFTERY |
| ReferenceSet | Reference set of covariates for plotting | PROC / PLOTS= |
| SimpleStatistics | Summary statistics of input continuous explanatory variables | PROC / SIMPLE |

## ODS Graphics

To request graphics with PROC PHREG, you must first enable ODS Graphics by specifying the `ods graphics on` statement. See Chapter 21, "Statistical Graphics Using ODS," for more information. You can reference every graph produced through ODS Graphics with a name. The names of the graphs that PROC PHREG generates are listed separately in Table 64.7 for the maximum likelihood analysis and in Table 64.8 for the Bayesian analysis.

**Table 64.7** ODS Graphics for a Maximum Likelihood Analysis Produced by PROC PHREG

| ODS Graph Name | Plot Description | Statement / Option |
|---|---|---|
| CumhazPlot | Cumulative hazard function plot | PROC / PLOTS=CUMHAZ |
| CumulativeResiduals | Cumulative martingale residual plot | ASSESS / VAR= |
| CumResidPanel | Panel plot of cumulative martingale residuals | ASSESS / VAR=, CRPANEL |
| MCFPlot | Mean cumulative function plot | PROC / PLOTS=MCF |
| ScoreProcess | Standardized score process plot | ASSESS / PH |
| SurvivalPlot | Survivor function plot | PROC / PLOTS=SURVIVAL |

**Table 64.8** ODS Graphics for a Bayesian Analysis Produced by PROC PHREG

| ODS Graph Name | Plot Description | Statement / Option |
|---|---|---|
| ADPanel | Autocorrelation function and density panel | BAYES / PLOTS=(AUTOCORR DENSITY) |
| AutocorrPanel | Autocorrelation function panel | BAYES / PLOTS= AUTOCORR |
| AutocorrPlot | Autocorrelation function plot | BAYES / PLOTS(UNPACK)=AUTOCORR |

**Table 64.8** *continued*

| ODS Graph Name | Plot Description | Statement / Option |
|---|---|---|
| CumhazPlot | Cumulative hazard function plot | PROC / PLOTS=CUMHAZ |
| DensityPanel | Density panel | BAYES / PLOTS=DENSITY |
| DensityPlot | Density plot | BAYES / PLOTS(UNPACK)=DENSITY |
| SurvivalPlot | Survivor function plot | PROC / PLOTS=SURVIVAL |
| TAPanel | Trace and autocorrelation function panel | BAYES / PLOTS=(TRACE AUTOCORR) |
| TADPanel | Trace, density, and autocorrelation function panel | BAYES / PLOTS=(TRACE AUTOCORR DENSITY) |
| TDPanel | Trace and density panel | BAYES / PLOTS=(TRACE DENSITY) |
| TracePanel | Trace panel | BAYES / PLOTS=TRACE |
| TracePlot | Trace plot | BAYES / PLOTS(UNPACK)=TRACE |

# Examples: PHREG Procedure

This section contains 14 examples of PROC PHREG applications. The first 12 examples use the classical method of maximum likelihood, while the last two examples illustrate the Bayesian methodology.

## Example 64.1: Stepwise Regression

Krall, Uthoff, and Harley (1975) analyzed data from a study on multiple myeloma in which researchers treated 65 patients with alkylating agents. Of those patients, 48 died during the study and 17 survived. The following DATA step creates the data set Myeloma. The variable Time represents the survival time in months from diagnosis. The variable VStatus consists of two values, 0 and 1, indicating whether the patient was alive or dead, respectively, at the end of the study. If the value of VStatus is 0, the corresponding value of Time is censored. The variables thought to be related to survival are LogBUN (log(BUN) at diagnosis), HGB (hemoglobin at diagnosis), Platelet (platelets at diagnosis: 0=abnormal, 1=normal), Age (age at diagnosis, in years), LogWBC (log(WBC) at diagnosis), Frac (fractures at diagnosis: 0=none, 1=present), LogPBM (log percentage of plasma cells in bone marrow), Protein (proteinuria at diagnosis), and SCalc (serum calcium at diagnosis). Interest lies in identifying important prognostic factors from these nine explanatory variables.

```
data Myeloma;
   input Time VStatus LogBUN HGB Platelet Age LogWBC Frac
         LogPBM Protein SCalc;
   label Time='Survival Time'
         VStatus='0=Alive 1=Dead';
   datalines;
 1.25  1  2.2175   9.4  1  67  3.6628  1  1.9542  12  10
```

```
 1.25  1  1.9395  12.0  1  38  3.9868  1  1.9542  20  18
 2.00  1  1.5185   9.8  1  81  3.8751  1  2.0000   2  15
 2.00  1  1.7482  11.3  0  75  3.8062  1  1.2553   0  12
 2.00  1  1.3010   5.1  0  57  3.7243  1  2.0000   3   9
 3.00  1  1.5441   6.7  1  46  4.4757  0  1.9345  12  10
 5.00  1  2.2355  10.1  1  50  4.9542  1  1.6628   4   9
 5.00  1  1.6812   6.5  1  74  3.7324  0  1.7324   5   9
 6.00  1  1.3617   9.0  1  77  3.5441  0  1.4624   1   8
 6.00  1  2.1139  10.2  0  70  3.5441  1  1.3617   1   8
 6.00  1  1.1139   9.7  1  60  3.5185  1  1.3979   0  10
 6.00  1  1.4150  10.4  1  67  3.9294  1  1.6902   0   8
 7.00  1  1.9777   9.5  1  48  3.3617  1  1.5682   5  10
 7.00  1  1.0414   5.1  0  61  3.7324  1  2.0000   1  10
 7.00  1  1.1761  11.4  1  53  3.7243  1  1.5185   1  13
 9.00  1  1.7243   8.2  1  55  3.7993  1  1.7404   0  12
11.00  1  1.1139  14.0  1  61  3.8808  1  1.2788   0  10
11.00  1  1.2304  12.0  1  43  3.7709  1  1.1761   1   9
11.00  1  1.3010  13.2  1  65  3.7993  1  1.8195   1  10
11.00  1  1.5682   7.5  1  70  3.8865  0  1.6721   0  12
11.00  1  1.0792   9.6  1  51  3.5051  1  1.9031   0   9
13.00  1  0.7782   5.5  0  60  3.5798  1  1.3979   2  10
14.00  1  1.3979  14.6  1  66  3.7243  1  1.2553   2  10
15.00  1  1.6021  10.6  1  70  3.6902  1  1.4314   0  11
16.00  1  1.3424   9.0  1  48  3.9345  1  2.0000   0  10
16.00  1  1.3222   8.8  1  62  3.6990  1  0.6990  17  10
17.00  1  1.2304  10.0  1  53  3.8808  1  1.4472   4   9
17.00  1  1.5911  11.2  1  68  3.4314  0  1.6128   1  10
18.00  1  1.4472   7.5  1  65  3.5682  0  0.9031   7   8
19.00  1  1.0792  14.4  1  51  3.9191  1  2.0000   6  15
19.00  1  1.2553   7.5  0  60  3.7924  1  1.9294   5   9
24.00  1  1.3010  14.6  1  56  4.0899  1  0.4771   0   9
25.00  1  1.0000  12.4  1  67  3.8195  1  1.6435   0  10
26.00  1  1.2304  11.2  1  49  3.6021  1  2.0000  27  11
32.00  1  1.3222  10.6  1  46  3.6990  1  1.6335   1   9
35.00  1  1.1139   7.0  0  48  3.6532  1  1.1761   4  10
37.00  1  1.6021  11.0  1  63  3.9542  0  1.2041   7   9
41.00  1  1.0000  10.2  1  69  3.4771  1  1.4771   6  10
41.00  1  1.1461   5.0  1  70  3.5185  1  1.3424   0   9
51.00  1  1.5682   7.7  0  74  3.4150  1  1.0414   4  13
52.00  1  1.0000  10.1  1  60  3.8573  1  1.6532   4  10
54.00  1  1.2553   9.0  1  49  3.7243  1  1.6990   2  10
58.00  1  1.2041  12.1  1  42  3.6990  1  1.5798  22  10
66.00  1  1.4472   6.6  1  59  3.7853  1  1.8195   0   9
67.00  1  1.3222  12.8  1  52  3.6435  1  1.0414   1  10
88.00  1  1.1761  10.6  1  47  3.5563  0  1.7559  21   9
89.00  1  1.3222  14.0  1  63  3.6532  1  1.6232   1   9
92.00  1  1.4314  11.0  1  58  4.0755  1  1.4150   4  11
 4.00  0  1.9542  10.2  1  59  4.0453  0  0.7782  12  10
 4.00  0  1.9243  10.0  1  49  3.9590  0  1.6232   0  13
 7.00  0  1.1139  12.4  1  48  3.7993  1  1.8573   0  10
 7.00  0  1.5315  10.2  1  81  3.5911  0  1.8808   0  11
 8.00  0  1.0792   9.9  1  57  3.8325  1  1.6532   0   8
12.00  0  1.1461  11.6  1  46  3.6435  0  1.1461   0   7
11.00  0  1.6128  14.0  1  60  3.7324  1  1.8451   3   9
```

```
12.00  0  1.3979   8.8  1  66  3.8388  1  1.3617   0   9
13.00  0  1.6628   4.9  0  71  3.6435  0  1.7924   0   9
16.00  0  1.1461  13.0  1  55  3.8573  0  0.9031   0   9
19.00  0  1.3222  13.0  1  59  3.7709  1  2.0000   1  10
19.00  0  1.3222  10.8  1  69  3.8808  1  1.5185   0  10
28.00  0  1.2304   7.3  1  82  3.7482  1  1.6721   0   9
41.00  0  1.7559  12.8  1  72  3.7243  1  1.4472   1   9
53.00  0  1.1139  12.0  1  66  3.6128  1  2.0000   1  11
57.00  0  1.2553  12.5  1  66  3.9685  0  1.9542   0  11
77.00  0  1.0792  14.0  1  60  3.6812  0  0.9542   0  12
;
```

The stepwise selection process consists of a series of alternating forward selection and backward elimination steps. The former adds variables to the model, while the latter removes variables from the model.

The following statements use PROC PHREG to produce a stepwise regression analyis. Stepwise selection is requested by specifying the SELECTION=STEPWISE option in the MODEL statement. The option SLENTRY=0.25 specifies that a variable has to be significant at the 0.25 level before it can be entered into the model, while the option SLSTAY=0.15 specifies that a variable in the model has to be significant at the 0.15 level for it to remain in the model. The DETAILS option requests detailed results for the variable selection process.

```
proc phreg data=Myeloma;
   model Time*VStatus(0)=LogBUN HGB Platelet Age LogWBC
                         Frac LogPBM Protein SCalc
                         / selection=stepwise slentry=0.25
                           slstay=0.15 details;
run;
```

Results of the stepwise regression analysis are displayed in Output 64.1.1 through Output 64.1.7.

Individual score tests are used to determine which of the nine explanatory variables is first selected into the model. In this case, the score test for each variable is the global score test for the model containing that variable as the only explanatory variable. Output 64.1.1 displays the chi-square statistics and the corresponding $p$-values. The variable LogBUN has the largest chi-square value (8.5164), and it is significant ($p$=0.0035) at the SLENTRY=0.25 level. The variable LogBUN is thus entered into the model.

**Output 64.1.1** Individual Score Test Results for All Variables

```
                        The PHREG Procedure

                         Model Information

         Data Set                  WORK.MYELOMA
         Dependent Variable        Time              Survival Time
         Censoring Variable        VStatus           0=Alive 1=Dead
         Censoring Value(s)        0
         Ties Handling             BRESLOW
```

**Output 64.1.1** *continued*

**Summary of the Number of Event and Censored Values**

| Total | Event | Censored | Percent Censored |
|---|---|---|---|
| 65 | 48 | 17 | 26.15 |

**Analysis of Effects Eligible for Entry**

| Effect | DF | Score Chi-Square | Pr > ChiSq |
|---|---|---|---|
| LogBUN | 1 | 8.5164 | 0.0035 |
| HGB | 1 | 5.0664 | 0.0244 |
| Platelet | 1 | 3.1816 | 0.0745 |
| Age | 1 | 0.0183 | 0.8924 |
| LogWBC | 1 | 0.5658 | 0.4519 |
| Frac | 1 | 0.9151 | 0.3388 |
| LogPBM | 1 | 0.5846 | 0.4445 |
| Protein | 1 | 0.1466 | 0.7018 |
| SCalc | 1 | 1.1109 | 0.2919 |

**Residual Chi-Square Test**

| Chi-Square | DF | Pr > ChiSq |
|---|---|---|
| 18.4550 | 9 | 0.0302 |

Output 64.1.2 displays the results of the first model. Since the Wald chi-square statistic is significant ($p = 0.0039$) at the SLSTAY=0.15 level, LogBUN stays in the model.

**Output 64.1.2** First Model in the Stepwise Selection Process

Step 1. Effect LogBUN is entered. The model contains the following effects:

LogBUN

**Convergence Status**

Convergence criterion (GCONV=1E-8) satisfied.

**Model Fit Statistics**

| Criterion | Without Covariates | With Covariates |
|---|---|---|
| -2 LOG L | 309.716 | 301.959 |
| AIC | 309.716 | 303.959 |
| SBC | 309.716 | 305.830 |

**Output 64.1.2** *continued*

**Testing Global Null Hypothesis: BETA=0**

| Test | Chi-Square | DF | Pr > ChiSq |
|---|---|---|---|
| Likelihood Ratio | 7.7572 | 1 | 0.0053 |
| Score | 8.5164 | 1 | 0.0035 |
| Wald | 8.3392 | 1 | 0.0039 |

**Analysis of Maximum Likelihood Estimates**

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| LogBUN | 1 | 1.74595 | 0.60460 | 8.3392 | 0.0039 | 5.731 |

The next step consists of selecting another variable to add to the model. Output 64.1.3 displays the chi-square statistics and $p$-values of individual score tests (adjusted for LogBUN) for the remaining eight variables. The score chi-square for a given variable is the value of the likelihood score test for testing the significance of the variable in the presence of LogBUN. The variable HGB is selected because it has the highest chi-square value (4.3468), and it is significant ($p = 0.0371$) at the SLENTRY=0.25 level.

**Output 64.1.3** Score Tests Adjusted for the Variable LogBUN

**Analysis of Effects Eligible for Entry**

| Effect | DF | Score Chi-Square | Pr > ChiSq |
|---|---|---|---|
| HGB | 1 | 4.3468 | 0.0371 |
| Platelet | 1 | 2.0183 | 0.1554 |
| Age | 1 | 0.7159 | 0.3975 |
| LogWBC | 1 | 0.0704 | 0.7908 |
| Frac | 1 | 1.0354 | 0.3089 |
| LogPBM | 1 | 1.0334 | 0.3094 |
| Protein | 1 | 0.5214 | 0.4703 |
| SCalc | 1 | 1.4150 | 0.2342 |

**Residual Chi-Square Test**

| Chi-Square | DF | Pr > ChiSq |
|---|---|---|
| 9.3164 | 8 | 0.3163 |

Output 64.1.4 displays the fitted model containing both LogBUN and HGB. Based on the Wald statistics, neither LogBUN nor HGB is removed from the model.

**Output 64.1.4** Second Model in the Stepwise Selection Process

```
Step 2. Effect HGB is entered. The model contains the following effects:

         LogBUN  HGB


                              Convergence Status

                Convergence criterion (GCONV=1E-8) satisfied.

                            Model Fit Statistics

                                 Without          With
                   Criterion    Covariates    Covariates

                   -2 LOG L        309.716       297.767
                   AIC             309.716       301.767
                   SBC             309.716       305.509

                   Testing Global Null Hypothesis: BETA=0

           Test                 Chi-Square        DF     Pr > ChiSq

           Likelihood Ratio        11.9493         2         0.0025
           Score                   12.7252         2         0.0017
           Wald                    12.1900         2         0.0023

                  Analysis of Maximum Likelihood Estimates

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

 LogBUN          1      1.67440      0.61209        7.4833        0.0062     5.336
 HGB             1     -0.11899      0.05751        4.2811        0.0385     0.888
```

Output 64.1.5 shows Step 3 of the selection process, in which the variable SCalc is added, resulting in the model with LogBUN, HGB, and SCalc as the explanatory variables. Note that SCalc has the smallest Wald chi-square statistic, and it is not significant ($p = 0.1782$) at the SLSTAY=0.15 level.

**Output 64.1.5** Third Model in the Stepwise Regression

```
Step 3. Effect SCalc is entered. The model contains the following effects:

         LogBUN  HGB  SCalc


                              Convergence Status

                Convergence criterion (GCONV=1E-8) satisfied.
```

**Output 64.1.5** *continued*

Model Fit Statistics

| Criterion | Without Covariates | With Covariates |
|---|---|---|
| -2 LOG L | 309.716 | 296.078 |
| AIC | 309.716 | 302.078 |
| SBC | 309.716 | 307.692 |

Testing Global Null Hypothesis: BETA=0

| Test | Chi-Square | DF | Pr > ChiSq |
|---|---|---|---|
| Likelihood Ratio | 13.6377 | 3 | 0.0034 |
| Score | 15.3053 | 3 | 0.0016 |
| Wald | 14.4542 | 3 | 0.0023 |

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| LogBUN | 1 | 1.63593 | 0.62359 | 6.8822 | 0.0087 | 5.134 |
| HGB | 1 | -0.12643 | 0.05868 | 4.6419 | 0.0312 | 0.881 |
| SCalc | 1 | 0.13286 | 0.09868 | 1.8127 | 0.1782 | 1.142 |

The variable SCalc is then removed from the model in a step-down phase in Step 4 (Output 64.1.6). The removal of SCalc brings the stepwise selection process to a stop in order to avoid repeatedly entering and removing the same variable.

**Output 64.1.6** Final Model in the Stepwise Regression

Step 4. Effect SCalc is removed. The model contains the following effects:

LogBUN HGB

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

| Criterion | Without Covariates | With Covariates |
|---|---|---|
| -2 LOG L | 309.716 | 297.767 |
| AIC | 309.716 | 301.767 |
| SBC | 309.716 | 305.509 |

**Output 64.1.6** *continued*

**Testing Global Null Hypothesis: BETA=0**

| Test | Chi-Square | DF | Pr > ChiSq |
|---|---|---|---|
| Likelihood Ratio | 11.9493 | 2 | 0.0025 |
| Score | 12.7252 | 2 | 0.0017 |
| Wald | 12.1900 | 2 | 0.0023 |

**Analysis of Maximum Likelihood Estimates**

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| LogBUN | 1 | 1.67440 | 0.61209 | 7.4833 | 0.0062 | 5.336 |
| HGB | 1 | -0.11899 | 0.05751 | 4.2811 | 0.0385 | 0.888 |

NOTE: Model building terminates because the effect to be entered is the effect that was removed in the last step.

The procedure also displays a summary table of the steps in the stepwise selection process, as shown in Output 64.1.7.

**Output 64.1.7** Model Selection Summary

**Summary of Stepwise Selection**

| Step | Effect Entered | Effect Removed | DF | Number In | Score Chi-Square | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|---|
| 1 | LogBUN | | 1 | 1 | 8.5164 | | 0.0035 |
| 2 | HGB | | 1 | 2 | 4.3468 | | 0.0371 |
| 3 | SCalc | | 1 | 3 | 1.8225 | | 0.1770 |
| 4 | | SCalc | 1 | 2 | | 1.8127 | 0.1782 |

The stepwise selection process results in a model with two explanatory variables, LogBUN and HGB.

## Example 64.2: Best Subset Selection

An alternative to stepwise selection of variables is best subset selection. This method uses the branch-and-bound algorithm of Furnival and Wilson (1974) to find a specified number of best models containing one, two, or three variables, and so on, up to the single model containing all of the explanatory variables. The criterion used to determine the “best” subset is based on the global score chi-square statistic. For two models A and B, each having the same number of explanatory variables, model A is considered to be better than model B if the global score chi-square statistic for A exceeds that for B.

In the following statements, best subset selection analysis is requested by specifying the SELECTION=SCORE option in the MODEL statement. The BEST=3 option requests the procedure to identify only the three best models for each size. In other words, PROC PHREG will list the three models having the highest score statistics of all the models possible for a given number of covariates.

```
proc phreg data=Myeloma;
   model Time*VStatus(0)=LogBUN HGB Platelet Age LogWBC
                         Frac LogPBM Protein SCalc
                         / selection=score best=3;
run;
```

Output 64.2.1 displays the results of this analysis. The number of explanatory variables in the model is given in the first column, and the names of the variables are listed on the right. The models are listed in descending order of their score chi-square values within each model size. For example, among all models containing two explanatory variables, the model that contains the variables LogBUN and HGB has the largest score value (12.7252), the model that contains the variables LogBUN and Platelet has the second-largest score value (11.1842), and the model that contains the variables LogBUN and SCalc has the third-largest score value (9.9962).

**Output 64.2.1** Best Variable Combinations

The PHREG Procedure

Regression Models Selected by Score Criterion

| Number of Variables | Score Chi-Square | Variables Included in Model |
|---|---|---|
| 1 | 8.5164 | LogBUN |
| 1 | 5.0664 | HGB |
| 1 | 3.1816 | Platelet |
| 2 | 12.7252 | LogBUN HGB |
| 2 | 11.1842 | LogBUN Platelet |
| 2 | 9.9962 | LogBUN SCalc |
| 3 | 15.3053 | LogBUN HGB SCalc |
| 3 | 13.9911 | LogBUN HGB Age |
| 3 | 13.5788 | LogBUN HGB Frac |
| 4 | 16.9873 | LogBUN HGB Age SCalc |
| 4 | 16.0457 | LogBUN HGB Frac SCalc |
| 4 | 15.7619 | LogBUN HGB LogPBM SCalc |
| 5 | 17.6291 | LogBUN HGB Age Frac SCalc |
| 5 | 17.3519 | LogBUN HGB Age LogPBM SCalc |
| 5 | 17.1922 | LogBUN HGB Age LogWBC SCalc |
| 6 | 17.9120 | LogBUN HGB Age Frac LogPBM SCalc |
| 6 | 17.7947 | LogBUN HGB Age LogWBC Frac SCalc |
| 6 | 17.7744 | LogBUN HGB Platelet Age Frac SCalc |
| 7 | 18.1517 | LogBUN HGB Platelet Age Frac LogPBM SCalc |
| 7 | 18.0568 | LogBUN HGB Age LogWBC Frac LogPBM SCalc |
| 7 | 18.0223 | LogBUN HGB Platelet Age LogWBC Frac SCalc |
| 8 | 18.3925 | LogBUN HGB Platelet Age LogWBC Frac LogPBM SCalc |
| 8 | 18.1636 | LogBUN HGB Platelet Age Frac LogPBM Protein SCalc |
| 8 | 18.1309 | LogBUN HGB Platelet Age LogWBC Frac Protein SCalc |
| 9 | 18.4550 | LogBUN HGB Platelet Age LogWBC Frac LogPBM Protein SCalc |

## Example 64.3: Modeling with Categorical Predictors

Consider the data for the Veterans Administration lung cancer trial presented in Appendix 1 of Kalbfleisch and Prentice (1980). In this trial, males with advanced inoperable lung cancer were randomized to a standard therapy and a test chemotherapy. The primary endpoint for the therapy comparison was time to death in days, represented by the variable Time. Negative values of Time are censored values. The data include information about a number of explanatory variables: Therapy (type of therapy: standard or test), Cell (type of tumor cell: adeno, large, small, or squamous), Prior (prior therapy: 0=no, 10=yes), Age (age, in years), Duration (months from diagnosis to randomization), and Kps (Karnofsky performance scale). A censoring indicator variable, Censor, is created from the data, with the value 1 indicating a censored time and the value 0 indicating an event time. The following DATA step saves the data in the data set VALung.

```
proc format;
   value yesno 0='no' 10='yes';
   run;

data VALung;
   drop check m;
   retain Therapy Cell;
   infile cards column=column;
   length Check $ 1;
   label Time='time to death in days'
         Kps='Karnofsky performance scale'
         Duration='months from diagnosis to randomization'
         Age='age in years'
         Prior='prior therapy'
         Cell='cell type'
         Therapy='type of treatment';
   format Prior yesno.;
   M=Column;
   input Check $ @@;
   if M>Column then M=1;
   if Check='s'|Check='t' then do;
      input @M Therapy $ Cell $;
      delete;
      end;
   else do;
      input @M Time Kps Duration Age Prior @@;
      Status=(Time>0);
      Time=abs(Time);
      end;
   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
;
```

The following statements use the PHREG procedure to fit the Cox proportional hazards model to these data. The variables Prior, Cell, and Therapy, which are categorical variables, are declared in the CLASS statement. By default, PROC PHREG parameterizes the CLASS variables by using the reference coding with the last category as the reference category. However, you can explicitly specify the reference category of your choice. Here, Prior=no is chosen as the reference category for prior therapy, Cell=large is chosen as the reference category for type of tumor cell, and Therapy=standard is chosen as the reference category for the type of therapy. In the MODEL statement, the term Prior|Therapy is just another way of specifying the main effects Prior, Therapy, and the Prior*Therapy interaction.

```
proc phreg data=VALung;
   class Prior(ref='no') Cell(ref='large') Therapy(ref='standard');
   model Time*Status(0) = Kps Duration Age Cell Prior|Therapy;
   run;
```

Coding of the CLASS variables is displayed in Output 64.3.1. There is one dummy variable for Prior and one for Therapy, since both variables are binary. The dummy variable has a value of 0 for the reference category (Prior=no, Therapy=standard). The variable Cell has four categories and is represented by three dummy variables. Note that the reference category, Cell=large, has a value of 0 for all three dummy variables.

**Output 64.3.1** Reference Coding of CLASS Variables

The PHREG Procedure

Class Level Information

| Class | Value | Design Variables | | |
|---|---|---|---|---|
| Prior | no | 0 | | |
| | yes | 1 | | |
| Cell | adeno | 1 | 0 | 0 |
| | large | 0 | 0 | 0 |
| | small | 0 | 1 | 0 |
| | squamous | 0 | 0 | 1 |
| Therapy | standard | 0 | | |
| | test | 1 | | |

The test results of individual model effects are shown in Output 64.3.2. There is a strong prognostic effect of Kps on patient's survivivorship ($p < 0.0001$), and the survival times for patients of different Cell types differ significantly ($p = 0.0003$). The Prior*Therapy interaction is marginally significant ($p$=0.0416)—that is, prior therapy might play a role in whether one treatment is more effective than the other.

**Output 64.3.2** Wald Tests of Individual Effects

Type 3 Tests

| Effect | DF | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|
| Kps | 1 | 35.5051 | <.0001 |
| Duration | 1 | 0.1159 | 0.7335 |
| Age | 1 | 1.9772 | 0.1597 |
| Cell | 3 | 18.5339 | 0.0003 |
| Prior | 1 | 2.5296 | 0.1117 |
| Therapy | 1 | 5.2349 | 0.0221 |
| Prior*Therapy | 1 | 4.1528 | 0.0416 |

In the Cox proportional hazards model, the effects of the covariates are to act multiplicatively on the hazard of the survival time, and therefore it is a little easier to interpret the corresponding hazard ratios than the regression parameters. For a parameter that corresponds to a continuous variable, the hazard ratio is the ratio of hazard rates for a increase of one unit of the variable. From Output 64.3.3, the hazard ratio estimate for Kps is 0.968, meaning that an increase of 10 units in

Karnofsky performance scale will shrink the hazard rate by $1 - (0.968)^{10}$=28%. For a CLASS variable parameter, the hazard ratio presented in the Output 64.3.3 is the ratio of the hazard rates between the given category and the reference category. The hazard rate of Cell=adeno is 219% that of Cell=large, the hazard rate of Cell=small is 62% that of Cell=large, and the hazard rate of Cell=squamous is only 66% that of Cell=large. Hazard ratios for Prior and Therapy are missing since the model contains the Prior*Therapy interaction. You can use the HAZARDRATIO statement to obtain the hazard ratios for a main effect in the presence of interaction as shown later in this example.

**Output 64.3.3** Parameters Estimates with Reference Coding

**Analysis of Maximum Likelihood Estimates**

| Parameter | | | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|---|
| Kps | | | 1 | -0.03300 | 0.00554 | 35.5051 | <.0001 |
| Duration | | | 1 | 0.00323 | 0.00949 | 0.1159 | 0.7335 |
| Age | | | 1 | -0.01353 | 0.00962 | 1.9772 | 0.1597 |
| Cell | adeno | | 1 | 0.78356 | 0.30382 | 6.6512 | 0.0099 |
| Cell | small | | 1 | 0.48230 | 0.26537 | 3.3032 | 0.0691 |
| Cell | squamous | | 1 | -0.40770 | 0.28363 | 2.0663 | 0.1506 |
| Prior | yes | | 1 | 0.45914 | 0.28868 | 2.5296 | 0.1117 |
| Therapy | test | | 1 | 0.56662 | 0.24765 | 5.2349 | 0.0221 |
| Prior*Therapy | yes | test | 1 | -0.87579 | 0.42976 | 4.1528 | 0.0416 |

**Analysis of Maximum Likelihood Estimates**

| Parameter | | | Hazard Ratio |
|---|---|---|---|
| Kps | | | 0.968 |
| Duration | | | 1.003 |
| Age | | | 0.987 |
| Cell | adeno | | 2.189 |
| Cell | small | | 1.620 |
| Cell | squamous | | 0.665 |
| Prior | yes | | . |
| Therapy | test | | . |
| Prior*Therapy | yes | test | . |

**Analysis of Maximum Likelihood Estimates**

| Parameter | | | Label |
|---|---|---|---|
| Kps | | | Karnofsky performance scale |
| Duration | | | months from diagnosis to randomization |
| Age | | | age in years |
| Cell | adeno | | cell type adeno |
| Cell | small | | cell type small |
| Cell | squamous | | cell type squamous |
| Prior | yes | | prior therapy yes |
| Therapy | test | | type of treatment test |
| Prior*Therapy | yes | test | prior therapy yes * type of treatment test |

The following PROC PHREG statements illustrate the use of the backward elimination process to identify the effects that affect the survivorship of the lung cancer patients. The option SELECTION=BACKWARD is specified to carry out the backward elimination. The option SLSTAY=0.1 specifies the significant level for retaining the effects in the model.

```
proc phreg data=VALung;
   class Prior(ref='no') Cell(ref='large') Therapy(ref='standard');
   model Time*Status(0) = Kps Duration Age Cell Prior|Therapy / selection=backward
                                                                slstay=0.1;
   run;
```

Results of the backward elimination process are summarized in Output 64.3.4. The effect Duration was eliminated first and was followed by Age.

**Output 64.3.4** Effects Eliminated from the Model

The PHREG Procedure

Summary of Backward Elimination

| Step | Effect Removed | DF | Number In | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|
| 1 | Duration | 1 | 6 | 0.1159 | 0.7335 |
| 2 | Age | 1 | 5 | 2.0458 | 0.1526 |

Summary of Backward Elimination

| Step | Effect Label |
|---|---|
| 1 | months from diagnosis to randomization |
| 2 | age in years |

Output 64.3.5 shows the Type 3 analysis of effects and the maximum likelihood estimates of the regression coefficients of the model. Without controlling for Age and Duration, KPS and Cell remain significant, but the Prior*Therapy interaction is less prominent than before ($p$=0.0871) though still significant at 0.1 level.

**Output 64.3.5** Type 3 Effects and Parameter Estimates for the Selected Model

Type 3 Tests

| Effect | DF | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|
| Kps | 1 | 35.9218 | <.0001 |
| Cell | 3 | 17.4134 | 0.0006 |
| Prior | 1 | 2.3113 | 0.1284 |
| Therapy | 1 | 3.8030 | 0.0512 |
| Prior*Therapy | 1 | 2.9269 | 0.0871 |

**Output 64.3.5** *continued*

**Analysis of Maximum Likelihood Estimates**

| Parameter | | | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|---|
| Kps | | | 1 | -0.03111 | 0.00519 | 35.9218 | <.0001 |
| Cell | adeno | | 1 | 0.74907 | 0.30465 | 6.0457 | 0.0139 |
| Cell | small | | 1 | 0.44265 | 0.26168 | 2.8614 | 0.0907 |
| Cell | squamous | | 1 | -0.41145 | 0.28309 | 2.1125 | 0.1461 |
| Prior | yes | | 1 | 0.41755 | 0.27465 | 2.3113 | 0.1284 |
| Therapy | test | | 1 | 0.45670 | 0.23419 | 3.8030 | 0.0512 |
| Prior*Therapy | yes | test | 1 | -0.69443 | 0.40590 | 2.9269 | 0.0871 |

**Analysis of Maximum Likelihood Estimates**

| Parameter | | | Hazard Ratio |
|---|---|---|---|
| Kps | | | 0.969 |
| Cell | adeno | | 2.115 |
| Cell | small | | 1.557 |
| Cell | squamous | | 0.663 |
| Prior | yes | | . |
| Therapy | test | | . |
| Prior*Therapy | yes | test | . |

**Analysis of Maximum Likelihood Estimates**

| Parameter | | | Label |
|---|---|---|---|
| Kps | | | Karnofsky performance scale |
| Cell | adeno | | cell type adeno |
| Cell | small | | cell type small |
| Cell | squamous | | cell type squamous |
| Prior | yes | | prior therapy yes |
| Therapy | test | | type of treatment test |
| Prior*Therapy | yes | test | prior therapy yes * type of treatment test |

Finally, the following statements refit the previous model and computes hazard ratios at settings beyond those displayed in the "Analysis of Maximum Likelihood Estimates" table. You can use either the HAZARDRATIO statement or the CONTRAST statement to obtain hazard ratios. Using the CONTRAST statement to compute hazard ratios for CLASS variables can be a daunting task unless you are familiar with the parameterization schemes (see the section "CLASS Variable Parameterization" on page 4574 for more information), but you have control over which specific hazard ratios you want to compute. HAZARDRATIO statements, on the other hand, are designed specifically to provide hazard ratios. They are easy to use and you can also request both the Wald confidence limits and the profile-likelihood confidence limits; the latter is not available for the CONTRAST statements. Three HAZARDRATIO statements are specified; each has the CL=BOTH option to request both the Wald confidence limits and the profile-likelihood limits. The first HAZARDRATIO statement, labeled 'H1', estimates the hazard ratio for an increase of 10 units in the KPS; the UNITS= option specifies the number of units increase. The second HAZARDRATIO statement, labeled 'H2' computes the hazard ratios for comparing any pairs of tumor Cell types. The third

HAZARDRATIO statement, labeled 'H3', compares the test therapy with the standard therapy. The DIFF=REF option specifies that each nonreference category is compared to the reference category. The purpose of using DIFF=REF here is to ensure that the hazard ratio is comparing the test therapy to the standard therapy instead of the other way around. Three CONTRAST statements, labeled 'C1', 'C2', and 'C3', parallel to the HAZARDRATIO statements 'H1', 'H2', and 'H3', respectively, are specified. The ESTIMATE=EXP option specifies that the linear predictors be estimated in the exponential scale, which are precisely the hazard ratios.

```
proc phreg data=VALung;
   class Prior(ref='no') Cell(ref='large') Therapy(ref='standard');
   model Time*Status(0) = Kps Cell Prior|Therapy;
   hazardratio 'H1' Kps / units=10 cl=both;
   hazardratio 'H2' Cell / cl=both;
   hazardratio 'H3' Therapy / diff=ref cl=both;
   contrast 'C1' Kps 10 / estimate=exp;
   contrast 'C2' cell 1  0  0, /* adeno vs large    */
                 cell 1 -1  0, /* adeno vs small    */
                 cell 1  0 -1, /* adeno vs squamous */
                 cell 0 -1  0, /* large vs small    */
                 cell 0  0 -1, /* large vs Squamous */
                 cell 0  1 -1  /* small vs squamous */
                  / estimate=exp;
   contrast 'C3' Prior 0 Therapy 1  Prior*Therapy 0,
                 Prior 0 Therapy 1  Prior*Therapy 1  / estimate=exp;
   run;
```

Output 64.3.6 displays the results of the three HAZARDRATIO statements in separate tables. Results of the three CONTRAST statements are shown in one table in Output 64.3.7. However, point estimates and the Wald confidence limits for the hazard ratio agree in between the two outputs.

**Output 64.3.6** Results from HAZARDRATIO Statements

The PHREG Procedure

H1: Hazard Ratios for Kps

| Description | Point Estimate | 95% Wald Confidence Limits | | 95% Profile Likelihood Confidence Limits | |
|---|---|---|---|---|---|
| Kps Unit=10 | 0.733 | 0.662 | 0.811 | 0.662 | 0.811 |

H2: Hazard Ratios for Cell

| Description | Point Estimate | 95% Wald Confidence Limits | | 95% Profile Likelihood Confidence Limits | |
|---|---|---|---|---|---|
| Cell adeno vs large | 2.115 | 1.164 | 3.843 | 1.162 | 3.855 |
| Cell adeno vs small | 1.359 | 0.798 | 2.312 | 0.791 | 2.301 |
| Cell adeno vs squamous | 3.192 | 1.773 | 5.746 | 1.770 | 5.768 |
| Cell large vs small | 0.642 | 0.385 | 1.073 | 0.380 | 1.065 |
| Cell large vs squamous | 1.509 | 0.866 | 2.628 | 0.863 | 2.634 |
| Cell small vs squamous | 2.349 | 1.387 | 3.980 | 1.399 | 4.030 |

**Output 64.3.6** *continued*

H3: Hazard Ratios for Therapy

| Description | Point Estimate | 95% Wald Confidence Limits | |
|---|---|---|---|
| Therapy test vs standard At Prior=no | 1.579 | 0.998 | 2.499 |
| Therapy test vs standard At Prior=yes | 0.788 | 0.396 | 1.568 |

H3: Hazard Ratios for Therapy

| 95% Profile Likelihood Confidence Limits | |
|---|---|
| 0.998 | 2.506 |
| 0.390 | 1.560 |

**Output 64.3.7** Results from CONTRAST Statements

Contrast Rows Estimation and Testing Results

| Contrast | Type | Row | Estimate | Standard Error | Alpha | Confidence Limits | |
|---|---|---|---|---|---|---|---|
| C1 | EXP | 1 | 0.7326 | 0.0380 | 0.05 | 0.6618 | 0.8111 |
| C2 | EXP | 1 | 2.1150 | 0.6443 | 0.05 | 1.1641 | 3.8427 |
| C2 | EXP | 2 | 1.3586 | 0.3686 | 0.05 | 0.7982 | 2.3122 |
| C2 | EXP | 3 | 3.1916 | 0.9575 | 0.05 | 1.7727 | 5.7462 |
| C2 | EXP | 4 | 0.6423 | 0.1681 | 0.05 | 0.3846 | 1.0728 |
| C2 | EXP | 5 | 1.5090 | 0.4272 | 0.05 | 0.8664 | 2.6282 |
| C2 | EXP | 6 | 2.3493 | 0.6318 | 0.05 | 1.3868 | 3.9797 |
| C3 | EXP | 1 | 1.5789 | 0.3698 | 0.05 | 0.9977 | 2.4985 |
| C3 | EXP | 2 | 0.7884 | 0.2766 | 0.05 | 0.3964 | 1.5680 |

Contrast Rows Estimation and Testing Results

| Contrast | Type | Row | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|
| C1 | EXP | 1 | 35.9218 | <.0001 |
| C2 | EXP | 1 | 6.0457 | 0.0139 |
| C2 | EXP | 2 | 1.2755 | 0.2587 |
| C2 | EXP | 3 | 14.9629 | 0.0001 |
| C2 | EXP | 4 | 2.8614 | 0.0907 |
| C2 | EXP | 5 | 2.1125 | 0.1461 |
| C2 | EXP | 6 | 10.0858 | 0.0015 |
| C3 | EXP | 1 | 3.8030 | 0.0512 |
| C3 | EXP | 2 | 0.4593 | 0.4980 |

## Example 64.4: Firth's Correction for Monotone Likelihood

In fitting the Cox regression model by maximizing the partial likelihood, the estimate of an explanatory variable X will be infinite if the value of X at each uncensored failure time is the largest of all the values of X in the risk set at that time (Tsiatis 1981; Bryson and Johnson 1981). You can exploit this information to artificially create a data set that has the condition of monotone likelihood for the Cox regression. The following DATA step modifies the Myeloma data in Example 64.1 to create a new explanatory variable, Contrived, which has the value 1 if the observed time is less than or equal to 65 and has the value 0 otherwise. The phenomenon of monotone likelihood will be demonstrated in the new data set Myeloma2.

```
data Myeloma2;
   set Myeloma;
   Contrived= (Time <= 65);
   run;
```

For illustration purposes, consider a Cox model with three explanatory variables, one of which is the variable Contrived. The following statements invoke PROC PHREG to perform the Cox regression. The IPRINT option is specified in the MODEL statement to print the iteration history of the optimization.

```
proc phreg data=Myeloma2;
   model Time*Vstatus(0)=LOGbun HGB Contrived / itprint;
   run;
```

The symptom of monotonity is demonstrated in Output 64.4.1. The log likelihood converges to the value of –136.56 while the coefficient for Contrived diverges. Although the Newton-Raphson optimization process did not fail, it is obvious that convergence is questionable. A close examination of the standard errors in the “Analyis of Maximum Likelihood Estimates” table reveals a very large value for the coefficient of Contrived. This is very typical of a diverged estimate.

**Output 64.4.1** Monotone Likelihood Behavior Displayed

**The PHREG Procedure**

**Maximum Likelihood Iteration History**

| Iter | Ridge | Log Likelihood | LogBUN | HGB | Contrived |
|---|---|---|---|---|---|
| 0 | 0 | -154.8579914384 | 0.0000000000 | 0.000000000 | 0.000000000 |
| 1 | 0 | -140.6934052686 | 1.9948819671 | -0.084318519 | 1.466331269 |
| 2 | 0 | -137.7841629416 | 1.6794678962 | -0.109067888 | 2.778361123 |
| 3 | 0 | -136.9711897754 | 1.7140611684 | -0.111564202 | 3.938095086 |
| 4 | 0 | -136.7078932606 | 1.7181735043 | -0.112273248 | 5.003053568 |
| 5 | 0 | -136.6164264879 | 1.7187547532 | -0.112369756 | 6.027435769 |
| 6 | 0 | -136.5835200895 | 1.7188294108 | -0.112382079 | 7.036444978 |
| 7 | 0 | -136.5715152788 | 1.7188392687 | -0.112383700 | 8.039763533 |
| 8 | 0 | -136.5671126045 | 1.7188405904 | -0.112383917 | 9.040984886 |
| 9 | 0 | -136.5654947987 | 1.7188407687 | -0.112383947 | 10.041434266 |
| 10 | 0 | -136.5648998913 | 1.7188407928 | -0.112383950 | 11.041599592 |
| 11 | 0 | -136.5646810709 | 1.7188407960 | -0.112383951 | 12.041660414 |
| 12 | 0 | -136.5646005760 | 1.7188407965 | -0.112383951 | 13.041682789 |
| 13 | 0 | -136.5645709642 | 1.7188407965 | -0.112383951 | 14.041691020 |
| 14 | 0 | -136.5645600707 | 1.7188407965 | -0.112383951 | 15.041694049 |
| 15 | 0 | -136.5645560632 | 1.7188407965 | -0.112383951 | 16.041695162 |
| 16 | 0 | -136.5645545889 | 1.7188407965 | -0.112383951 | 17.041695572 |

**Output 64.4.1** *continued*

**Analysis of Maximum Likelihood Estimates**

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| LogBUN | 1 | 1.71884 | 0.58376 | 8.6697 | 0.0032 | 5.578 |
| HGB | 1 | -0.11238 | 0.06090 | 3.4053 | 0.0650 | 0.894 |
| Contrived | 1 | 17.04170 | 1080 | 0.0002 | 0.9874 | 25183399 |

Next, the Firth correction was applied as shown in the following statements. Also, the profile-likelihood confidence limits for the hazard ratios are requested by using the RISKLIMITS=PL option.

```
proc phreg data=Myeloma2;
   model Time*Vstatus(0)=LogBUN HGB Contrived /
      firth risklimits=pl itprint;
   run;
```

PROC PHREG uses the penalized likelihood maximum to obtain a finite estimate for the coefficient of Contrived (Output 64.4.2). The much preferred profile-likelihood confidence limits, as shown in (Heinze and Schemper 2001), are also displayed.

**Output 64.4.2** Convergence Obtained with the Firth Correction

**The PHREG Procedure**

**Maximum Likelihood Iteration History**

| Iter | Ridge | Log Likelihood | LogBUN | HGB | Contrived |
|---|---|---|---|---|---|
| 0 | 0 | -150.7361197494 | 0.0000000000 | 0.000000000 | 0.0000000000 |
| 1 | 0 | -136.9933949142 | 2.0262484120 | -0.086519138 | 1.4338859318 |
| 2 | 0 | -134.5796594364 | 1.6770836974 | -0.109172604 | 2.6221444778 |
| 3 | 0 | -134.1572923217 | 1.7163408994 | -0.111166227 | 3.4458043289 |
| 4 | 0 | -134.1229607193 | 1.7209210332 | -0.112007726 | 3.7923555412 |
| 5 | 0 | -134.1228364805 | 1.7219588214 | -0.112178328 | 3.8174197804 |
| 6 | 0 | -134.1228355256 | 1.7220081673 | -0.112187764 | 3.8151642206 |

**Output 64.4.2** *continued*

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|
| LogBUN | 1 | 1.72201 | 0.58379 | 8.7008 | 0.0032 |
| HGB | 1 | -0.11219 | 0.06059 | 3.4279 | 0.0641 |
| Contrived | 1 | 3.81516 | 1.55812 | 5.9955 | 0.0143 |

Analysis of Maximum Likelihood Estimates

| Parameter | Hazard Ratio | 95% Hazard Ratio Profile Likelihood Confidence Limits | |
|---|---|---|---|
| LogBUN | 5.596 | 1.761 | 17.231 |
| HGB | 0.894 | 0.794 | 1.007 |
| Contrived | 45.384 | 5.406 | 6005.404 |

## Example 64.5: Conditional Logistic Regression for m:n Matching

Conditional logistic regression is used to investigate the relationship between an outcome and a set of prognostic factors in matched case-control studies. The outcome is whether the subject is a case or a control. If there is only one case and one control, the matching is 1:1. The *m:n* matching refers to the situation in which there is a varying number of cases and controls in the matched sets. You can perform conditional logistic regression with the PHREG procedure by using the discrete logistic model and forming a stratum for each matched set. In addition, you need to create dummy survival times so that all the cases in a matched set have the same event time value, and the corresponding controls are censored at later times.

Consider the following set of low infant birth-weight data extracted from Appendix 1 of Hosmer and Lemeshow (1989). These data represent 189 women, of whom 59 had low-birth-weight babies and 130 had normal-weight babies. Under investigation are the following risk factors: weight in pounds at the last menstrual period (LWT), presence of hypertension (HT), smoking status during pregnancy (Smoke), and presence of uterine irritability (UI). For HT, Smoke, and UI, a value of 1 indicates a "yes" and a value of 0 indicates a "no." The woman's age (Age) is used as the matching variable. The SAS data set LBW contains a subset of the data corresponding to women between the ages of 16 and 32.

```
data LBW;
   input id Age Low LWT Smoke HT UI @@;
   Time=2-Low;
   datalines;
 25  16   1   130   0  0  0    143  16   0   110   0  0  0
166  16   0   112   0  0  0    167  16   0   135   1  0  0
189  16   0   135   1  0  0    206  16   0   170   0  0  0
216  16   0    95   0  0  0     37  17   1   130   1  0  1
```

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 45 | 17 | 1 | 110 | 1 | 0 | 0 | 68 | 17 | 1 | 120 | 1 | 0 | 0 |
| 71 | 17 | 1 | 120 | 0 | 0 | 0 | 83 | 17 | 1 | 142 | 0 | 1 | 0 |
| 93 | 17 | 0 | 103 | 0 | 0 | 0 | 113 | 17 | 0 | 122 | 1 | 0 | 0 |
| 116 | 17 | 0 | 113 | 0 | 0 | 0 | 117 | 17 | 0 | 113 | 0 | 0 | 0 |
| 147 | 17 | 0 | 119 | 0 | 0 | 0 | 148 | 17 | 0 | 119 | 0 | 0 | 0 |
| 180 | 17 | 0 | 120 | 1 | 0 | 0 | 49 | 18 | 1 | 148 | 0 | 0 | 0 |
| 50 | 18 | 1 | 110 | 1 | 0 | 0 | 89 | 18 | 0 | 107 | 1 | 0 | 1 |
| 100 | 18 | 0 | 100 | 1 | 0 | 0 | 101 | 18 | 0 | 100 | 1 | 0 | 0 |
| 132 | 18 | 0 | 90 | 1 | 0 | 1 | 133 | 18 | 0 | 90 | 1 | 0 | 1 |
| 168 | 18 | 0 | 229 | 0 | 0 | 0 | 205 | 18 | 0 | 120 | 1 | 0 | 0 |
| 208 | 18 | 0 | 120 | 0 | 0 | 0 | 23 | 19 | 1 | 91 | 1 | 0 | 1 |
| 33 | 19 | 1 | 102 | 0 | 0 | 0 | 34 | 19 | 1 | 112 | 1 | 0 | 1 |
| 85 | 19 | 0 | 182 | 0 | 0 | 1 | 96 | 19 | 0 | 95 | 0 | 0 | 0 |
| 97 | 19 | 0 | 150 | 0 | 0 | 0 | 124 | 19 | 0 | 138 | 1 | 0 | 0 |
| 129 | 19 | 0 | 189 | 0 | 0 | 0 | 135 | 19 | 0 | 132 | 0 | 0 | 0 |
| 142 | 19 | 0 | 115 | 0 | 0 | 0 | 181 | 19 | 0 | 105 | 0 | 0 | 0 |
| 187 | 19 | 0 | 235 | 1 | 1 | 0 | 192 | 19 | 0 | 147 | 1 | 0 | 0 |
| 193 | 19 | 0 | 147 | 1 | 0 | 0 | 197 | 19 | 0 | 184 | 1 | 1 | 0 |
| 224 | 19 | 0 | 120 | 1 | 0 | 0 | 27 | 20 | 1 | 150 | 1 | 0 | 0 |
| 31 | 20 | 1 | 125 | 0 | 0 | 1 | 40 | 20 | 1 | 120 | 1 | 0 | 0 |
| 44 | 20 | 1 | 80 | 1 | 0 | 1 | 47 | 20 | 1 | 109 | 0 | 0 | 0 |
| 51 | 20 | 1 | 121 | 1 | 0 | 1 | 60 | 20 | 1 | 122 | 1 | 0 | 0 |
| 76 | 20 | 1 | 105 | 0 | 0 | 0 | 87 | 20 | 0 | 105 | 1 | 0 | 0 |
| 104 | 20 | 0 | 120 | 0 | 0 | 1 | 146 | 20 | 0 | 103 | 0 | 0 | 0 |
| 155 | 20 | 0 | 169 | 0 | 0 | 1 | 160 | 20 | 0 | 141 | 0 | 0 | 1 |
| 172 | 20 | 0 | 121 | 1 | 0 | 0 | 177 | 20 | 0 | 127 | 0 | 0 | 0 |
| 201 | 20 | 0 | 120 | 0 | 0 | 0 | 211 | 20 | 0 | 170 | 1 | 0 | 0 |
| 217 | 20 | 0 | 158 | 0 | 0 | 0 | 20 | 21 | 1 | 165 | 1 | 1 | 0 |
| 28 | 21 | 1 | 200 | 0 | 0 | 1 | 30 | 21 | 1 | 103 | 0 | 0 | 0 |
| 52 | 21 | 1 | 100 | 0 | 0 | 0 | 84 | 21 | 1 | 130 | 1 | 1 | 0 |
| 88 | 21 | 0 | 108 | 1 | 0 | 1 | 91 | 21 | 0 | 124 | 0 | 0 | 0 |
| 128 | 21 | 0 | 185 | 1 | 0 | 0 | 131 | 21 | 0 | 160 | 0 | 0 | 0 |
| 144 | 21 | 0 | 110 | 1 | 0 | 1 | 186 | 21 | 0 | 134 | 0 | 0 | 0 |
| 219 | 21 | 0 | 115 | 0 | 0 | 0 | 42 | 22 | 1 | 130 | 1 | 0 | 1 |
| 67 | 22 | 1 | 130 | 1 | 0 | 0 | 92 | 22 | 0 | 118 | 0 | 0 | 0 |
| 98 | 22 | 0 | 95 | 0 | 1 | 0 | 137 | 22 | 0 | 85 | 1 | 0 | 0 |
| 138 | 22 | 0 | 120 | 0 | 1 | 0 | 140 | 22 | 0 | 130 | 1 | 0 | 0 |
| 161 | 22 | 0 | 158 | 0 | 0 | 0 | 162 | 22 | 0 | 112 | 1 | 0 | 0 |
| 174 | 22 | 0 | 131 | 0 | 0 | 0 | 184 | 22 | 0 | 125 | 0 | 0 | 0 |
| 204 | 22 | 0 | 169 | 0 | 0 | 0 | 220 | 22 | 0 | 129 | 0 | 0 | 0 |
| 17 | 23 | 1 | 97 | 0 | 0 | 1 | 59 | 23 | 1 | 187 | 1 | 0 | 0 |
| 63 | 23 | 1 | 120 | 0 | 0 | 0 | 69 | 23 | 1 | 110 | 1 | 0 | 0 |
| 82 | 23 | 1 | 94 | 1 | 0 | 0 | 130 | 23 | 0 | 130 | 0 | 0 | 0 |
| 139 | 23 | 0 | 128 | 0 | 0 | 0 | 149 | 23 | 0 | 119 | 0 | 0 | 0 |
| 164 | 23 | 0 | 115 | 1 | 0 | 0 | 173 | 23 | 0 | 190 | 0 | 0 | 0 |
| 179 | 23 | 0 | 123 | 0 | 0 | 0 | 182 | 23 | 0 | 130 | 0 | 0 | 0 |
| 200 | 23 | 0 | 110 | 0 | 0 | 0 | 18 | 24 | 1 | 128 | 0 | 0 | 0 |
| 19 | 24 | 1 | 132 | 0 | 1 | 0 | 29 | 24 | 1 | 155 | 1 | 0 | 0 |
| 36 | 24 | 1 | 138 | 0 | 0 | 0 | 61 | 24 | 1 | 105 | 1 | 0 | 0 |
| 118 | 24 | 0 | 90 | 1 | 0 | 0 | 136 | 24 | 0 | 115 | 0 | 0 | 0 |
| 150 | 24 | 0 | 110 | 0 | 0 | 0 | 156 | 24 | 0 | 115 | 0 | 0 | 0 |
| 185 | 24 | 0 | 133 | 0 | 0 | 0 | 196 | 24 | 0 | 110 | 0 | 0 | 0 |
| 199 | 24 | 0 | 110 | 0 | 0 | 0 | 225 | 24 | 0 | 116 | 0 | 0 | 0 |
| 13 | 25 | 1 | 105 | 0 | 1 | 0 | 15 | 25 | 1 | 85 | 0 | 0 | 1 |

```
 24   25   1   115   0  0  0    26   25   1    92   1  0  0
 32   25   1    89   0  0  0    46   25   1   105   0  0  0
103   25   0   118   1  0  0   111   25   0   120   0  0  1
120   25   0   155   0  0  0   121   25   0   125   0  0  0
169   25   0   140   0  0  0   188   25   0    95   1  0  1
202   25   0   241   0  1  0   215   25   0   120   0  0  0
221   25   0   130   0  0  0    35   26   1   117   1  0  0
 54   26   1    96   0  0  0    75   26   1   154   0  1  0
 77   26   1   190   1  0  0    95   26   0   113   1  0  0
115   26   0   168   1  0  0   154   26   0   133   1  0  0
218   26   0   160   0  0  0    16   27   1   150   0  0  0
 43   27   1   130   0  0  1   125   27   0   124   1  0  0
  4   28   1   120   1  0  1    79   28   1    95   1  0  0
105   28   0   120   1  0  0   109   28   0   120   0  0  0
112   28   0   167   0  0  0   151   28   0   140   0  0  0
159   28   0   250   1  0  0   212   28   0   134   0  0  0
214   28   0   130   0  0  0    10   29   1   130   0  0  1
 94   29   0   123   1  0  0   114   29   0   150   0  0  0
123   29   0   140   1  0  0   190   29   0   135   0  0  0
191   29   0   154   0  0  0   209   29   0   130   1  0  0
 65   30   1   142   1  0  0    99   30   0   107   0  0  1
141   30   0    95   1  0  0   145   30   0   153   0  0  0
176   30   0   110   0  0  0   195   30   0   137   0  0  0
203   30   0   112   0  0  0    56   31   1   102   1  0  0
107   31   0   100   0  0  1   126   31   0   215   1  0  0
163   31   0   150   1  0  0   222   31   0   120   0  0  0
 22   32   1   105   1  0  0   106   32   0   121   0  0  0
134   32   0   132   0  0  0   170   32   0   134   1  0  0
175   32   0   170   0  0  0   207   32   0   186   0  0  0
;
```

The variable Low is used to determine whether the subject is a case (Low=1, low-birth-weight baby) or a control (Low=0, normal-weight baby). The dummy time variable Time takes the value 1 for cases and 2 for controls.

The following statements produce a conditional logistic regression analysis of the data. The variable Time is the response, and Low is the censoring variable. Note that the data set is created so that all the cases have the same event time and the controls have later censored times. The matching variable Age is used in the STRATA statement so that each unique age value defines a stratum. The variables LWT, Smoke, HT, and UI are specified as explanatory variables. The TIES=DISCRETE option requests the discrete logistic model.

```
proc phreg data=LBW;
   model Time*Low(0)= LWT Smoke HT UI / ties=discrete;
   strata Age;
run;
```

The procedure displays a summary of the number of event and censored observations for each stratum. These are the number of cases and controls for each matched set shown in Output 64.5.1.

**Output 64.5.1** Summary of Number of Case and Controls

The PHREG Procedure

Model Information

| | |
|---|---|
| Data Set | WORK.LBW |
| Dependent Variable | Time |
| Censoring Variable | Low |
| Censoring Value(s) | 0 |
| Ties Handling | DISCRETE |

Summary of the Number of Event and Censored Values

| Stratum | Age | Total | Event | Censored | Percent Censored |
|---|---|---|---|---|---|
| 1 | 16 | 7 | 1 | 6 | 85.71 |
| 2 | 17 | 12 | 5 | 7 | 58.33 |
| 3 | 18 | 10 | 2 | 8 | 80.00 |
| 4 | 19 | 16 | 3 | 13 | 81.25 |
| 5 | 20 | 18 | 8 | 10 | 55.56 |
| 6 | 21 | 12 | 5 | 7 | 58.33 |
| 7 | 22 | 13 | 2 | 11 | 84.62 |
| 8 | 23 | 13 | 5 | 8 | 61.54 |
| 9 | 24 | 13 | 5 | 8 | 61.54 |
| 10 | 25 | 15 | 6 | 9 | 60.00 |
| 11 | 26 | 8 | 4 | 4 | 50.00 |
| 12 | 27 | 3 | 2 | 1 | 33.33 |
| 13 | 28 | 9 | 2 | 7 | 77.78 |
| 14 | 29 | 7 | 1 | 6 | 85.71 |
| 15 | 30 | 7 | 1 | 6 | 85.71 |
| 16 | 31 | 5 | 1 | 4 | 80.00 |
| 17 | 32 | 6 | 1 | 5 | 83.33 |
| Total | | 174 | 54 | 120 | 68.97 |

Results of the conditional logistic regression analysis are shown in Output 64.5.2. Based on the Wald test for individual variables, the variables LWT, Smoke, and HT are statistically significant while UI is marginal.

The hazard ratios, computed by exponentiating the parameter estimates, are useful in interpreting the results of the analysis. If the hazard ratio of a prognostic factor is larger than 1, an increment in the factor increases the hazard rate. If the hazard ratio is less than 1, an increment in the factor decreases the hazard rate. Results indicate that women were more likely to have low-birth-weight babies if they were underweight in the last menstrual cycle, were hypertensive, smoked during pregnancy, or suffered uterine irritability.

**Output 64.5.2** Conditional Logistic Regression Analysis for the Low-Birth-Weight Study

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

**Output 64.5.2** *continued*

```
                          Model Fit Statistics

                                      Without           With
                   Criterion        Covariates      Covariates

                   -2 LOG L            159.069         141.108
                   AIC                 159.069         149.108
                   SBC                 159.069         157.064

                   Testing Global Null Hypothesis: BETA=0

           Test                 Chi-Square       DF     Pr > ChiSq

           Likelihood Ratio        17.9613        4         0.0013
           Score                   17.3152        4         0.0017
           Wald                    15.5577        4         0.0037

                  Analysis of Maximum Likelihood Estimates

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

LWT             1        -0.01498       0.00706        4.5001        0.0339     0.985
Smoke           1         0.80805       0.36797        4.8221        0.0281     2.244
HT              1         1.75143       0.73932        5.6120        0.0178     5.763
UI              1         0.88341       0.48032        3.3827        0.0659     2.419
```

For matched case-control studies with one case per matched set (1:$n$ matching), the likelihood function for the conditional logistic regression reduces to that of the Cox model for the continuous time scale. For this situation, you can use the default TIES=BRESLOW.

## Example 64.6: Model Using Time-Dependent Explanatory Variables

Time-dependent variables can be used to model the effects of subjects transferring from one treatment group to another. One example of the need for such strategies is the Stanford heart transplant program. Patients are accepted if physicians judge them suitable for heart transplant. Then, when a donor becomes available, physicians choose transplant recipients according to various medical criteria. A patient's status can be changed during the study from waiting for a transplant to being a transplant recipient. Transplant status can be defined by the time-dependent covariate function $z = z(t)$ as

$$z(t) = \begin{cases} 0 & \text{if the patient has not received the transplant at time } t \\ 1 & \text{if the patient has received the transplant at time } t \end{cases}$$

The Stanford heart transplant data that appear in Crowley and Hu (1977) consist of 103 patients, 69 of whom received transplants. The data are saved in a SAS data set called Heart in the following DATA step. For each patient in the program, there is a birth date (Bir_Date), a date of acceptance

(Acc_Date), and a date last seen (Ter_Date). The survival time (Time) in days is defined as Time = Ter_Date − Acc_Date. The survival time is said to be uncensored (Status=1) or censored (Status=0), depending on whether Ter_Date is the date of death or the closing date of the study. The age, in years, at acceptance into the program is Acc_Age = (Acc_Date − Bir_Date) / 365. Previous open-heart surgery for each patient is indicated by the variable PrevSurg. For each transplant recipient, there is a date of transplant (Xpl_Date) and three measures (NMismatch, Antigen, Mismatch) of tissue-type mismatching. The waiting period (WaitTime) in days for a transplant recipient is calculated as WaitTime = Xpl_Date − Acc_Date, and the age (in years) at transplant is Xpl_Age = (Xpl_Date − Bir_Date) / 365. For those who do not receive heart transplants, the WaitTime, Xpl_Age, NMismatch, Antigen, and Mismatch variables contain missing values.

The input data contain dates that have a two-digit year representation. The SAS option YEARCUTOFF=1900 is specified to ensure that a two-digit year xx is year 19xx.

```
options yearcutoff=1900;
data Heart;
   input ID
         @5  Bir_Date mmddyy8.
         @14 Acc_Date mmddyy8.
         @23 Xpl_Date mmddyy8.
         @32 Ter_Date mmddyy8.
         @41 Status 1.
         @43 PrevSurg 1.
         @45 NMismatch 1.
         @47 Antigen 1.
         @49 Mismatch 4.
         @54 Reject 1.
         @56 NotTyped $1.;
   label Bir_Date ='Date of birth'
         Acc_Date ='Date of acceptance'
         Xpl_Date ='Date of transplant'
         Ter_Date ='Date last seen'
         Status   =  'Dead=1 Alive=0'
         PrevSurg ='Previous surgery'
         NMismatch= 'No of mismatches'
         Antigen  = 'HLA-A2 antigen'
         Mismatch ='Mismatch score'
         NotTyped = 'y=not tissue-typed';
   Time= Ter_Date - Acc_Date;
   Acc_Age=int( (Acc_Date - Bir_Date)/365 );
   if ( Xpl_Date ne .) then do;
      WaitTime= Xpl_Date - Acc_Date;
      Xpl_Age= int( (Xpl_Date - Bir_Date)/365 );
   end;
   datalines;
  1 01 10 37 11 15 67          01 03 68 1 0
  2 03 02 16 01 02 68          01 07 68 1 0
  3 09 19 13 01 06 68 01 06 68 01 21 68 1 0 2 0 1.11 0
  4 12 23 27 03 28 68 05 02 68 05 05 68 1 0 3 0 1.66 0
  5 07 28 47 05 10 68          05 27 68 1 0
  6 11 08 13 06 13 68          06 15 68 1 0
  7 08 29 17 07 12 68 08 31 68 05 17 70 1 0 4 0 1.32 1
```

```
 8 03 27 23 08 01 68          09 09 68 1 0
 9 06 11 21 08 09 68          11 01 68 1 0
10 02 09 26 08 11 68 08 22 68 10 07 68 1 0 2 0 0.61 1
11 08 22 20 08 15 68 09 09 68 01 14 69 1 0 1 0 0.36 0
12 07 09 15 09 17 68          09 24 68 1 0
13 02 22 14 09 19 68 10 05 68 12 08 68 1 0 3 0 1.89 1
14 09 16 14 09 20 68 10 26 68 07 07 72 1 0 1 0 0.87 1
15 12 04 14 09 27 68          09 27 68 1 1
16 05 16 19 10 26 68 11 22 68 08 29 69 1 0 2 0 1.12 1
17 06 29 48 10 28 68          12 02 68 1 0
18 12 27 11 11 01 68 11 20 68 12 13 68 1 0 3 0 2.05 0
19 10 04 09 11 18 68          12 24 68 1 0
20 10 19 13 01 29 69 02 15 69 02 25 69 1 0 3 1 2.76 1
21 09 29 25 02 01 69 02 08 69 11 29 71 1 0 2 0 1.13 1
22 06 05 26 03 18 69 03 29 69 05 07 69 1 0 3 0 1.38 1
23 12 02 10 04 11 69 04 13 69 04 13 71 1 0 3 0 0.96 1
24 07 07 17 04 25 69 07 16 69 11 29 69 1 0 3 1 1.62 1
25 02 06 36 04 28 69 05 22 69 04 01 74 0 0 2 0 1.06 0
26 10 18 38 05 01 69          03 01 73 0 0
27 07 21 60 05 04 69          01 21 70 1 0
28 05 30 15 06 07 69 08 16 69 08 17 69 1 0 2 0 0.47 0
29 02 06 19 07 14 69          08 17 69 1 0
30 09 20 24 08 19 69 09 03 69 12 18 71 1 0 4 0 1.58 1
31 10 04 14 08 23 69          09 07 69 1 0
32 04 02 05 08 29 69 09 14 69 11 13 69 1 0 4 0 0.69 1
33 01 01 21 11 27 69 01 16 70 04 01 74 0 0 3 0 0.91 0
34 05 24 29 12 12 69 01 03 70 04 01 74 0 0 2 0 0.38 0
35 08 04 26 01 21 70          02 01 70 1 0
36 05 01 21 04 04 70 05 19 70 07 12 70 1 0 2 0 2.09 1
37 10 24 08 04 25 70 05 13 70 06 29 70 1 0 3 1 0.87 1
38 11 14 28 05 05 70 05 09 70 05 09 70 1 0 3 0 0.87 0
39 11 12 19 05 20 70 05 21 70 07 11 70 1 0                 y
40 11 30 21 05 25 70 07 04 70 04 01 74 0 1 4 0 0.75 0
41 04 30 25 08 19 70 10 15 70 04 01 74 0 1 2 0 0.98 0
42 03 13 34 08 21 70          08 23 70 1 0
43 06 01 27 10 22 70          10 23 70 1 1
44 05 02 28 11 30 70          01 08 71 1 1
45 10 30 34 01 05 71 01 05 71 02 18 71 1 0 1 0 0.0  0
46 06 01 22 01 10 71 01 11 71 10 01 73 1 1 2 0 0.81 1
47 12 28 23 02 02 71 02 22 71 04 14 71 1 0 3 0 1.38 1
48 01 23 15 02 05 71          02 13 71 1 0
49 06 21 34 02 15 71 03 22 71 04 01 74 0 1 4 0 1.35 0
50 03 28 25 02 15 71 05 08 71 10 21 73 1 1                 y
51 06 29 22 03 24 71 04 24 71 01 02 72 1 0 4 1 1.08 1
52 01 24 30 04 25 71          08 04 71 1 0
53 02 27 24 07 02 71 08 11 71 01 05 72 1 0                 y
54 09 16 23 07 02 71          07 04 71 1 0
55 02 24 19 08 09 71 08 18 71 10 08 71 1 0 2 0 1.51 1
56 12 05 32 09 03 71 11 08 71 04 01 74 0 0 4 0 0.98 0
57 06 08 30 09 13 71          02 08 72 1 0
58 09 17 23 09 23 71 10 13 71 08 30 72 1 1 2 1 1.82 1
59 05 12 30 09 29 71 12 15 71 04 01 74 0 1 2 0 0.19 0
60 10 29 22 11 18 71 11 20 71 01 24 72 1 0 3 0 0.66 1
61 05 12 19 12 04 71          12 05 71 1 0
```

```
 62 08 01 32 12 09 71          02 15 72 1 0
 63 04 15 39 12 12 71 01 07 72 04 01 74 0 0 3 1 1.93 0
 64 04 09 23 02 01 72 03 04 72 09 06 73 1 1 1 0 0.12 0
 65 11 19 20 03 06 72 03 17 72 05 22 72 1 0 2 0 1.12 1
 66 01 02 19 03 20 72          04 20 72 1 0
 67 09 03 52 03 23 72 05 18 72 01 01 73 1 0 3 0 1.02 0
 68 01 10 27 04 07 72 04 09 72 06 13 72 1 0 3 1 1.68 1
 69 06 05 24 06 01 72 06 10 72 04 01 74 0 0 2 0 1.20 0
 70 06 17 19 06 17 72 06 21 72 07 16 72 1 0 3 1 1.68 1
 71 02 22 25 07 21 72 08 20 72 04 01 74 0 0 3 0 0.97 0
 72 11 22 45 08 14 72 08 17 72 04 01 74 0 0 3 1 1.46 0
 73 05 13 16 09 11 72 10 07 72 12 09 72 1 0 3 1 2.16 1
 74 07 20 43 09 18 72 09 22 72 10 04 72 1 0 1 0 0.61 0
 75 07 25 20 09 29 72          09 30 72 1 0
 76 09 03 20 10 04 72 11 18 72 04 01 74 0 1 3 1 1.70 0
 77 08 27 31 10 06 72          10 26 72 1 0
 78 02 20 24 11 03 72 05 31 73 04 01 74 0 0 3 0 0.81 0
 79 02 18 19 11 30 72 02 04 73 03 05 73 1 0 2 0 1.08 1
 80 06 27 26 12 06 72 12 31 72 04 01 74 0 1 3 0 1.41 0
 81 02 21 20 01 12 73 01 17 73 04 01 74 0 0 4 1 1.94 0
 82 08 19 42 11 01 71          01 01 73 0 0
 83 10 04 19 01 24 73 02 24 73 04 13 73 1 0 4 0 3.05 0
 84 05 13 30 01 30 73 03 07 73 12 29 73 1 0 4 0 0.60 1
 85 02 13 25 02 06 73          02 10 73 1 0
 86 03 30 24 03 01 73 03 08 73 04 01 74 0 0 3 1 1.44 0
 87 12 19 26 03 21 73 05 19 73 07 08 73 1 0 2 0 2.25 1
 88 11 16 18 03 28 73 04 27 73 04 01 74 0 0 3 0 0.68 0
 89 03 19 22 04 05 73 08 21 73 10 28 73 1 0 4 1 1.33 1
 90 03 25 21 04 06 73 09 12 73 10 08 73 1 1 3 1 0.82 0
 91 09 08 25 04 13 73          03 18 74 1 0
 92 05 03 28 04 27 73 03 02 74 04 01 74 0 0 1 0 0.16 0
 93 10 10 25 07 11 73 08 07 73 04 01 74 0 0 2 0 0.33 0
 94 11 11 29 09 14 73 09 17 73 02 25 74 1 1 3 0 1.20 1
 95 06 11 33 09 22 73 09 23 73 10 07 73 1 0                y
 96 02 09 47 10 04 73 10 16 73 04 01 74 0 0 2 0 0.46 0
 97 04 11 50 11 22 73 12 12 73 04 01 74 0 0 3 1 1.78 0
 98 04 28 45 12 14 73 03 19 74 04 01 74 0 0 4 1 0.77 0
 99 02 24 24 12 25 73          01 14 74 1 0
100 01 31 39 02 22 74 03 31 74 04 01 74 0 1 3 0 0.67 0
101 08 25 24 03 02 74          04 01 74 0 0
102 10 30 33 03 22 74          04 01 74 0 0
103 05 20 28 09 13 67          09 18 67 1 0
   ;
```

Crowley and Hu (1977) have presented a number of analyses to assess the effects of various explanatory variables on the survival of patients. This example fits two of the models that they have considered.

The first model consists of two explanatory variables—the transplant status and the age at acceptance. The transplant status (XStatus) is a time-dependent variable defined by the programming statements between the MODEL statement and the RUN statement. The XStatus variable takes the value 1 or 0 at time $t$ (measured from the date of acceptance), depending on whether or not the patient has received a transplant at that time. Note that the value of XStatus changes for subjects in each risk set (subjects still alive just before each distinct event time); therefore, the variable cannot be created in the DATA step. The variable Acc_Age, which is not time dependent, accounts for the possibility that pretransplant risks vary with age. The following statements fit this model:

```
proc phreg data= Heart;
   model Time*Status(0)= XStatus Acc_Age;
   if (WaitTime = . or Time < WaitTime) then XStatus=0.;
   else  XStatus= 1.0;
run;
```

Results of this analysis are shown in Output 64.6.1. Transplantation appears to be associated with a slight decrease in risk, although the effect is not significant ($p = 0.8261$). The age at acceptance as a pretransplant risk factor adds significantly to the model ($p = 0.0289$). The risk increases significantly with age at acceptance.

**Output 64.6.1** Heart Transplant Study Analysis I

```
                         The PHREG Procedure

                          Model Information

            Data Set                 WORK.HEART
            Dependent Variable       Time
            Censoring Variable       Status         Dead=1 Alive=0
            Censoring Value(s)       0
            Ties Handling            BRESLOW

                   Number of Observations Read         103
                   Number of Observations Used         103

          Summary of the Number of Event and Censored Values

                                                   Percent
                    Total        Event     Censored     Censored

                      103           75           28        27.18

                          Convergence Status

            Convergence criterion (GCONV=1E-8) satisfied.

                        Model Fit Statistics

                                   Without           With
                   Criterion     Covariates     Covariates

                   -2 LOG L         596.651        591.292
                   AIC              596.651        595.292
                   SBC              596.651        599.927
```

**Output 64.6.1** *continued*

**Testing Global Null Hypothesis: BETA=0**

| Test | Chi-Square | DF | Pr > ChiSq |
|---|---|---|---|
| Likelihood Ratio | 5.3593 | 2 | 0.0686 |
| Score | 4.8093 | 2 | 0.0903 |
| Wald | 4.7999 | 2 | 0.0907 |

**Analysis of Maximum Likelihood Estimates**

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| XStatus | 1 | -0.06720 | 0.30594 | 0.0482 | 0.8261 | 0.935 |
| Acc_Age | 1 | 0.03158 | 0.01446 | 4.7711 | 0.0289 | 1.032 |

The second model consists of three explanatory variables—the transplant status, the transplant age, and the mismatch score. Four transplant recipients who were not typed have no Mismatch values; they are excluded from the analysis by the use of a WHERE clause. The transplant age (XAge) and the mismatch score (XScore) are also time dependent and are defined in a fashion similar to that of XStatus. While the patient is waiting for a transplant, XAge and XScore have a value of 0. After the patient has migrated to the recipient population, XAge takes on the value of Xpl_Age (transplant age for the recipient), and XScore takes on the value of Mismatch (a measure of the degree of dissimilarity between donor and recipient). The following statements fit this model:

```
proc phreg data= Heart;
   model Time*Status(0)= XStatus XAge XScore;
   where NotTyped ^= 'y';
   if (WaitTime = . or Time < WaitTime) then do;
      XStatus=0.;
      XAge=0.;
      XScore= 0.;
   end;
   else do;
      XStatus= 1.0;
      XAge= Xpl_Age;
      XScore= Mismatch;
   end;
run;
```

Results of the analysis are shown in Output 64.6.2. Note that only 99 patients are included in this analysis, instead of 103 patients as in the previous analysis, since four transplant recipients who were not typed are excluded. The variable XAge is statistically significant ($p = 0.0143$), with a hazard ratio exceeding 1. Therefore, patients who had a transplant at younger ages lived longer than those who received a transplant later in their lives. The variable XScore has only minimal effect on the survival ($p = 0.1121$).

**Output 64.6.2** Heart Transplant Study Analysis II

```
                          The PHREG Procedure

                           Model Information

             Data Set                    WORK.HEART
             Dependent Variable          Time
             Censoring Variable          Status          Dead=1 Alive=0
             Censoring Value(s)          0
             Ties Handling               BRESLOW

                    Number of Observations Read           99
                    Number of Observations Used           99

               Summary of the Number of Event and Censored Values

                                                          Percent
                     Total        Event     Censored     Censored

                        99           71           28        28.28

                           Convergence Status

                 Convergence criterion (GCONV=1E-8) satisfied.

                         Model Fit Statistics

                                      Without            With
                    Criterion      Covariates      Covariates

                    -2 LOG L          561.680         551.874
                    AIC               561.680         557.874
                    SBC               561.680         564.662

                  Testing Global Null Hypothesis: BETA=0

            Test                 Chi-Square       DF     Pr > ChiSq

            Likelihood Ratio         9.8059        3         0.0203
            Score                    9.0521        3         0.0286
            Wald                     9.0554        3         0.0286

                 Analysis of Maximum Likelihood Estimates

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

  XStatus       1      -3.19837      1.18746       7.2547        0.0071      0.041
  XAge          1       0.05544      0.02263       6.0019        0.0143      1.057
  XScore        1       0.44490      0.28001       2.5245        0.1121      1.560
```

## Example 64.7: Time-Dependent Repeated Measurements of a Covariate

Repeated determinations can be made during the course of a study of variables thought to be related to survival. Consider an experiment to study the dosing effect of a tumor-promoting agent. Forty-five rodents initially exposed to a carcinogen were randomly assigned to three dose groups. After the first death of an animal, the rodents were examined every week for the number of papillomas. Investigators were interested in determining the effects of dose on the carcinoma incidence after adjusting for the number of papillomas.

The input data set TUMOR consists of the following 19 variables:

- ID (subject identification)
- Time (survival time of the subject)
- Dead (censoring status where 1=dead and 0=censored)
- Dose (dose of the tumor-promoting agent)
- P1–P15 (number of papillomas at the 15 times that animals died. These 15 death times are weeks 27, 34, 37, 41, 43, 45, 46, 47, 49, 50, 51, 53, 65, 67, and 71. For instance, subject 1 died at week 47; it had no papilloma at week 27, five papillomas at week 34, six at week 37, eight at week 41, and 10 at weeks 43, 45, 46, and 47. For an animal that died before week 71, the number of papillomas is missing for those times beyond its death.)

The following SAS statements create the data set TUMOR:

```
data Tumor;
   infile datalines missover;
   input ID Time Dead Dose P1-P15;
   label ID='Subject ID';
   datalines;
 1 47 1  1.0  0  5  6  8 10 10 10 10
 2 71 1  1.0  0  0  0  0  0  0  0  0  1  1  1  1 1 1 1
 3 81 0  1.0  0  1  1  1  1  1  1  1  1  1  1  1 1 1 1
 4 81 0  1.0  0  0  0  0  0  0  0  0  0  0  0  0 0 0 0
 5 81 0  1.0  0  0  0  0  0  0  0  0  0  0  0  0 0 0 0
 6 65 1  1.0  0  0  0  1  1  1  1  1  1  1  1  1 1
 7 71 0  1.0  0  0  0  0  0  0  0  0  0  0  0  0 0 0 0
 8 69 0  1.0  0  0  0  0  0  0  0  0  0  0  0  0 0 0
 9 67 1  1.0  0  0  1  1  2  2  2  2  3  3  3  3 3 3
10 81 0  1.0  0  0  0  0  0  0  0  0  0  0  0  0 0 0 0
11 37 1  1.0  9  9  9
12 81 0  1.0  0  0  0  0  0  0  0  0  0  0  0  0 0 0 0
13 77 0  1.0  0  0  0  0  1  1  1  1  1  1  1  1 1 1 1
14 81 0  1.0  0  0  0  0  0  0  0  0  0  0  0  0 0 0 0
15 81 0  1.0  0  0  0  0  0  0  0  0  0  0  0  0 0 0 0
16 54 0  2.5  0  1  1  1  2  2  2  2  2  2  2  2
17 53 0  2.5  0  0  0  0  0  0  0  0  0  0  0  0
18 38 0  2.5  5 13 14
```

```
19 54 0  2.5  2  6  6  6  6  6  6  6  6  6  6  6
20 51 1  2.5 15 15 15 16 16 17 17 17 17 17 17
21 47 1  2.5 13 20 20 20 20 20 20 20
22 27 1  2.5 22
23 41 1  2.5  6 13 13 13
24 49 1  2.5  0  3  3  3  3  3  3  3  3
25 53 0  2.5  0  0  1  1  1  1  1  1  1  1  1  1
26 50 1  2.5  0  0  2  3  4  6  6  6  6  6
27 37 1  2.5  3 15 15
28 49 1  2.5  2  3  3  3  3  4  4  4  4
29 46 1  2.5  4  6  7  9  9  9  9
30 48 0  2.5 15 26 26 26 26 26 26 26
31 54 0 10.0 12 14 15 15 15 15 15 15 15 15 15 15
32 37 1 10.0 12 16 17
33 53 1 10.0  3  6  6  6  6  6  6  6  6  6  6  6
34 45 1 10.0  4 12 15 20 20 20
35 53 0 10.0  6 10 13 13 13 15 15 15 15 15 15 20
36 49 1 10.0  0  2  2  2  2  2  2  2  2
37 39 0 10.0  7  8  8
38 27 1 10.0 17
39 49 1 10.0  0  6  9 14 14 14 14 14 14
40 43 1 10.0 14 18 20 20 20
41 28 0 10.0  8
42 34 1 10.0 11 18
43 45 1 10.0 10 12 16 16 16 16
44 37 1 10.0  0  1  1
45 43 1 10.0  9 19 19 19 19
;
```

The number of papillomas (NPap) for each animal in the study was measured repeatedly over time. One way of handling time-dependent repeated measurements in the PHREG procedure is to use programming statements to capture the appropriate covariate values of the subjects in each risk set. In this example, NPap is a time-dependent explanatory variable with values that are calculated by means of the programming statements shown in the following SAS statements:

```
proc phreg data=Tumor;
   model Time*Dead(0)=Dose NPap;
   array pp{*} P1-P14;
   array tt{*} t1-t15;
    t1=27;  t2=34;  t3=37;  t4=41;  t5=43;
    t6=45;  t7=46;  t8=47;  t9=49; t10=50;
   t11=51; t12=53; t13=65; t14=67; t15=71;
   if Time <  tt[1]  then NPap=0;
   else if time >= tt[15] then NPap=P15;
   else do i=1 to dim(pp);
      if tt[i] <= Time < tt[i+1] then NPap= pp[i];
   end;
run;
```

At each death time, the NPap value of each subject in the risk set is recalculated to reflect the actual number of papillomas at the given death time. For instance, subject one in the data set Tumor was in the risk sets at weeks 27 and 34; at week 27, the animal had no papilloma, while at week 34, it had five papillomas. Results of the analysis are shown in Output 64.7.1. After the number of papillomas is adjusted for, the dose effect of the tumor-promoting agent is not statistically significant.

**Output 64.7.1** Cox Regression Analysis on the Survival of Rodents

The PHREG Procedure

Model Information

| | |
|---|---|
| Data Set | WORK.TUMOR |
| Dependent Variable | Time |
| Censoring Variable | Dead |
| Censoring Value(s) | 0 |
| Ties Handling | BRESLOW |

| | |
|---|---|
| Number of Observations Read | 45 |
| Number of Observations Used | 45 |

Summary of the Number of Event and Censored Values

| Total | Event | Censored | Percent Censored |
|---|---|---|---|
| 45 | 25 | 20 | 44.44 |

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

| Criterion | Without Covariates | With Covariates |
|---|---|---|
| -2 LOG L | 166.793 | 143.269 |
| AIC | 166.793 | 147.269 |
| SBC | 166.793 | 149.707 |

Testing Global Null Hypothesis: BETA=0

| Test | Chi-Square | DF | Pr > ChiSq |
|---|---|---|---|
| Likelihood Ratio | 23.5243 | 2 | <.0001 |
| Score | 28.0498 | 2 | <.0001 |
| Wald | 21.1646 | 2 | <.0001 |

**Output 64.7.1** *continued*

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| Dose | 1 | 0.06885 | 0.05620 | 1.5010 | 0.2205 | 1.071 |
| NPap | 1 | 0.11714 | 0.02998 | 15.2705 | <.0001 | 1.124 |

Another way to handle time-dependent repeated measurements in the PHREG procedure is to use the counting process style of input. Multiple records are created for each subject, one record for each distinct pattern of the time-dependent measurements. Each record contains a T1 value and a T2 value representing the time interval (T1,T2] during which the values of the explanatory variables remain unchanged. Each record also contains the censoring status at T2.

One advantage of using the counting process formulation is that you can easily obtain various residuals and influence statistics that are not available when programming statements are used to compute the values of the time-dependent variables. On the other hand, creating multiple records for the counting process formulation requires extra effort in data manipulation.

Consider a counting process style of input data set named Tumor1. It contains multiple observations for each subject in the data set Tumor. In addition to variables ID, Time, Dead, and Dose, four new variables are generated:

- T1 (left endpoint of the risk interval)
- T2 (right endpoint of the risk interval)
- NPap (number of papillomas in the time interval (T1,T2])
- Status (censoring status at T2)

For example, five observations are generated for the rodent that died at week 47 and that had no papilloma at week 27, five papillomas at week 34, six at week 37, eight at week 41, and 10 at weeks 43, 45, 46, and 47. The values of T1, T2, NPap, and Status for these five observations are (0,27,0,0), (27,34,5,0), (34,37,6,0), (37,41,8,0), and (41,47,10,1). Note that the variables ID, Time, and Dead are not needed for the estimation of the regression parameters, but they are useful for plotting the residuals.

The following SAS statements create the data set Tumor1:

```
data Tumor1(keep=ID Time Dead Dose T1 T2 NPap Status);
   array pp{*} P1-P14;
   array qq{*} P2-P15;
   array tt{1:15} _temporary_
      (27 34 37 41 43 45 46 47 49 50 51 53 65 67 71);
   set Tumor;
   T1 = 0;
   T2 = 0;
   Status = 0;
   if ( Time = tt[1] ) then do;
      T2 = tt[1];
      NPap = p1;
      Status = Dead;
      output;
   end;
   else do _i_=1 to dim(pp);
      if ( tt[_i_] = Time ) then do;
         T2= Time;
         NPap = pp[_i_] ;
         Status = Dead;
         output;
      end;
      else if (tt[_i_]  < Time ) then do;
         if (pp[_i_]  ^= qq[_i_] ) then do;
            if qq[_i_]  = . then T2= Time;
            else                 T2= tt[_i_] ;
            NPap= pp[_i_] ;
            Status= 0;
            output;
            T1 = T2;
         end;
      end;
   end;
   if ( Time >= tt[15] ) then do;
      T2 = Time;
      NPap = P15;
      Status = Dead;
      output;
   end;
   run;
```

In the following SAS statements, the counting process MODEL specification is used. The DFBETA statistics are output to a SAS data set named Out1. Note that Out1 contains multiple observations for each subject—that is, one observation for each risk interval (T1,T2].

```
proc phreg data=Tumor1;
   model (T1,T2)*Status(0)=Dose NPap;
   output out=Out1 resmart=Mart dfbeta=db1-db2;
   id ID Time Dead;
run;
```

The output from PROC PHREG (not shown) is identical to Output 64.7.1 except for the “Summary of the Number of Event and Censored Values” table. The number of event observations remains unchanged between the two specifications of PROC PHREG, but the number of censored observations differs due to the splitting of each subject’s data into multiple observations for the counting process style of input.

Next, the MEANS procedure sums up the component statistics for each subject and outputs the results to a SAS data set named Out2:

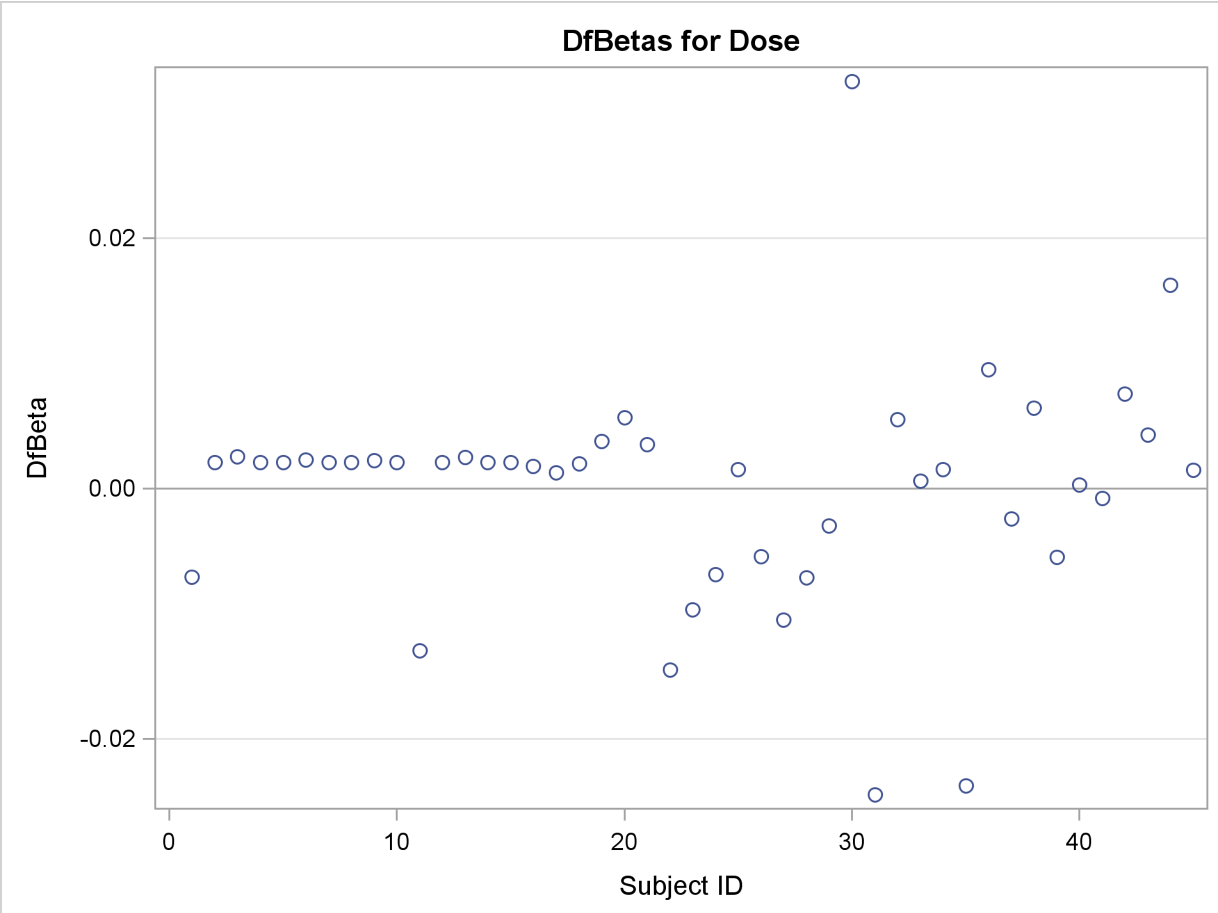
```
proc means data=Out1 noprint;
   by ID Time Dead;
   var Mart db1-db2;
   output out=Out2 sum=Mart db_Dose db_NPap;
run;
```

Finally, DFBETA statistics are plotted against subject ID for easy identification of influential points:

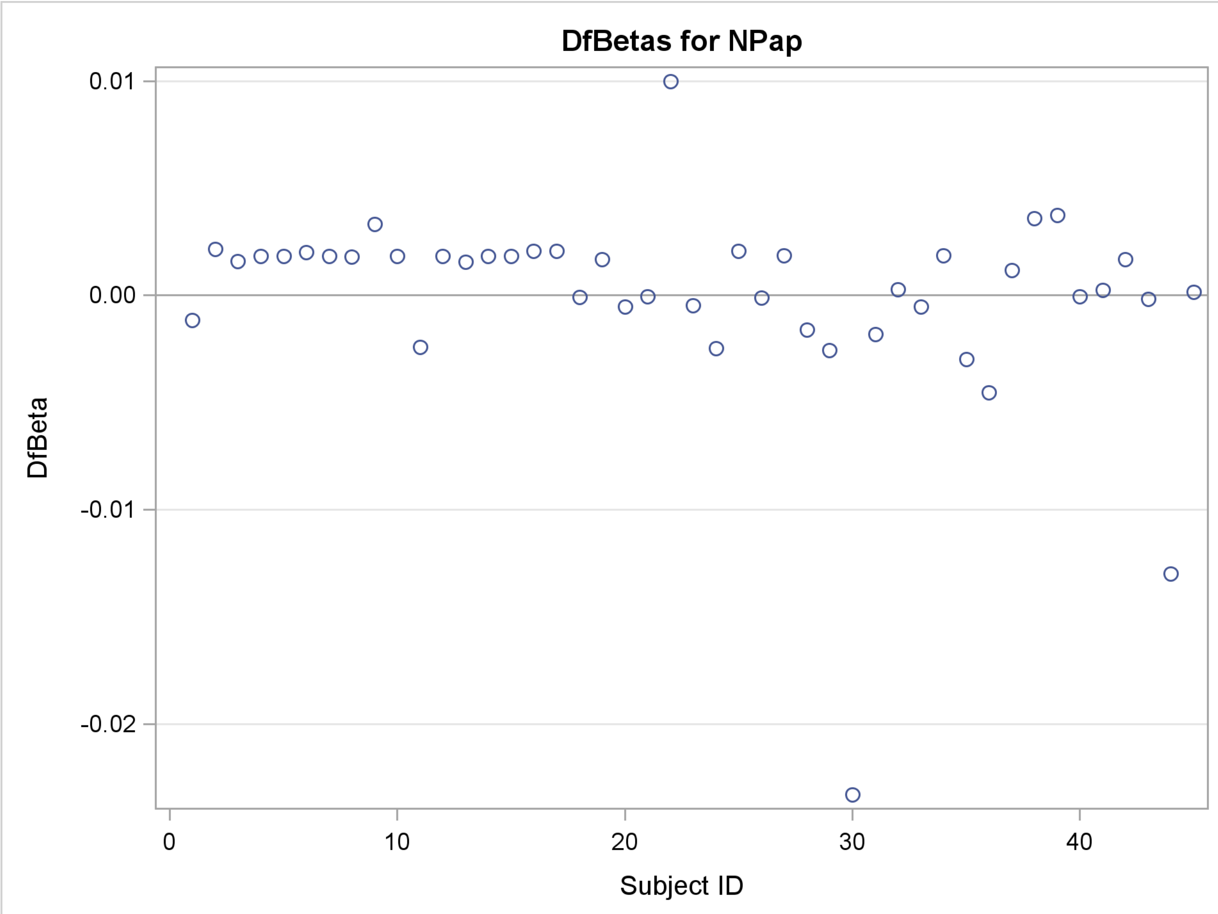
```
title 'DfBetas for Dose';
proc sgplot data=Out2;
   yaxis label="DfBeta" grid;
   refline 0 / axis=y;
   scatter y=db_Dose x=ID;
   run;
title 'DfBetas for NPap';
proc sgplot data=Out2;
   yaxis label="DfBeta" grid;
   refline 0 / axis=y;
   scatter y=db_NPap x=ID;
   run;
```

The plots of the DFBETA statistics are shown in Output 64.7.2 and Output 64.7.3. Subject 30 appears to have a large influence on both the Dose and NPap coefficients. Subjects 31 and 35 have considerable influences on the DOSE coefficient, while subjects 22 and 44 have rather large influences on the NPap coefficient.

**Output 64.7.2** Plot of DFBETA Statistic for DOSE versus Subject Number



**Output 64.7.3** Plot of DFBETA Statistic for NPAP versus Subject Number



## Example 64.8: Survivor Function Estimates for Specific Covariate Values

You might want to use your regression analysis results to generate predicted survival curves for subjects not in the study. The COVARIATES= data set in the BASELINE statement enables you to specify the sets of covariate values for the prediction. By using the PLOTS= option in the PROC PHREG statement, you can display a survival curve for each row of covariates in the COVARIATES= data set. You can elect to output the predicted survival curves in a SAS data set by using just the BASELINE statement. This example illustrates these two tasks by using the Myeloma data in Example 64.1.

In Example 64.1, variables LogBUN and HGB were identified as the most important prognostic factors for the myeloma data. Two sets of covariates for predicting the survivor function are saved in the data set Inrisks in the following DATA step. Also created in this data set is the variable Id, whose values will be used in identifying the covariate sets in the survival plot.

```
data Inrisks;
   length Id $20;
   input LogBUN HGB Id $12-31;
   datalines;
1.00 10.0  logBUN=1.0 HGB=10
1.80 12.0  logBUN=1.8 HGB=12
;
```

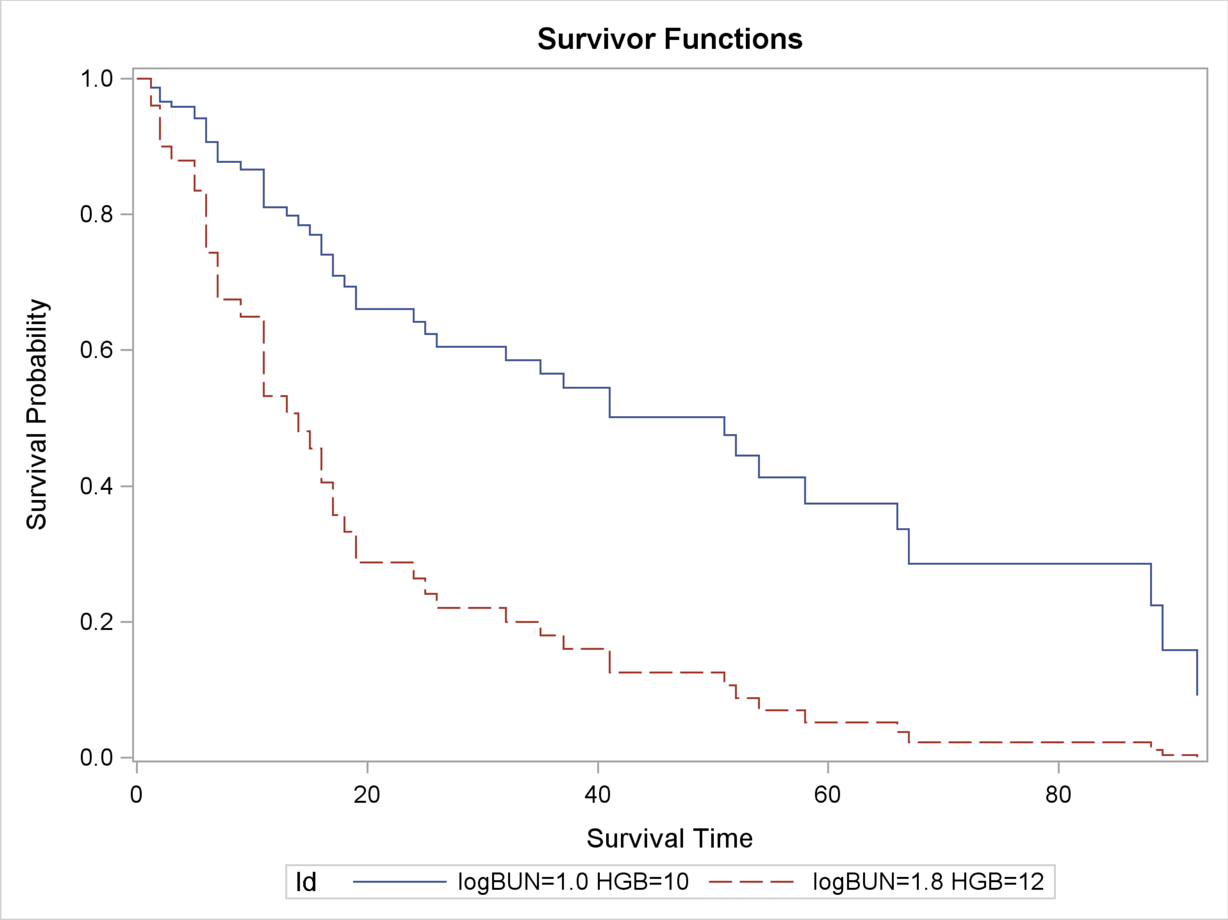
The following statements produce the plot in Output 64.8.1 and create the BASELINE data set Pred1:

```
ods graphics on;
proc phreg data=Myeloma plots(overlay)=survival;
   model Time*VStatus(0)=LogBUN HGB;
   baseline covariates=Inrisks out=Pred1 survival=_all_ / rowid=Id;
run;
ods graphics off;
```

The COVARIATES= option in the BASELINE statement specifies the data set that contains the set of covariates of interest. The PLOTS= option in the PROC PHREG statement creates the survivor plot. The OVERLAY suboption overlays the two curves in the same plot. If the OVERLAY suboption is not specified, each curve is displayed in a separate plot. The ROWID= option in the BASELINE statement specifies that the values of the variable Id in the COVARIATES= data set be used to identify the curves in the plot. The SURVIVAL=_ALL_ option in the BASELINE statement requests that the estimated survivor function, standard error, and lower and upper confidence limits for the survivor function be output into the SAS data set specified in the OUT= option.

The survival Plot (Output 64.8.1) contains two curves, one for each of row of covariates in the data set Inrisks.

**Output 64.8.1** Estimated Survivor Function Plot




Finally, PROC PRINT is used to print out the observations in the data set Pred1 for the realization LogBUN=1.00 and HGB=10.0:

```
proc print data=Pred1(where=(logBUN=1 and HGB=10));
run;
```

As shown in Output 64.8.2, there are 32 observations representing the survivor function for the realization LogBUN=1.00 and HGB=10.0. The first observation has survival time 0 and survivor function estimate 1.0. Each of the remaining 31 observations represents a distinct event time in the input data set Myeloma. These observations are presented in ascending order of the event times. Note that all the variables in the COVARIATE=InRisks data set are included in the OUT=Pred1 data set. Likewise, you can print out the observations that represent the survivor function for the realization LogBUN=1.80 and HGB=12.0.

**Output 64.8.2** Survivor Function Estimates for LogBUN=1.0 and HGB=10.0

| Obs | Id | Log BUN | HGB | Time | Survival | StdErr Survival | Lower Survival | Upper Survival |
|---|---|---|---|---|---|---|---|---|
| 1 | logBUN=1.0 HGB=10 | 1 | 10 | 0.00 | 1.00000 | . | . | . |
| 2 | logBUN=1.0 HGB=10 | 1 | 10 | 1.25 | 0.98678 | 0.01043 | 0.96655 | 1.00000 |
| 3 | logBUN=1.0 HGB=10 | 1 | 10 | 2.00 | 0.96559 | 0.01907 | 0.92892 | 1.00000 |
| 4 | logBUN=1.0 HGB=10 | 1 | 10 | 3.00 | 0.95818 | 0.02180 | 0.91638 | 1.00000 |
| 5 | logBUN=1.0 HGB=10 | 1 | 10 | 5.00 | 0.94188 | 0.02747 | 0.88955 | 0.99729 |
| 6 | logBUN=1.0 HGB=10 | 1 | 10 | 6.00 | 0.90635 | 0.03796 | 0.83492 | 0.98389 |
| 7 | logBUN=1.0 HGB=10 | 1 | 10 | 7.00 | 0.87742 | 0.04535 | 0.79290 | 0.97096 |
| 8 | logBUN=1.0 HGB=10 | 1 | 10 | 9.00 | 0.86646 | 0.04801 | 0.77729 | 0.96585 |
| 9 | logBUN=1.0 HGB=10 | 1 | 10 | 11.00 | 0.81084 | 0.05976 | 0.70178 | 0.93686 |
| 10 | logBUN=1.0 HGB=10 | 1 | 10 | 13.00 | 0.79800 | 0.06238 | 0.68464 | 0.93012 |
| 11 | logBUN=1.0 HGB=10 | 1 | 10 | 14.00 | 0.78384 | 0.06515 | 0.66601 | 0.92251 |
| 12 | logBUN=1.0 HGB=10 | 1 | 10 | 15.00 | 0.76965 | 0.06779 | 0.64762 | 0.91467 |
| 13 | logBUN=1.0 HGB=10 | 1 | 10 | 16.00 | 0.74071 | 0.07269 | 0.61110 | 0.89781 |
| 14 | logBUN=1.0 HGB=10 | 1 | 10 | 17.00 | 0.71005 | 0.07760 | 0.57315 | 0.87966 |
| 15 | logBUN=1.0 HGB=10 | 1 | 10 | 18.00 | 0.69392 | 0.07998 | 0.55360 | 0.86980 |
| 16 | logBUN=1.0 HGB=10 | 1 | 10 | 19.00 | 0.66062 | 0.08442 | 0.51425 | 0.84865 |
| 17 | logBUN=1.0 HGB=10 | 1 | 10 | 24.00 | 0.64210 | 0.08691 | 0.49248 | 0.83717 |
| 18 | logBUN=1.0 HGB=10 | 1 | 10 | 25.00 | 0.62360 | 0.08921 | 0.47112 | 0.82542 |
| 19 | logBUN=1.0 HGB=10 | 1 | 10 | 26.00 | 0.60523 | 0.09136 | 0.45023 | 0.81359 |
| 20 | logBUN=1.0 HGB=10 | 1 | 10 | 32.00 | 0.58549 | 0.09371 | 0.42784 | 0.80122 |
| 21 | logBUN=1.0 HGB=10 | 1 | 10 | 35.00 | 0.56534 | 0.09593 | 0.40539 | 0.78840 |
| 22 | logBUN=1.0 HGB=10 | 1 | 10 | 37.00 | 0.54465 | 0.09816 | 0.38257 | 0.77542 |
| 23 | logBUN=1.0 HGB=10 | 1 | 10 | 41.00 | 0.50178 | 0.10166 | 0.33733 | 0.74639 |
| 24 | logBUN=1.0 HGB=10 | 1 | 10 | 51.00 | 0.47546 | 0.10368 | 0.31009 | 0.72901 |
| 25 | logBUN=1.0 HGB=10 | 1 | 10 | 52.00 | 0.44510 | 0.10522 | 0.28006 | 0.70741 |
| 26 | logBUN=1.0 HGB=10 | 1 | 10 | 54.00 | 0.41266 | 0.10689 | 0.24837 | 0.68560 |
| 27 | logBUN=1.0 HGB=10 | 1 | 10 | 58.00 | 0.37465 | 0.10891 | 0.21192 | 0.66232 |
| 28 | logBUN=1.0 HGB=10 | 1 | 10 | 66.00 | 0.33626 | 0.10980 | 0.17731 | 0.63772 |
| 29 | logBUN=1.0 HGB=10 | 1 | 10 | 67.00 | 0.28529 | 0.11029 | 0.13372 | 0.60864 |
| 30 | logBUN=1.0 HGB=10 | 1 | 10 | 88.00 | 0.22412 | 0.10928 | 0.08619 | 0.58282 |
| 31 | logBUN=1.0 HGB=10 | 1 | 10 | 89.00 | 0.15864 | 0.10317 | 0.04435 | 0.56750 |
| 32 | logBUN=1.0 HGB=10 | 1 | 10 | 92.00 | 0.09180 | 0.08545 | 0.01481 | 0.56907 |

## Example 64.9: Analysis of Residuals

Residuals are used to investigate the lack of fit of a model to a given subject. You can obtain martingale and deviance residuals for the Cox proportional hazards regression analysis by requesting that they be included in the OUTPUT data set. You can plot these statistics and look for outliers.

Consider the stepwise regression analysis performed in Example 64.1. The final model included variables LogBUN and HGB. You can generate residual statistics for this analysis by refitting the model containing those variables and including an OUTPUT statement as in the following invocation of PROC PHREG. The keywords XBETA, RESMART, and RESDEV identify new variables that contain the linear predictor scores $\mathbf{z}'_j\hat{\boldsymbol{\beta}}$, martingale residuals, and deviance residuals. These variables are xb, mart, and dev, respectively.

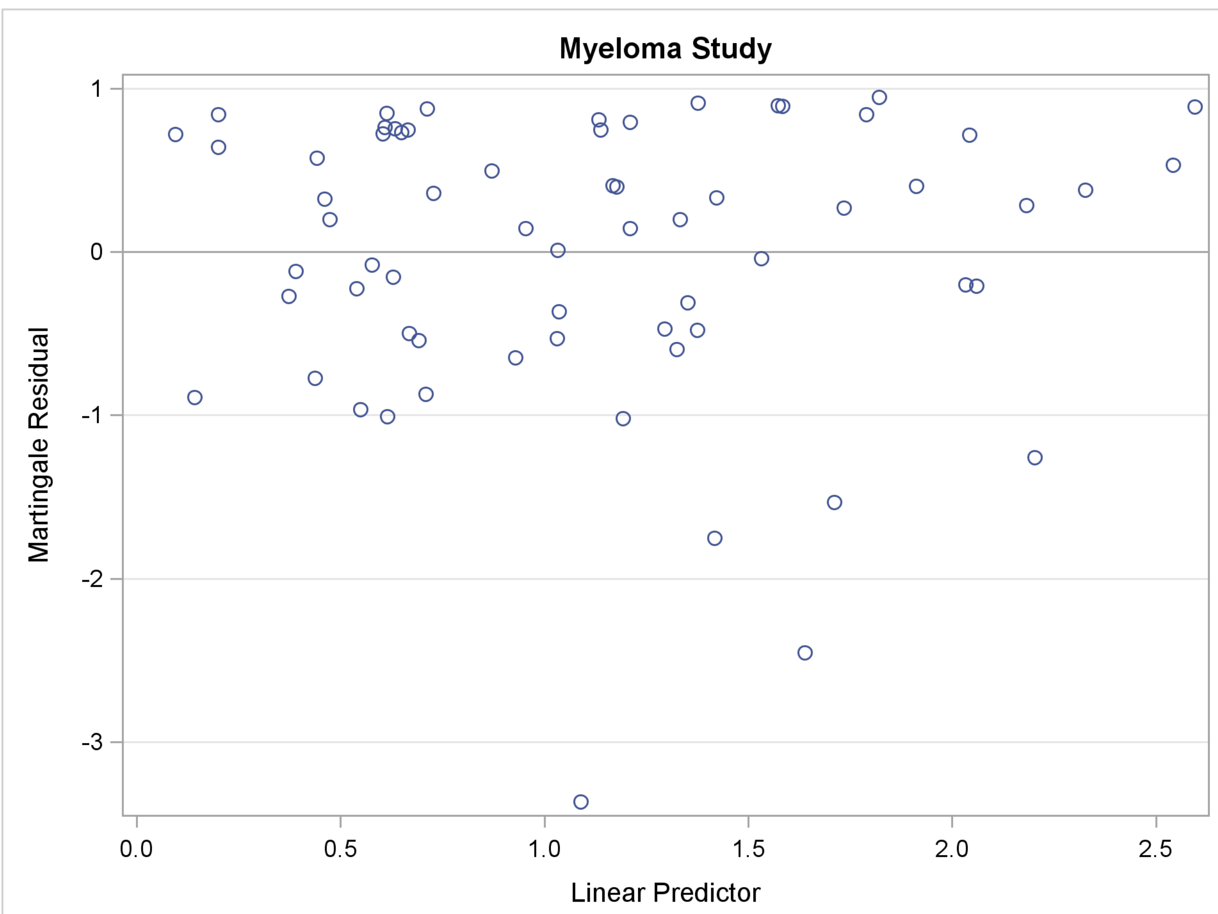
```
proc phreg data=Myeloma noprint;
   model Time*Vstatus(0)=LogBUN HGB;
   output out=Outp xbeta=Xb resmart=Mart resdev=Dev;
run;
```

The following statements plot the residuals against the linear predictor scores:

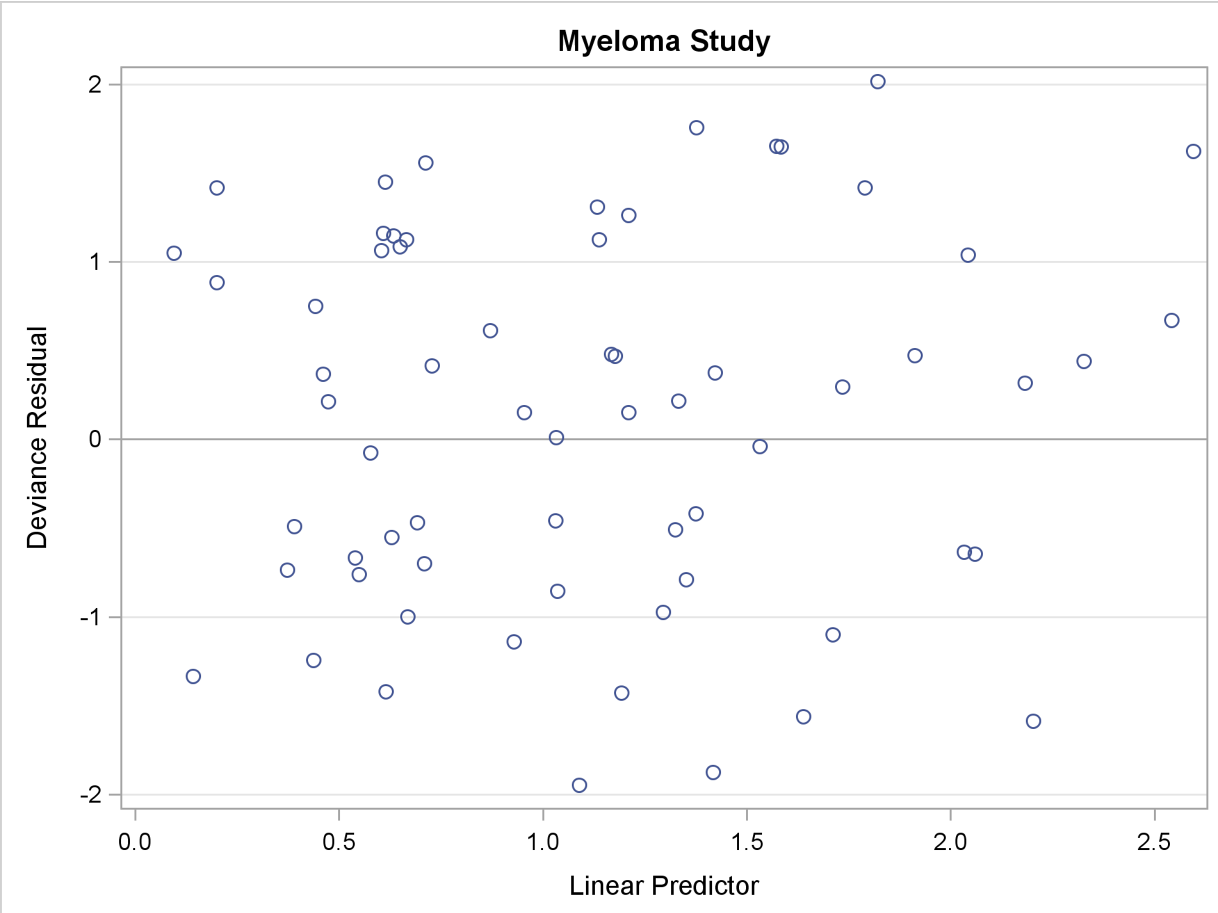
```
title "Myeloma Study";
proc sgplot data=Outp;
   yaxis grid;
   refline 0 / axis=y;
   scatter y=Mart x=Xb;
   run;
proc sgplot data=Outp;
   yaxis grid;
   refline 0 / axis=y;
   scatter y=Dev x=Xb;
   run;
```

The resulting plots are shown in Output 64.9.1 and Output 64.9.2. The martingale residuals are skewed because of the single event setting of the Cox model. The martingale residual plot shows an isolation point (with linear predictor score 1.09 and martingale residual −3.37), but this observation is no longer distinguishable in the deviance residual plot. In conclusion, there is no indication of a lack of fit of the model to individual observations.

**Output 64.9.1** Martingale Residual Plot



**Output 64.9.2** Deviance Residual Plot




## Example 64.10: Analysis of Recurrent Events Data

Recurrent events data consist of times to a number of repeated events for each sample unit—for example, times of recurrent episodes of a disease in patients. Various ways of analyzing recurrent events data are described in the section "Analysis of Multivariate Failure Time Data" on page 4595. The bladder cancer data listed in Wei, Lin, and Weissfeld (1989) are used here to illustrate these methods.

The data consist of 86 patients with superficial bladder tumors, which were removed when the patients entered the study. Of these patients, 48 were randomized into the placebo group, and 38 were randomized into the group receiving thiotepa. Many patients had multiple recurrences of tumors during the study, and new tumors were removed at each visit. The data set contains the first four recurrences of the tumor for each patient, and each recurrence time was measured from the patient's entry time into the study.

The data consist of the following eight variables:

- Trt, treatment group (1=placebo and 2=thiotepa)
- Time, follow-up time
- Number, number of initial tumors
- Size, initial tumor size
- T1, T2, T3, and T4, times of the four potential recurrences of the bladder tumor. A patient with only two recurrences has missing values in T3 and T4.

In the data set Bladder, four observations are created for each patient, one for each of the four potential tumor recurrences. In addition to values of Trt, Number, and Size for the patient, each observation contains the following variables:

- ID, patient's identification (which is the sequence number of the subject)
- Visit, visit number (with value $k$ for the $k$th potential tumor recurrence)
- TStart, time of the $(k-1)$th recurrence for Visit=$k$, or the entry time 0 if VISIT=1, or the follow-up time if the $(k-1)$th recurrence does not occur
- TStop, time of the $k$th recurrence if Visit=$k$ or follow-up time if the $k$th recurrence does not occur
- Status, event status of TStop (1=recurrence and 0=censored)

For instance, a patient with only one recurrence time at month 6 who was followed until month 10 will have values for Visit, TStart, TStop, and Status of (1,0,6,1), (2,6,10,0), (3,10,10,0), and (4,10,10,0), respectively. The last two observations are redundant for the intensity model and the proportional means model, but they are important for the analysis of the marginal Cox models. If the follow-up time is beyond the time of the fourth tumor recurrence, it is tempting to create a fifth observation with the time of the fourth tumor recurrence as the TStart value, the follow-up time as the TStop value, and a Status value of 0. However, Therneau and Grambsch (2000, Section 8.5) have warned against incorporating such observations into the analysis.

The following SAS statements create the data set Bladder:

```
data Bladder;
   keep ID TStart TStop Status Trt Number Size Visit;
   retain ID TStart 0;
   array tt T1-T4;
   infile datalines missover;
   input Trt Time Number Size T1-T4;
   ID + 1;
   TStart=0;
   do over tt;
      Visit=_i_;
      if tt = . then do;
         TStop=Time;
         Status=0;
      end;
      else do;
         TStop=tt;
         Status=1;
      end;
      output;
      TStart=TStop;
   end;
   if (TStart < Time) then delete;
   datalines;
1       0       1      1
1       1       1      3
1       4       2      1
1       7       1      1
1       10      5      1
1       10      4      1    6
1       14      1      1
1       18      1      1
1       18      1      3    5
1       18      1      1    12   16
1       23      3      3
1       23      1      3    10   15
1       23      1      1    3    16   23
1       23      3      1    3    9    21
1       24      2      3    7    10   16   24
1       25      1      1    3    15   25
1       26      1      2
1       26      8      1    1
1       26      1      4    2    26
1       28      1      2    25
1       29      1      4
1       29      1      2
1       29      4      1
1       30      1      6    28   30
1       30      1      5    2    17   22
1       30      2      1    3    6    8    12
1       31      1      3    12   15   24
1       32      1      2
1       34      2      1
```

```
1        36       2       1
1        36       3       1    29
1        37       1       2
1        40       4       1    9    17   22   24
1        40       5       1    16   19   23   29
1        41       1       2
1        43       1       1    3
1        43       2       6    6
1        44       2       1    3    6    9
1        45       1       1    9    11   20   26
1        48       1       1    18
1        49       1       3
1        51       3       1    35
1        53       1       7    17
1        53       3       1    3    15   46   51
1        59       1       1
1        61       3       2    2    15   24   30
1        64       1       3    5    14   19   27
1        64       2       3    2    8    12   13
2        1        1       3
2        1        1       1
2        5        8       1    5
2        9        1       2
2        10       1       1
2        13       1       1
2        14       2       6    3
2        17       5       3    1    3    5    7
2        18       5       1
2        18       1       3    17
2        19       5       1    2
2        21       1       1    17   19
2        22       1       1
2        25       1       3
2        25       1       5
2        25       1       1
2        26       1       1    6    12   13
2        27       1       1    6
2        29       2       1    2
2        36       8       3    26   35
2        38       1       1
2        39       1       1    22   23   27   32
2        39       6       1    4    16   23   27
2        40       3       1    24   26   29   40
2        41       3       2
2        41       1       1
2        43       1       1    1    27
2        44       1       1
2        44       6       1    2    20   23   27
2        45       1       2
2        46       1       4    2
2        46       1       4
2        49       3       3
2        50       1       1
2        50       4       1    4    24   47
```

```
2        54       3      4
2        54       2      1    38
2        59       1      3
;
run;
```

First, consider fitting the intensity model (Andersen and Gill 1982) and the proportional means model (Lin et al. 2000). The counting process style of input is used in the PROC PHREG specification. For the proportional means model, inference is based on the robust sandwich covariance estimate, which is requested by the COVB(AGGREGATE) option in the PROC PHREG statement. The COVM option is specified for the analysis of the intensity model to use the model-based covariance estimate. Note that some of the observations in the data set Bladder have a degenerated interval of risk. The presence of these observations does not affect the results of the analysis since none of these observations are included in any of the risk sets. However, the procedure will run more efficiently without these observations; consequently, in the following SAS statements, the WHERE clause is used to eliminate these redundant observations:

```
title 'Intensity Model and Proportional Means Model';
proc phreg data=Bladder covm covs(aggregate);
   model (TStart, TStop) * Status(0) = Trt Number Size;
   id id;
   where TStart < TStop;
run;
```

Results of fitting the intensity model and the proportional means model are shown in Output 64.10.1 and Output 64.10.2, respectively. The robust sandwich standard error estimate for Trt is larger than its model-based counterpart, rendering the effect of thiotepa less significant in the proportional means model ($p$=0.0747) than in the intensity model ($p$=0.0215).

**Output 64.10.1** Analysis of the Intensity Model

**Intensity Model and Proportional Means Model**

**The PHREG Procedure**

**Analysis of Maximum Likelihood Estimates with Model-Based Variance Estimate**

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| Trt | 1 | -0.45979 | 0.19996 | 5.2873 | 0.0215 | 0.631 |
| Number | 1 | 0.17165 | 0.04733 | 13.1541 | 0.0003 | 1.187 |
| Size | 1 | -0.04256 | 0.06903 | 0.3801 | 0.5375 | 0.958 |

**Output 64.10.2** Analysis of the Proportional Means Model

Analysis of Maximum Likelihood Estimates with Sandwich Variance Estimate

| Parameter | DF | Parameter Estimate | Standard Error | StdErr Ratio | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|---|
| Trt | 1 | -0.45979 | 0.25801 | 1.290 | 3.1757 | 0.0747 | 0.631 |
| Number | 1 | 0.17165 | 0.06131 | 1.296 | 7.8373 | 0.0051 | 1.187 |
| Size | 1 | -0.04256 | 0.07555 | 1.094 | 0.3174 | 0.5732 | 0.958 |

Next, consider the conditional models of Prentice, Williams, and Peterson (1981). In the PWP models, the risk set for the $(k+1)$th recurrence is restricted to those patients who have experienced the first $k$ recurrences. For example, a patient who experienced only one recurrence is an event observation for the first recurrence; this patient is a censored observation for the second recurrence and should not be included in the risk set for the third or fourth recurrence. The following DATA step eliminates those observations that should not be in the risk sets, forming a new input data set (named Bladder2) for fitting the PWP models. The variable Gaptime, representing the gap times between successive recurrences, is also created.

```
data Bladder2(drop=LastStatus);
    retain LastStatus;
    set Bladder;
    by ID;
    if first.id then LastStatus=1;
    if (Status=0 and LastStatus=0) then delete;
    LastStatus=Status;
    Gaptime=Tstop-Tstart;
run;
```

The following statements fit the PWP total time model. The variables Trt1, Trt2, Trt3, and Trt4 are visit-specific variables for Trt; the variables Number1, Number2, Numvber3, and Number4 are visit-specific variables for Number; and the variables Size1, Size2, Size3, and Size4 are visit-specific variables for Size.

```
title 'PWP Total Time Model with Noncommon Effects';
proc phreg data=Bladder2;
   model (TStart,Tstop) * Status(0) = Trt1-Trt4 Number1-Number4
                                      Size1-Size4;
   Trt1= Trt * (Visit=1);
   Trt2= Trt * (Visit=2);
   Trt3= Trt * (Visit=3);
   Trt4= Trt * (Visit=4);
   Number1= Number * (Visit=1);
   Number2= Number * (Visit=2);
   Number3= Number * (Visit=3);
   Number4= Number * (Visit=4);
   Size1= Size * (Visit=1);
   Size2= Size * (Visit=2);
   Size3= Size * (Visit=3);
   Size4= Size * (Visit=4);
   strata Visit;
run;
```

Results of the analysis of the PWP total time model are shown in Output 64.10.3. There is no significant treatment effect on the total time in any of the four tumor recurrences.

**Output 64.10.3** Analysis of the PWP Total Time Model with Noncommon Effects

PWP Total Time Model with Noncommon Effects

The PHREG Procedure

Summary of the Number of Event and Censored Values

| Stratum | Visit | Total | Event | Censored | Percent Censored |
|---|---|---|---|---|---|
| 1 | 1 | 85 | 47 | 38 | 44.71 |
| 2 | 2 | 46 | 29 | 17 | 36.96 |
| 3 | 3 | 27 | 22 | 5 | 18.52 |
| 4 | 4 | 20 | 14 | 6 | 30.00 |
| Total | | 178 | 112 | 66 | 37.08 |

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| Trt1 | 1 | -0.51757 | 0.31576 | 2.6868 | 0.1012 | 0.596 |
| Trt2 | 1 | -0.45967 | 0.40642 | 1.2792 | 0.2581 | 0.631 |
| Trt3 | 1 | 0.11700 | 0.67183 | 0.0303 | 0.8617 | 1.124 |
| Trt4 | 1 | -0.04059 | 0.79251 | 0.0026 | 0.9592 | 0.960 |
| Number1 | 1 | 0.23605 | 0.07607 | 9.6287 | 0.0019 | 1.266 |
| Number2 | 1 | -0.02044 | 0.09052 | 0.0510 | 0.8213 | 0.980 |
| Number3 | 1 | 0.01219 | 0.18208 | 0.0045 | 0.9466 | 1.012 |
| Number4 | 1 | 0.18915 | 0.24443 | 0.5989 | 0.4390 | 1.208 |
| Size1 | 1 | 0.06790 | 0.10125 | 0.4498 | 0.5024 | 1.070 |
| Size2 | 1 | -0.15425 | 0.12300 | 1.5728 | 0.2098 | 0.857 |
| Size3 | 1 | 0.14891 | 0.26299 | 0.3206 | 0.5713 | 1.161 |
| Size4 | 1 | 0.0000732 | 0.34297 | 0.0000 | 0.9998 | 1.000 |

The following statements fit the PWP gap-time model:

```
title 'PWP Gap-Time Model with Noncommon Effects';
proc phreg data=Bladder2;
   model Gaptime * Status(0) = Trt1-Trt4 Number1-Number4
                               Size1-Size4;
   Trt1= Trt * (Visit=1);
   Trt2= Trt * (Visit=2);
   Trt3= Trt * (Visit=3);
   Trt4= Trt * (Visit=4);
   Number1= Number * (Visit=1);
   Number2= Number * (Visit=2);
   Number3= Number * (Visit=3);
   Number4= Number * (Visit=4);
   Size1= Size * (Visit=1);
   Size2= Size * (Visit=2);
   Size3= Size * (Visit=3);
```

```
      Size4= Size * (Visit=4);
      strata Visit;
   run;
```

Results of the analysis of the PWP gap-time model are shown in Output 64.10.4. Note that the regression coefficients for the first tumor recurrence are the same as those of the total time model, since the total time and the gap time are the same for the first recurrence. There is no significant treatment effect on the gap times for any of the four tumor recurrences.

**Output 64.10.4** Analysis of the PWP Gap-Time Model with Noncommon Effects

PWP Gap-Time Model with Noncommon Effects

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| Trt1 | 1 | -0.51757 | 0.31576 | 2.6868 | 0.1012 | 0.596 |
| Trt2 | 1 | -0.25911 | 0.40511 | 0.4091 | 0.5224 | 0.772 |
| Trt3 | 1 | 0.22105 | 0.54909 | 0.1621 | 0.6873 | 1.247 |
| Trt4 | 1 | -0.19498 | 0.64184 | 0.0923 | 0.7613 | 0.823 |
| Number1 | 1 | 0.23605 | 0.07607 | 9.6287 | 0.0019 | 1.266 |
| Number2 | 1 | -0.00571 | 0.09667 | 0.0035 | 0.9529 | 0.994 |
| Number3 | 1 | 0.12935 | 0.15970 | 0.6561 | 0.4180 | 1.138 |
| Number4 | 1 | 0.42079 | 0.19816 | 4.5091 | 0.0337 | 1.523 |
| Size1 | 1 | 0.06790 | 0.10125 | 0.4498 | 0.5024 | 1.070 |
| Size2 | 1 | -0.11636 | 0.11924 | 0.9524 | 0.3291 | 0.890 |
| Size3 | 1 | 0.24995 | 0.23113 | 1.1695 | 0.2795 | 1.284 |
| Size4 | 1 | 0.03557 | 0.29043 | 0.0150 | 0.9025 | 1.036 |

You can fit the PWP total time model with common effects by using the following SAS statements. However, the analysis is not shown here.

```
title2 'PWP Total Time Model with Common Effects';
proc phreg data=Bladder2;
   model (tstart,tstop) * status(0) = Trt Number Size;
   strata Visit;
run;
```

You can fit the PWP gap-time model with common effects by using the following statements. Again, the analysis is not shown here.

```
title2 'PWP Gap Time Model with Common Effects';
proc phreg data=Bladder2;
   model Gaptime * Status(0) = Trt Number Size;
   strata Visit;
run;
```

Recurrent events data are a special case of multiple events data in which the recurrence times are regarded as multivariate failure times and the marginal approach of Wei, Lin, and Weissfeld (1989) can be used. WLW fits a Cox model to each of the component times and makes statistical inference of the regression parameters based on a robust sandwich covariance matrix estimate. No specific correlation structure is imposed on the multivariate failure times. For the $k$th marginal model, let $\boldsymbol{\beta}_k$ denote the row vector of regression parameters, let $\hat{\boldsymbol{\beta}}_k$ denote the maximum likelihood estimate of $\boldsymbol{\beta}_k$, let $\hat{\mathbf{A}}_k$ denote the covariance matrix obtained by inverting the observed information matrix, and let $\mathbf{R}_k$ denote the matrix of score residuals. WLW showed that the joint distribution of $(\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_4)'$ can be approximated by a multivariate normal distribution with mean vector $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_4)'$ and robust covariance matrix

$$\begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} & \mathbf{V}_{13} & \mathbf{V}_{14} \\ \mathbf{V}_{21} & \mathbf{V}_{22} & \mathbf{V}_{23} & \mathbf{V}_{24} \\ \mathbf{V}_{31} & \mathbf{V}_{32} & \mathbf{V}_{33} & \mathbf{V}_{34} \\ \mathbf{V}_{41} & \mathbf{V}_{42} & \mathbf{V}_{43} & \mathbf{V}_{44} \end{pmatrix}$$

with the submatrix $\mathbf{V}_{ij}$ given by

$$\mathbf{V}_{ij} = \hat{\mathbf{A}}_i(\mathbf{R}_i'\mathbf{R}_j)\hat{\mathbf{A}}_j$$

In this example, there are four marginal proportional hazards models, one for each potential recurrence time. Instead of fitting one model at a time, you can fit all four marginal models in one analysis by using the STRATA statement and model-specific covariates as in the following statements. Using Visit as the STRATA variable on the input data set Bladder, PROC PHREG simultaneously fits all four marginal models, one for each Visit value. The COVS(AGGREGATE) option is specified to compute the robust sandwich variance estimate by summing up the score residuals for each distinct pattern of ID value. The TEST statement TREATMENT is used to perform the global test of no treatment effect for each tumor recurrence, the AVERAGE option is specified to estimate the parameter for the common treatment effect, and the E option displays the optimal weights for the common treatment effect.

```
title 'Wei-Lin-Weissfeld Model';
proc phreg data=Bladder covs(aggregate);
   model TStop*Status(0)=Trt1-Trt4 Number1-Number4 Size1-Size4;
   Trt1= Trt * (Visit=1);
   Trt2= Trt * (Visit=2);
   Trt3= Trt * (Visit=3);
   Trt4= Trt * (Visit=4);
   Number1= Number * (Visit=1);
   Number2= Number * (Visit=2);
   Number3= Number * (Visit=3);
   Number4= Number * (Visit=4);
   Size1= Size * (Visit=1);
   Size2= Size * (Visit=2);
   Size3= Size * (Visit=3);
   Size4= Size * (Visit=4);
   strata Visit;
   id ID;
   TREATMENT: test trt1,trt2,trt3,trt4/average e;
   run;
```

Out of the 86 patients, 47 patients have only one tumor recurrence, 29 patients have two recurrences, 22 patients have three recurrences, and 14 patients have four recurrences (Output 64.10.5). Parameter estimates for the four marginal models are shown in Output 64.10.6. The 4 DF Wald test (Output 64.10.7) indicates a lack of evidence of a treatment effect in any of the four recurrences ($p$=0.4105). The optimal weights for estimating the parameter of the common treatment effect are 0.67684, 0.25723, –0.07547, and 0.14140 for Trt1, Trt2, Trt3, and Trt4, respectively, which gives a parameter estimate of –0.5489 with a standard error estimate of 0.2853. A more sensitive test for a treatment effect is the 1 DF test based on this common parameter; however, there is still insufficient evidence for such effect at the 0.05 level ($p$=0.0543).

**Output 64.10.5** Summary of Bladder Tumor Recurrences in 86 Patients

**Wei-Lin-Weissfeld Model**

**The PHREG Procedure**

**Summary of the Number of Event and Censored Values**

| Stratum | Visit | Total | Event | Censored | Percent Censored |
|---|---|---|---|---|---|
| 1 | 1 | 86 | 47 | 39 | 45.35 |
| 2 | 2 | 86 | 29 | 57 | 66.28 |
| 3 | 3 | 86 | 22 | 64 | 74.42 |
| 4 | 4 | 86 | 14 | 72 | 83.72 |
| Total | | 344 | 112 | 232 | 67.44 |

**Output 64.10.6** Analysis of Marginal Cox Models

**Analysis of Maximum Likelihood Estimates**

| Parameter | DF | Parameter Estimate | Standard Error | StdErr Ratio | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|---|
| Trt1 | 1 | -0.51762 | 0.30750 | 0.974 | 2.8336 | 0.0923 | 0.596 |
| Trt2 | 1 | -0.61944 | 0.36391 | 0.926 | 2.8975 | 0.0887 | 0.538 |
| Trt3 | 1 | -0.69988 | 0.41516 | 0.903 | 2.8419 | 0.0918 | 0.497 |
| Trt4 | 1 | -0.65079 | 0.48971 | 0.848 | 1.7661 | 0.1839 | 0.522 |
| Number1 | 1 | 0.23599 | 0.07208 | 0.947 | 10.7204 | 0.0011 | 1.266 |
| Number2 | 1 | 0.13756 | 0.08690 | 0.946 | 2.5059 | 0.1134 | 1.147 |
| Number3 | 1 | 0.16984 | 0.10356 | 0.984 | 2.6896 | 0.1010 | 1.185 |
| Number4 | 1 | 0.32880 | 0.11382 | 0.909 | 8.3453 | 0.0039 | 1.389 |
| Size1 | 1 | 0.06789 | 0.08529 | 0.842 | 0.6336 | 0.4260 | 1.070 |
| Size2 | 1 | -0.07612 | 0.11812 | 0.881 | 0.4153 | 0.5193 | 0.927 |
| Size3 | 1 | -0.21131 | 0.17198 | 0.943 | 1.5097 | 0.2192 | 0.810 |
| Size4 | 1 | -0.20317 | 0.19106 | 0.830 | 1.1308 | 0.2876 | 0.816 |

**Output 64.10.7** Tests of Treatment Effects

Wei-Lin-Weissfeld Model

The PHREG Procedure

Linear Coefficients for Test TREATMENT

| Parameter | Row 1 | Row 2 | Row 3 | Row 4 | Average Effect |
|---|---|---|---|---|---|
| Trt1 | 1 | 0 | 0 | 0 | 0.67684 |
| Trt2 | 0 | 1 | 0 | 0 | 0.25723 |
| Trt3 | 0 | 0 | 1 | 0 | -0.07547 |
| Trt4 | 0 | 0 | 0 | 1 | 0.14140 |
| Number1 | 0 | 0 | 0 | 0 | 0.00000 |
| Number2 | 0 | 0 | 0 | 0 | 0.00000 |
| Number3 | 0 | 0 | 0 | 0 | 0.00000 |
| Number4 | 0 | 0 | 0 | 0 | 0.00000 |
| Size1 | 0 | 0 | 0 | 0 | 0.00000 |
| Size2 | 0 | 0 | 0 | 0 | 0.00000 |
| Size3 | 0 | 0 | 0 | 0 | 0.00000 |
| Size4 | 0 | 0 | 0 | 0 | 0.00000 |
| CONSTANT | 0 | 0 | 0 | 0 | 0.00000 |

Test TREATMENT Results

| Wald Chi-Square | DF | Pr > ChiSq |
|---|---|---|
| 3.9668 | 4 | 0.4105 |

Average Effect for Test TREATMENT

| Estimate | Standard Error | z-Score | Pr > \|z\| |
|---|---|---|---|
| -0.5489 | 0.2853 | -1.9240 | 0.0543 |

## Example 64.11: Analysis of Clustered Data

When experimental units are naturally or artificially clustered, failure times of experimental units within a cluster are correlated. Lee, Wei, and Amato (1992) estimate the regression parameters in the Cox model by the maximum partial likelihood estimates under an independent working assumption and use a robust sandwich covariance matrix estimate to account for the intracluster dependence. A subset of data from the Diabetic Retinopathy Study (DRS) is used to illustrate the methodology as in Lin (1994).

The following DATA step creates the data set Blind that represents 197 diabetic patients who have a high risk of experiencing blindness in both eyes as defined by DRS criteria. One eye of each patient is treated with laser photocoagulation. The hypothesis of interest is whether the laser treatment delays the occurrence of blindness. Since juvenile and adult diabetes have very different courses, it is also desirable to examine how the age of onset of diabetes might affect the time of blindness. Since there are no biological differences between the left eye and the right eye, it is natural to assume a common baseline hazard function for the failure times of the left and right eyes.

Each patient is a cluster that contributes two observations to the input data set, one for each eye. The following variables are in the input data set Blind:

- ID, patient's identification
- Time, failure time
- Status, event indicator (0=censored and 1=uncensored)
- Treatment, treatment received (1=laser photocoagulation and 0=otherwise)
- DiabeticType, type of diabetes (0=juvenile onset with age of onset at 20 or under, and 1= adult onset with age of onset over 20)

```
data Blind;
   input ID Time Status DiabeticType Treatment @@;
   datalines;
   5 46.23 0 1 1    5 46.23 0 1 0   14 42.50 0 0 1   14 31.30 1 0 0
  16 42.27 0 0 1   16 42.27 0 0 0   25 20.60 0 0 1   25 20.60 0 0 0
  29 38.77 0 0 1   29  0.30 1 0 0   46 65.23 0 0 1   46 54.27 1 0 0
  49 63.50 0 0 1   49 10.80 1 0 0   56 23.17 0 0 1   56 23.17 0 0 0
  61  1.47 0 0 1   61  1.47 0 0 0   71 58.07 0 1 1   71 13.83 1 1 0
 100 46.43 1 1 1  100 48.53 0 1 0  112 44.40 0 1 1  112  7.90 1 1 0
 120 39.57 0 1 1  120 39.57 0 1 0  127 30.83 1 1 1  127 38.57 1 1 0
 133 66.27 0 1 1  133 14.10 1 1 0  150 20.17 1 0 1  150  6.90 1 0 0
 167 58.43 0 1 1  167 41.40 1 1 0  176 58.20 0 0 1  176 58.20 0 0 0

... more lines ...

1727 49.97 0 1 1 1727  2.90 1 1 0 1746 45.90 0 0 1 1746  1.43 1 0 0
1749 41.93 0 1 1 1749 41.93 0 1 0
;
run;
```

As a preliminary analysis, PROC FREQ is used to break down the numbers of blindness in the control and treated eyes:

```
proc freq data=Blind;
   table Treatment*Status;
   run;
```

By the end of the study, 54 treated eyes and 101 untreated eyes have developed blindness (Output 64.11.1).

**Output 64.11.1** Breakdown of Blindness in the Control and Treated Groups

```
                    Wei-Lin-Weissfeld Model

                      The FREQ Procedure

                 Table of Treatment by Status

           Treatment     Status

           Frequency|
           Percent  |
           Row Pct  |
           Col Pct  |       0|       1|  Total
           ---------+--------+--------+
                  0 |     96 |    101 |    197
                    |  24.37 |  25.63 |  50.00
                    |  48.73 |  51.27 |
                    |  40.17 |  65.16 |
           ---------+--------+--------+
                  1 |    143 |     54 |    197
                    |  36.29 |  13.71 |  50.00
                    |  72.59 |  27.41 |
                    |  59.83 |  34.84 |
           ---------+--------+--------+
           Total         239      155      394
                       60.66    39.34   100.00
```

The analysis of Lee, Wei, and Amato (1992) can be carried out by the following PROC PHREG specification. The explanatory variables in this Cox model are Treatment, DiabeticType, and the Treatment × DiabeticType interaction. The COVS(AGGREGATE) is specified to compute the robust sandwich covariance matrix estimate.

```
proc phreg data=Blind covs(aggregate) namelen=22;
   model Time*Status(0)=Treatment DiabeticType Treatment*DiabeticType;
   id ID;
   run;
```

The robust standard error estimates are smaller than the model-based counterparts (Output 64.11.2), since the ratio of the robust standard error estimate relative to the model-based estimate is less than 1 for each variable. Laser photocoagulation appears to be effective ($p$=0.0217) in delaying the occurrence of blindness. The effect is much more prominent for adult-onset diabetes than for juvenile-onset diabetes.

**Output 64.11.2** Inference Based on the Robust Sandwich Covariance

```
                          Wei-Lin-Weissfeld Model

                            The PHREG Procedure

                   Analysis of Maximum Likelihood Estimates

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

Treatment               1    -0.42467     0.18497  0.850     5.2713     0.0217
DiabeticType            1     0.34084     0.19558  0.982     3.0371     0.0814
Treatment*DiabeticType  1    -0.84566     0.30353  0.865     7.7622     0.0053

                  Analysis of Maximum Likelihood Estimates

                                  Hazard
        Parameter                  Ratio    Label

        Treatment                    .
        DiabeticType                 .
        Treatment*DiabeticType       .       Treatment * DiabeticType
```

## Example 64.12: Model Assessment Using Cumulative Sums of Martingale Residuals

The Mayo liver disease example of Lin, Wei, and Ying (1993) is reproduced here to illustrate the checking of the functional form of a covariate and the assessment of the proportional hazards assumption. The data represent 418 patients with primary biliary cirrhosis (PBC), among whom 161 had died as of the date of data listing. A subset of the variables is saved in the SAS data set Liver. The data set contains the following variables:

- Time, follow-up time, in years
- Status, event indicator with value 1 for death time and value 0 for censored time
- Age, age in years from birth to study registration
- Albumin, serum albumin level, in gm/dl
- Bilirubin, serum bilirubin level, in mg/dl
- Edema, edema presence
- Protime, prothrombin time, in seconds

The following statements create the data set Liver:

```
data Liver;
   input Time Status Age Albumin Bilirubin Edema Protime @@;
   label Time="Follow-up Time in Years";
   Time= Time / 365.25;
   datalines;
  400 1 58.7652 2.60 14.5 1.0 12.2 4500 0 56.4463 4.14  1.1 0.0 10.6
 1012 1 70.0726 3.48  1.4 0.5 12.0 1925 1 54.7406 2.54  1.8 0.5 10.3
 1504 0 38.1054 3.53  3.4 0.0 10.9 2503 1 66.2587 3.98  0.8 0.0 11.0
 1832 0 55.5346 4.09  1.0 0.0  9.7 2466 1 53.0568 4.00  0.3 0.0 11.0
 2400 1 42.5079 3.08  3.2 0.0 11.0   51 1 70.5599 2.74 12.6 1.0 11.5
 3762 1 53.7139 4.16  1.4 0.0 12.0  304 1 59.1376 3.52  3.6 0.0 13.6
 3577 0 45.6893 3.85  0.7 0.0 10.6 1217 1 56.2218 2.27  0.8 1.0 11.0

... more lines ...

 1103 0 39.0000 3.83  0.9 0.0 11.2 1055 0 57.0000 3.42  1.6 0.0  9.9
  691 0 58.0000 3.75  0.8 0.0 10.4  976 0 53.0000 3.29  0.7 0.0 10.6
;
run;
```

Consider fitting a Cox model for the survival time of the PCB patients with the covariates Bilirubin, log(Protime), log(Albumin), Age, and Edema. The log transform, which is often applied to blood chemistry measurements, is deliberately not employed for Bilirubin. It is of interest to assess the functional form of the variable Bilirubin in the Cox model. The specifications are as follows:

```
ods graphics on;
proc phreg data=Liver;
   model Time*Status(0)=Bilirubin logProtime logAlbumin Age Edema;
   logProtime=log(Protime);
   logAlbumin=log(Albumin);
   assess var=(Bilirubin) /  resample seed=7548;
   run;
ods graphics off;
```

The ASSESS statement creates a plot of the cumulative martingale residuals against the values of the covariate Bilirubin, which is specified in the VAR= option. The RESAMPLE option computes the $p$-value of a Kolmogorov-type supremum test based on a sample of 1,000 simulated residual patterns.

Parameter estimates of the model fit are shown in Output 64.12.1. The plot in Output 64.12.2 displays the observed cumulative martingale residual process for Bilirubin together with 20 simulated realizations from the null distribution. This graphical display is requested by specifying the **ods graphics on** statement and the ASSESS statement. It is obvious that the observed process is atypical compared to the simulated realizations. Also, none of the 1,000 simulated realizations has an absolute maximum exceeding that of the observed cumulative martingale residual process. Both the graphical and numerical results indicate that a transform is deemed necessary for Bilirubin in the model.

**Output 64.12.1** Cox Model with Bilirubin as a Covariate

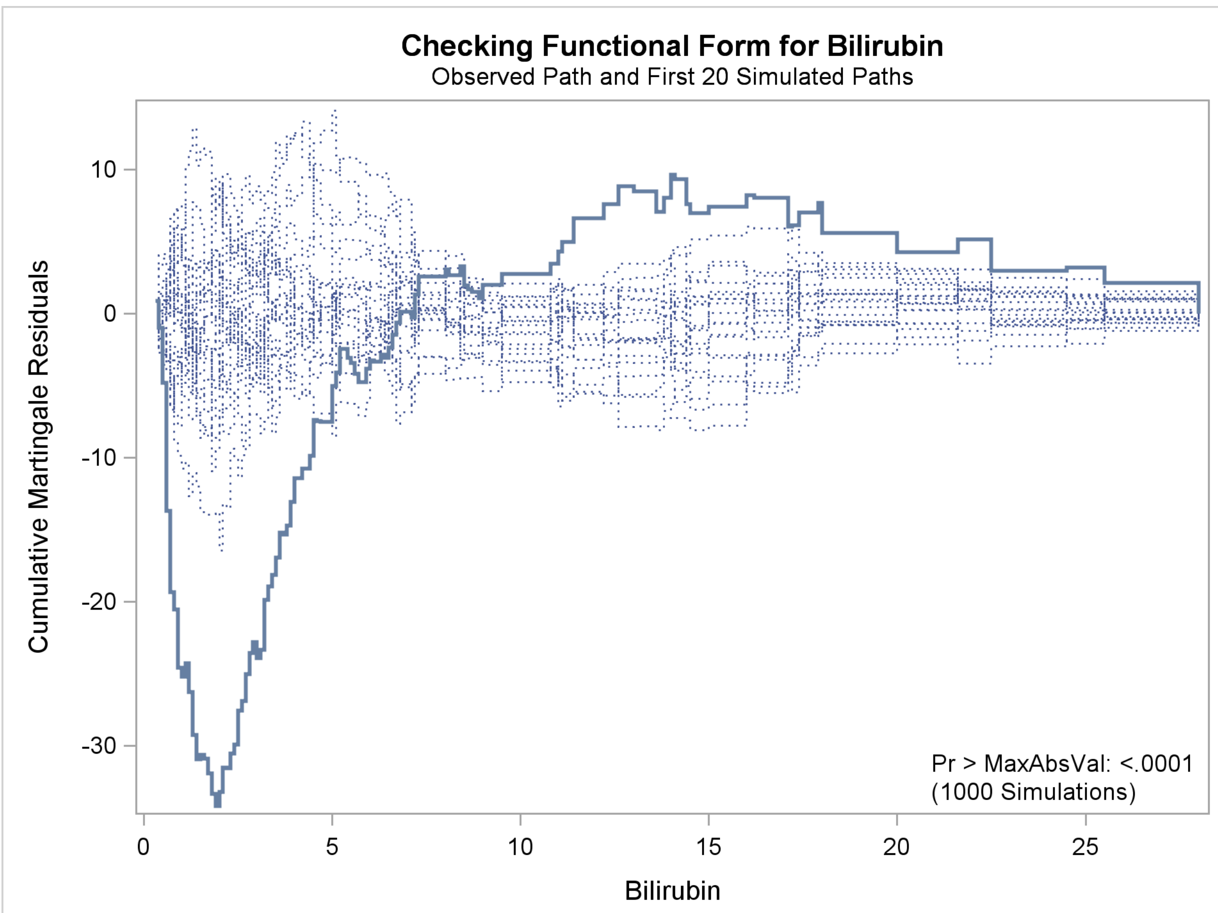
Wei-Lin-Weissfeld Model

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| Bilirubin | 1 | 0.11733 | 0.01298 | 81.7567 | <.0001 | 1.124 |
| logProtime | 1 | 2.77581 | 0.71482 | 15.0794 | 0.0001 | 16.052 |
| logAlbumin | 1 | -3.17195 | 0.62945 | 25.3939 | <.0001 | 0.042 |
| Age | 1 | 0.03779 | 0.00805 | 22.0288 | <.0001 | 1.039 |
| Edema | 1 | 0.84772 | 0.28125 | 9.0850 | 0.0026 | 2.334 |

**Output 64.12.2** Cumulative Martingale Residuals vs Bilirubin



The cumulative martingale residual plots in Output 64.12.3 provide guidance in suggesting a more appropriate functional form for a covariate. The four curves were created from simple forms of misspecification by using 1,000 simulated times from a exponential model with 20% censoring. The true and fitted models are shown in Table 64.9. The following statements produce Output 64.12.3.

```
data sim(drop=tmp);
   p = 1 / 91;
   seed = 1;
   do n = 1 to 10000;
      x1 = rantbl( seed, p, p, p, p, p, p, p, p, p, p,
                         p, p, p, p, p, p, p, p, p, p,
                         p, p, p, p, p, p, p, p, p, p,
                         p, p, p, p, p, p, p, p, p, p,
                         p, p, p, p, p, p, p, p, p, p,
                         p, p, p, p, p, p, p, p, p, p,
                         p, p, p, p, p, p, p, p, p, p,
                         p, p, p, p, p, p, p, p, p, p,
                         p, p, p, p, p, p, p, p, p, p );

      x1 = 1 + ( x1 - 1 ) / 10;
      x2= x1 * x1;
      x3= x1 * x2;
      status= rantbl(seed, .8);
      tmp= log(1-ranuni(seed));
      t1= -exp(-log(x1)) * tmp;
      t2= -exp(-.1*(x1+x2)) * tmp;
      t3= -exp(-.01*(x1+x2+x3)) * tmp;
      tt= -exp(-(x1>5)) * tmp;
      output;
   end;
   run;
proc sort data=sim;
   by x1;
   run;
proc phreg data=sim noprint;
   model t1*status(2)=x1;
   output out=out1 resmart=resmart;
   run;
proc phreg data=sim noprint;
   model t2*status(2)=x1;
   output out=out2 resmart=resmart;
   run;

proc phreg data=sim noprint;
   model t3*status(2)=x1 x2;
   output out=out3 resmart=resmart;
   run;

proc phreg data=sim noprint;
   model tt*status(2)=x1;
   output out=out4 resmart=resmart;
   run;
```

```
data out1(keep=x1 cresid1);
   retain cresid1 0;
   set out1;
   by x1;
   cresid1 + resmart;
   if last.x1  then output;
   run;
data out2(keep=x1 cresid2);
   retain cresid2 0;
   set out2;
   by x1;
   cresid2 + resmart;
   if last.x1  then output;
   run;
data out3(keep=x1 cresid3);
   retain cresid3 0;
   set out3;
   by x1;
   cresid3 + resmart;
   if last.x1  then output;
   run;
data out4(keep=x1 cresid4);
   retain cresid4 0;
   set out4;
   by x1;
   cresid4 + resmart;
   if last.x1  then output;
   run;
data all;
   set out1;
   set out2;
   set out3;
   set out4;
   run;

proc template;
   define statgraph MisSpecification;
      BeginGraph;
         entrytitle "Covariate Misspecification";
         layout lattice / columns=2 rows=2 columndatarange=unionall;

            columnaxes;
               columnaxis / display=(ticks tickvalues label) label="x";
               columnaxis / display=(ticks tickvalues label) label="x";
            endcolumnaxes;

            cell;
               cellheader;
                  entry "(a) Data: log(X), Model: X";
               endcellheader;
               layout overlay / xaxisopts=(display=none)
                                yaxisopts=(label="Cumulative Residual");
                  seriesplot y=cresid1 x=x1 / lineattrs=GraphFit;
               endlayout;
```

```
         endcell;

         cell;
            cellheader;
               entry "(b) Data: X*X, Model: X";
            endcellheader;
            layout overlay / xaxisopts=(display=none)
                             yaxisopts=(label=" ");
               seriesplot y=cresid2 x=x1 / lineattrs=GraphFit;
            endlayout;
         endcell;

         cell;
             cellheader;
                entry "(c) Data: X*X*X, Model: X*X";
             endcellheader;
             layout overlay / xaxisopts=(display=none)
                              yaxisopts=(label="Cumulative Residual");
                seriesplot y=cresid3 x=x1 / lineattrs=GraphFit;
             endlayout;
         endcell;

         cell;
            cellheader;
               entry "(d) Data: I(X>5), Model: X";
            endcellheader;
            layout overlay / xaxisopts=(display=none)
                             yaxisopts=(label=" ");
               seriesplot y=cresid4 x=x1 / lineattrs=GraphFit;
            endlayout;
         endcell;

      endlayout;
   EndGraph;
 end;

proc sgrender data=all template="MisSpecification";
   run;
```

**Output 64.12.3** Typical Cumulative Residual Plot Patterns

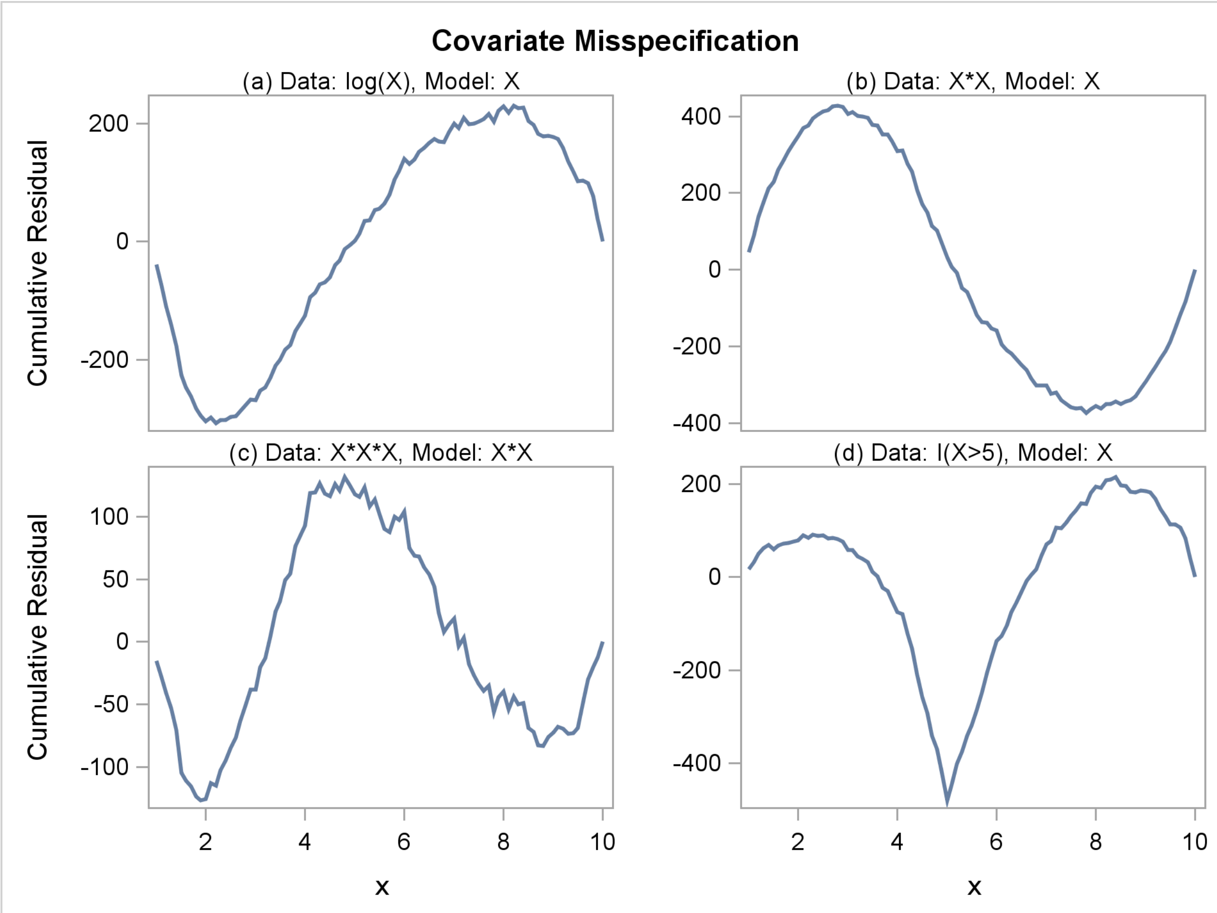



**Table 64.9** Model Misspecifications

| Plot | Data | Fitted Model |
|---|---|---|
| (a) | $\log(X)$ | $X$ |
| (b) | $\{X, X^2\}$ | $X$ |
| (c) | $\{X, X^2, X^3\}$ | $\{X, X^2\}$ |
| (d) | $I(X > 5)$ | $X$ |

The curve of observed cumulative martingale residuals in Output 64.12.2 most resembles the behavior of the curve in plot (a) of Output 64.12.3, indicating that log(Bilirubin) might be a more appropriate term in the model than Bilirubin.

Next, the analysis of the natural history of the PBC is repeated with log(Bilirubin) replacing Bilirubin, and the functional form of log(Bilirubin) is assessed. Also assessed is the proportional hazards assumption for the Cox model. The analysis is carried out by the following statements:

```
ods graphics on;
proc phreg data=Liver;
   model Time*Status(0)=logBilirubin logProtime logAlbumin Age Edema;
   logBilirubin=log(Bilirubin);
   logProtime=log(Protime);
   logAlbumin=log(Albumin);
   assess var=(logBilirubin) ph / crpanel resample seed=19;
   run;
ods html close;
```

The SEED= option specifies a integer seed for generating random numbers. The CRPANEL option in the ASSESS statement requests a panel of four plots. Each plot displays the observed cumulative martingale residual process along with two simulated realizations. The PH option checks the proportional hazards assumption of the model by plotting the observed standardized score process with 20 simulated realizations for each covariate in the model.

Output 64.12.4 displays the parameter estimates of the fitted model. The cumulative martingale residual plots in Output 64.12.5 and Output 64.12.6 show that the observed martingale residual process is more typical of the simulated realizations. The *p*-value for the Kolmogorov-type supremum test based on 1,000 simulations is 0.052, indicating that the log transform is a much improved functional form for Bilirubin.

**Output 64.12.4** Model with log(Bilirubin) as a Covariate

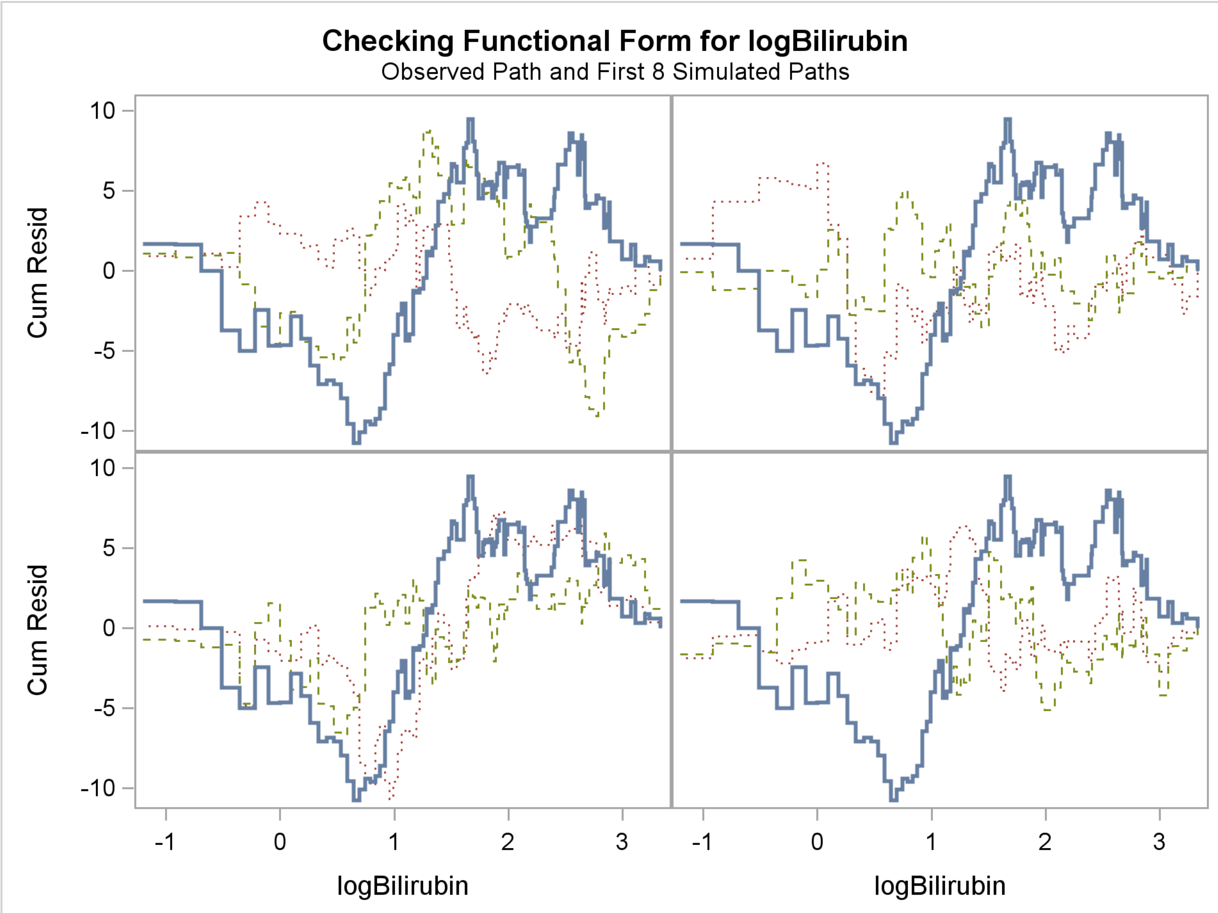
**Wei-Lin-Weissfeld Model**

**The PHREG Procedure**

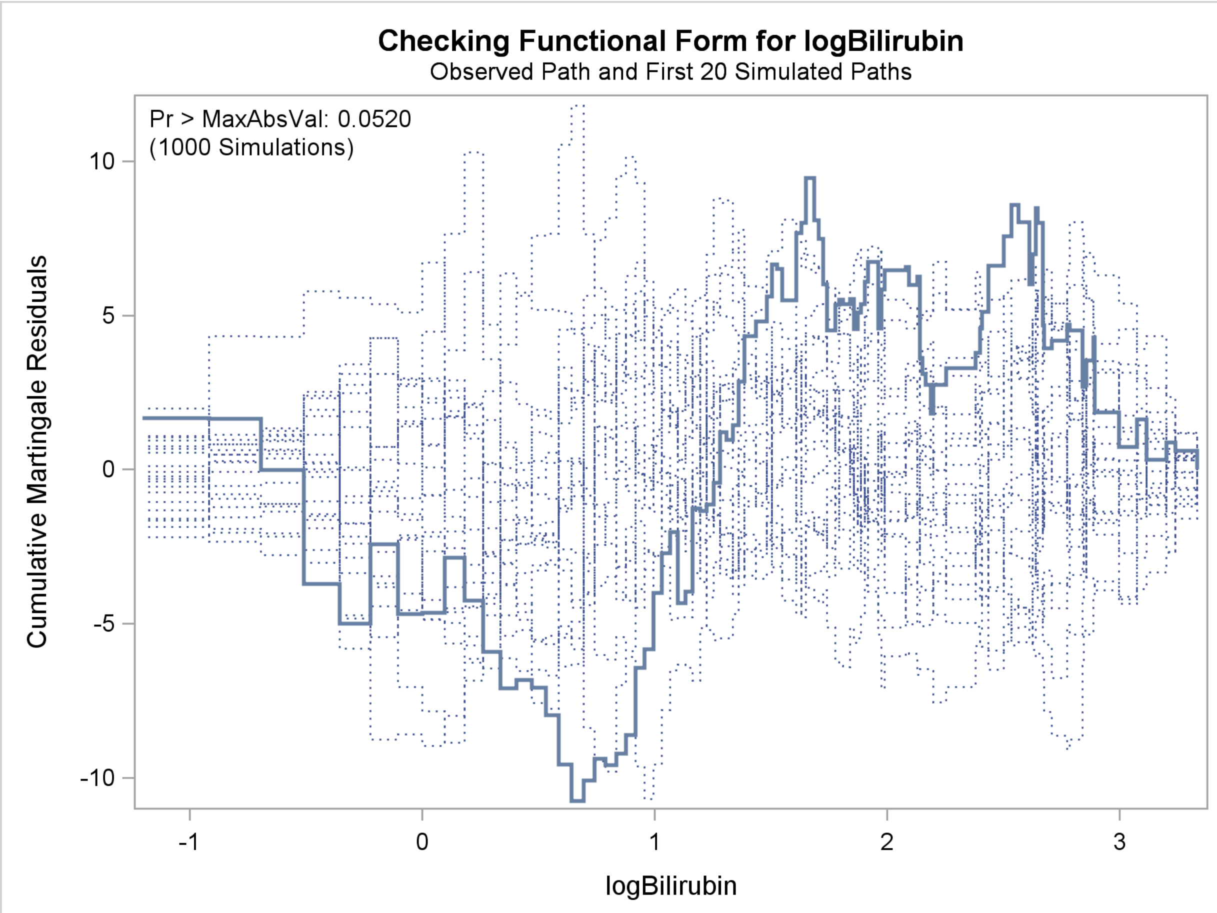
**Analysis of Maximum Likelihood Estimates**

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| logBilirubin | 1 | 0.87072 | 0.08263 | 111.0484 | <.0001 | 2.389 |
| logProtime | 1 | 2.37789 | 0.76674 | 9.6181 | 0.0019 | 10.782 |
| logAlbumin | 1 | -2.53264 | 0.64819 | 15.2664 | <.0001 | 0.079 |
| Age | 1 | 0.03940 | 0.00765 | 26.5306 | <.0001 | 1.040 |
| Edema | 1 | 0.85934 | 0.27114 | 10.0447 | 0.0015 | 2.362 |

**Output 64.12.5** Panel Plot of Cumulative Martingale Residuals versus log(Bilirubin)

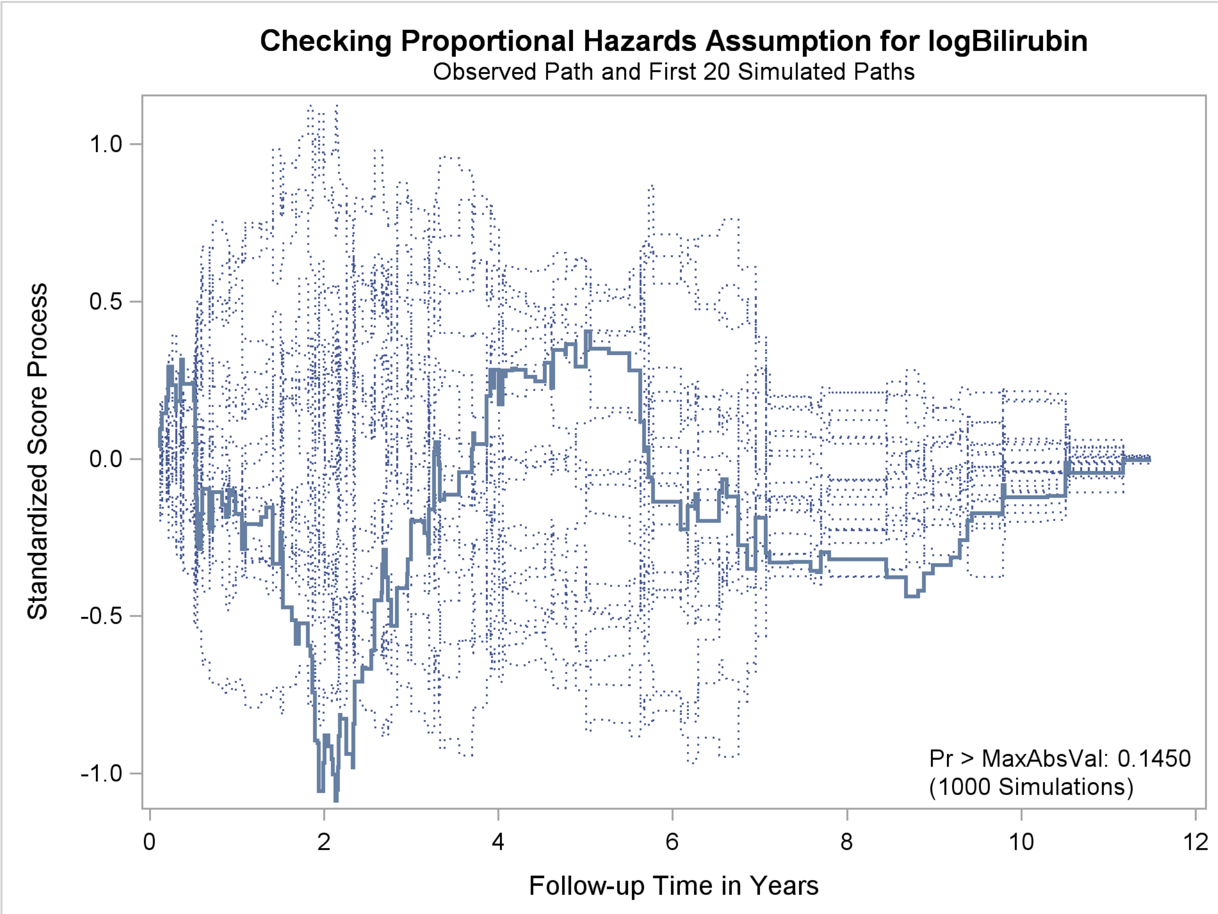


**Output 64.12.6** Cumulative Martingale Residuals versus log(Bilirubin)



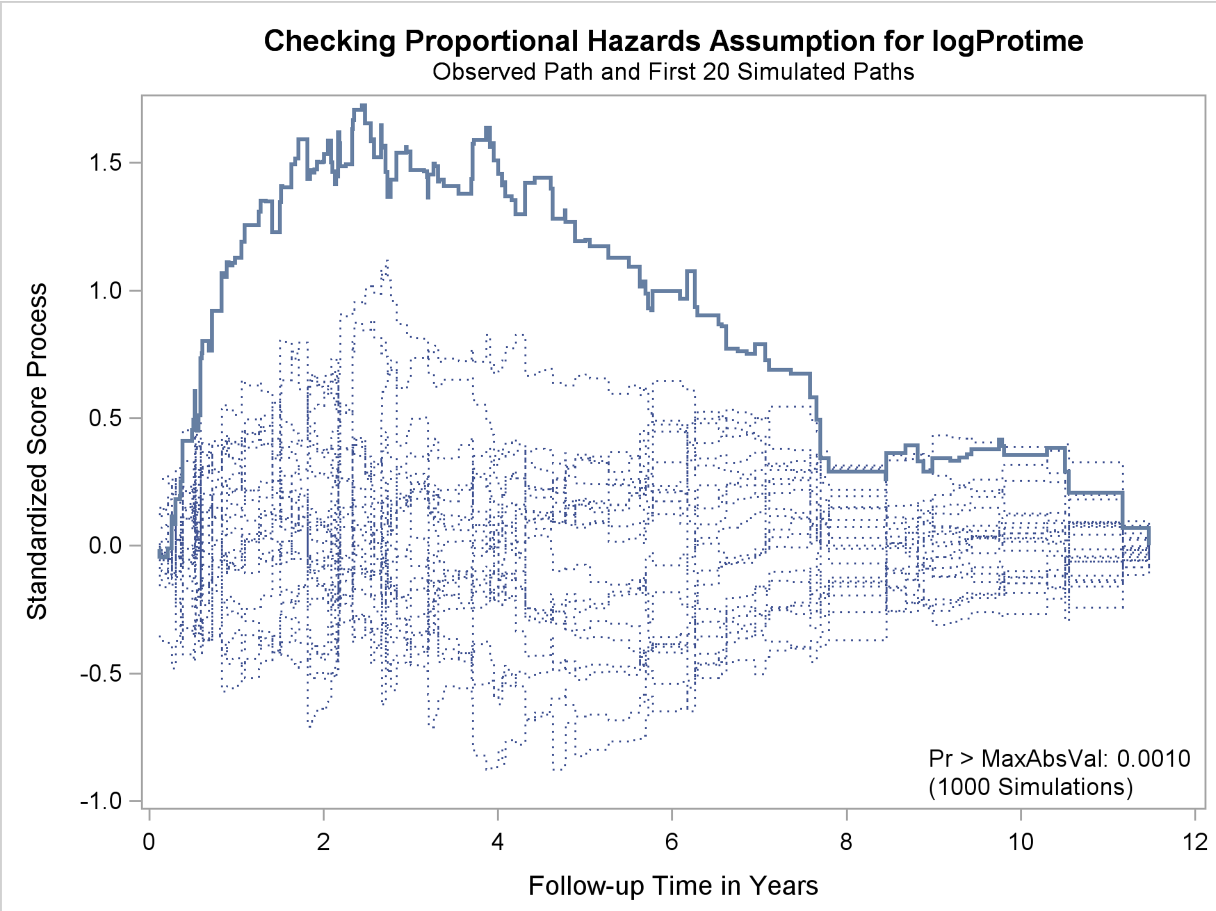

Output 64.12.7 and Output 64.12.8 display the results of proportional hazards assumption assessment for log(Bilirubin) and log(Protime), respectively. The latter plot reveals nonproportional hazards for log(Protime).

**Output 64.12.7** Standardized Score Process for log(Bilirubin)




[

**Output 64.12.8** Standardized Score Process for log(Protime)




Plots for log(Albumin), Age, and Edema are not shown here. The Kolmogorov-type supremum test results for all the covariates are shown in Output 64.12.9. In addition to log(Protime), the proportional hazards assumption appears to be violated for Edema.

**Output 64.12.9** Kolmogorov-Type Supremum Tests for Proportional Hazards Assumption

**Supremum Test for Proportionals Hazards Assumption**

| Variable | Maximum Absolute Value | Replications | Seed | Pr > MaxAbsVal |
|---|---|---|---|---|
| logBilirubin | 1.0880 | 1000 | 19 | 0.1450 |
| logProtime | 1.7243 | 1000 | 19 | 0.0010 |
| logAlbumin | 0.8443 | 1000 | 19 | 0.4330 |
| Age | 0.7387 | 1000 | 19 | 0.4620 |
| Edema | 1.4350 | 1000 | 19 | 0.0330 |

## Example 64.13: Bayesian Analysis of the Cox Model

This example illustrates the use of an informative prior. Hazard ratios, which are transformations of the regression parameters, are useful for interpreting survival models. This example also demonstrates the use of the HAZARDRATIO statement to obtain customized hazard ratios.

Consider the VALung data set in Example 64.3. In this example, the Cox model is used for the Bayesian analysis. The parameters are the coefficients of the continuous explanatory variables (Kps, Duration, and Age) and the coefficients of the design variables for the categorical explanatory variables (Prior, Cell, and Therapy). You use the CLASS statement in PROC PHREG to specify the categorical variables and their reference levels. Using the default reference parameterization, the design variables for the categorical variables are Prioryes (for Prior with Prior='no' as reference), Celladeno, Cellsmall, Cellsquamous (for Cell with Cell='large' as reference), and Therapytest (for Therapy='standard' as reference).

Consider the explanatory variable Kps. The Karnofsky performance scale index enables patients to be classified according to their functional impairment. The scale can range from 0 to 100—0 for dead, and 100 for a normal, healthy person with no evidence of disease. Recall that a flat prior was used for the regression coefficient in the example in the section "Bayesian Analysis" on page 4525. A flat prior on the Kps coefficient implies that the coefficient is as likely to be 0.1 as it is to be –100000. A coefficient of –5 means that a decrease of 20 points in the scale increases the hazard by $e^{-20\times-5}$(=2.68 $\times 10^{43}$)-fold, which is a rather unreasonable and unrealistic expectation for the effect of the Karnofsky index, much less than the value of –100000. Suppose you have a more realistic expectation: the effect is somewhat small and is more likely to be negative than positive, and a decrease of 20 points in the Karnofsky index will change the hazard from 0.9-fold (some minor positive effect) to 4-fold (a large negative effect). You can convert this opinion to a more informative prior on the Kps coefficient $\boldsymbol{\beta}_1$. Mathematically,

$$0.9 < e^{-20\boldsymbol{\beta}_1} < 4$$

which is equivalent to

$$-0.0693 < \boldsymbol{\beta}_1 < 0.0053$$

This becomes the plausible range that you believe the Kps coefficient can take. Now you can find a normal distribution that best approximates this belief by placing the majority of the prior distribution mass within this range. Assuming this interval is $\mu \pm 2\sigma$, where $\mu$ and $\sigma$ are the mean and standard deviation of the normal prior, respectively, the hyperparameters $\mu$ and $\sigma$ are computed as follows:

$$\begin{aligned}\mu &= \frac{-0.0693 + 0.0053}{2} = -0.032 \\ \sigma &= \frac{0.0053 - (-0.0693)}{4} = 0.0186\end{aligned}$$

Note that a normal prior distribution with mean –0.0320 and standard deviation 0.0186 indicates that you believe, before looking at the data, that a decrease of 20 points in the Karnofsky index will probably change the hazard rate by 0.9-fold to 4-fold. This does not rule out the possibility that the Kps coefficient can take a more extreme value such as –5, but the probability of having such extreme values is very small.

Assume the prior distributions are independent for all the parameters. For the coefficient of Kps, you use a normal prior distribution with mean –0.0320 and variance $0.0186^2$(=0.00035). For other parameters, you resort to using a normal prior distribution with mean 0 and variance $10^6$, which is fairly noninformative. Means and variances of these independent normal distributions are saved in the data set Prior as follows:

```
proc format;
   value yesno 0='no' 10='yes';
   run;

data VALung;
   drop check m;
   retain Therapy Cell;
   infile cards column=column;
   length Check $ 1;
   label Time='time to death in days'
         Kps='Karnofsky performance scale'
         Duration='months from diagnosis to randomization'
         Age='age in years'
         Prior='prior therapy'
         Cell='cell type'
         Therapy='type of treatment';
   format Prior yesno.;
   M=Column;
   input Check $ @@;
   if M>Column then M=1;
   if Check='s'|Check='t' then do;
      input @M Therapy $ Cell $;
      delete;
      end;
   else do;
      input @M Time Kps Duration Age Prior @@;
      Status=(Time>0);
      Time=abs(Time);
      end;
   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
;

data Prior;
   input _TYPE_ $ Kps Duration Age Prioryes Celladeno Cellsmall
         Cellsquamous Therapytest;
   datalines;
   Mean -0.0320 0 0 0 0 0 0 0
   Var  0.00035 1e6 1e6 1e6 1e6 1e6 1e6 1e6
   ;
run;
```

In the following BAYES statement, COEFFPRIOR=NORMAL(INPUT=Prior) specifies the normal prior distribution for the regression coefficients with details contained in the data set Prior. Summary statistics of the posterior distribution are produced by default. Autocorrelations and effective sample size are requested as convergence diagnostics as well as the trace plots for visual analysis. For comparisons of hazards, three HAZARDRATIO statements are specified—one for the variable Therapy, one for the variable Age, and one for the variable Cell.

```
ods graphics on;
proc phreg data=VALung;
   class Prior(ref='no') Cell(ref='large') Therapy(ref='standard');
   model Time*Status(0) = Kps Duration Age Prior Cell Therapy;
   bayes seed=1 coeffprior=normal(input=Prior) diagnostic=(autocorr ess)
         plots=trace;
   hazardratio 'Hazard Ratio Statement 1' Therapy;
   hazardratio 'Hazard Ratio Statement 2' Age / unit=10;
   hazardratio 'Hazard Ratio Statement 3' Cell;
run;
ods graphics off;
```

This analysis generates a posterior chain of 10,000 iterations after 2,000 iterations of burn-in, as depicted in Output 64.13.1.

**Output 64.13.1** Model Information

The PHREG Procedure

Bayesian Analysis

Model Information

| | | |
|---|---|---|
| Data Set | WORK.VALUNG | |
| Dependent Variable | Time | time to death in days |
| Censoring Variable | Status | |
| Censoring Value(s) | 0 | |
| Model | Cox | |
| Ties Handling | BRESLOW | |
| Burn-In Size | 2000 | |
| MC Sample Size | 10000 | |
| Thinning | 1 | |

Output 64.13.2 displays the names of the parameters and their corresponding effects and categories.

**Output 64.13.2** Parameter Names

Regression Parameter Information

| Parameter | Effect | Prior | Cell | Therapy |
|---|---|---|---|---|
| Kps | Kps | | | |
| Duration | Duration | | | |
| Age | Age | | | |
| Prioryes | Prior | yes | | |
| Celladeno | Cell | | adeno | |
| Cellsmall | Cell | | small | |
| Cellsquamous | Cell | | squamous | |
| Therapytest | Therapy | | | test |

PROC PHREG computes the maximum likelihood estimates of regression parameters (Output 64.13.3). These estimates are used as the starting values for the simulation of posterior samples.

**Output 64.13.3** Parameter Estimates

Maximum Likelihood Estimates

| Parameter | DF | Estimate | Standard Error | 95% Confidence Limits | |
|---|---|---|---|---|---|
| Kps | 1 | -0.0326 | 0.00551 | -0.0434 | -0.0218 |
| Duration | 1 | -0.00009 | 0.00913 | -0.0180 | 0.0178 |
| Age | 1 | -0.00855 | 0.00930 | -0.0268 | 0.00969 |
| Prioryes | 1 | 0.0723 | 0.2321 | -0.3826 | 0.5273 |
| Celladeno | 1 | 0.7887 | 0.3027 | 0.1955 | 1.3819 |
| Cellsmall | 1 | 0.4569 | 0.2663 | -0.0650 | 0.9787 |
| Cellsquamous | 1 | -0.3996 | 0.2827 | -0.9536 | 0.1544 |
| Therapytest | 1 | 0.2899 | 0.2072 | -0.1162 | 0.6961 |

Output 64.13.4 displays the independent normal prior for the analysis.

**Output 64.13.4** Coefficient Prior

Independent Normal Prior for Regression Coefficients

| Parameter | Mean | Precision |
|---|---|---|
| Kps | -0.032 | 2857.143 |
| Duration | 0 | 1E-6 |
| Age | 0 | 1E-6 |
| Prioryes | 0 | 1E-6 |
| Celladeno | 0 | 1E-6 |
| Cellsmall | 0 | 1E-6 |
| Cellsquamous | 0 | 1E-6 |
| Therapytest | 0 | 1E-6 |

Fit statistics are displayed in Output 64.13.5. These statistics are useful for variable selection.

**Output 64.13.5** Fit Statistics

Fit Statistics

| | |
|---|---|
| DIC (smaller is better) | 966.260 |
| pD (Effective Number of Parameters) | 7.934 |

Summary statistics of the posterior samples are shown in Output 64.13.6 and Output 64.13.7. These results are quite comparable to the classical results based on maximizing the likelihood as shown in Output 64.13.3, since the prior distribution for the regression coefficients is relatively flat.

**Output 64.13.6** Summary Statistics

The PHREG Procedure

Bayesian Analysis

Posterior Summaries

| Parameter | N | Mean | Standard Deviation | Percentiles 25% | 50% | 75% |
|---|---|---|---|---|---|---|
| Kps | 10000 | -0.0326 | 0.00523 | -0.0362 | -0.0326 | -0.0291 |
| Duration | 10000 | -0.00159 | 0.00954 | -0.00756 | -0.00093 | 0.00504 |
| Age | 10000 | -0.00844 | 0.00928 | -0.0147 | -0.00839 | -0.00220 |
| Prioryes | 10000 | 0.0742 | 0.2348 | -0.0812 | 0.0737 | 0.2337 |
| Celladeno | 10000 | 0.7881 | 0.3065 | 0.5839 | 0.7876 | 0.9933 |
| Cellsmall | 10000 | 0.4639 | 0.2709 | 0.2817 | 0.4581 | 0.6417 |
| Cellsquamous | 10000 | -0.4024 | 0.2862 | -0.5927 | -0.4025 | -0.2106 |
| Therapytest | 10000 | 0.2892 | 0.2038 | 0.1528 | 0.2893 | 0.4240 |

**Output 64.13.7** Interval Statistics

Posterior Intervals

| Parameter | Alpha | Equal-Tail Interval | | HPD Interval | |
|---|---|---|---|---|---|
| Kps | 0.050 | -0.0429 | -0.0222 | -0.0433 | -0.0226 |
| Duration | 0.050 | -0.0220 | 0.0156 | -0.0210 | 0.0164 |
| Age | 0.050 | -0.0263 | 0.00963 | -0.0265 | 0.00941 |
| Prioryes | 0.050 | -0.3936 | 0.5308 | -0.3832 | 0.5384 |
| Celladeno | 0.050 | 0.1879 | 1.3920 | 0.1764 | 1.3755 |
| Cellsmall | 0.050 | -0.0571 | 1.0167 | -0.0888 | 0.9806 |
| Cellsquamous | 0.050 | -0.9687 | 0.1635 | -0.9641 | 0.1667 |
| Therapytest | 0.050 | -0.1083 | 0.6930 | -0.1284 | 0.6710 |

With autocorrelations retreating quickly to 0 (Output 64.13.8) and large effective sample sizes (Output 64.13.9), both diagnostics indicate a reasonably good mixing of the Markov chain. The trace plots in Output 64.13.10 also confirm the convergence of the Markov chain.

**Output 64.13.8** Autocorrelation Diagnostics

**The PHREG Procedure**

**Bayesian Analysis**

**Posterior Autocorrelations**

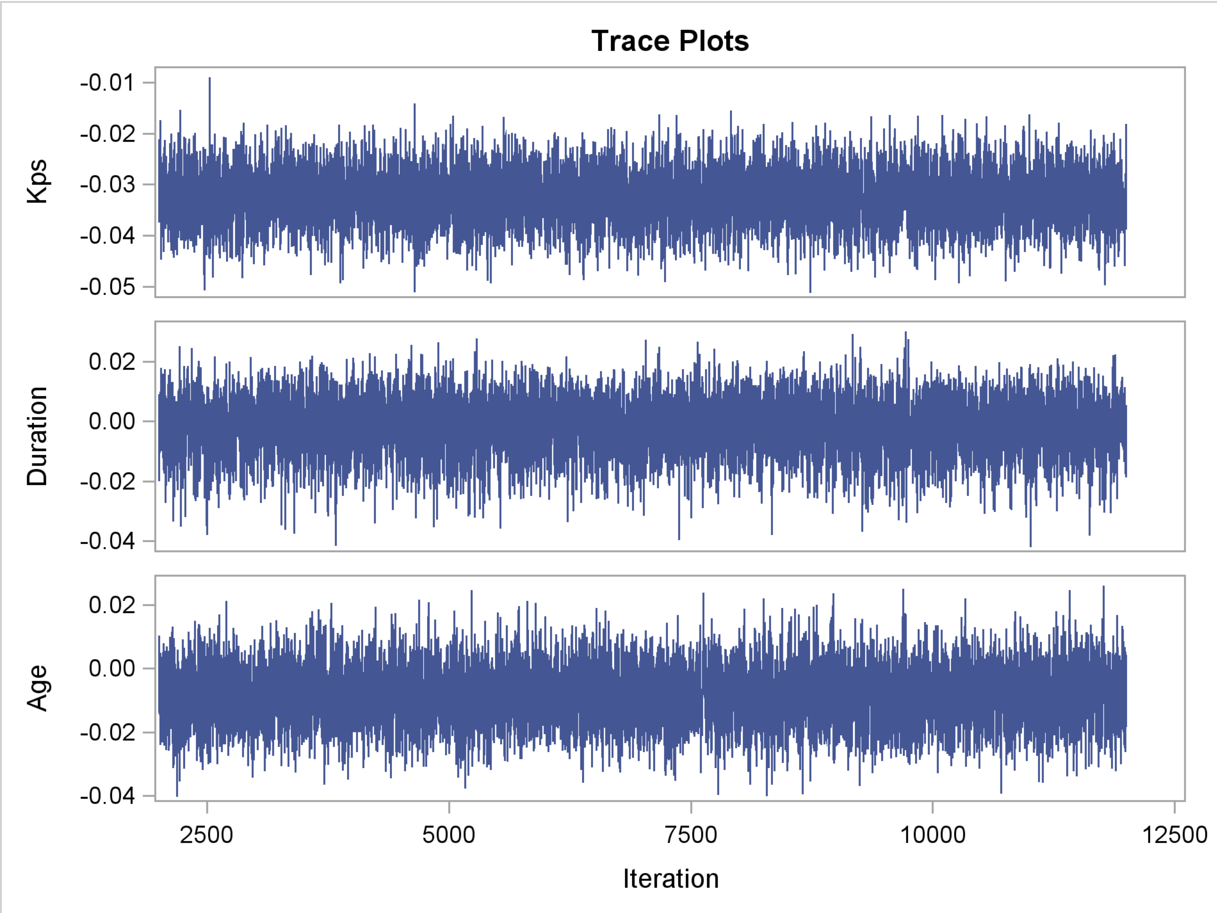
| Parameter | Lag 1 | Lag 5 | Lag 10 | Lag 50 |
|---|---|---|---|---|
| Kps | 0.1442 | -0.0016 | 0.0096 | -0.0013 |
| Duration | 0.2672 | -0.0054 | -0.0004 | -0.0011 |
| Age | 0.1374 | -0.0044 | 0.0129 | 0.0084 |
| Prioryes | 0.2507 | -0.0271 | -0.0012 | 0.0004 |
| Celladeno | 0.4160 | 0.0265 | -0.0062 | 0.0190 |
| Cellsmall | 0.5055 | 0.0277 | -0.0011 | 0.0271 |
| Cellsquamous | 0.3586 | 0.0252 | -0.0044 | 0.0107 |
| Therapytest | 0.2063 | 0.0199 | -0.0047 | -0.0166 |

**Output 64.13.9** Effective Sample Size Diagnostics

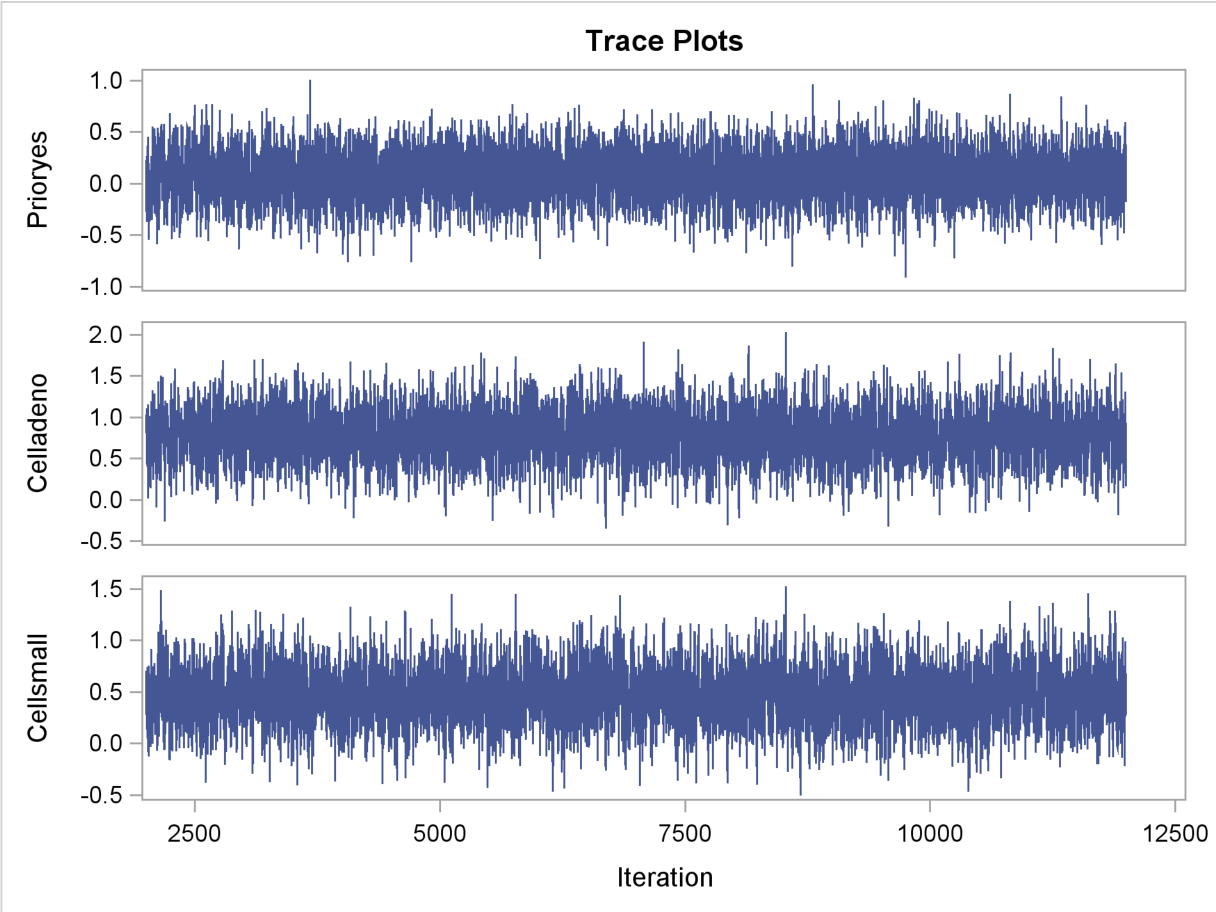
**Effective Sample Sizes**

| Parameter | ESS | Correlation Time | Efficiency |
|---|---|---|---|
| Kps | 7046.7 | 1.4191 | 0.7047 |
| Duration | 5790.0 | 1.7271 | 0.5790 |
| Age | 7426.1 | 1.3466 | 0.7426 |
| Prioryes | 6102.2 | 1.6388 | 0.6102 |
| Celladeno | 3673.4 | 2.7223 | 0.3673 |
| Cellsmall | 3346.4 | 2.9883 | 0.3346 |
| Cellsquamous | 4052.8 | 2.4674 | 0.4053 |
| Therapytest | 6870.8 | 1.4554 | 0.6871 |

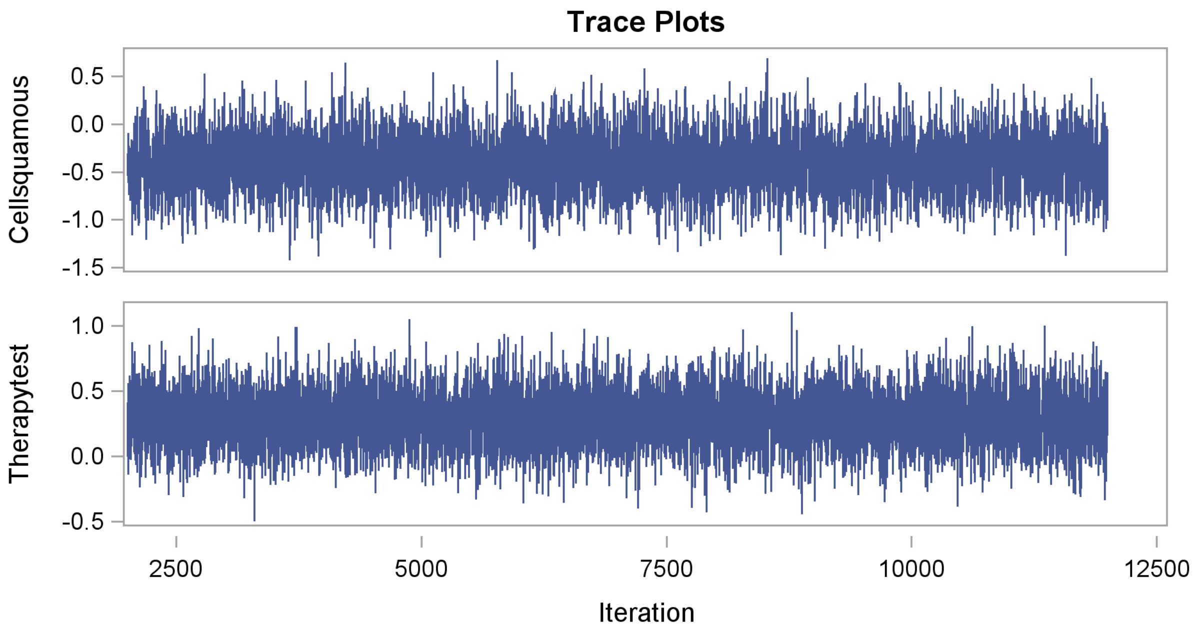
**Output 64.13.10** Trace Plots



**Output 64.13.10** *continued*



**Output 64.13.10** *continued*




The first HAZARDRATIO statement compares the hazards between the standard therapy and the test therapy. Summaries of the posterior distribution of the corresponding hazard ratio are shown in Output 64.13.11. There is a 95% chance that the hazard ratio of standard therapy versus test therapy lies between 0.5 and 1.1.

**Output 64.13.11** Hazard Ratio for Treatment

**Hazard Ratio Statement 1: Hazard Ratios for Therapy**

| Description | N | Mean | Standard Deviation |
|---|---|---|---|
| Therapy standard vs test | 10000 | 0.7645 | 0.1573 |

**Hazard Ratio Statement 1: Hazard Ratios for Therapy**

| Quantiles 25% | Quantiles 50% | Quantiles 75% | 95% Equal-Tail Interval | | 95% HPD Interval | |
|---|---|---|---|---|---|---|
| 0.6544 | 0.7488 | 0.8583 | 0.5001 | 1.1143 | 0.4788 | 1.0805 |

The second HAZARDRATIO statement assesses the change of hazards for an increase in Age of 10 years. Summaries of the posterior distribution of the corresponding hazard ratio are shown in Output 64.13.12.

**Output 64.13.12** Hazard Ratio for Age

**Hazard Ratio Statement 2: Hazard Ratios for Age**

| Description | N | Mean | Standard Deviation | Quantiles 25% | 50% | 75% |
|---|---|---|---|---|---|---|
| Age Unit=10 | 10000 | 0.9230 | 0.0859 | 0.8635 | 0.9195 | 0.9782 |

**Hazard Ratio Statement 2: Hazard Ratios for Age**

| 95% Equal-Tail Interval | | 95% HPD Interval | |
|---|---|---|---|
| 0.7685 | 1.1011 | 0.7650 | 1.0960 |

The third HAZARDRATIO statement compares the changes of hazards between two types of cells. For four types of cells, there are six different pairs of cell comparisons. The results are shown in Output 64.13.13.

**Output 64.13.13** Hazard Ratios for Cell

**Hazard Ratio Statement 3: Hazard Ratios for Cell**

| Description | N | Mean | Standard Deviation |
|---|---|---|---|
| Cell adeno vs large | 10000 | 2.3048 | 0.7224 |
| Cell adeno vs small | 10000 | 1.4377 | 0.4078 |
| Cell adeno vs squamous | 10000 | 3.4449 | 1.0745 |
| Cell large vs small | 10000 | 0.6521 | 0.1780 |
| Cell large vs squamous | 10000 | 1.5579 | 0.4548 |
| Cell small vs squamous | 10000 | 2.4728 | 0.7081 |

**Hazard Ratio Statement 3: Hazard Ratios for Cell**

| Quantiles 25% | 50% | 75% | 95% Equal-Tail Interval | | 95% HPD Interval | |
|---|---|---|---|---|---|---|
| 1.7929 | 2.1982 | 2.7000 | 1.2067 | 4.0227 | 1.0053 | 3.7057 |
| 1.1522 | 1.3841 | 1.6704 | 0.7930 | 2.3999 | 0.7309 | 2.2662 |
| 2.6789 | 3.2941 | 4.0397 | 1.8067 | 5.9727 | 1.6303 | 5.5946 |
| 0.5264 | 0.6325 | 0.7545 | 0.3618 | 1.0588 | 0.3331 | 1.0041 |
| 1.2344 | 1.4955 | 1.8089 | 0.8492 | 2.6346 | 0.7542 | 2.4575 |
| 1.9620 | 2.3663 | 2.8684 | 1.3789 | 4.1561 | 1.2787 | 3.9263 |

## Example 64.14: Bayesian Analysis of Piecewise Exponential Model

This example illustrates using a piecewise exponential model in a Bayesian analysis. Consider the Rats data set in the section "Getting Started: PHREG Procedure" on page 4520. In the following statements, PROC PHREG is used to carry out a Bayesian analysis for the piecewise exponential model. In the BAYES statement, the option PIECEWISE stipulates a piecewise exponential model, and PIECEWISE=HAZARD requests that the constant hazards be modeled in the original scale. By default, eight intervals of constant hazards are used, and the intervals are chosen such that each has roughly the same number of events.

```
   data Rats;
      label Days  ='Days from Exposure to Death';
      input Days Status Group @@;
      datalines;
   143 1 0   164 1 0   188 1 0   188 1 0
   190 1 0   192 1 0   206 1 0   209 1 0
   213 1 0   216 1 0   220 1 0   227 1 0
   230 1 0   234 1 0   246 1 0   265 1 0
   304 1 0   216 0 0   244 0 0   142 1 1
   156 1 1   163 1 1   198 1 1   205 1 1
   232 1 1   232 1 1   233 1 1   233 1 1
   233 1 1   233 1 1   239 1 1   240 1 1
   261 1 1   280 1 1   280 1 1   296 1 1
   296 1 1   323 1 1   204 0 1   344 0 1
   ;
   run;

proc phreg data=Rats;
   model Days*Status(0)=Group;
   bayes seed=1 piecewise=hazard;
run;
```

The "Model Information" table in Output 64.14.1 shows that the piecewise exponential model is being used.

**Output 64.14.1** Model Information

```
                              The PHREG Procedure

                               Bayesian Analysis

                               Model Information

 Data Set                    WORK.RATS
 Dependent Variable          Days                       Days from Exposure to Death
 Censoring Variable          Status
 Censoring Value(s)          0
 Model                       Piecewise Exponential
 Burn-In Size                2000
 MC Sample Size              10000
 Thinning                    1
```

By default the time axis is partitioned into eight intervals of constant hazard. Output 64.14.2 details the number of events and observations in each interval. Note that the constant hazard parameters are named Lambda1,..., Lambda8. You can supply your own partition by using the INTERVALS= suboption within the PIECEWISE=HAZARD option.

**Output 64.14.2** Interval Partition

**Constant Hazard Time Intervals**

| Interval [Lower, | Upper) | N | Event | Hazard Parameter |
|---|---|---|---|---|
| 0 | 176 | 5 | 5 | Lambda1 |
| 176 | 201.5 | 5 | 5 | Lambda2 |
| 201.5 | 218 | 7 | 5 | Lambda3 |
| 218 | 232.5 | 5 | 5 | Lambda4 |
| 232.5 | 233.5 | 4 | 4 | Lambda5 |
| 233.5 | 253.5 | 5 | 4 | Lambda6 |
| 253.5 | 288 | 4 | 4 | Lambda7 |
| 288 | Infty | 5 | 4 | Lambda8 |

The model parameters consist of the eight hazard parameters Lambda1, ..., Lambda8, and the regression coefficient Group. The maximum likelihood estimates are displayed in Output 64.14.3. Again, these estimates are used as the starting values for simulation of the posterior distribution.

**Output 64.14.3** Maximum Likelihood Estimates

**Maximum Likelihood Estimates**

| Parameter | DF | Estimate | Standard Error | 95% Confidence Limits | |
|---|---|---|---|---|---|
| Lambda1 | 1 | 0.000953 | 0.000443 | 0.000084 | 0.00182 |
| Lambda2 | 1 | 0.00794 | 0.00371 | 0.000672 | 0.0152 |
| Lambda3 | 1 | 0.0156 | 0.00734 | 0.00120 | 0.0300 |
| Lambda4 | 1 | 0.0236 | 0.0115 | 0.00112 | 0.0461 |
| Lambda5 | 1 | 0.3669 | 0.1959 | -0.0172 | 0.7509 |
| Lambda6 | 1 | 0.0276 | 0.0148 | -0.00143 | 0.0566 |
| Lambda7 | 1 | 0.0262 | 0.0146 | -0.00233 | 0.0548 |
| Lambda8 | 1 | 0.0545 | 0.0310 | -0.00626 | 0.1152 |
| Group | 1 | -0.6223 | 0.3468 | -1.3020 | 0.0573 |

Without using the PRIOR= suboption within the PIECEWISE=HAZARD option to specify the prior of the hazard parameters, the default is to use the noninformative and improper prior displayed in Output 64.14.4.

**Output 64.14.4** Hazard Prior

**Improper Prior for Hazards**

| Parameter | Prior |
|---|---|
| Lambda1 | 1 / Lambda1 |
| Lambda2 | 1 / Lambda2 |
| Lambda3 | 1 / Lambda3 |
| Lambda4 | 1 / Lambda4 |
| Lambda5 | 1 / Lambda5 |
| Lambda6 | 1 / Lambda6 |
| Lambda7 | 1 / Lambda7 |
| Lambda8 | 1 / Lambda8 |

The noninformative uniform prior is used for the regression coefficient Group (Output 64.14.5), as in the section "Bayesian Analysis" on page 4525.

**Output 64.14.5** Coefficient Prior

**Uniform Prior for Regression Coefficients**

| Parameter | Prior |
|---|---|
| Group | Constant |

Summary statistics for all model parameters are shown in Output 64.14.6 and Output 64.14.7.

**Output 64.14.6** Summary Statistics

**The PHREG Procedure**

**Bayesian Analysis**

**Posterior Summaries**

| Parameter | N | Mean | Standard Deviation | Percentiles 25% | Percentiles 50% | Percentiles 75% |
|---|---|---|---|---|---|---|
| Lambda1 | 10000 | 0.000945 | 0.000444 | 0.000624 | 0.000876 | 0.00118 |
| Lambda2 | 10000 | 0.00782 | 0.00363 | 0.00519 | 0.00724 | 0.00979 |
| Lambda3 | 10000 | 0.0155 | 0.00735 | 0.0102 | 0.0144 | 0.0195 |
| Lambda4 | 10000 | 0.0236 | 0.0116 | 0.0152 | 0.0217 | 0.0297 |
| Lambda5 | 10000 | 0.3634 | 0.1965 | 0.2186 | 0.3266 | 0.4685 |
| Lambda6 | 10000 | 0.0278 | 0.0153 | 0.0166 | 0.0249 | 0.0356 |
| Lambda7 | 10000 | 0.0265 | 0.0151 | 0.0157 | 0.0236 | 0.0338 |
| Lambda8 | 10000 | 0.0558 | 0.0323 | 0.0322 | 0.0488 | 0.0721 |
| Group | 10000 | -0.6154 | 0.3570 | -0.8569 | -0.6186 | -0.3788 |

**Output 64.14.7** Interval Statistics

Posterior Intervals

| Parameter | Alpha | Equal-Tail Interval | | HPD Interval | |
|---|---|---|---|---|---|
| Lambda1 | 0.050 | 0.000289 | 0.00199 | 0.000208 | 0.00182 |
| Lambda2 | 0.050 | 0.00247 | 0.0165 | 0.00194 | 0.0152 |
| Lambda3 | 0.050 | 0.00484 | 0.0331 | 0.00341 | 0.0301 |
| Lambda4 | 0.050 | 0.00699 | 0.0515 | 0.00478 | 0.0462 |
| Lambda5 | 0.050 | 0.0906 | 0.8325 | 0.0541 | 0.7469 |
| Lambda6 | 0.050 | 0.00676 | 0.0654 | 0.00409 | 0.0580 |
| Lambda7 | 0.050 | 0.00614 | 0.0648 | 0.00421 | 0.0569 |
| Lambda8 | 0.050 | 0.0132 | 0.1368 | 0.00637 | 0.1207 |
| Group | 0.050 | -1.3190 | 0.0893 | -1.3379 | 0.0652 |

The default diagnostics—namely, lag1, lag5, lag10, lag50 autocorrelations (Output 64.14.8), the Geweke diagnostics (Output 64.14.9), and the effective sample size diagnostics (Output 64.14.10)—show a good mixing of the Markov chain.

**Output 64.14.8** Autocorrelations

The PHREG Procedure

Bayesian Analysis

Posterior Autocorrelations

| Parameter | Lag 1 | Lag 5 | Lag 10 | Lag 50 |
|---|---|---|---|---|
| Lambda1 | 0.0705 | 0.0015 | 0.0017 | -0.0076 |
| Lambda2 | 0.0909 | 0.0206 | -0.0013 | -0.0039 |
| Lambda3 | 0.0861 | -0.0072 | 0.0011 | 0.0002 |
| Lambda4 | 0.1447 | -0.0023 | 0.0081 | 0.0082 |
| Lambda5 | 0.1086 | 0.0072 | -0.0038 | -0.0028 |
| Lambda6 | 0.1281 | 0.0049 | -0.0036 | 0.0048 |
| Lambda7 | 0.1925 | -0.0011 | 0.0094 | -0.0011 |
| Lambda8 | 0.2128 | 0.0322 | -0.0042 | -0.0045 |
| Group | 0.5638 | 0.0410 | -0.0003 | -0.0071 |

**Output 64.14.9** Geweke Diagnostics

**Geweke Diagnostics**

| Parameter | z | Pr > \|z\| |
|---|---|---|
| Lambda1 | -0.0705 | 0.9438 |
| Lambda2 | -0.4936 | 0.6216 |
| Lambda3 | 0.5751 | 0.5652 |
| Lambda4 | 1.0514 | 0.2931 |
| Lambda5 | 0.8910 | 0.3729 |
| Lambda6 | 0.2976 | 0.7660 |
| Lambda7 | 1.6543 | 0.0981 |
| Lambda8 | 0.6686 | 0.5038 |
| Group | -1.2621 | 0.2069 |

**Output 64.14.10** Effective Sample Size

**Effective Sample Sizes**

| Parameter | ESS | Correlation Time | Efficiency |
|---|---|---|---|
| Lambda1 | 7775.3 | 1.2861 | 0.7775 |
| Lambda2 | 6874.8 | 1.4546 | 0.6875 |
| Lambda3 | 7655.7 | 1.3062 | 0.7656 |
| Lambda4 | 6337.1 | 1.5780 | 0.6337 |
| Lambda5 | 6563.3 | 1.5236 | 0.6563 |
| Lambda6 | 6720.8 | 1.4879 | 0.6721 |
| Lambda7 | 5968.7 | 1.6754 | 0.5969 |
| Lambda8 | 5137.2 | 1.9466 | 0.5137 |
| Group | 2980.4 | 3.3553 | 0.2980 |

# References


Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1992), *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.

Andersen, P. K. and Gill, R. D. (1982), "Cox's Regression Model Counting Process: A Large Sample Study," *Annals of Statistics*, 10, 1100–1120.

Binder, D. A. (1992), "Fitting Cox's Proportional Hazards Models from Survey Data," *Biometrika*, 79, 139–147.

Breslow, N. E. (1972), "Discussion of Professor Cox's Paper," *J. Royal Stat. Soc. B*, 34, 216–217.

Breslow, N. E. (1974), "Covariance Analysis of Censored Survival Data," *Biometrics*, 30, 89–99.

Bryson, M. C. and Johnson, M. E. (1981), "The Incidence of Monotone Likelihood in the Cox Model," *Technometrics*, 23(4), 381–383.

Cain, K. C. and Lange, N. T. (1984), "Approximate Case Influence for the Proportional Hazards Regression Model with Censored Data," *Biometrics*, 40, 493–499.

Cox, D. R. (1972), "Regression Models and Life Tables," *Journal of the Royal Statistical Society, Series B*, 20, 187–220, with discussion.

Cox, D. R. (1975), "Partial Likelihood," *Biometrika*, 62, 269–276.

Crowley, J. and Hu, M. (1977), "Covariance Analysis of Heart Transplant Survival Data," *Journal of the American Statistical Association*, 72, 27–36.

DeLong, D. M., Guirguis, G. H., and So, Y. C. (1994), "Efficient Computation of Subset Selection Probabilities with Application to Cox Regression," *Biometrika*, 81, 607–611.

Efron, B. (1977), "The Efficiency of Cox's Likelihood Function for Censored Data," *Journal of the American Statistical Association*, 72, 557–565.

Firth, D. (1993), "Bias Reduction of Maximum Likelihood Estimates," *Biometrika*, 80, 27–38.

Fleming, T. R. and Harrington, D. (1991), *Counting Processes and Survival Analysis*, New York: John Wiley & Sons.

Furnival, G. M. and Wilson, R. W. (1974), "Regression by Leaps and Bounds," *Technometrics*, 16, 499–511.

Gail, M. H., Lubin, J. H., and Rubinstein, L. V. (1981), "Likelihood Calculations for Matched Case-Control Studies and Survival Studies with Tied Death Times," *Biometrika*, 68, 703–707.

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995), "Adaptive Rejection Metropolis Sampling with Gibbs Sampling," *Applied Statistics*, 44, 455–472.

Gilks, W. R. and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41, 337–348.

Grambsch, P. M. and Therneau, T. M. (1994), "Proportional Hazards Tests and Diagnostics Based on Weighted Residuals," *Biometrika*, 81, 515–526.

Heinze, G. (1999), *The Application of Firth's Procedure to Cox and Logistic Regression*, Technical Report 10/1999, update in January 2001, Section of Clinical Biometrics, Department of Medical Computer Sciences, University of Vienna.

Heinze, G. and Schemper, M. (2001), "A Solution to the Problem of Monotone Likelihood in Cox Regression," *Biometrics*, 51, 114–119.

Hosmer, D. W., Jr. and Lemeshow, S. (1989), *Applied Logistic Regression*, New York: John Wiley & Sons.

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001), *Bayesian Survival Analysis*, New York: Springer-Verlag.

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

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. (1998), "Markov Chain Monte Carlo in Practice: A Roundtable Discussion," *The American Statistician*, 52, 93–100.

Krall, J. M., Uthoff, V. A., and Harley, J. B. (1975), "A Step-up Procedure for Selecting Variables Associated with Survival," *Biometrics*, 31, 49–57.

Lawless, J. F. (2003), *Statistical Model and Methods for Lifetime Data*, Second Edition, New York: John Wiley & Sons.

Lawless, J. F. and Nadeau, C. (1995), "Some Simple Robust Methods for the Analysis of Recurrent Events," *Technometrics*, 37, 158–168.

Lee, E. W., Wei, L. J., and Amato, D. (1992), "Cox-Type Regression Analysis for Large Numbers of Small Groups of Correlated Failure Time Observations," 237–247, Netherlands: Kluwer Academic.

Lin, D., Wei, L. J., Yang, I., and Ying, Z. (2000), "Semiparametric Regression for the Mean and Rate Functions of Recurrent Events," *Journal of the Royal Statistical Society, Series B*, 62, 711–730.

Lin, D., Wei, L. J., and Ying, Z. (1993), "Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals," *Biometrika*, 80, 557–572.

Lin, D. Y. (1994), "Cox Regression Analysis of Multivariate Failure Time Data: The Marginal Approach," *Statistics in Medicine*, 13, 2233–2247.

Lin, D. Y. and Wei, L. J. (1989), "The Robust Inference for the Proportional Hazards Model," *Journal of the American Statistical Association*, 84, 1074–1078.

Nelson, W. (2002), *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications*, ASA-SIAM Series on Statistics and Applied Probability.

Pepe, M. S. and Cai, J. (1993), "Some Graphical Displays and Marginal Regression Analyses for Recurrent Failure Times and Time Dependent Covariates," *Journal of the American Statistical Association*, 88, 881–820.

Pettitt, A. N. and Bin Daud, I. (1989), "Case-Weighted Measures of Influence for Proportional Hazards Regression," *Applied Statistics*, 38, 313–329.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981), "On the Regression Analysis of Multivariate Failure Time Data," *Biometrika*, 68, 373–379.

Reid, N. and Crèpeau, H. (1985), "Influence Functions for Proportional Hazards Regression," *Biometrika*, 72, 1–9.

Sinha, D., Ibrahim, J. G., and Chen, M. H. (2003), "A Bayesian Justification of Cox's Partial Likelihood," *Biometrika*, 90, 629–641.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society, Series B*, 64(4), 583–616, with discussion.

Spiekerman, C. F. and Lin, D. Y. (1998), "Marginal Regression Models for Multivariate Failure Time Data," *Journal of American Statistical Association*, 93, 1164–1175.

Therneau, T. M. (1994), *A Package for Survival Analysis in S*, Technical Report 53, Section of Biostatistics, Mayo Clinic, Rochester, MN.

Therneau, T. M. and Grambsch, P. M. (2000), *Modeling Survival Data: Extending the Cox Model*, New York: Springer-Verlag.

Tsiatis, A. (1981), "A Large Sample Study of the Estimates for the Integrated Hazard Function in Cox's Regression Model for Survival Data," *Annals of Statistics*, 9, 93–108.

Venzon, D. J. and Moolgavkar, S. H. (1988), "A Method for Computing Profile-Likelihood Based Confidence Intervals," *Applied Statistics*, 37, 87–94.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989), "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distribution," *Journal of the American Statistical Association*, 84, 1065–1073.

# Subject Index

# Syntax Index

## Your Turn

We welcome your feedback.

- If you have comments about this book, please send them to **yourturn@sas.com**. Include the full title and page numbers (if applicable).
- If you have comments about the software, please send them to **suggest@sas.com**.

# SAS® Publishing Delivers!

**Whether you are new to the work force or an experienced professional, you need to distinguish yourself in this rapidly changing and competitive job market. SAS® Publishing provides you with a wide range of resources to help you set yourself apart. Visit us online at support.sas.com/bookstore.**

## SAS® Press

Need to learn the basics? Struggling with a programming problem? You'll find the expert answers that you need in example-rich books from SAS Press. Written by experienced SAS professionals from around the world, SAS Press books deliver real-world insights on a broad range of topics for all skill levels.

**support.sas.com/saspress**

## SAS® Documentation

To successfully implement applications using SAS software, companies in every industry and on every continent all turn to the one source for accurate, timely, and reliable information: SAS documentation. We currently produce the following types of reference documentation to improve your work experience:

- Online help that is built into the software.
- Tutorials that are integrated into the product.
- Reference documentation delivered in HTML and PDF – **free** on the Web.
- Hard-copy books.

**support.sas.com/publishing**

## SAS® Publishing News

Subscribe to SAS Publishing News to receive up-to-date information about all new SAS titles, author podcasts, and new Web site features via e-mail. Complete instructions on how to subscribe, as well as access to past issues, are available at our Web site.

**support.sas.com/spn**

THE POWER TO KNOW.