

Paper 253-2010

## An Overview of the CLASS, CONTRAST, and HAZARDRATIO Statements in the SAS® 9.2 PHREG Procedure

Paul T. Savarese and Michael J. Patetta, SAS Institute Inc., Cary, NC

### ABSTRACT

The PHREG procedure fits a number of models collectively known as Cox regression models, including the well-known Cox proportional hazards model. This paper provides an overview of several new features, including three new statements (CLASS, CONTRAST, and HAZARDRATIO) in PROC PHREG. The emphasis is on illustrative examples of comparisons for main effects and interaction models via the new statements. The paper concentrates on common CLASS variable parameterization methods such as reference coding and GLM coding. Caveats regarding CLASS variables and time (including time-dependent covariates) are also discussed. This paper is intended for an intermediate-level audience that has some familiarity with Cox regression models and PROC PHREG.

### INTRODUCTION

PROC PHREG fits Cox regression models, including the well-known Cox proportional hazards (PH) model from which the procedure derives its name. A typical formulation of the Cox PH model of the hazard function  $h(t)$  is as follows:

$$h_i(t) = h_0(t) \exp(X_i'\beta)$$

In this equation,  $h_0(t)$  is an unknown and unspecified baseline hazard.  $X_i$  is a vector of explanatory variables (often called *covariates*) for the  $i$ th individual, and  $\beta$  is a vector of unknown regression coefficients.  $X_i'\beta$  is also known as the **linear predictor** of the form  $\beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$ . To estimate  $\beta$ , Cox (1972, 1975) introduced the **partial likelihood function**, which eliminates the unknown baseline  $h_0(t)$  and accounts for censored survival times. Cox's proportional hazards model is widely used in the analysis of survival data to explain the effect of the explanatory variables on hazard rates. In the analysis of the proportional hazards model, the hazard ratios (HRs) that are associated with each effect in the model are of particular interest.

In SAS 9.2, PROC PHREG has undergone significant additions, not the least of which is the new CLASS, CONTRAST, and HAZARDRATIO statements. These statements aid in modeling categorical variables, specifying interactions, testing hypothesis, and generating custom hazard ratios. The HAZARDRATIO statement is particularly well suited for producing hazard ratios for interaction models. This paper discusses each of these statements, including selected elements of their syntax and aspects of their use. Code examples are also illustrated, with a primary concentration on the construction of hazard ratios. The paper also presents results from various data-analytic scenarios. The next section introduces an acute myocardial infarction data set that is used throughout the remainder of this paper.

### ACUTE MYOCARDIAL INFARCTION DATA

A study was conducted to examine heart attacks among patients that were admitted to hospitals in the Worcester, Massachusetts metropolitan area. The main goal of this study was to examine the survival times of patients following hospital admission for acute myocardial infarction (AMI). This paper uses a subset of the data, taken from an example in Hosmer and Lemshew (1999), to illustrate the use of the CLASS, CONTRAST, and HAZARDRATIO statements. In the examples, this data set is referred to as the AMI data set.

The variables in the AMI data set are as follows:

- **DAYS**—survival time in days following hospital admission for an AMI
- **STATUS**—a censoring indicator (0=alive (censored); 1=died (event))
- **AGE**—age in years upon hospital admission
- **GROUP**—the group (A, B, or C) into which each AMI is categorized (Group A=Q-wave; Group B=Not Q-wave; Group C=Indeterminate)
- **SEX**—the sex (**F**emale or **M**ale) of the patient

## THE CLASS STATEMENT

Beginning with SAS 9.2, the CLASS statement is available in PROC PHREG and enables convenient handling of categorical variables. The statement's syntax and usage are similar to those for the CLASS statement in the LOGISTIC procedure, although with different defaults. The CLASS statement, in general, makes it unnecessary for you to manually code dummy variables to represent the levels of categorical variables. However, there are still cases in which you might want to code dummy variables.

### CLASS STATEMENT SYNTAX

The CLASS statement, which must precede the MODEL statement in PROC PHREG, names the categorical variables that you want to use in your analysis. The syntax for the CLASS statement is as follows:

```
CLASS variable <(individual options)> <variable <(individual options)> . . . > </ global options >;
```

The specific CLASS variable options that are used in the PROC PHREG examples in this paper are as follows:

- **ORDER=DATA | FORMATTED | FREQ | INTERNAL**—specifies the sorting order for the levels of the categorical variables and determines which parameters correspond to each category in the data. ORDER=FORMATTED is the default option setting.
- **PARAM=keyword**—specifies the dummy variable coding scheme (*parameterization method*) that is used to create a design matrix of values that represent the levels of the CLASS variable. Valid values for *keyword* are as follows:
 

○ GLM	○ ORTHEFFECT
○ REFERENCE   REF	○ ORTHORDINAL
○ EFFECT	○ ORTHPOLY
○ ORDINAL	○ ORTHREF
○ POLYNOMIAL   POLY	

PARAM=REF is the default option setting.

- **REF='level' | keyword**—specifies the reference level for the hazard ratios of the CLASS variables that have PARAM=EFFECT or PARAM=REF coding schemes. For an individual variable option, which is listed in parenthesis after a CLASS variable, you can use the REF= option to specify an explicit *level* for the reference category. Also, for either an individual variable option or a global option—after the forward slash (/)—you can use either the FIRST or the LAST keyword.
  - FIRST selects the first-ordered category as the reference level.
  - LAST selects the last-ordered category as the reference level.

For details about all of the options that are available in the CLASS statement, see "The PHREG Procedure" in the SAS/STAT® 9.2 User's Guide, Second Edition (SAS Institute Inc. 2009).

### CLASS Statement Variable Parameterization Methods: GLM and REF Coding

Both PARAM=GLM and PARAM=REF cause PROC PHREG to construct a design matrix of binary zero-one (0/1) variables to represent the levels of the categorical variables that are listed in the CLASS statement. The following two examples illustrate this concept using the GLM coding method (PARAM=GLM) and the REF coding method (PARAM=REF).

**Example 1: PARAM=GLM**

```
proc phreg data=sasuser.ami;
  class group sex / param=glm order=internal;
  model days*status(0) = group sex group*sex / ties=Efron;
run;
```

This PHREG procedure generates the following Class Level Information table:

Class Level Information				
Class	Value	Design Variables		
group	A	1	0	0
	B	0	1	0
	C	0	0	1
sex	Female	1	0	
	Male	0	1	

For PARAM=GLM, you do NOT control the reference category by virtue of the REF= option. Instead, the last ordered category is always used as the reference category. Thus, to change the reference category with GLM coding, you must recode or format the data so that the desired reference level is the last sorted level. In GLM coding, for a CLASS variable with *c* levels, the design matrix has *c* columns that are over-parameterized (not full rank). In the Parameter Estimates table (not shown here) that PROC PHREG creates by default, the beta coefficients estimate the difference in the effect of each non-reference level compared to the reference (last) level. The last level is also referred to as the *omitted level*, and it appears with degrees-of-freedom (DF)=0 and Parameter Estimate=0.

**Example 2: PARAM=REF**

```
proc phreg data=sasuser.ami;
  class group(ref='C') sex(ref='Female') / param=ref order=internal;
  model days*status(0)=group sex group*sex / ties=Efron;
run;
```

The resulting Class Level Information table for this example is as follows:

Class Level Information				
Class	Value	Design Variables		
group	A	1	0	
	B	0	1	
	C	0	0	
sex	Female	0		
	Male	1		

PARAM=REF is similar to GLM coding. However, for a CLASS variable with *c* levels, the design matrix has *c*-1 columns (full rank). You control the omitted category via the REF= option, which does not appear in the Parameter Estimates table. Again, the beta coefficients estimate the difference in the effect of each non-reference level compared to the effect of the reference level. All things being equal, the parameter estimates are essentially the same with either PARAM=GLM or PARAM=REF given the same class-level ordering and reference category. Thus, in many cases, using either PARAM=GLM or PARAM=REF is a matter of choice. However, PARAM=GLM might be more familiar to you because its construct is similar to other procedures (such as the GLM, LIFEREG, and MIXED procedures) that also use GLM coding.

**Using Selected Options in the CLASS and MODEL Statements**

This section presents several examples of CLASS statement syntax that are selected to give you an idea of how the various individual variable options and global options are specified.

The advent of the CLASS statement also enables new syntax in the MODEL statement for specification of interaction terms. New MODEL statement syntax for interaction models includes the vertical bar operator (|) and the at-sign operator (@). One or both of these operators can be used in the MODEL statement as shortcuts to represent interaction models with compact notation. The vertical bar operator separates variables for which you want all possible interactions; the at-sign operator followed by an integer stipulates the highest order of interaction terms to be included in the model. See [Example 5](#) and [Example 6](#) for two representative examples of the vertical bar notation and the at-sign notation along with their equivalent expansions (without using the shortcut notation).

For the following examples, consider two categorical variables with no formatting applied: the variable GROUP with levels (values) **A**, **B**, and **C** and the variable SEX with levels **Male** and **Female**.

**Example 1**

Default (global option)

```
class group (ref='A') sex (ref='Female') / param=ref order=internal;
```

Because the default setting is PARAM=REF, this option is not necessary in this statement. However, it is generally a good idea to include default settings as good programming etiquette. Note also that the REF= option in this example does not apply if the statement uses PARAM=GLM.

**Example 2**

```
class group (param=ref ref='A') sex (param=ref ref='Female') / order=internal;
```

This specification is equivalent to that shown in the previous example. In this case, the PARAM= option is used as an individual variable rather than a global option. In general, it is recommended that you use one parameterization scheme for all of the CLASS variables.

**Example 3**

```
class group sex / param=glm order=internal;
```

In this example, the GLM parameterization enforces the last levels of GROUP and SEX as the reference categories. Hence, the reference levels here are GROUP='C' and SEX='Male'.

**Example 4**

```
class group sex / ref=last;
```

This CLASS statement achieves essentially the same reference levels (that is, the last levels) as shown in Example 3, which uses the GLM coding method. If formats are applied to the GROUP or SEX variables, the levels for those variables are ordered by the formatted values because ORDER=FORMATTED is the default setting.

**Example 5**

Consider the following MODEL statement that uses the vertical bar operator:

```
model time*censor(1)=X1 | X2 | X3;
```

The previous statement is equivalent to the following expanded MODEL statement with all possible interactions:

```
model time*censor(1) = X1 X2 X3 X1*X2 X1*X3 X2*X3 X1*X2*X3;
```

**Example 6**

Consider the following MODEL statement using both the vertical bar and the at-sign operators:

```
model time*censor(1)=X1 | X2 | X3 @ 2;
```

The previous statement is equivalent to the following expanded MODEL statement, which contains only two-factor interactions:

```
model time*censor(1)=X1 X2 X3 X1*X2 X1*X3 X2*X3;
```

**THE CONTRAST STATEMENT**

In syntax and function, the CONTRAST statement behaves in a similar way to the CONTRAST statements of procedures such as GLM, LOGISTIC, and MIXED. This statement enables you to specify one or more linear combinations of the parameters to test against zero. Formally, the CONTRAST statement enables you to specify a matrix (or vector), **L**, of contrast coefficients for testing the hypothesis  $H_0: \mathbf{L}\boldsymbol{\beta} = 0$ .

## CONTRAST STATEMENT SYNTAX

The syntax for the CONTRAST statement is as follows:

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

The elements in this statement are defined as follows:

- **label**—specifies a text label that is used to identify the contrast on the procedure output. This label, which is required, is useful in identifying the contrast results in the output. ✓
- **row-description**—specifies effects in the following form: ✓

```
effect values <, ... .effect values>
```

In this syntax, *effect* corresponds to an effect or a term in the model.

- **values**—specifies the elements of the **L** matrix (or vector) of contrast coefficients that correspond to *effect* in the row description. To correctly specify the contrast coefficients, it is important to know the order of the parameters that are associated with each effect as well as the reference category for correctly specifying the contrast coefficients. The Class Level Information table aids in verifying the reference level and gives the ordering of the values (levels) of the CLASS variables as specified by the ORDER= option. ✓

Two particularly valuable CONTRAST statement options are the ESTIMATE= and E options. **Note:** You specify these options after a forward slash (/).

- **ESTIMATE=keyword**—specifies that each contrast (each row of **Lβ**) or each exponentiated contrast be estimated and tested. You can estimate the individual contrast, the exponentiated contrast, or both by choosing one of the following *keywords*: ✓
  - PARM—estimates the contrast itself (the linear combination of parameters in the contrast).
  - EXP—exponentiates the contrast that is to be estimated (typically used for custom hazard ratios).
  - BOTH—estimates both the contrast and the exponentiated contrast.
- **E**—specifies that the contrast coefficients (the **L** matrix) be displayed. This option can help you to verify the correspondence of contrast coefficients with model parameters. ✓

For details about all options that are available in the CONTRAST statement, see "The PHREG Procedure" in the SAS/STAT<sup>®</sup> 9.2 User's Guide, Second Edition (SAS Institute Inc. 2009).

## CONTRAST STATEMENT EXAMPLES

This section presents several PROC PHREG examples in which the Parameter Estimates table does not provide desired hazard ratios. You can use CONTRAST statements to produce custom hazard ratios in this situation. You can also use the HAZARDRATIO statement, which is subsequently discussed. However, there are situations in which the HAZARDRATIO statement cannot be used. In such situations, you must use the CONTRAST statement in order to yield the desired hazard ratios. The following examples illustrate how to determine the CONTRAST coefficients that lead to specific custom hazard ratios.

This section presents a general representation of the log hazard function and the design matrix for each model. The examples illustrate the use of the log hazard and design matrix for determining the appropriate set of contrast coefficients **L**. You should be able to use the techniques that are illustrated in order to generate contrast coefficients for custom hazard ratios in your own models.

Two types of interaction models are examined in this section: Class-Variable-by-Class-Variable and Class-Variable-by-Continuous-Variable interaction models. Regarding custom hazard ratios, customers most commonly ask about these two models when they contact SAS Technical Support. The models that are illustrated provide a foundation for extending the methods to other modeling situations.

The PROC PHREG procedure code in the following examples fits an interaction model that involves the CLASS variables GROUP and SEX from the AMI data set. This procedure uses the PARAM=REF coding that is shown in "The CLASS Statement" section. In this section, Example 1 refers back to the PARAM=REF coding that is illustrated in "The CLASS Statement." Example 2 presents a Class-Variable-by-Continuous-Variable interaction model and discusses various custom hazard ratios for categorical and continuous variables that are involved in the interaction. Example 3 concerns a custom hazard ratio that cannot be computed with the new HAZARDRATIO statement and, therefore, must be computed via the CONTRAST statement. Finally, Example 4 briefly discusses a multiple degrees-of-freedom contrast.

**Example 1: Contrasts in a Class-Variable-by-Class-Variable Interaction Model**

```

proc phreg data=sasuser.ami;
  class group(ref='C') sex(ref='Female') / param=ref order=internal;
  model days*status(0) = group sex group*sex / ties=Efron;

  * Contrasts among groups for SEX=Male and SEX=Female. *;
  contrast 'AvsB sex=Male'    group 1 -1 sex 0 group*sex 1 -1 / estimate=exp e;
  contrast 'AvsC sex=Male'    group 1 0 sex 0 group*sex 1 0 / estimate=exp e;
  contrast 'BvsC sex=Male'    group 0 1 sex 0 group*sex 0 1 / estimate=exp e;

  contrast 'AvsB sex=Female'  group 1 -1 sex 0 group*sex 0 0 / estimate=exp e;
  contrast 'AvsC sex=Female'  group 1 0 sex 0 group*sex 0 0 / estimate=exp e;
  contrast 'BvsC sex=Female'  group 0 1 sex 0 group*sex 0 0 / estimate=exp e;

  * Contrasts among sexes for groups A, B, and C. *;
  contrast 'Male vs Female group=A' group 0 0 sex 1 group*sex 1 0 /
    estimate=exp e;
  contrast 'Male vs Female group=B' group 0 0 sex 1 group*sex 0 1 /
    estimate=exp e;
  contrast 'Male vs Female group=C' group 0 0 sex 1 group*sex 0 0 /
    estimate=exp e;

  * Contrast that cannot be replicated with a HAZARDRATIO statement. This *;
  * contrast compares group A with the average of groups B and C for males. *;
  contrast 'A vs (B+C)/2 at sex=Male' group 1 -0.5 sex 0 group*sex 1 -0.5 /
    estimate=exp e;
run;

```

Selected columns from the contrasts that are generated by the PHREG procedure are shown below. For space considerations, the **Row**, **Alpha**, and **Standard Error** columns are not shown.

**Contrast Estimate Table 1 (selected output)**

Contrast Rows Estimation and Testing Results						
Contrast	Type	Estimate	Confidence Limits		Wald Chi-Square	Pr > ChiSq
AvsB sex=Male	EXP	1.3314	0.9195	1.9279	2.2964	0.1297
AvsC sex=Male	EXP	0.1183	0.0369	0.3796	12.8781	0.0003
BvsC sex=Male	EXP	0.0889	0.0271	0.2911	15.9842	<.0001
AvsB sex=Female	EXP	1.2471	0.8617	1.8049	1.3708	0.2417
AvsC sex=Female	EXP	0.6055	0.1477	2.4815	0.4861	0.4857
BvsC sex=Female	EXP	0.4855	0.1179	1.9991	1.0014	0.3170
Male vs Female group=A	EXP	0.6655	0.4819	0.9192	6.1075	0.0135
Male vs Female group=B	EXP	0.6234	0.4132	0.9405	5.0729	0.0243
Male vs Female group=C	EXP	3.4058	0.5667	20.4668	1.7939	0.1805
A vs (B+C)/2 at sex=Male	EXP	0.3969	0.2115	0.7446	8.2850	0.0040

Contrast Estimate Table 1 is produced by the CONTRAST statements that are shown in the previous PROC PHREG code. Each CONTRAST statement specifies a single degrees-of-freedom contrast  $L\beta$  for which  $L$  is a vector of contrast coefficients instead of a matrix. In particular, the table exhibits the value of the exponentiated contrast,  $\exp(L\beta)$ , via the ESTIMATE=EXP option.

The computed value of the exponentiated contrast in the **Estimate** column gives the hazard ratio for the comparison of interest. A required part of the CONTRAST statement syntax is the contrast label. This label is the quoted string in the CONTRAST statement that describes the specific contrast (for example, "**AvsB sex=Male**"). The table also contains the standard error of the hazard ratio estimate and confidence limits on the hazard ratio. The Wald chi-square statistic and the  $p$ -value test the hypothesis  $H_0: L\beta=0$ . If necessary, you can determine the parameter estimate and standard error underlying the Wald chi-square with the ESTIMATE=PARM option or the ESTIMATE=BOTH option in the CONTRAST statement. Because this example focuses on the hazard ratios, it uses ESTIMATE=EXP exclusively.

Selected columns from the parameter estimates that are generated by the PHREG procedure are shown in the following table. The **Standard Error** column is not shown due to space considerations.

**Parameter Estimates Table 1 (selected output)**

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
group A	1	-0.50178	0.4861	0.4857	.	group A
group B	1	-0.72261	1.0014	0.3170	.	group B
sex Male	1	1.22548	1.7939	0.1805	.	sex Male
group*sex A	Male 1	-1.63266	3.0820	0.0792	.	group A * sex Male
group*sex B	Male 1	-1.69806	3.2715	0.0705	.	group B * sex Male

Parameter Estimates Table 1, produced by default from the PHREG code shown in "Example 1: Contrasts in a Class-Variable-by-Class-Variable Interaction Model," reveals that the hazard ratios for any effects involved in interactions are, by design, not given.

**Custom Hazard Ratios**

In the Parameter Estimates Table 1, the lack of hazard ratios for effects that are involved in interaction terms occurs because their hazard ratios depend on the level of the other variable(s) in the interaction. **Therefore, you must create custom hazard ratios with the help of either the CONTRAST statement or the HAZARDRATIO statement.** Custom hazard ratios are, like any contrast, specific linear combinations of the parameter estimates. Carefully considered custom hazard ratios enable you to make useful comparisons, particularly with respect to effects that are involved in interactions.

As stated earlier, you can produce custom hazard ratios with CONTRAST statements by appealing to the log hazard function and the design matrix, which are discussed in the sections that follow. The E option that is specified in the CONTRAST statements in PROC PHREG produce Contrast Coeff tables (not shown) that reveal the order and names of the underlying parameters. The order of the effects in the Contrast Coeff tables agrees with the Parameter Estimates Table 1. For the model in Example 1, the CONTRAST statement's E option gives the order and names of the underlying parameters as groupA, groupB, sexMale, groupAsexMale, and groupBsexMale, which corresponds with the values in the Parameter Estimates Table 1. The E option shows the correspondence of the contrast coefficients that is specified in the CONTRAST statement with the underlying parameters. Thus, the Contrast Coeff table(s) can serve as a useful check to ensure that the intended linear combination of parameters is being tested.

In order to specify a complete set of contrast coefficients for the model in Example 1, you must specify two coefficients for group (one for groupA and one for groupB), one coefficient for sex (for sexMale); and two coefficients for the group\*sex interaction (one for groupAsexMale and one for groupBsexMale), in that order. This order is maintained in the following sections that consider the log hazard and design matrix.

**Log Hazard**

Regarding the log hazard and effects that are produced by the CLASS and MODEL statements in the PROC PHREG example code, the expression of the Cox model,  $h_i(t) = h_0(t) \exp(X_i'\beta)$  is as follows:

$$\log h(t) = \log h_0(t) + \beta_1 * \text{groupA} + \beta_2 * \text{groupB} + \beta_3 * \text{sexMale} + \beta_4 * \text{groupA} * \text{sexMale} + \beta_5 * \text{groupB} * \text{sexMale}$$

The estimated coefficients for this model are shown below and correspond to the values in [Parameter Estimates Table 1](#):

$$\beta_1 = -0.50178 \quad \beta_2 = -0.72261 \quad \beta_3 = 1.22548 \quad \beta_4 = -1.63266 \quad \beta_5 = -1.69806$$

**Design Matrix**

The design matrix shows the parameterization for the variables that appear in the model. For CLASS variables, the parameterization depends on the PARAM= and REF= options in the CLASS statement. The CLASS statement in this example creates binary (0/1) dummy variables to represent the levels of the CLASS variables. The estimated coefficients of these binary dummy variables are found, along with the coefficients of any other terms in the model, by Cox' method of partial likelihood; the values of the estimated coefficients appear in [Parameter Estimates Table 1](#).

CLASS variable parameterizations are given in the [Class Level Information](#) table that is produced by the CLASS statement in the PHREG output. (Several examples of the Class Level Information table are presented in the section "The CLASS Statement.") **Parameterizations for interaction terms are found, obviously enough, by multiplying the design variable values of the terms involved in the interaction.**



The design matrix for the model in this example, shown below, takes into account the PARAM=REF and REF= specifications in the CLASS statement and, naturally, the specification of the MODEL statement. The reference levels are REF='C' for GROUP and REF='Female' for SEX, which are associated with zero (0) values of the binary dummy variables for GROUP and SEX, respectively, in the design matrix. This follows the pattern that is revealed in the [Class Level Information](#) table illustrated previously in the section "[CLASS Statement Variable Parameterization Methods: GLM and REF Coding](#)." You can refer to specific rows of the design matrix when you are crafting custom hazard ratios for various group and sex comparisons.

**Design Matrix 1**

Variables in the Model		Design Matrix				
GROUP	SEX	$\beta_1$ groupA	$\beta_2$ groupB	$\beta_3$ sexMale	$\beta_4$ groupAsexMale	$\beta_5$ groupBsexMale
A	Male	1	0	1	1	0
A	Female	1	0	0	0	0
B	Male	0	1	1	0	1
B	Female	0	1	0	0	0
C	Male	0	0	1	0	0
C	Female	0	0	0	0	0

**Custom Hazard Ratios for Group Comparisons**

This section illustrates how you develop contrast coefficients for group comparisons. This is illustrated by considering the ratio of hazards for specific group comparisons and transforming them to the log hazard ratio, which is equivalent to subtracting two representations of the appropriate log hazard function. In this manner, you manually verify several of the hazard ratio estimates that are shown in [Contrast Estimate Table 1](#). This section also explains how to subtract specific rows of the design matrix to form the correct contrast coefficients for the selected comparisons. Interpretations of the numeric values of the custom hazard ratios are also given.

**Hazard Ratio for Group A versus Group B with SEX=Male**

Because the variable GROUP is involved in an interaction with the variable SEX, group comparisons must take into account the value of SEX. The log hazard ratio (log HR) is as follows:

$$\log HR = \log [ h(t | \text{group=A, sex=Male}) ] - \log [ h(t | \text{group=B, sex=Male}) ]$$

Substituting the corresponding values from the design matrix into the log hazard gives the following (apart from  $\log h_0(t)$ , which always subtracts out):

$$\begin{aligned} \log HR &= [ \beta_1*(1) + \beta_2*(0) + \beta_3*(1) + \beta_4*(1) + \beta_5*(0) ] - [ \beta_1*(0) + \beta_2*(1) + \beta_3*(1) + \beta_4*(0) + \beta_5*(1) ] \\ &= [ \beta_1 + \beta_3 + \beta_4 ] - [ \beta_2 + \beta_3 + \beta_5 ] \\ &= [ \beta_1 - \beta_2 + \beta_4 - \beta_5 ] \end{aligned}$$

Hence, the hazard ratio for group A versus group B for males is  $HR(\text{AvsB at sex=Male}) = \exp [ \beta_1 - \beta_2 + \beta_4 - \beta_5 ]$

Noting which coefficients are involved in the hazard ratio, you can easily determine the **L vector of contrast coefficients**. Hence,  $L = \{ 1 -1 0 1 -1 \}$  for which  $L\beta$  returns the linear combination  $\beta_1 - \beta_2 + \beta_4 - \beta_5$ . The L vector corresponds to the contrast coefficients that are used in the following CONTRAST statement:

```
contrast 'AvsB sex=Male' group 1 -1 sex 0 group*sex 1 -1 / estimate=exp;
```

This statement corresponds to the hazard ratio **1.3314** in [Contrast Estimate Table 1](#). The hazard ratio is obtained mathematically by substituting the parameter estimates into the hazard ratio formula, as follows:

$$\begin{aligned} HR(\text{AvsB at sex=Male}) &= \exp [ \beta_1 - \beta_2 + \beta_4 - \beta_5 ] \\ &= \exp [ (-0.50178) - (-0.72261) + (-1.63266) - (-1.69806) ] \\ &= \exp [ 0.28623 ] \\ &= 1.3314 \end{aligned}$$

This formula gives the same hazard ratio estimate as does the CONTRAST statement above. The interpretation is that there is approximately a 33% higher chance of death after hospitalization for an acute myocardial infarction for men in group A (Q-wave) as opposed to men in group B (Not Q-wave).

You might have realized that the set of contrast coefficients achieved here correspond to subtracting the appropriate rows of the design matrix. In this case, for  $HR(\text{AvsB at sex=Male})$  rows are subtracted that correspond to  $GROUP=A, SEX=Male$  and  $GROUP=B, SEX=Male$  in the design matrix, as shown in the following table:



GROUP	SEX	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
A	Male	1	0	1	1	0
B	Male	0	1	1	0	1
<b>subtraction</b>		1	-1	0	1	-1

Subtracting these two rows in the design matrix gives our previous  $L$  vector:  $L = \{1 \ -1 \ 0 \ 1 \ -1\}$ .

#### **Hazard Ratio for Group B versus Group C with SEX=Female**

You can use similar constructs to determine the contrast coefficients for any group comparisons given the value of SEX. For example, the contrast coefficients for group B versus group C at SEX=Female involves subtracting the following two rows of the design matrix;

GROUP	SEX	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
B	Female	0	1	0	0	0
C	Female	0	0	0	0	0
<b>subtraction</b>		0	1	0	0	0

Subtracting these two rows in the design matrix gives the appropriate  $L$  vector:  $L = \{0 \ 1 \ 0 \ 0 \ 0\}$ , yielding or  $L\beta = \beta_2$ . The CONTRAST statement in this case is as follows:

```
contrast 'BvsC sex=Female' group 0 1 sex 0 group*sex 0 0 / estimate=exp;
```

This statement corresponds to the hazard ratio 0.4855 that is shown in [Contrast Estimate Table 1](#), which agrees with exponentiating the coefficient  $\beta_2$ , as shown here:

$$HR(\text{BvsC at sex=Male}) = \exp[\beta_2] = \exp[-0.72261] = .4855$$

#### **Custom Hazard Ratios for Sex Comparisons**

This section illustrates how you develop contrast coefficients for a sex comparison in the same manner as shown for a group comparison in the last two sections. This section also shows how to develop custom hazard ratios via the log hazard function and, equivalently, via the design matrix.

#### **Hazard Ratio for Male versus Female for GROUP=A**

This example applies the same techniques for sex comparisons (for example, to male versus female) for group A. The log hazard ratio is as follows:

$$\log HR = \log [h(t | \text{group}=A, \text{sex}=Male)] - \log [h(t | \text{group}=A, \text{sex}=Female)]$$

Substituting the corresponding values from the design matrix into the log hazard gives the following (apart from  $\log h_0(t)$ , which subtracts out):

$$\begin{aligned} \log HR &= [\beta_1*(1) + \beta_2*(0) + \beta_3*(1) + \beta_4*(1) + \beta_5*(0)] - [\beta_1*(1) + \beta_2*(0) + \beta_3*(0) + \beta_4*(0) + \beta_5*(0)] \\ &= [\beta_1 + \beta_3 + \beta_4] - [\beta_1] \\ &= [\beta_3 + \beta_4] \end{aligned}$$

Hence, the hazard ratio for males versus females for group A is  $HR(\text{MvsF at group}=A) = \exp[\beta_3 + \beta_4]$ .

Noting the linear combination of coefficients in the hazard ratio, you again can easily determine the  $L$  vector of contrast coefficients  $L = \{0 \ 0 \ 1 \ 1 \ 0\}$  that is used in the CONTRAST statement:

```
contrast 'Male vs Female group=A' group 0 0 sex 1 group*sex 1 0 / estimate=exp;
```

This statement corresponded to the hazard ratio 0.6655 that is shown in [Contrast Estimate Table 1](#).

If you substitute the parameter estimates into the hazard ratio formula, as shown in the following expression, you obtain the same hazard ratio as that obtained with the CONTRAST statement above.

$HR(\text{Male vs Female at group}=A) = \exp[\beta_3 + \beta_4] = \exp[1.22548 + (-1.63266)] = \exp[-0.40718] = 0.6655$ .  
The interpretation of the hazard ratio can be expressed in several equivalent ways. One interpretation is that men in group A (Q-wave) have approximately 67% of the risk of death following hospitalization for an AMI as do females in group A. Similarly, there is a 33%  $[(HR-1)*100 = (0.6655 - 1)*100 = -33\%]$  reduction in the hazard of AMI in men in group A as compared to women in group A. Also, when the hazard ratio is less than 1, it is often instructive to consider the reciprocal of the hazard ratio. In this case,  $1/HR = 1/0.6655 = 1.5$ , which means that women in group A have 1.5 times the hazard of death following hospitalization due to an AMI as do men in group A.

As has been shown in previous scenarios, the set of contrast coefficients here corresponds to subtracting the appropriate rows of the design matrix. In this case, for HR(MvsF at group=A), you subtract the rows that correspond to GROUP=A, SEX=Male and GROUP=A, SEX=Female in the design matrix, as shown here:

GROUP	SEX	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
A	Male	1	0	1	1	0
A	Female	1	0	0	0	0
subtraction		0	0	1	1	0

Subtracting these two rows gives the proper contrast coefficient vector:  $L = \{0 \ 0 \ 1 \ 1 \ 0\}$ .

### Example 2: Contrasts in a Class-Variable-by-Continuous-Variable Interaction Model

Example 2 focuses on a model involving a continuous variable that interacts with a CLASS variable. Again using the AMI data, this section considers a model with a continuous variable AGE in a model with main effects GROUP and SEX, and it adds an AGE\*SEX interaction. The PROC PHREG code for this example is as follows:

```
proc phreg data=sasuser.ami;
  class group(ref='C') sex(ref='Female') / param=ref order=internal;
  model days*status(0) = age group sex age*sex / ties=Efron;

  * Sex contrasts for AGE=25,45,65,85. *;
  contrast 'sex MvsF at age=25' age 0 group 0 0 sex 1 age*sex 25 /
    estimate=exp e;

  contrast 'sex MvsF at age=45' age 0 group 0 0 sex 1 age*sex 45 /
    estimate=exp e;
  contrast 'sex MvsF at age=65' age 0 group 0 0 sex 1 age*sex 65 /
    estimate=exp e;
  contrast 'sex MvsF at age=85' age 0 group 0 0 sex 1 age*sex 85 /
    estimate=exp e;

  * Group contrasts do not depend on age or sex because the variable GROUP *;
  * does not interact with either the AGE or the SEX variables. *;
  contrast 'group AvsB' age 0 group 1 -1 sex 0 age*sex 0 / estimate=exp e;
  contrast 'group AvsC' age 0 group 1 0 sex 0 age*sex 0 / estimate=exp e;
  contrast 'group BvsC' age 0 group 0 1 sex 0 age*sex 0 / estimate=exp e;

  * HR for continuous variable AGE depends on units and sex *;
  * HR for a change of k units in age for sex=Male *;
  * HR for age=x+k vs age=x for sex=Male *;
  * This example uses k=1 and k=10. *;
  hazardratio '1 unit change in age for males' age / at (sex='Male') units=1;
  hazardratio '10 unit change in age for males' age / at (sex='Male') units=10;

  * The hazard ratio for a 1-unit change in AGE for SEX=Male via the *;
  * CONTRAST statement. *;
  contrast 'HR for age at sex=M' age 1 group 0 0 sex 0 age*sex 1 / estimate=exp e;
run;
```

The output from this PHREG procedure is as follows. The **Standard Error** column is not shown due to space considerations.

#### Parameter Estimates Table 2 (selected output)

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
age	1	0.00817	14.2495	0.0002	.	
group	A	0.46098	4.5541	0.0328	0.374	group A
group	B	0.46439	7.2361	0.0071	0.287	group B
sex	Male	0.79265	4.1046	0.0428	.	sex Male
age*sex	Male	0.01088	3.6862	0.0549	.	sex Male * age

Parameter Estimates Table 2 naturally confirms that group effects are not involved in interactions, because hazard ratios are obtained for group A and group B. This is a consequence of the model where the variable GROUP is not involved in any interactions. Hence, when you consider hazard ratios that involve age and sex effects (which are involved in an interaction with each other), you do not need to specify group effects. In this particular model, group effects cancel out of hazard ratios for sex at specific ages.

In Contrast Estimate Table 2, only selected columns are shown. For space considerations, the **Row**, **Alpha**, and **Standard Error** columns are not shown.

#### Contrast Estimate Table 2 (selected output)

Contrast Rows Estimation and Testing Results						
Contrast	Type	Estimate	Confidence	Limits	Wald	
					Chi-Square	Pr>ChiSq
sex MvsF at age=25	EXP	0.3384	0.1207	0.9491	4.2407	0.0395
sex MvsF at age=45	EXP	0.5139	0.2746	0.9620	4.3315	0.0374
sex MvsF at age=65	EXP	0.7805	0.5818	1.0472	2.7309	0.0984
sex MvsF at age=85	EXP	1.1854	0.8114	1.7318	0.7731	0.3793
group AvsB	EXP	1.3040	1.0045	1.6929	3.9742	0.0462
group AvsC	EXP	0.3739	0.1515	0.9229	4.5541	0.0328
group BvsC	EXP	0.2867	0.1154	0.7125	7.2361	0.0071
HR for age at sex=M	EXP	1.0531	1.0382	1.0683	50.2965	<.0001

#### Log Hazard

The expression of the underlying model in terms of the log hazard and effects that are produced by the CLASS and MODEL statements is as follows:

$$\log h(t) = \log h_0(t) + \beta_1 * \text{age} + \beta_2 * \text{groupA} + \beta_3 * \text{groupB} + \beta_4 * \text{sexMale} + \beta_5 * \text{age} * \text{sexMale}$$

From Parameter Estimates Table 2, the coefficient estimates are as follows:

$$\beta_1 = 0.03086 \quad \beta_2 = -0.98374 \quad \beta_3 = -1.24921 \quad \beta_4 = -1.60589 \quad \beta_5 = 0.02089$$

#### Design Matrix

Recall that the reference levels are **C** for GROUP and **Female** for SEX, which are associated with zero (0) values of the dummy variables for GROUP and SEX, respectively, in the design matrix. The continuous variable age is represented as having an arbitrary value  $x$ .

#### Design Matrix 2

Variables in the Model			Design Matrix				
AGE	GROUP	SEX	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
			age	groupA	groupB	sexMale	sexMaleage
X	A	Male	x	1	0	1	X
X	A	Female	x	1	0	0	0
X	B	Male	x	0	1	1	X
X	B	Female	x	0	1	0	0
X	C	Male	x	0	0	1	X
X	C	Female	x	0	0	0	0

#### Custom Hazard Ratios for Sex Comparisons

Because the CLASS variable SEX is involved in an interaction with the continuous variable AGE, you do not get hazard ratios for SEX (or AGE) in the [Parameter Estimates Table 2](#). Again, you must produce custom hazard ratios for SEX. This next section illustrates how you develop contrast coefficients for sex comparisons. As in other cases, this case also uses the technique of subtracting representations of the appropriate log hazard functions. This section also demonstrates that subtracting specific rows of the design matrix yields appropriate contrast coefficients.

**Hazard Ratio for Male versus Female at AGE=x**

The log hazard ratio is as follows:

$$\log \text{HR} = \log [ h(t | \text{age}=x, \text{sex}=\text{Male}) ] - \log [ h(t | \text{age}=x, \text{sex}=\text{Female}) ]$$

Substituting corresponding values from the design matrix for AGE and SEX into the log hazard, without specifying values of the GROUP dummy variables (because they are not involved here), results in the following expressions (apart from  $\log h_0(t)$ ):

$$\begin{aligned} \log \text{HR} &= [ \beta_1^*(x) + \beta_2^*\text{groupA} + \beta_3^*\text{groupB} + \beta_4^*(1) + \beta_5^*(x)^*(1) ] - \\ &\quad [ \beta_1^*(x) + \beta_2^*\text{groupA} + \beta_3^*\text{groupB} + \beta_4^*(0) + \beta_5^*(x)^*(0) ] \\ &= [ \beta_1^*(x) + \beta_4 + \beta_5^*(x) ] - [ \beta_1^*(x) ] \\ &= [ \beta_4 + \beta_5^*(x) ] \end{aligned}$$

Hence, the hazard ratio for males versus females for a given AGE=x is  $\text{HR}(\text{MvsF at age}=x) = \exp [ \beta_4 + \beta_5^*(x) ]$ .

Noting the linear combination of coefficients in the hazard ratio, the **L** vector of contrast coefficients is  $\mathbf{L} = \{ 0 \ 0 \ 0 \ 1 \ x \}$ , which is used in the CONTRAST statements for sex comparisons (which depend on the AGE value x). For example, the CONTRAST statement comparing males and females at age x=25 relates to contrast coefficients  $\mathbf{L} = \{ 0 \ 0 \ 0 \ 1 \ 25 \}$ :

```
contrast 'sex MvsF at age=25' age 0 group 0 0 sex 1 age*sex 25 / estimate=exp;
```

This statement corresponds to the hazard ratio **0.3384** in [Contrast Estimate Table 2](#). Exponentiating the derived linear combination at AGE=25 yields the following expression:

$$\text{HR}(\text{male vs female at age}=25) = \exp [ \beta_4 + \beta_5^*(25) ] = \exp [ -1.60589 + (0.02089*25) ] = \exp [ -1.08364 ] = 0.3384$$

This value indicates that for individuals of age 25, the hazard of death following hospitalization for an AMI in men is just over one-third of that for women.

As seen in the previous scenarios, the set of contrast coefficients here corresponds to subtracting the appropriate rows of the design matrix. In this case, for  $\text{HR}(\text{MvsF at age}=x)$ , you subtract the rows that correspond to SEX=Male, AGE=x and SEX=Female, AGE=x in the design matrix. In this case, because GROUP is not involved in the hazard ratio, you can use rows that correspond to any particular group. However, the two rows that you select must correspond to the same group, as shown below, which has two rows that both correspond to GROUP=A:

AGE	GROUP	SEX	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
X	A	Male	x	1	0	1	x
X	A	Female	x	1	0	0	0
<b>subtraction</b>			0	0	0	1	x

Thus, subtracting any two rows corresponding to **Male** and **Female** for the same group results in the proper contrast coefficient vector:  $\mathbf{L} = \{ 0 \ 0 \ 0 \ 1 \ x \}$ , where x represents the value of AGE for the comparison.

**Custom Hazard Ratios for Group Comparisons**

Because the CLASS variable GROUP is not involved in an interaction, CONTRAST statements for GROUP are not covered here. Default hazard ratios for Group A and Group B, each versus the reference category Group C, are given in [Parameter Estimates Table 2](#). Using the techniques that were discussed previously, you can find custom hazard ratios for additional group comparisons.

**Custom Hazard Ratios for Age (Continuous Variable)**

This next section sets up custom hazard ratios for AGE. Typically, the hazard ratio for a continuous covariate such as AGE is the change in the hazard for a 1-unit increase in the covariate, holding all other covariates constant. Thus, you should typically consider a hazard ratio for AGE+1 versus AGE. This example generalizes to a k-unit increase in age for a given level of sex because age is involved in an interaction with sex.

**Hazard Ratio for a k-Unit Change in AGE for SEX=Male**

The log hazard ratio is as follows:

$$\log \text{HR} = \log [ h(t | \text{age}=x+k, \text{sex}=\text{Male}) ] - \log [ h(t | \text{age}=x, \text{sex}=\text{Male}) ]$$

Substituting corresponding values from the design matrix for AGE and SEX into the log hazard, holding values of the group dummy variables constant (because they are not involved here), results in the following expression (apart from  $\log h_0(t)$ ):

$$\begin{aligned} \log \text{HR} &= [\beta_1*(x+k) + \beta_2*\text{groupA} + \beta_3*\text{groupB} + \beta_4*(1) + \beta_5*(x+k)*(1)] - \\ &\quad [\beta_1*(x) + \beta_2*\text{groupA} + \beta_3*\text{groupB} + \beta_4*(1) + \beta_5*(x)*(1)] \\ &= [\beta_1*(x+k) + \beta_5*(x+k)] - [\beta_1*(x) + \beta_5*(x)] \\ &= [\beta_1*(k) + \beta_5*(k)] \\ &= [\beta_1 + \beta_5] * k \end{aligned}$$

Hence, the hazard ratio for a k-unit change in age for males is the following:

$$\text{HR}(k\text{-unit change in age for sex=M}) = \exp [ (\beta_1 + \beta_5)*k ] = \{ \exp [ \beta_1 + \beta_5 ] \}^{**k}$$

You can write a CONTRAST statement for a 1-unit increase and then exponentiate the point estimate to get the appropriate hazard ratio for a k-unit increase. Given the linear combination of coefficients in the hazard ratio, the contrast coefficient vector is  $\mathbf{L} = \{ 1 \ 0 \ 0 \ 0 \ 1 \}$ , which corresponds to the following CONTRAST statement:

```
contrast 'HR for age at sex=M' age 1 group 0 0 sex 0 age*sex 1 / estimate=exp;
```

This statement yields the hazard ratio **1.0531** in [Contrast Estimate Table 2](#). Note that this corresponds to the following expression:

$$\exp [ \beta_1 + \beta_5 ] = \exp [ 0.03086 + 0.02089 ] = \exp [ 0.05175 ] = 1.0531$$

This expression confirms the CONTRAST statement results. Now, exponentiating the 1-unit hazard ratio to the power 10 yields the following:

$$\text{HR}(k=10 \text{ unit change in age for sex=M}) = [1.0531]^{**10} = 1.678$$

So, for each 10-year increase in age for men, the hazard of death following hospitalization for AMI increases 1.7 times. Note that you can much more easily obtain this value by using a HAZARDRATIO statement with the UNITS= option:

```
hazardratio '10 unit change in age for males' age / at (sex='Male') units=10;
```

The section "[The HAZARDRATIO Statement](#)" further demonstrates the convenience of the HAZARDRATIO statement. However, the next section considers a hazard ratio calculation that cannot be performed with a HAZARDRATIO statement.

### Example 3: Custom Hazard Ratio That Requires Using the CONTRAST Statement

As mentioned previously, there are situations in which the HAZARDRATIO statement cannot be used and in which you must rely on the CONTRAST statement. This example refers back to the interaction model used in [Example 1](#) that involves GROUP, SEX, and GROUP\*SEX with respect to the AMI data.

#### *Hazard ratio Calculation for Group A versus the Average of Groups B and C for SEX=Male*

This example computes the hazard for Group A versus the average hazard for Groups B and C combined. This calculation cannot be done in a HAZARDRATIO statement and must be done in a CONTRAST statement. There is no facility in the HAZARDRATIO statement for considering averages of this kind. In this example, the average of groups B and C is expressed as  $(B+C)/2$  or  $0.5*(B+C)$ , and the example again uses the model that involves the GROUP\*SEX interaction:

$$\log h(t) = \log h_0(t) + \beta_1*\text{groupA} + \beta_2*\text{groupB} + \beta_3*\text{sexMale} + \beta_4*\text{groupA*sexMale} + \beta_5*\text{groupB*sexMale}$$

This expression yields the estimated coefficients  $\beta_1 = -0.50178$ ,  $\beta_2 = -0.72261$ ,  $\beta_3 = 1.22548$ ,  $\beta_4 = -1.63266$ , and  $\beta_5 = -1.69806$ . The log hazard ratio for group A versus the average of groups B and C for SEX=Male is as follows:

$$\log \text{HR} = \log [ h(t \mid \text{group=A, sex=Male}) ] - \log [ h(t \mid 0.5(\text{group=B} + \text{group=C}), \text{sex=Male}) ]$$

The relevant rows of the design matrix that are used from [Example 1](#) are the following:

Variables in the Model		Design Matrix				
		$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
GROUP	SEX	groupA	groupB	sexMale	groupAsexMale	groupBsexMale
A	Male	1	0	1	1	0
B	Male	0	1	1	0	1
C	Male	0	0	1	0	0

For these rows, you can determine the coefficients for 0.5(B+C) for SEX=Male by adding the rows for Group B/SEX=Male and Group C/SEX=Male and dividing by two (multiplying by 0.5):

Variables in the Model		Design Matrix				
		$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
GROUP	SEX	groupA	groupB	sexMale	groupAsexMale	groupBsexMale
A	Male	1	0	1	1	0
0.5(B+C)	Male	0	0.5	1	0	0.5
<b>subtraction</b>		1	-0.5	0	1	-0.5

Subtracting the two resulting rows above gives the vector of contrast coefficients  $L = \{ 1 \ -0.5 \ 0 \ 1 \ -0.5 \}$ , which leads to the following CONTRAST statement:

```
contrast 'A vs (B+C)/2 at sex=Male' group 1 -0.5 sex 0 group*sex 1 -0.5 /
estimate=exp e;
```

The log HR =  $[\beta_1 - (0.5)*\beta_2 + \beta_4 - (0.5)*\beta_5]$  and the hazard ratio is found, as always, by exponentiating the following linear combination:

```
HR(A vs (B+C)/2 at sex=M) = exp [  $\beta_1 - (0.5)*\beta_2 + \beta_4 - (0.5)*\beta_5$  ]
" = exp [ (-0.50178) - (0.5)*(-0.72261) + (-1.63266) - (0.5)*(-1.69806) ]
" = exp [-0.92411]
" = 0.3969
```

This value agrees with the CONTRAST statement results in [Contrast Estimate Table 1](#) (from [Example 1](#)).

#### Example 4: Multiple Degrees-of-Freedom Contrasts

As stipulated earlier, multiple degrees-of-freedom (df) hypotheses of  $H_0: L\beta=0$  are tested by specifying multiple rows of contrast coefficients where  $L$  is now a matrix instead of a vector. These contrast coefficients are separated by commas in a CONTRAST statement. Consider the following CONTRAST statements:

```
contrast 'AvsB sex=Male' group 1 -1 sex 0 group*sex 1 -1 / estimate=exp;
contrast 'AvsC sex=Male' group 1 0 sex 0 group*sex 1 0 / estimate=exp;
contrast 'BvsC sex=Male' group 0 1 sex 0 group*sex 0 1 / estimate=exp;
```

If you want to simultaneously test group A versus group B, group A versus group C, and group B versus group C for males, you can combine the individual contrasts for these hypotheses into one multiple degrees-of-freedom contrast, as shown here:

```
contrast 'AvsB, AvsC, BvsC for sex=Male'
group 1 -1 sex 0 group*sex 1 -1,
group 1 0 sex 0 group*sex 1 0,
group 0 1 sex 0 group*sex 0 1 / estimate=exp;
```

Because the rows of this CONTRAST statement are linearly dependent, (that is, row1 = row 2 - row3), the test has two degrees of freedom (df) in Contrast Test Table 1, which follows. The hypothesis tested is that, for males, the hazard for group A is equivalent to the hazard for group B, and that of group A is equivalent to group C (by dropping row 3). Likewise, you obtain the same test by dropping row 2, which tests that, for males, the hazard for group A is equivalent to the hazard for group B, and that the hazard for group B is equivalent to group C. Therefore, you are testing that, for males, A equals B and B equals C, implying that that A equals C, which requires only two degrees of freedom. As a result, the hypothesis tested is still a simultaneous test of the three groups for males.

Results for all of these contrasts are shown in Contrast Test Table 1 and Contrast Estimates Table 3 as follows:

**Contrast Test Table 1**

Contrast Test Results			
Contrast	DF	Wald	
		Chi-Square	Pr > ChiSq
AvsB sex=Male	1	2.2964	0.1297
AvsC sex=Male	1	12.8781	0.0003
BvsC sex=Male	1	15.9842	<.0001
AvsB, AvsC, BvsC for sex=Male	2	16.4508	0.0003

Only selected columns are shown in Contrast Estimates Table 3. For space considerations, the **Row**, **Alpha**, and **Standard Error** columns are not shown.

**Contrast Estimates Table 3**

Contrast Rows Estimation and Testing Results						
Contrast	Type	Estimate	Wald			
			Confidence Limits	Chi-Square	Pr > ChiSq	
AvsB sex=Male	EXP	1.3314	0.9195	1.9279	2.2964	0.1297
AvsC sex=Male	EXP	0.1183	0.0369	0.3796	12.8781	0.0003
BvsC sex=Male	EXP	0.0889	0.0271	0.2911	15.9842	<.0001
AvsB, AvsC, BvsC for sex=Male	EXP	1.3314	0.9195	1.9279	2.2964	0.1297
AvsB, AvsC, BvsC for sex=Male	EXP	0.1183	0.0369	0.3796	12.8781	0.0003
AvsB, AvsC, BvsC for sex=Male	EXP	0.0889	0.0271	0.2911	15.9842	<.0001

Contrast Test Table 1 provides the results of all contrast tests—both single and multiple degrees-of-freedom Wald chi-square tests. Contrast Estimates Table 3 displays results for each individual row of every contrast. The individual row tests in [Contrast Estimate Table 1](#) for the multiple degrees-of-freedom CONTRAST statement is the same as the individual row tests for the single degrees-of-freedom CONTRAST statements.

## THE HAZARDRATIO STATEMENT

Beginning with SAS 9.2, the HAZARDRATIO statement is entirely new to PROC PHREG, and it is unlike any other statement in SAS software. The HAZARDRATIO statement is especially suited to produce custom hazard ratios for interactions and also for multiple unit changes in continuous covariates.

### HAZARDRATIO STATEMENT SYNTAX

The syntax for the HAZARDRATIO statement is as follows:

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

The elements of the statement are defined as follows:

- **label**—specifies a text label that is used to identify the hazard ratio results in the procedure output. This label is optional. However, it serves a very useful purpose, and its use is recommended.
- **variable**—specifies a variable in the model for which custom hazard ratios are desired (the **variable of interest**).

Three selected HAZARDRATIO statement options that are used in this paper are the DIFF=, AT, and UNITS= options.

- **DIFF=keyword**—specifies which hazard ratios are computed for *variable* (when it is a CLASS variable) by choosing one of the following keywords:
  - ALL requests comparisons among all levels of the CLASS variable.
  - REF requests comparisons of each level with the reference level of the CLASS variable. ✓

The DIFF= option is ignored if *variable* is a continuous variable.



- **AT (variable=list . . . variable=list)**—specifies the variables that interact with the variable of interest and lists the values of the interacting variable at which hazard ratios are to be computed. Hazard ratios are computed at each value in the list if a list of values is given. If the AT variable is a CLASS variable, then *list* must be a list of one or more quoted strings that correspond to the levels of the CLASS variable. These strings are case sensitive. As an option, you can specify the keyword ALL or REF. For interacting CLASS variables, AT (variable=ALL) is the default setting, and hazard ratios are computed at each level of the interacting CLASS variable. AT (variable=REF) also applies to interacting CLASS variables and specifies that hazard ratios be computed at only the reference level of the interacting CLASS variable. If the interacting CLASS variable is formatted, then *list* must contain quoted strings of the formatted values. For interacting continuous variables, *list* is a list of one or more numeric values. If *list* is not specified, the default behavior is to compute the hazard ratio at the mean value of the interacting continuous variable. ✓
- **UNITS=value**—specifies the units of change for customized hazard ratios when the variable of interest is a continuous variable. UNITS=1 is the default setting. The UNITS= option is ignored if *variable* (the variable of interest) is a CLASS variable.

For complete details about all options available in the HAZARDRATIO statement, see "The PHREG Procedure" in the SAS/STAT<sup>®</sup> 9.2 User's Guide, Second Edition (SAS Institute Inc. 2009).

### Custom Hazard Ratios via the HAZARDRATIO Statement

The HAZARDRATIO statement enables you to produce hazard ratios for any variable in the model at customized values. Earlier in this paper, it is noted that to correctly specify CONTRAST statements, you must be aware of the CLASS variable parameterization. This is even more necessary when interactions are involved. The HAZARDRATIO statement syntax simplifies this process significantly. It involves only the variable name of the effect of interest plus knowledge of the levels of the other effects that are involved in interactions with the given effect. For example, you can use the following statements to compute hazard ratios that compare all groups in the AMI data set, given an interaction model with GROUP and SEX (for example, MODEL DAYS\*STATUS(0) = GROUP SEX GROUP\*SEX);

```
hazardratio 'Group comparisons at sex=Male' group / diff=all at (sex='Male');
hazardratio 'Group comparisons at sex=Female' group / diff=all at (sex='Female');
```

All you need to know here is that GROUP interacted with SEX and what the specific levels are for SEX. The two HAZARDRATIO statements above produce the following output:

#### Group Comparisons (output)

Group comparisons at sex=Male: Hazard Ratios for GROUP			
Description	Point Estimate	95% Wald Confidence Limits	
group A vs B At sex=Male	1.331	0.919	1.928
group A vs C At sex=Male	0.118	0.037	0.380
group B vs C At sex=Male	0.089	0.027	0.291

  

Group comparisons at sex=Female: Hazard Ratios for GROUP			
Description	Point Estimate	95% Wald Confidence Limits	
group A vs B At sex=Female	1.247	0.862	1.805
group A vs C At sex=Female	0.605	0.148	2.482
group B vs C At sex=Female	0.485	0.118	1.999

Note that the values for **Point Estimate** in this output are the estimated hazard ratios. Recall from [Example 1](#) in the section "[CONTRAST Statement Examples](#)" that the reference parameterization for GROUP and SEX required the following six CONTRAST statements to find the same custom hazard ratios:

```
contrast 'AvsB sex=Male' group 1 -1 sex 0 group*sex 1 -1 / estimate=exp;
contrast 'AvsC sex=Male' group 1 0 sex 0 group*sex 1 0 / estimate=exp;
contrast 'BvsC sex=Male' group 0 1 sex 0 group*sex 0 1 / estimate=exp;

contrast 'AvsB sex=Female' group 1 -1 sex 0 group*sex 0 0 / estimate=exp;
contrast 'AvsC sex=Female' group 1 0 sex 0 group*sex 0 0 / estimate=exp;
contrast 'BvsC sex=Female' group 0 1 sex 0 group*sex 0 0 / estimate=exp;
```

However, as noted previously, there are custom hazard ratios that cannot be addressed via the HAZARDRATIO statement and for which a CONTRAST statement must be used.

In general, you can compute the hazard ratio by exponentiating the difference of the log hazard between any two covariate profiles. This is simply the hazard for one covariate profile divided by the hazard for another covariate profile. You can use these computations created via the HAZARDRATIO statement regardless of the CLASS variable parameterization or the form of the model (that is, main effect, interaction, or nested models), which makes use of the HAZARDRATIO statement compelling.

#### Using the UNITS= Option to Compute Hazard Ratios

Typically hazard ratios represent the change in the hazard (of the event) for a 1-unit increase in the covariate of interest, while holding all other covariates constant. For CLASS covariates with GLM or REF parameterizations, a 1-unit increase relates to a change in category (for example, SEX=Male versus SEX=Female). For a continuous covariate, such as age in years, the hazard ratio for a 1-unit (1-year) increase in age might not be informative, whereas a hazard ratio for a 5-year or a 10-year increase in age might be of interest. The UNITS= option conveniently enables you to compute hazard ratios for a specified increase in a continuous covariate instead of the default 1-unit increase. Mathematically, the hazard ratio for a change in a continuous covariate  $X$  from  $a$  to  $b$  is estimated by raising the hazard ratio estimate for a 1-unit change in  $X$  to the power of  $k = b - a$ . Hence, if  $X$  has coefficient  $\beta_1$ , then the general hazard ratio for a  $k$ -unit increase in  $X$  is as follows (assuming no interactions):

$$HR(X) = \exp [(b-a) \beta_1] = [\exp(\beta_1)]^{b-a} = [\exp(\beta_1)]^k$$

In the AMI example, the hazard ratio for a 10-year increase in age can be found using the UNITS= option as shown in the following statement:

```
hazardratio '10 year increase in age' age / units=10; ✓
```

The UNITS= option is ignored for CLASS variables.

#### Interpretation of the Hazard Ratio

The hazard ratio represents the change in the hazard of the event with respect to a change in a covariate. For a CLASS variable, the hazard ratio corresponds to a change in category. For example, in the [previous HAZARDRATIO statement example](#), the hazard ratio for group A versus group B at sex=Male is 1.331. This means the hazard of the event (death) is 1.331 times larger for a male individual in group A (Q-wave) as compared to a male individual in group B (Not Q-wave), while holding all other covariates constant (if there are any present). For females, the group A versus group B hazard ratio is also greater than 1; for example, a hazard ratio of 1.247 gives evidence for a higher hazard and lower survival in group A. Apparently, for both males and females, being in group B rather than A is advantageous. However, notice that the 95% confidence interval includes one (1), which corresponds to the hazard ratio being not significantly different from 1 at the alpha=.05 level.

Continuing with this example, the hazard ratio for group A versus group C at sex=Male is 0.118 and the 95% confidence limits indicate that the hazard ratio is significantly different from 1. This means that the hazard of the event (death) is only .118 times larger for a male individual in group A (Q-wave) as compared to a male individual in group C (Indeterminate), while holding all other covariates constant. Taking the reciprocal of .118 (for example,  $1/.118 = 8.475$ ), illustrates that the reverse hazard ratio (that is, for group C versus group A at SEX=Male) is rather large. This means that the hazard of death is nearly 8.5 times higher for males in group C as opposed to those in group A. For males, it is clearly advantageous to be in group A (Q-wave) rather than group C (Indeterminate).

To aid in the interpretation of the hazard ratio, it might be helpful to compute the percent change in the hazard by subtracting 1 from the HR and multiplying by 100. For example, if the covariate is AGE and the hazard ratio is 1.05, then  $(HR - 1) * 100 = 5$  percent. So, for each one-year increase in age, the hazard of the event increases by 5%. For a two-year increase in AGE, you square the hazard ratio. For a five-year increase in age, you raise the hazard ratio to the fifth power. This is essentially the functionality that is provided by the UNITS= option.

#### Example 1: Custom Hazard Ratios in a Class-Variable-by-Class-Variable Interaction Model

This example continues with the model involving GROUP, SEX, and GROUP\*SEX that is considered [at the beginning of this section](#). The following PROC PHREG syntax contains additional HAZARDRATIO statements and their equivalent CONTRAST statement counterparts. (Many of the CONTRAST statement results are given in previous sections of this paper.)

```
proc phreg data=sasuser.ami;
  class group(ref='C') sex(ref='Female') / param=ref order=internal;
  model days*status(0) = group sex group*sex / Ties=Efron;

  * Group comparisons at SEX. To obtain all group comparisons, use DIFF=ALL. *;
  hazardratio 'Group comparisons at SEX=Male' group / diff=all at (sex='Male');
  hazardratio 'Group comparisons at SEX=Female' group / diff=all at (sex='Female');
```

```

* Contrasts among groups for sex=Male and sex=Female. These statements give
* the same hazard ratios as the HAZARDRATIO statements above.
contrast 'AvsB sex=Male'    group 1 -1  sex 0  group*sex 1 -1 / estimate=exp;
contrast 'AvsC sex=Male'    group 1  0  sex 0  group*sex 1  0 / estimate=exp;
contrast 'BvsC sex=Male'    group 0  1  sex 0  group*sex 0  1 / estimate=exp;
contrast 'AvsB sex=Female'  group 1 -1  sex 0  group*sex 0  0 / estimate=exp;
contrast 'AvsC sex=Female'  group 1  0  sex 0  group*sex 0  0 / estimate=exp;
contrast 'BvsC sex=Female'  group 0  1  sex 0  group*sex 0  0 / estimate=exp;

* Sex comparisons at GROUP. DIFF=REF gives the hazard ratio for males
* versus females.
hazardratio 'Sex comparisons at group=A'  sex / diff=ref at (group='A');
hazardratio 'Sex comparisons at group=B'  sex / diff=ref at (group='B');
hazardratio 'Sex comparisons at group=C'  sex / diff=ref at (group='C');

* Contrasts among sexes for Groups A, B, and C. These statements give the
* same hazard ratio as the HAZARDRATIO statements above.
contrast 'male vs female group=A'  group 0  0  sex 1  group*sex 1  0 / estimate=exp;
contrast 'male vs female group=B'  group 0  0  sex 1  group*sex 0  1 / estimate=exp;
contrast 'male vs female group=C'  group 0  0  sex 1  group*sex 0  0 / estimate=exp;

* A contrast that cannot be replicated with a HAZARDRATIO statement
* Compare group A with the average of groups B and C for males.
contrast 'A vs (B+C)/2 at sex=Male'  group 1 -0.5  sex 0  group*sex 1 -0.5 /
estimate=exp;
run;

```

As explained at the beginning of ["Custom Hazard Ratios via the HAZARDRATIO Statement,"](#) it takes six CONTRAST statements to produce the same results as two HAZARDRATIO statements for all group comparisons. Subsequently, the PROC PHREG code above uses three HAZARDRATIO statements for comparing the effect of sex on the hazard at the three levels of group. Considering these sex comparisons, the following hazard ratio table is produced, which agrees with the three corresponding CONTRAST statements (shown in ["Example 1: Contrasts in a Class-Variable-by-Class-Variable Interaction Model"](#)):

#### Hazard Ratios Table 1 (output)

Sex comparisons at GROUP=A: Hazard Ratios for SEX			
Description	Point Estimate	95% Wald Confidence Limits	
sex male vs female At group=A	0.666	0.482	0.919
Sex comparisons at GROUP=B: Hazard Ratios for SEX			
Description	Point Estimate	95% Wald Confidence Limits	
sex male vs female At group=B	0.623	0.413	0.941
Sex comparisons at group=C: Hazard Ratios for SEX			
Description	Point Estimate	95% Wald Confidence Limits	
sex male vs female At group=C	3.406	0.567	20.467

#### The DIFF=REF Option versus the DIFF=ALL Option in Example 1

These options refer to the variable of interest, not the AT variable. DIFF=REF gives comparisons versus the reference level of the variable of interest. DIFF=ALL gives comparisons of all *combinations* of levels of the variable of interest in order (with the ORDER= option, which is specified in the CLASS statement). For example, using DIFF=ALL for GROUP={A,B,C} gives the combinations A versus B, A versus C, and B versus C. Using DIFF=REF for GROUP={A,B,C} gives hazard ratios of A versus C and B versus C because C is the reference level for group. Using DIFF=ALL for SEX={Female, Male} gives comparisons of Female versus Male, assuming the levels of SEX are ordered by ORDER=INTERNAL in the CLASS statement.

Because SEX=Female is specified as the reference category, DIFF=REF is used for the three HAZARDRATIO statements that involve sex to provoke Male versus Female comparisons for each group. In any event, hazard ratios for Male versus Female and Female versus Male are just reciprocals of one another and convey the same information about changes in the hazard of the event.

**Example 2: Custom Hazard Ratios in a Class-Variable-by-Continuous Variable Interaction Model**

This section considers the same model here as in the CONTRAST statement section with AGE, GROUP, and SEX effects plus an AGE\*SEX interaction. With simpler and more intuitive syntax, this example illustrates HAZARDRATIO statements that produce hazard ratios equivalent to those produced by the CONTRAST statements. (CONTRAST statement results are given in the previous section).

```
proc phreg data=sasuser.ami;
  class group(ref='C') sex(ref='Female') / param=ref order=internal;
  model days*status(0) = age group sex age*sex / ties=Efron;

  * Hazard ratio for SEX (Male versus Female) with comparisons at AGE=x, *;
  * and DIFF-REF gives a hazard ratio for males versus females. *;
  hazardratio 'sex Male vs Female at age=25' sex / diff=ref at (age=25);
  hazardratio 'sex Male vs Female at age=45' sex / diff=ref at (age=45);
  hazardratio 'sex Male vs Female at age=65' sex / diff=ref at (age=65);
  hazardratio 'sex Male vs Female at age=85' sex / diff=ref at (age=85);

  * Sex contrasts for AGE=25,45,65,85. These CONTRAST statements give the *;
  * hazard ratio as the HAZARDRATIO statements above. *;
  contrast 'sex MvsF at age=25' age 0 group 0 0 sex 1 age*sex 25 /
    estimate=exp;
  contrast 'sex MvsF at age=45' age 0 group 0 0 sex 1 age*sex 45 /
    estimate=exp;
  contrast 'sex MvsF at age=65' age 0 group 0 0 sex 1 age*sex 65 /
    estimate=exp;
  contrast 'sex MvsF at age=85' age 0 group 0 0 sex 1 age*sex 85 /
    estimate=exp;

  * Hazard ratios for group comparisons do not depend on age. *
  hazardratio 'group comparisons' group / diff=all;

  * Group contrasts do not depend on AGE or SEX because GROUP does not *;
  * interact with either variable. These CONTRAST statements give the *;
  * hazard ratio as the single HAZARDRATIO statement above. *;
  contrast 'group AvsB' age 0 group 1 -1 sex 0 age*sex 0 /
    estimate=exp;
  contrast 'group AvsC' age 0 group 1 0 sex 0 age*sex 0 /
    estimate=exp;
  contrast 'group BvsC' age 0 group 0 1 sex 0 age*sex 0 /
    estimate=exp;

  * Hazard ratio for the continuous variable AGE depends on units and sex. *;
  * Hazard ratio for a change of k units in AGE for sex=Male. *;
  * Hazard ratio for age=x+k versus age=x for sex=Male and sex=Female. *;
  * For an example, this code uses k=10 for the hazard ratio. *;
  hazardratio '1 unit change in age for males' age / at (sex='Male') units=1;
  hazardratio '10 unit change in age for males' age / at (sex='Male') units=10;
  hazardratio '1 unit change in age for females' age / at (sex='Female') units=1;
  hazardratio '10 unit change in age for females' age / at (sex='Female') units=10;

  * Hazard ratio for a 1-unit change in age for sex=Male via the *;
  * following CONTRAST statement. *;
  contrast 'HR for age at sex=M' age 1 group 0 0 sex 0 age*sex 1 / estimate=exp;

  * Hazard ratio for a 1-unit change in age for sex=Female via *;
  * the following CONTRAST statement. *;
  contrast 'HR for age at sex=F' age 1 group 0 0 sex 0 age*sex 0 / estimate=exp;
run;
```

This code produces the hazard ratios table shown below. In the table, extraneous spaces between lines of output are compressed for illustration purposes and some additional titles are inserted for segregation of the output into cogent units.

**Hazard Ratios Table 2 (output)****Sex Comparisons (MvsF) for Various Ages**

sex Male vs Female at AGE=25: Hazard Ratios for SEX

Description	Point Estimate	95% Wald Confidence Limits
sex male vs female At age=25	0.338	0.121 0.949

sex Male vs Female at AGE=45: Hazard Ratios for SEX

Description	Point Estimate	95% Wald Confidence Limits
sex male vs female At age=45	0.514	0.275 0.962

sex Male vs Female at age=65: Hazard Ratios for SEX

Description	Point Estimate	95% Wald Confidence Limits
sex male vs female At age=65	0.781	0.582 1.047

sex Male vs Female at age=85: Hazard Ratios for SEX

Description	Point Estimate	95% Wald Confidence Limits
sex male vs female At age=85	1.185	0.811 1.732

**Group Comparisons (Do Not Depend on Age)**

group comparisons: Hazard Ratios for group

Description	Point Estimate	95% Wald Confidence Limits
group A vs B	1.304	1.004 1.693
group A vs C	0.374	0.151 0.923
group B vs C	0.287	0.115 0.712

In the section ["Example 2: Custom Hazard Ratios in a Class-Variable-by-Continuous Variable Interaction Model,"](#) the PROC PHREG code also produces the following output. The 1-unit and 10-unit hazard ratios in the output below illustrate that the 10-unit hazard ratio is just the 1-unit hazard ratio raised to the 10<sup>th</sup> power, as shown previously. In verifying such calculations manually, it is a good idea to keep a few more significant digits to avoid round-off issues.

**Hazard Ratios Table 2: 1-Unit and 10-Unit Changes in Age for Males and Females**

1-unit change in age for males: Hazard Ratios for AGE

Description	Point Estimate	95% Wald Confidence Limits
age Unit=1 At sex=Male	1.053	1.038 1.068

10 unit change in age for males: Hazard Ratios for AGE

Description	Point Estimate	95% Wald Confidence Limits
age Unit=10 At sex=Male	1.678	1.454 1.936

1-unit change in age for females: Hazard Ratios for AGE

Description	Point Estimate	95% Wald Confidence Limits
age Unit=1 At sex=Female	1.031	1.015 1.048

10-unit change in age for females: Hazard Ratios for AGE

Description	Point Estimate	95% Wald Confidence Limits
age Unit=10 At sex=Female	1.362	1.160 1.598

It is obvious from these examples that, for producing custom hazard ratios, the HAZARDRATIO statement provides a simple and intuitive alternative to the CONTRAST statement.

## CAVEATS WITH RESPECT TO TIME AND TIME-DEPENDENT COVARIATES INVOLVING CLASS EFFECTS

There are stipulations regarding the use of CLASS variables and programming statements that came about with the introduction of the experimental TPHREG procedure in SAS 9.1. These issues carry over into SAS9.2 where PROC PHREG now incorporates the CLASS and CONTRAST statements.

**Note:** PROC TPHREG has been phased out and is no longer available.

For more information about this topic, see "Clarification of the Time and CLASS Variables Usage" in the section "The PHREG Procedure" in the *SAS/STAT<sup>®</sup> 9.2 User's Guide, Second Edition* (SAS Institute Inc 2009). This section in the documentation explains the following caveats.

- **Time dependency for effects that involve time and CLASS variables in the MODEL statement:**

When the time variable is used explicitly in an explanatory effect involving a CLASS variable in the MODEL statement, the effect is NOT time dependent. In the following specification, T is the time variable, but the explanatory effect T\*A is not a time-dependent covariate:

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

The effect T\*A is not time-dependent, which can be shown by demonstrating that the results (not shown) are identical to the following procedure step:

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

In this procedure, MirrorT is an exact copy of T. The effect MirrorT\*A is not time dependent, so neither is T\*A.

- **CLASS variable creation or modification:**

You cannot create or modify the values of a CLASS variable with programming statements inside the PROC PHREG step. The CLASS variable(s) must exist and have values in the DATA= data set. Then the CLASS statement creates binary dummy variables to represent the levels of the categorical variable (as in any other procedure with a CLASS statement). If you try to incorporate or create a CLASS variable with programming statements, the values of the CLASS variable remain unchanged and the results of the PHREG step are the same as without the modification.

- **Exact values of CLASS variables that are used in programming statements:**

While you cannot create or modify CLASS variable values with programming statements, you can still use them in programming statements, as shown here:

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

The section "The PHREG Procedure: Clarification of the Time and CLASS Variables Usage" in the *SAS/STAT<sup>®</sup> 9.2 User's Guide, Second Edition* states the following:

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. (SAS Institute Inc. 2009)

What this means is that the dummy coded binary values that represent A are NOT used in producing X. For example, suppose A is a numeric categorical variable with values {3, 5, 7}. These exact values, then, are used in producing X rather than using the dummy coded representations of these values (via the PARAM= option). Of course, this means that if A is a character-valued categorical variable with values {'Dog', 'Cat', 'Bird'}, then the programming statement  $X=T*A$  produces an error.

Chapter 64, "The PHREG Procedure", in the *SAS/STAT<sup>®</sup> 9.2 User's Guide, Second Edition* (SAS Institute Inc. 2009) illustrates the correct way to use the values of a CLASS variable in an interaction (for example, for purposes of checking the proportional hazards assumption where the exact values of A are numeric {1, 2}, as shown in this procedure):

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

The Boolean parenthetical expressions (A=1) and (A=2) resolve to a numeric one (1) if true and a numeric zero (0) if false. This construct is a convenient way to use the exact values of a CLASS variable in programming statements.

In summary, the idea is that in order to use categorical variables in programming statements, it might be necessary to use manually coded, binary dummy variables to represent the levels of the categorical variables instead of using the CLASS statement.

As an alternative to programming statements, if you want to have a categorical variable as a time-dependent covariate, you might consider using the counting-process style of input that involves the MODEL (start, stop) syntax. In the counting-process style of input, you set up start and stop times for each interval of time over which the covariates are constant. The time-dependent covariates are given on each observation in a prior DATA step. See the section "Counting Process Style of Input" in Chapter 64 of the *SAS/STAT<sup>®</sup> 9.2 User's Guide, Second Edition* (SAS Institute Inc. 2009) for more information about the counting-process syntax. It is still recommended that you use manually coded dummy variables to represent the levels of the categorical variable(s). However, an advantage of using the counting-process syntax over programming statements is that you can examine the prior data set easily to determine whether the coding of the time-dependent covariates is as you want it.

## CONCLUSION

There are many new features in PROC PHREG in SAS9.2. As illustrated in this paper, the advent of the CLASS, CONTRAST, and HAZARDRATIO statements are significant additions to the PHREG procedure's capability and flexibility in fitting and analyzing Cox regression models. The CLASS statement simplifies parameterization for categorical covariates, including control of the hazard ratio reference level. The CONTRAST statement is useful for testing single and multiple degrees-of-freedom hypotheses of model effects and in producing custom hazard ratios. The HAZARDRATIO statement greatly simplifies the process of producing hazard ratios in Cox models. Finally, the CONTRAST statement can be used to generate custom hazard ratios in situations in which the HAZARDRATIO statement is not suitable.

## REFERENCES

- Cox, David.R. 1972. "Regression Models and Life Tables," *Journal of the Royal Statistical Society, Series B* (Methodological), 34: 187–220, with discussion.
- Cox, David. R. (1975), "Partial Likelihood," *Biometrika*, 62: 269–276.
- Hosmer, Jr., David W, and Stanley Lemeshow. 1999. *Applied Survival Analysis: Regression Modeling of Time to Event Data*, New York, NY: John Wiley & Sons, Inc.
- SAS Institute Inc. 2009. "The PHREG Procedure". *SAS/STAT<sup>®</sup> 9.2 User's Guide, Second Edition*. Cary, NC: SAS Institute Inc. Available at [support.sas.com/documentation/cdl/en/statug/63033/PDF/default/statug.pdf](http://support.sas.com/documentation/cdl/en/statug/63033/PDF/default/statug.pdf).



## ACKNOWLEDGEMENTS

I would like to recognize the assistance of my co-author, Michael J. Patetta of the SAS Education and Training Division for providing a data set to use throughout this paper, providing ideas for the paper, and editing the manuscript. My colleague in the SAS Research and Development Division, Dr. Ying So, has provided inestimable service and mentoring, and I gratefully acknowledge his invaluable assistance. I would also like to acknowledge the encouragement and support of the management staff and my colleagues in the SAS Technical Support Division; in particular, my manager Phil Gibbs. Lastly, I would like to identify the inestimable aid of the SAS Technical Support Writing and Editing staff, with special recognition and thanks to Susan Berry for her able and persistent prosecution.

## RECOMMENDED READING

Allison, Paul D. 1995. *Survival Analysis Using the SAS® System: A Practical Guide*. Cary, NC: SAS Institute Inc. (**Note:** a second edition is under development and expected to be available soon.)

Klein, John P, and Melvin L. Moeschberger. 1997. *Survival Analysis: Techniques for Censored and Truncated Data*. New York, NY: Springer-Verlag.

Kleinbaum, David G, and Mitchel Klein. 2005. *Survival Analysis: A Self-Learning Text, Second Edition*. New York, NY: Springer-Verlag.

Patetta, Michael J. 2009. "Survival Analysis Using the Proportional Hazards Model". Course Notes. Cary, NC: SAS Institute Inc. Available at [www.sas.com/apps/pubscat/bookdetails.jsp?catid=1&pc=E1330](http://www.sas.com/apps/pubscat/bookdetails.jsp?catid=1&pc=E1330).

SAS Institute Inc. 2009. "The PHREG Procedure". *SAS/STAT® 9.2 User's Guide, Second Edition*. Cary, NC: SAS Institute Inc. Available at [support.sas.com/documentation/cdl/en/statug/63033/PDF/default/statug.pdf](http://support.sas.com/documentation/cdl/en/statug/63033/PDF/default/statug.pdf).

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Paul T. Savarese, Ph.D.  
Sr. Technical Support Statistician  
Technical Support Division  
SAS Campus Drive  
SAS Institute Inc.  
Cary, NC 27513  
[support@sas.com](mailto:support@sas.com)

Michael J. Patetta, M.S.  
Statistical Instructor  
Education Division  
SAS Campus Drive  
SAS Institute Inc.  
Cary, NC 27513  
(919) 531-7316  
[mike.patetta@sas.com](mailto:mike.patetta@sas.com)

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

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