

sample j is smaller than in sample $j + 1$, so the hypothesis $H_{0j} : b_j(t) = b_{j+1}(t)$ is rejected in favor of $H_{Aj} : b_j(t) < b_{j+1}(t)$ when $[Z_j(\tau) - Z_{j+1}(\tau)]/[\hat{\sigma}_{jj} + \hat{\sigma}_{j+1j+1} - 2\hat{\sigma}_{jj+1}]^{1/2}$ is smaller than the α th lower percentile of a standard normal. Use the information in Example 7.6 and this statistic to make the multiple comparisons.

- 7.11** The data on laryngeal cancer patients was collected over the period 1970–1978. It is possible that the therapy used to treat laryngeal cancer may have changed over this nine year period. To adjust for this possible confounding fact, test the hypothesis of no difference in survival between patients with different stages of disease against a global alternative using a test which stratifies on the cancer being diagnosed prior to 1975 or not. Also perform a separate test of the hypothesis of interest in each stratum.
- 7.12** (a) Repeat Exercise 3 using the log-rank version of the Renyi statistic.
(b) Repeat Exercise 4 using the Gehan version of the Renyi statistic.
- 7.13** In Table 1.3 of section 1.5, the data on time to death for breast cancer-patients who were classed as lymph node negative by standard light microscopy (SLM) or by immunohistochemical (IH) examination of their lymph nodes is reported. Test the hypothesis that there is no difference in survival between these two groups using
(a) the log-rank test,
(b) the Renyi statistic based on the log-rank test,
(c) the Cramer-von Mises statistic, and
(d) the weighted difference in the Kaplan–Meier statistic W_{KM} .
- 7.14** Repeat Exercise 7 using
(a) the Renyi statistic based on the log-rank test,
(b) the Cramer-von Mises statistic, and
(c) the weighted difference in the Kaplan–Meier statistic W_{KM} .
- 7.15** Using the data of section 1.3,
(a) compare the three survival functions for ALL, AML low-risk, and AML high-risk at one year;
(b) perform pairwise multiple comparisons for the three groups employing the Bonferroni correction for multiple tests.

8

Semiparametric Proportional Hazards Regression with Fixed Covariates

8.1 Introduction

Often one is interested in comparing two or more groups of times-to-event. If the groups are similar, except for the treatment under study, then, the nonparametric methods of Chapter 7 may be used directly. More often than not, the subjects in the groups have some additional characteristics that may affect their outcome. For example, subjects may have demographic variables recorded, such as age, gender, socioeconomic status, or education; behavioral variables, such as dietary habits, smoking history, physical activity level, or alcohol consumption; or physiological variables, such as blood pressure, blood glucose levels, hemoglobin levels, or heart rate. Such variables may be used as covariates (explanatory variables, confounders, risk factors, independent variables) in explaining the response (dependent) variable. After

adjustment for these potential explanatory variables, the comparison of survival times between groups should be less biased and more precise than a simple comparison.

Another important problem is to predict the distribution of the time to some event from a set of explanatory variables. Here, the interest is in predicting risk factors for the event of interest. Statistical strategies for prediction are similar to those utilized in ordinary regression. However, the details for regression techniques in survival studies are unique.

In section 2.6, we introduced models which allow us to quantify the relationship between the time to event and a set of explanatory variables. In this chapter, we will consider in more detail the widely used multiplicative hazards model due to Cox (1972), often called the proportional hazards model.

As before, let X denote the time to some event. Our data, based on a sample of size n , consists of the triple $(T_j, \delta_j, \mathbf{Z}_j(t))$, $j = 1, \dots, n$ where T_j is the time on study for the j th patient, δ_j is the event indicator for the j th patient ($\delta_j = 1$ if the event has occurred and $\delta_j = 0$ if the lifetime is right-censored) and $\mathbf{Z}_j(t) = (Z_{j1}(t), \dots, Z_{jp}(t))'$ is the vector of covariates or risk factors for the j th individual at time t which may affect the survival distribution of X . Here the $Z_{jk}(t)$'s, $k = 1, \dots, p$, may be time-dependent covariates whose value changes over time, such as current disease status, serial blood pressure measurements, etc., or they may be constant (or fixed) values known at time 0, such as sex, treatment group, race, initial disease state, etc. In this chapter, we shall consider the fixed-covariate case where $\mathbf{Z}_j(t) = \mathbf{Z}_j = (Z_{j1}, \dots, Z_{jp})'$, and the former situation involving time-dependent covariates will be treated in Chapter 9.

Let $h(t | \mathbf{Z})$ be the hazard rate at time t for an individual with risk vector \mathbf{Z} . The basic model due to Cox (1972) is as follows:

$$h(t | \mathbf{Z}) = h_0(t)c(\boldsymbol{\beta}'\mathbf{Z}) \quad (8.1.1)$$

where $h_0(t)$ is an arbitrary baseline hazard rate, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is a parameter vector, and $c(\boldsymbol{\beta}'\mathbf{Z})$ is a known function. This is called a semi-parametric model because a parametric form is assumed only for the covariate effect. The baseline hazard rate is treated nonparametrically. Because $h(t | \mathbf{Z})$ must be positive, a common model for $c(\boldsymbol{\beta}'\mathbf{Z})$ is

$$c(\boldsymbol{\beta}'\mathbf{Z}) = \exp(\boldsymbol{\beta}'\mathbf{Z}) = \exp\left(\sum_{k=1}^p \beta_k Z_k\right)$$

yielding

$$h(t | \mathbf{Z}) = h_0(t) \exp(\boldsymbol{\beta}'\mathbf{Z}) = h_0(t) \exp\left(\sum_{k=1}^p \beta_k Z_k\right) \quad (8.1.2)$$

and, thus, the logarithm of $h(t | \mathbf{Z})/h_0(t)$ is $\sum_{k=1}^p \beta_k Z_k$ in the spirit of the usual linear models formulation for the effects of covariates. The coding of factors and their interaction effects follows the usual rules for linear models. For example, if a factor has four levels, three indicator (or dummy) variables may be constructed to model the effect of the factor. An interaction between two or more factors may be examined by constructing new variables which are the product of the variables associated with the individual factors as is commonly done in other (least squares or logistic) regression contexts. One needs to take care in interpreting coefficients so constructed.

The Cox model is often called a proportional hazards model because, if we look at two individuals with covariate values \mathbf{Z} and \mathbf{Z}^* , the ratio of their hazard rates is

$$\frac{h(t | \mathbf{Z})}{h(t | \mathbf{Z}^*)} = \frac{h_0(t) \exp\left[\sum_{k=1}^p \beta_k Z_k\right]}{h_0(t) \exp\left[\sum_{k=1}^p \beta_k Z_k^*\right]} = \exp\left[\sum_{k=1}^p \beta_k (Z_k - Z_k^*)\right] \quad (8.1.3)$$

which is a constant. So, the hazard rates are proportional. The quantity (8.1.3) is called the relative risk (hazard ratio) of an individual with risk factor \mathbf{Z} having the event as compared to an individual with risk factor \mathbf{Z}^* . In particular, if Z_1 indicates the treatment effect ($Z_1 = 1$ if treatment and $Z_1 = 0$ if placebo) and all other covariates have the same value, then, $h(t | \mathbf{Z})/h(t | \mathbf{Z}^*) = \exp(\beta_1)$, is the risk of having the event if the individual received the treatment relative to the risk of having the event should the individual have received the placebo.

In section 8.2 coding of both quantitative and qualitative covariates and a discussion of their interpretation is presented. Typically the goal of an investigation is to make an inference about $\boldsymbol{\beta}$ in a global sense, as discussed in sections 8.3 (for distinct event time data) and 8.4 (when ties are present), or, more often than not, to make an inference about a subset of $\boldsymbol{\beta}$ (called a local test) as discussed in section 8.5. Sometimes an investigator would like to treat a continuous covariate as binary. An example of such a covariate might be blood pressure, which is, in theory, a continuous variate; but a researcher might want to classify a patient as being normotensive or hypertensive. The rationale and details of the methodology of discretizing a continuous covariate are provided in section 8.6.

In section 8.7 these techniques are used to build the most appropriate model for survival. Inference for $\boldsymbol{\beta}$ in these sections is based on a partial or conditional likelihood rather than a full likelihood approach. In these analyses, the baseline hazard, $h_0(t)$, is treated as a nuisance parameter function. Sometimes, however, one is interested in estimating the survival function for a patient with a certain set of conditions and characteristics. This is accomplished by utilizing the results described in section 8.8.

8.2 Coding Covariates

In general regression analyses one may have either quantitative or qualitative independent variables. The dependent variable, in the context of this book, is a quantitative variable, the time-to-event, along with an indicator variable which indicates whether or not the event of interest occurred. As indicated in section 8.1, the independent variables may be quantitative—such as blood pressure, blood glucose levels, age, heart rate, or waiting time until a transplant—or they may be qualitative—such as gender, smoking behavior, stage of disease, or the presence or absence of any particular characteristic which the researcher wishes to investigate. Qualitative variables can be used in a regression analysis, just as quantitative variables can be used; however, more care needs to be taken in the manner they are coded and interpreted. Usually, independent variables are known at the start of the study. They are called fixed time covariates. Occasionally independent variables may change values after the study starts and are known as time-dependent covariates. It is extremely important to make this distinction since the methods of analyses differ substantially for time-dependent covariates. First, we shall discuss fixed time covariates. Time-dependent covariates are discussed in Chapter 9.

There are many ways of coding qualitative variables. For dichotomous variables, like gender, the obvious way is to code one of the genders as 1, the other as 0. Which way we do this coding is arbitrary and immaterial. The interpretation of the results, of course, will depend on the way the coding is actually done. For example, if we code the gender variable as $Z_1 = 1$ if male, 0 if female, the hazard rate for males will be $h(t | Z) = h_0(t)\exp(\beta_1)$, and for females will be $h(t | Z) = h_0(t)\exp(0) = h_0(t)$. Here the natural logarithm of the ratio of the hazard function for males relative to the hazard function for females is β_1 , and the ratio of the hazard functions for males relative to females (the relative risk) will be $\exp(\beta_1)$. The variable Z_1 is called an indicator (or dummy) variable since it indicates to which group the subject under consideration belongs. If we had coded another variable as $Z_2 = 1$ if female, 0 if male, then the hazard rate for females would have been $h(t | Z) = h_0(t)\exp(\beta_2)$ and for males will be $h(t | Z) = h_0(t)\exp(0) = h_0(t)$. Here the natural logarithm of the ratio of the hazard function for females relative to the hazard function for males is β_2 , and the ratio of the hazard functions for females relative to males (the relative risk) will be $\exp(\beta_2) = 1/\exp(\beta_1) = \exp(-\beta_1)$. Either way the coding is performed, the interpretation will lead to the same conclusion. Consider the coding for an example which will be used in a subsequent section.

EXAMPLE 8.1

In section 1.5 we introduced a study designed to determine if female breast cancer patients, originally classified as lymph-node-negative by standard light microscopy (SLM), could be more accurately classified by immunohistochemical (IH) examination of their lymph nodes with an anticytokeratin monoclonal antibody cocktail. The data for 45 female breast cancer patients with negative axillary lymph nodes and a minimum 10-year follow-up were selected from The Ohio State University Hospitals Cancer Registry. Of the 45 patients, 9 were immunoperoxidase positive and the remaining 36 still remained negative.

In this example we wish to perform a proportional hazards regression with immunoperoxidase status as the single covariate in the model. We adopt the usual regression formulation of a dichotomous independent variable and construct a dummy (or indicator) variable as follows.

Let $Z = 1$ if the patient is immunoperoxidase positive, 0 otherwise. The model is $h(t | Z) = h_0(t)\exp(\beta Z)$, where $h_0(t)$ is an arbitrary baseline hazard rate and β is the regression coefficient for Z . The ratio of the hazard functions for patient being immunoperoxidase positive relative to the patient being immunoperoxidase negative (the relative risk) will be $\exp(\beta)$. In a later example in section 8.3, the estimate of β , denoted by b , is determined to be 0.9802. Thus the relative risk of dying for an immunoperoxidase-positive patient relative to an immunoperoxidase-negative patient is $\exp(0.9802) = 2.67$. That is, a patient who is immunoperoxidase positive is 2.67 times more likely to die than an immunoperoxidase-negative patient.

When the qualitative variables (sometimes called factors) have more than two categories, there is more choice in the coding schemes. For example, when coding a factor, sometimes termed a "risk group," which has three categories, we utilize a simple extension of the indicator variable coding scheme described above. In particular, we code two indicator variables as

$$\begin{aligned} Z_1 &= 1 \text{ if the subject is in category 1, 0 otherwise,} \\ Z_2 &= 1 \text{ if the subject is in category 2, 0 otherwise.} \end{aligned} \quad (8.2.1)$$

One might be tempted to make a third category $Z_3 = 1$ if the subject is in category 3, 0 otherwise; but to do this would make the three variables $Z_i (i = 1, 2, 3)$ dependent upon each other. This can be seen because if you know the value of Z_1 and the value of Z_2 , then you would know the value of Z_3 . This is contrary to the principle of multiple regression, where you wish to introduce independent variables into the model. The independent variables may be correlated but they should not be completely dependent, since this introduces an undesirable complexity in analysis and interpretation.

There are many other ways to code the categories so as to obtain two "independent" variables which perform the same test of $\beta_1 = \beta_2 = 0$, but we shall not dwell on them, except to say that the interpretation must be carefully understood. Instead we will elaborate a bit more on the coding scheme discussed above. Consider a three level factor, such as race (black, white, Hispanic), using the coding as in (8.2.1)

$$Z_1 = 1 \quad \text{if the subject is black, 0 if otherwise,}$$

$$Z_2 = 1 \quad \text{if the subject is white, 0 otherwise,}$$

The hazard rate, in general, is $h(t | Z) = h_0(t) \exp\{\sum_{k=1}^2 \beta_k Z_k\}$ and, in particular, the hazard rates for blacks, whites, and Hispanics, respectively, is as follows:

$$h(t | Z_1 = 1, Z_2 = 0) = h_0(t) \exp(\beta_1),$$

$$h(t | Z_1 = 0, Z_2 = 1) = h_0(t) \exp(\beta_2),$$

$$h(t | Z_1 = 0, Z_2 = 0) = h_0(t). \quad (8.2.2)$$

Here we can see that the risk of the events occurring among blacks relative to the risk of the events occurring among Hispanics is $\exp(\beta_1)$, the risk of the events occurring among whites relative to the risk of the events occurring among Hispanics is $\exp(\beta_2)$, and the risk of the events occurring among blacks relative to the risk of the events occurring among whites is $\exp(\beta_1 - \beta_2)$.

A note of caution is in order here. If the independent variable is strictly categorical with more than two groups, then it would be inappropriate to code the variable as a single covariate. Suppose we have a categorical covariate with $k (> 2)$ categories and we define a single covariate $Z = i$, if the individual belongs to category i , $i = 1, \dots, k$. The proportional hazards model assumes that the relative risk of an event for an individual in category i as compared to an individual in category $i - 1$ is e^β for any $i = 2, \dots, k$.

For example, suppose we code the patient's race as 1 if black, 2 if white, and 3 if Hispanic. A consequence of this model is that the following relationships between the relative risks must hold:

$$RR(\text{White/Black}) = RR(\text{Hispanic/White}) = e^\beta$$

and

$$RR(\text{Hispanic/Black}) = e^{2\beta}$$

relationships which are not likely to be true.

EXAMPLE 8.2

In section 1.8, a study of 90 males diagnosed with cancer of the larynx was described. In addition to the outcome variable, time from first treatment until either death or the end of the study, the independent

variables, patient's age (in years) at the time of diagnosis and the stage of the patient's cancer, were recorded. A basic test for trend on stage was performed in section 7.4.

Here we wish to illustrate the coding of the variable stage of disease in preparation for performing a proportional hazards regression test with only stage in the model. Since stage has four levels, we adopt the usual indicator variable coding methodology as in (8.2.1) and construct the dummy (or indicator) variables as follows.

$$\text{Let } Z_1 = 1 \quad \text{if the patient is in stage II, 0 otherwise,}$$

$$Z_2 = 1 \quad \text{if the patient is in Stage III, 0 otherwise,}$$

and

$$Z_3 = 1 \quad \text{if the patient is in Stage IV, 0 otherwise.} \quad (8.2.3)$$

This places the patient with Stage I cancer in the referent group; i.e., such a patient will have $Z_1 = Z_2 = Z_3 = 0$. Usually the coding is accomplished so that the referent group is expected to be at either extreme of risk from a subject matter point of view.

In section 8.4 we shall see that $b_1 = 0.06576$, $b_2 = 0.61206$, $b_3 = 1.172284$. The full model for this situation is

$$h(t | Z) = h_0(t) \exp\{\beta'Z\} = h_0(t) \exp\{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3\}.$$

Thus the estimated relative risks of dying for patients with Stage II, III, and IV disease relative to Stage I disease is $\exp(0.06576) = 1.068$, $\exp(0.61206) = 1.844$, and $\exp(1.172284) = 5.600$, respectively. One may also calculate the relative risk of dying for patients for Stage III disease relative to patients for Stage II disease as $\exp(0.61206 - 0.06576) = 1.727$.

A basic test for trend was performed on the data of section 1.8 in Example 7.6 of section 7.4. Since the scores test in the proportional hazards model is identical to the log rank test, when there are no ties (see Practical Note 3 in section 8.3), one could approximate the test for trend in Example 7.6 by taking the stage variable as a continuous variable (stage = 1, 2, 3, 4). The scores test in this proportional hazards model has a chi-squared of 13.64 with 1 degree of freedom, a result consistent with what we found in Example 7.6. As discussed earlier, the estimate of β must be interpreted with caution since it assumes equal relative risk between adjacent stages of disease.

On the other hand, if an independent variable is continuous, such as age, then it would be appropriate to code the variable as a single covariate. In this case, the exponentiated coefficient, e^β , for the variable

$$Z = \text{age (in years)}$$

would be the relative risk of an event for an individual of age i years compared to an individual of age $i - 1$ years. Sometimes we wish to speak of the relative risk of the event for an individual 10 years older than another individual. In this case, the ratio of the hazard (or risk) of the individual 10 years older compared to the hazard (or risk) of the referent individual would be relative risk = $\exp(10\beta)$. There may be other covariates in the model, in which case, the coefficient is termed a partial regression coefficient. Such partial regression coefficients relate the relationship of that variable, say, age, to the outcome variable, time-to-event, controlling for all other variables in the model. Tests based on partial regression coefficients utilize local tests as described in section 8.5. The results from a data set analyzed in section 8.5 are used to illustrate the interpretation of such parameters below.

EXAMPLE 8.2

(continued) Continuing the examination of the data set in section 1.8, we will introduce the age covariate, Z_4 = age of the patient, in addition to the stage indicator variables defined in (8.2.3). The model then is

$$\begin{aligned} b(t | \mathbf{Z}) &= b_0(t) \exp\{\beta' \mathbf{Z}\} \\ &= b_0(t) \exp\{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4\}. \end{aligned} \quad (8.2.4)$$

Here the natural logarithm of the ratio of the hazard function for a 50-year-old individual with Stage IV disease relative to the hazard function for a 40-year-old individual with Stage IV disease is $10\beta_4$; i.e., the relative risk for a 50-year-old patient compared to a 40-year-old (both with Stage IV disease) is $\exp(10\beta_4)$, since the stage of disease parameter will cancel out in the proportional hazards model.

The estimates of the parameters are obtained in section 8.5 as

$$b_1 = 0.1386, b_2 = 0.6383, b_3 = 1.6931, \text{ and } b_4 = 0.0189. \quad (8.2.5)$$

Thus the relative risk for a 50-year-old patient compared to a 40-year-old (both with Stage IV disease) is $\exp(10b_4) = 1.2$. Another way of stating the interpretation of a partial relative risk is that a 50-year-old patient has a probability of dying 1.2 times greater than the probability of dying for a 40-year-old patient with the same stage of disease.

Factors such as gender, age, race, or stage of disease taken individually are often referred to as main effects, i.e., their relationship with the time-to-event outcome is tested for statistical significance as if their relationship does not depend on other factors. An important concept in regression is the consideration of the effect of one factor in the presence of another factor. This concept is termed interaction.

As in other (least squares or logistic) regression contexts, interaction effects between variables may exist and these effects may be very important. An interaction effect exists if the relative risk for two levels of

one factor differs for distinct levels of a second factor. Consider modeling a clinical trial of two treatments based on a proportional hazards model with covariate coded as $Z_1 = 1$ for treatment 1, 0 for treatment 2. Here $\exp(\beta_1)$ is the risk of the first treatment relative to the second. Suppose there is the potential for males and females to respond differently to the two treatments so the relative risk of treatment 1 compared to treatment 2 may depend on sex. As usual we code sex as $Z_2 = 1$ if male and 0 if female. Interactions are formed by multiplying the independent variables of the two individual factors, termed main effects, together. That is, a third variable $Z_3 = Z_1 \times Z_2$ will be created. Here, the exponential of the coefficient of Z_3 , the product of the treatment and gender covariate, is the excess relative risk of treatment 1 compared to treatment 2 for males compared to females. Now the full model will be

$$b(t | \mathbf{Z}) = b_0(t) \exp\{\beta' \mathbf{Z}\} = b_0(t) \exp\{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3\}. \quad (8.2.6)$$

The relative risk of treatment 1 compared to treatment 2 for males is $\exp\{\beta_1 + \beta_3\}$, while for females it is $\exp\{\beta_1\}$. If $\beta_3 = 0$, then the relative risk of treatment 1 compared to treatment 2 will be identical for the two sexes.

The following example illustrates the construction of the interaction of two categorical variables.

EXAMPLE 8.3

In section 1.7 a data set of 863 kidney transplant patients with data on race (white, black) and gender is described. In this study there were 432 white males, 92 black males, 280 white females, and 59 black females. Again, there are various coding options. First, one may treat this study as a four-group problem as we have done in Example 8.2. The three indicator variables may be defined in any desirable way but usually one wants either the best or the worst survival group as the referent group. For example, we may code

$$Z_1 = 1 \text{ if the subject is a black male, 0 otherwise,}$$

$$Z_2 = 1 \text{ if the subject is a white male, 0 otherwise,}$$

and

$$Z_3 = 1 \text{ if the subject is a black female, 0 otherwise.}$$

Here the referent group is being a white female. Again the full model will be

$$b(t | \mathbf{Z}) = b_0(t) \exp\{\beta' \mathbf{Z}\} = b_0(t) \exp\{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3\}.$$

The estimates of the parameters are obtained in section 8.5 as

$$b_1 = 0.1596, b_2 = 0.2484, b_3 = 0.6567.$$

Thus the relative risks for black male, white male, and black female relative to white female are 1.17, 1.28, 1.93, respectively.

Alternatively, we may code the variables as two main effect terms, race and gender, and an interaction term. For example,

$$Z_1 = 1 \text{ if the subject is a female, } 0 \text{ otherwise,}$$

$$Z_2 = 1 \text{ if the subject is black, } 0 \text{ otherwise,}$$

and

$$Z_3 = Z_1 \times Z_2 = 1 \text{ if the subject is a black female, } 0 \text{ otherwise.}$$

Again the full model will be

$$b(t | Z) = h_0(t) \exp\{\beta'Z\} = h_0(t) \exp\{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3\}.$$

Note that the parameters β_i will have a different interpretation. The estimates of the parameters are obtained in section 8.5 as $b_1 = -0.2484$, $b_2 = -0.0888$, $b_3 = 0.7455$. Here the interest will center on the interaction term β_3 , which will be tested in section 8.5. Here, the exponential of the coefficient of Z_3 , the product of the treatment and gender covariate, is the excess relative risk of being black for females compared to males, $\exp(0.7455) = 2.11$. It is also instructive to see that the relative risks for black male, white male, and black female relative to white female are $\exp(-0.0888 - (-0.2484)) = 1.17$, $\exp(0 - (-0.2484)) = 1.28$, $\exp(-0.2484 - 0.0888 + 0.7455 - (-0.2484)) = 1.93$, respectively, just as we obtained for the earlier coding. These are two different coding schemes; the first treats the samples as four groups and the second treats the samples as a 2×2 factorial, where interest may center on the interaction between gender and race. The interpretation of the two coding schemes are equivalent in that they lead to the same relative risks and the same likelihood.

The following example illustrates the construction of the interaction of a continuous variable and a categorical variable.

EXAMPLE 8.2

(continued) Consider two of the factors, namely age and stage of disease, in the data introduced in section 1.8. As usual, Z_i , $i = 1, 2, 3$ are defined as before in (8.2.3) and Z_4 will be the age of the patient. The interaction between age and stage will involve three product terms, namely, $Z_5 = Z_1 Z_4$; $Z_6 = Z_2 Z_4$ and $Z_7 = Z_3 Z_4$. Thus, for a 50-year-old man with Stage II cancer, the three interaction variables will take on the following values: $Z_5 = Z_1 Z_4 = (1)(50) = 50$; $Z_6 = Z_2 Z_4 = (0)(50) = 0$ and $Z_7 = Z_3 Z_4 = (0)(50) = 0$. Other combinations of age and stage can be appropriately formed. Now the full model will be

$$b(t | Z) = h_0(t) \exp\{\beta'Z\} \\ = h_0(t) \exp\{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + \beta_5 Z_5 + \beta_6 Z_6 + \beta_7 Z_7\}. \quad (8.2.7)$$

8.3 Partial Likelihoods for Distinct-Event Time Data

The null hypothesis of no interaction between age and stage may be written as $H_0 : \beta_5 = \beta_6 = \beta_7 = 0$ vs. the alternate hypothesis, which will be the negation of the null. This example will be considered in much more detail in section 8.5. The estimates of the parameters and their interpretation will be delayed until that discussion.

As indicated earlier, our data is based on a sample of size n consisting of the triple (T_j, δ_j, Z_j) , $j = 1, \dots, n$. We assume that censoring is noninformative in that, given Z_j , the event and censoring time for the j th patient are independent. Suppose that there are no ties between the event times. Let $t_1 < t_2 < \dots < t_D$ denote the ordered event times and $Z_{(t)k}$ be the k th covariate associated with the individual whose failure time is t_i . Define the risk set at time t_i , $R(t_i)$, as the set of all individuals who are still under study at a time just prior to t_i . The partial likelihood (see Theoretical Notes 1 and 2), based on the hazard function as specified by (8.1.2), is expressed by

$$L(\beta) = \prod_{i=1}^D \frac{\exp[\sum_{k=1}^p \beta_k Z_{(t)k}]}{\sum_{j \in R(t_i)} \exp[\sum_{k=1}^p \beta_k Z_{jk}]} \quad (8.3.1)$$

This is treated as a usual likelihood, and inference is carried out by usual means. It is of interest to note that the numerator of the likelihood depends only on information from the individual who experiences the event, whereas the denominator utilizes information about all individuals who have not yet experienced the event (including some individuals who will be censored later).

Let $LL(\beta) = \ln[L(\beta)]$. Then, after a bit of algebra, we can write $LL(\beta)$ as

$$LL(\beta) = \sum_{i=1}^D \sum_{k=1}^p \beta_k Z_{(t)k} - \sum_{i=1}^D \ln \left[\sum_{j \in R(t_i)} \exp \left(\sum_{k=1}^p \beta_k Z_{jk} \right) \right] \quad (8.3.2)$$

The (partial) maximum likelihood estimates are found by maximizing (8.3.1), or, equivalently, (8.3.2). The efficient score equations are found by taking partial derivatives of (8.3.2) with respect to the β 's as follows. Let $U_b(\beta) = \delta LL(\beta) / \delta \beta_b$, $b = 1, \dots, p$.

Then,

$$U_b(\beta) = \sum_{i=1}^D Z_{(i)b} - \sum_{i=1}^D \frac{\sum_{j \in R(t_i)} Z_{jb} \exp[\sum_{k=1}^p \beta_k Z_{jk}]}{\sum_{j \in R(t_i)} \exp[\sum_{k=1}^p \beta_k Z_{jk}]} \quad (8.3.3)$$

The information matrix is the negative of the matrix of second derivatives of the log likelihood and is given by $\mathbf{I}(\beta) = [I_{gb}(\beta)]_{p \times p}$ with the (g, h) th element given by

$$I_{gb}(\beta) = \sum_{i=1}^D \frac{\sum_{j \in R(t_i)} Z_{jg} Z_{jb} \exp[\sum_{k=1}^p \beta_k Z_{jk}]}{\sum_{j \in R(t_i)} \exp[\sum_{k=1}^p \beta_k Z_{jk}]} - \sum_{i=1}^D \left[\frac{\sum_{j \in R(t_i)} Z_{jg} \exp(\sum_{k=1}^p \beta_k Z_{jk})}{\sum_{j \in R(t_i)} \exp(\sum_{k=1}^p \beta_k Z_{jk})} \right] \left[\frac{\sum_{j \in R(t_i)} Z_{jb} \exp(\sum_{k=1}^p \beta_k Z_{jk})}{\sum_{j \in R(t_i)} \exp(\sum_{k=1}^p \beta_k Z_{jk})} \right] \quad (8.3.4)$$

The (partial) maximum likelihood estimates are found by solving the set of p nonlinear equations $U_b(\beta) = 0, b = 1, \dots, p$. This can be done numerically, as shown in Appendix A, using a Newton-Raphson technique (or some other iterative method), with (8.3.3) and (8.3.4). Most major software packages will perform this iterative maximization. Note that (8.3.2) does not depend upon the baseline hazard rate $h_0(x)$, so that inferences may be made on the effects of the explanatory variables without knowing $h_0(x)$.

There are three main tests (described in more detail in Appendix B) for hypotheses about regression parameters β . Let $\mathbf{b} = (b_1, \dots, b_p)'$ denote the (partial) maximum likelihood estimates of β and let $\mathbf{I}(\beta)$ be the $p \times p$ information matrix evaluated at β . The first test is the usual test based on the asymptotic normality of the (partial) maximum likelihood estimates, referred to as Wald's test. It is based on the result that, for large samples, \mathbf{b} has a p -variate normal distribution with mean β and variance-covariance estimated by $I^{-1}(\mathbf{b})$. A test of the global hypothesis of $H_0 : \beta = \beta_0$ is

$$X_{\text{W}}^2 = (\mathbf{b} - \beta_0)' \mathbf{I}(\mathbf{b})(\mathbf{b} - \beta_0) \quad (8.3.5)$$

which has a chi-squared distribution with p degrees of freedom if H_0 is true for large samples.

The second test is the likelihood ratio test of the hypothesis of $H_0 : \beta = \beta_0$ and uses

$$X_{\text{LR}}^2 = 2[LL(\mathbf{b}) - LL(\beta_0)] \quad (8.3.6)$$

which has a chi-squared distribution with p degrees of freedom under H_0 for large n .

The third test is the scores test. It is based on the efficient scores, $\mathbf{U}(\beta) = (U_1(\beta), \dots, U_p(\beta))'$ where $U_b(\beta)$ is defined by (8.3.3). For large samples, $\mathbf{U}(\beta)$ is asymptotically p -variate normal with mean $\mathbf{0}$ and covariance $\mathbf{I}(\beta)$ when H_0 is true. Thus a test of $H_0 : \beta = \beta_0$ is

$$X_{\text{SC}}^2 = \mathbf{U}(\beta_0)' \mathbf{I}^{-1}(\beta_0) \mathbf{U}(\beta_0) \quad (8.3.7)$$

which has a large-sample chi-squared distribution with p degrees of freedom under H_0 .

EXAMPLE 8.1

(continued) In section 1.5, we introduced a study designed to determine if female breast-cancer patients, originally classified as lymph node negative by standard light microscopy (SLM), could be more accurately classified by immunohistochemical (IH) examination of their lymph nodes with an anticytokeratin, monoclonal antibody cocktail.

In this example, we wish to perform a proportional hazards regression with immunoperoxidase status as the single covariate in the model. We adopt the usual regression formulation of a dichotomous independent variable and construct a dummy (or indicator) variable as follows.

Let $Z = 1$ if the patient is immunoperoxidase positive, 0 otherwise. The model is $h(t | Z) = h_0(t) \exp(\beta Z)$, where $h_0(t)$ is an arbitrary baseline hazard rate and β is the regression coefficient.

For this model, $\sum_{i=1}^D Z_{(i)} = d_1$, the number of deaths in the immunoperoxidase positive sample, and $\sum_{j \in R(t_i)} \exp(\beta Z_j) = Y_{0i} + Y_{1i} e^\beta$, where Y_{0i} (Y_{1i}) is the number of individuals at risk in the immunoperoxidase negative (positive) sample at time t_i . From (8.3.2)–(8.3.4),

$$LL(\beta) = \beta d_1 - \sum_{i=1}^D \ln[Y_{0i} + Y_{1i} e^\beta],$$

$$U(\beta) = d_1 - \sum_{i=1}^D \frac{Y_{1i} e^\beta}{Y_{0i} + Y_{1i} e^\beta},$$

and

$$I(\beta) = \sum_{i=1}^D \left[\frac{Y_{1i} e^\beta}{(Y_{0i} + Y_{1i} e^\beta)^2} - \frac{Y_{1i}^2 e^{2\beta}}{(Y_{0i} + Y_{1i} e^\beta)^3} \right].$$

The simplest test of the hypothesis that $\beta = 0$ is the score test. In this case,

$$U(0) = d_1 - \sum_{i=1}^D \frac{Y_{1i}}{Y_{0i} + Y_{1i}},$$

and

$$I(0) = \sum_{i=1}^D \frac{Y_{1i}}{(Y_{0i} + Y_{1i})} - \frac{Y_{1i}^2}{(Y_{0i} + Y_{1i})^2} = \sum_{i=1}^D \frac{Y_{1i} Y_{0i}}{(Y_{0i} + Y_{1i})^2}.$$

Note that, in this case, where there are no ties between the death times, the score statistic $X_{sc}^2 = U(0)^2/I(0)$ is precisely the two-sample, log-rank test presented in section 7.3. In this example, $U(0) = 4.19$ and $I(0) = 3.19$ so the value of $X_{sc}^2 = 5.49$ which has a p -value of 0.019, when compared to the chi-squared distribution, with one degree of freedom.

To obtain the estimate of β , the likelihood is maximized by a numerical technique. Routines for this maximization are available in most statistical packages. Using the Newton-Raphson algorithm described in Appendix A, we start with an initial guess at β of $b_0 = 0$ and compute an updated estimate at the m th stage by $b_m = b_{m-1} + U(b_{m-1})/I(b_{m-1})$. The iterative process is declared converged when the relative change in log likelihoods between successive steps is less than 0.0001. Here we have the results of three iterations:

m	b_{m-1}	$LL(b_{m-1})$	$U(b_{m-1})$	$I(b_{m-1})$	$b_m = b_{m-1} +$		
					$\frac{U(b_{m-1})}{I(b_{m-1})}$	$LL(b_m)$	$\frac{LL(b_m) - LL(b_{m-1})}{ LL(b_{m-1}) }$
1	0	-83.7438	4.1873	3.1912	1.3121	-81.8205	0.0230
2	1.3121	-81.8205	-1.8382	5.7494	0.9924	-81.5210	0.0037
3	0.9924	-81.5210	-0.0646	5.3084	0.9802	-81.5206	<0.0001

The Newton-Raphson algorithm converges after three steps.

To test the hypothesis $H_0 : \beta = 0$ using the likelihood ratio test, $X_{LR}^2 = 2\{LL(0.9802) - LL(0)\} = 2\{-81.52 - (-83.74)\} = 4.44$ which has a p -value of 0.035. To perform the Wald test we first estimate the standard error of our estimate of β as $SE(b) = 1/I(0.9802)^{1/2} = 1/5.2871^{1/2} = 0.4349$. The Wald test is $(0.9802 - 0)^2/(0.4349)^2 = 5.08$, which has a p -value of 0.024.

The exponential of b gives the estimated relative risk, which in this example is $e^{0.9802} = 2.67$. This number tells us that a patient, who is immunoperoxidase positive, is 2.67 times more likely to die than an immunoperoxidase negative patient. Using the asymptotic normality of b , a 95% confidence interval for the relative risk is $\exp(0.9802 \pm 1.96 \times 0.4349) = (1.14, 6.25)$.

Practical Notes

1. Algorithms for the estimation of β in the Cox regression model are available in many statistical packages. The procedure PHREG in SAS and coxph in S-Plus provide estimates of β , its standard error and the Wald, score and likelihood ratio tests of the global hypothesis of no covariate effect. A Newton-Raphson algorithm is used to estimate β with 0 as an initial value.
2. If a covariate is perfectly correlated with the event times, that is, the covariates are ordered with $Z_{(1)k} \leq Z_{(2)k} \leq \dots \leq Z_{(D)k}$ (or $Z_{(1)k} \geq Z_{(2)k} \geq \dots \geq Z_{(D)k}$) the (partial) maximum likelihood estimate of β_k will be ∞ (or $-\infty$). When declaring convergence of a numerical maximization routine based on differences in likelihoods at successive iterations, one should carefully check that successive values of the estimates are close to each other as well to avoid this problem.
3. If there are no ties between the event times, the scores test in the proportional hazards model is identical to the log-rank test.
4. Empirical studies have shown that the convergence rate of the likelihood ratio and Wald tests are similar. The score test converges less rapidly to the limiting chi-squared distribution.
5. The tests performed in this section have assumed that the hazard rates are proportional. They, indeed, are but we shall present tools for checking this assumption in Chapters 9 and 11.

Theoretical Notes

1. The probability that an individual dies at time t_i with covariates $\mathbf{Z}_{(i)}$, given one of the individuals in $R(t_i)$ dies at this time, is given by

$$\begin{aligned} &P[\text{individual dies at } t_i \mid \text{one death at } t_i] \\ &= \frac{P[\text{individual dies at } t_i \mid \text{survival to } t_i]}{P[\text{one death at } t_i \mid \text{survival to } t_i]} \\ &= \frac{h(t_i \mid \mathbf{Z}_{(i)})}{\sum_{j \in R(t_i)} h(t_i \mid \mathbf{Z}_j)} = \frac{b_0(t_i) \exp\{\boldsymbol{\beta}' \mathbf{Z}_{(i)}\}}{\sum_{j \in R(t_i)} b_0(t_i) \exp\{\boldsymbol{\beta}' \mathbf{Z}_j\}} \\ &= \frac{\exp\{\boldsymbol{\beta}' \mathbf{Z}_{(i)}\}}{\sum_{j \in R(t_i)} \exp\{\boldsymbol{\beta}' \mathbf{Z}_j\}}. \end{aligned}$$

The partial likelihood is formed by multiplying these conditional probabilities over all deaths, so we have the likelihood function

$$L(\boldsymbol{\beta}) = \prod_{i=1}^D \frac{\exp\{\boldsymbol{\beta}' \mathbf{Z}_{(i)}\}}{\sum_{j \in R(t_i)} \exp\{\boldsymbol{\beta}' \mathbf{Z}_j\}} \quad \text{as in (8.3.1).}$$

2. The Cox partial likelihood can be derived as a profile likelihood from the full censored-data likelihood, as discussed by Johansen (1983). Here, we start with the complete censored-data likelihood, which, by the discussion in section 3.5, is given by

$$\begin{aligned} L(\boldsymbol{\beta}, b_0(t)) &= \prod_{j=1}^n b(T_j | \mathbf{Z}_j)^{\delta_j} S(T_j | \mathbf{Z}_j) \\ &= \prod_{j=1}^n b_0(T_j)^{\delta_j} [\exp(\boldsymbol{\beta}'\mathbf{Z}_j)]^{\delta_j} \exp(-H_0(T_j) \exp(\boldsymbol{\beta}'\mathbf{Z}_j)). \end{aligned} \tag{8.3.8}$$

Now, fix $\boldsymbol{\beta}$ and consider maximizing this likelihood as a function of $b_0(t)$ only. The function to be maximized is

$$L_{\boldsymbol{\beta}}(b_0(t)) = \left[\prod_{i=1}^D b_0(t_i) \exp(\boldsymbol{\beta}'\mathbf{Z}_{(i)}) \right] \exp \left[- \sum_{j=1}^n H_0(T_j) \exp(\boldsymbol{\beta}'\mathbf{Z}_j) \right]. \tag{8.3.9}$$

This function is maximal when $b_0(t) = 0$ except for times at which the events occurs. Let $b_{0i} = b_0(t_i)$, $i = 1, \dots, D$ so $H_0(T_j) = \sum_{t_i \leq T_j} b_{0i}$. Thus, after some simplification, (8.3.9) can be written as

$$L_{\boldsymbol{\beta}}(b_{01}, \dots, b_{0D}) \propto \prod_{i=1}^D b_{0i} \exp \left[-b_{0i} \sum_{j \in R(t_i)} \exp(\boldsymbol{\beta}'\mathbf{Z}_j) \right]$$

and the profile maximum likelihood estimator of b_{0i} is given by

$$\hat{b}_{0i} = \frac{1}{\sum_{j \in R(t_i)} \exp(\boldsymbol{\beta}'\mathbf{Z}_j)}.$$

Combining these estimates yields an estimate of $H_0(t)$ given by

$$\hat{H}_0(t) = \sum_{t_i \leq t} \frac{1}{\sum_{j \in R(t_i)} \exp(\boldsymbol{\beta}'\mathbf{Z}_j)}.$$

This is Breslow's estimator of the baseline cumulative hazard rate in the case of, at most, one death at any time and is discussed in more detail in section 8.8. Substituting $\hat{H}_0(t)$ in (8.3.8) and simplifying yields a profile likelihood proportional to the partial likelihood of Eq. (8.3.1).

8.4 Partial Likelihoods When Ties Are Present

In section 8.3, we presented the partial likelihood for the proportional hazards regression problem when there are no ties between the event times. Often, due to the way times are recorded, ties between event times are found in the data. Alternate partial likelihoods have been provided by a variety of authors when there are ties between event times.

Let $t_1 < t_2 < \dots < t_D$ denote the D distinct, ordered, event times. Let d_i be the number of deaths at t_i and \mathbb{D}_i the set of all individuals who die at time t_i . Let \mathbf{s}_i be the sum of the vectors \mathbf{Z}_j over all individuals who die at t_i . That is $\mathbf{s}_i = \sum_{j \in \mathbb{D}_i} \mathbf{Z}_j$. Let R_i be the set of all individuals at risk just prior to t_i .

There are several suggestions for constructing the partial likelihood when there are ties among the event times. The first, due to Breslow (1974), arises naturally from the profile likelihood construction discussed in Theoretical Note 2 of the previous section. The partial likelihood is expressed as

$$L_1(\boldsymbol{\beta}) = \prod_{i=1}^D \frac{\exp(\boldsymbol{\beta}'\mathbf{s}_i)}{\left[\sum_{j \in R_i} \exp(\boldsymbol{\beta}'\mathbf{Z}_j) \right]^{d_i}}. \tag{8.4.1}$$

This likelihood considers each of the d_i events at a given time as distinct, constructs their contribution to the likelihood function, and obtains the contribution to the likelihood by multiplying over all events at time t_i . When there are few ties, this approximation works quite well, and this likelihood is implemented in most statistical packages.

Efron (1977) suggests a partial likelihood of

$$L_2(\boldsymbol{\beta}) = \prod_{i=1}^D \frac{\exp(\boldsymbol{\beta}'\mathbf{s}_i)}{\prod_{j=1}^{d_i} \left[\sum_{k \in R_i} \exp(\boldsymbol{\beta}'\mathbf{Z}_k) - \frac{t-1}{d_i} \sum_{k \in \mathbb{D}_i} \exp(\boldsymbol{\beta}'\mathbf{Z}_k) \right]}, \tag{8.4.2}$$

which is closer to the correct partial likelihood based on a discrete hazard model than Breslow's likelihood. When the number of ties is small, Efron's and Breslow's likelihoods are quite close.

The third partial likelihood due to Cox (1972) is based on a discrete-time, hazard-rate model. This likelihood is constructed by assuming a logistic model for the hazard rate, that is, if we let $h(t | \mathbf{Z})$ denote the conditional death probability in the interval $(t, t + 1)$ given survival to the start of the interval and if we assume

$$\frac{h(t | \mathbf{Z})}{1 - h(t | \mathbf{Z})} = \frac{b_0(t)}{1 - b_0(t)} \exp(\boldsymbol{\beta}'\mathbf{Z}),$$

then, this likelihood is the proper partial likelihood. To construct the likelihood, let Q_i denote the set of all subsets of d_i individuals who

could be selected from the risk set R_i . Each element of Q_i is a d_i -tuple of individuals who could have been one of the d_i failures at time t_i . Let $q = (q_1, \dots, q_{d_i})$ be one of these elements of Q_i and define $\mathbf{s}_q^* = \sum_{j=1}^{d_i} \mathbf{Z}_{qj}$. Then, the discrete log likelihood is given by

$$L_3(\beta) = \prod_{i=1}^D \frac{\exp(\beta' \mathbf{s}_i)}{\sum_{q \in Q_i} \exp(\beta' \mathbf{s}_q^*)} \quad (8.4.3)$$

When there are no ties between the event times, this likelihood and Breslow's and Efron's likelihoods reduce to the partial likelihood in the previous section.

EXAMPLE 8.4

A study to assess the time to first exit-site infection (in months) in patients with renal insufficiency was introduced in section 1.4. Forty-three patients utilized a surgically placed catheter and 76 patients utilized a percutaneous placement of their catheter. Catheter failure was the primary reason for censoring. To apply a proportional hazards regression, let $Z = 1$ if the patient has a percutaneous placement of the catheter, and 0 otherwise.

To see how the three likelihoods differ, consider the contribution of the six deaths at time 0.5. All six deaths have $Z = 1$, and there are 76 patients at risk with $Z = 1$ and 43 patients at risk with $Z = 0$. The contribution to the likelihood is

$$\begin{aligned} \text{Breslow: } & \frac{\exp(6\beta)}{[43 + 76 \exp(\beta)]^6}, \\ \text{Efron: } & \frac{\exp(6\beta)}{\prod_{j=1}^6 [76e^\beta + 43 - \frac{j-1}{6}(6e^\beta)]}, \\ \text{Discrete: } & \frac{\exp(6\beta)}{\binom{43}{6} + \binom{43}{5} \binom{76}{1} e^\beta + \binom{43}{4} \binom{76}{2} e^{2\beta} + \binom{43}{3} \binom{76}{3} e^{3\beta} + \binom{43}{2} \binom{76}{4} e^{4\beta} + \binom{43}{1} \binom{76}{5} e^{5\beta} + \binom{76}{6} e^{6\beta}} \end{aligned}$$

Using the three likelihoods, we have the following results:

	Breslow's Likelihood (8.4.1)	Efron's Likelihood (8.4.2)	Discrete Likelihood (8.4.3)
Initial likelihood	-104.4533	-104.2319	-94.1869
Final likelihood	-103.2285	-103.0278	-92.9401
b	-0.6182	-0.6126	-0.6294
SE(b)	0.3981	0.3979	0.4019
Relative risk, e^b	0.539	0.542	0.553
Score test	$X^2 = 2.49$ ($p = 0.115$)	$X^2 = 2.44$ ($p = 0.117$)	$X^2 = 2.53$ ($p = 0.112$)
Wald test	$X^2 = 2.41$ ($p = 0.121$)	$X^2 = 2.37$ ($p = 0.124$)	$X^2 = 2.45$ ($p = 0.117$)
Likelihood ratio test	$X^2 = 2.45$ ($p = 0.118$)	$X^2 = 2.41$ ($p = 0.121$)	$X^2 = 2.49$ ($p = 0.114$)

Here, we see that, assuming the proportional hazards model is appropriate, for any of the three likelihoods the regression coefficient is not statistically significant and the relative risk is about 0.54. As we shall test formally in Chapter 9, the proportional hazards assumption is not appropriate for this data. Thus the relative risk is not constant but depends upon time, and the reported relative risk of 0.54 is not correct. Furthermore, a potentially significant result could be overlooked because the proportional hazards assumption is not satisfied. This has implications for procedures used in model building, which will be discussed in section 8.7. As a graphical check of the proportional hazards assumption, we compute the Nelson-Aalen estimator of the cumulative hazard rate for each treatment. If the proportional hazards model is correct, then, we should have $H(t | Z = 1) = e^\beta H(t | Z = 0)$, so that a plot of $\ln[\hat{H}(t | Z = 1)] - \ln[\hat{H}(t | Z = 0)]$ versus t should be roughly equal to β . The plot for this data set, shown in Figure 8.1, strongly suggests

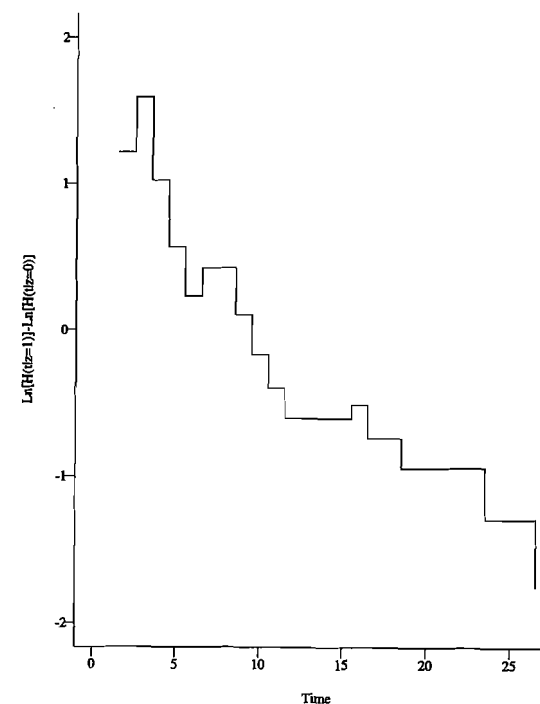


Figure 8.1 Graphical check of the proportional hazards assumption for the renal insufficiency study.

nonproportional hazards. Other graphical checks of the proportional hazards assumption are discussed in section 11.4.

EXAMPLE 8.2

(continued) In section 1.8, a study of 90 males diagnosed with cancer of the larynx was described. In addition to the outcome variable, time from first treatment until either death or the end of the study, the independent variables, patient's age (in years) at the time of diagnosis and the stage of the patient's cancer, were recorded. A basic test for trend on stage was performed in section 7.4.

Here, we shall perform a global proportional hazards regression test with only stage in the model. Because stage has four levels, we adopt the usual indicator variable coding methodology as in (8.2.3). The maximum likelihood parameter estimates, b_i ($i = 1, \dots, 4$), (and their corresponding standard errors) are 0.0658 (0.4584), 0.612 (0.3552), and 1.723 (0.4197), respectively. It follows that the relative risks, $RR(\text{Stage II}/\text{Stage I}) = 1.07$, $RR(\text{Stage III}/\text{Stage I}) = 1.84$, and $RR(\text{Stage IV}/\text{Stage I}) = 5.60$.

The global likelihood ratio, Wald, and score chi-squared (with three degrees of freedom) statistics for stage are 16.26 (p -value = 0.001), 18.95 (p -value = 0.0003), and 22.46 (p -value = 0.0001), respectively, using Breslow's method of handling ties. All three tests suggest that the survival rates are different for, at least, one stage of cancer. In the next section, we shall consider local tests which provide information on which stages differ in survival.

The following example illustrates another example of performing a global test for different groups and will be followed up in the next section with a local test for interaction.

EXAMPLE 8.3

(continued) In section 1.7 a data set of 863 kidney transplant patients with data on race (white, black) and gender is described. In this study there were 432 white males, 92 black males, 280 white females, and 59 black females. Again, there are various coding options, as described in section 8.2. First, one may treat this study as a four-group problem. The three indicator variables have been defined in the usual way as described in section 8.2 as

$$Z_1 = 1 \quad \text{if the subject is a black male, 0 otherwise}$$

$$Z_2 = 1 \quad \text{if the subject is a white male, 0 otherwise}$$

and

$$Z_3 = 1 \quad \text{if the subject is a black female, 0 otherwise.}$$

Here the referent group is being a white female. Again, the full model will be

$$h(t | Z) = b_0(t) \exp\{\beta'Z\} = b_0(t) \exp\{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3\}.$$

The maximum likelihood estimates of the parameters are obtained as

$$b_1 = 0.160, b_2 = 0.248, b_3 = 0.657.$$

Thus the relative risks for black male, white male, and black female relative to white female are 1.17, 1.28, 1.93, respectively. The global likelihood ratio, Wald, and score chi-squared (with 3 degrees of freedom) statistics for groups are 4.37 (p -value = 0.22), 4.64 (p -value = 0.20), and 4.74 (p -value = 0.19), respectively, using the Breslow method of handling ties. All three tests suggest the survival rates are not different for the four groups of subjects. In the next section we shall consider local tests which provide information on testing for an interaction between race and gender.

Practical Notes

1. SAS PHREG uses Breslow's likelihood as a default and allows the user to specify that calculations be carried out using either the discrete or Efron likelihood. SAS also allows the user to specify an "exact" likelihood based on a generalized rank statistic derivation of the likelihood (see Kalbfleisch and Prentice (1980) for details). This likelihood requires a bit more computer time to implement and gives results quite close to the discrete likelihood.
2. The S-Plus function `coxph` uses Efron's likelihood as a default when there are ties between the event times. Breslow's likelihood and the exact likelihood are also available.

8.5 Local Tests

Often, one is interested in testing a hypothesis about a subset of the β 's. The hypothesis is then $H_0: \beta_1 = \beta_{10}$, where $\beta = (\beta_1', \beta_2')'$. Here β_1 is a $q \times 1$ vector of the β 's of interest and β_2 is the vector of the remaining $p - q$ β 's.

The Wald test of $H_0: \beta_1 = \beta_{10}$ is based on the maximum partial likelihood estimators of β . Let $\mathbf{b} = (\mathbf{b}_1', \mathbf{b}_2')'$ be the maximum partial likelihood estimator of β . Suppose we partition the information matrix \mathbf{I} as

$$\mathbf{I} = \begin{pmatrix} \mathbf{I}_{11} & \mathbf{I}_{12} \\ \mathbf{I}_{21} & \mathbf{I}_{22} \end{pmatrix},$$

where \mathbf{I}_{11} (\mathbf{I}_{22}) is the $q \times q$ [$(p - q) \times (p - q)$] submatrix of second partial derivatives of the minus log likelihood with respect to β_1 (β_2) and \mathbf{I}_{12} and \mathbf{I}_{21} , the matrices of mixed second partial derivatives. The Wald test statistic is

$$X_w^2 = (\mathbf{b}_1 - \beta_{10})' [\mathbf{I}^{11}(\mathbf{b})]^{-1} (\mathbf{b}_1 - \beta_{10}) \quad (8.5.1)$$

where $\mathbf{I}^{11}(\mathbf{b})$ is the upper $q \times q$ submatrix of $\mathbf{I}^{-1}(\mathbf{b})$ (see Appendix B). For large samples, this statistic has a chi-squared distribution with q degrees of freedom under H_0 .

Let $\mathbf{b}_2(\beta_{10})$ be the (partial) maximum likelihood estimates of β_2 based on the log likelihood (8.3.2) with the first q β 's fixed at a value β_{10} . The likelihood ratio test of $H_0 : \beta_1 = \beta_{10}$ is expressed by

$$X_{LR}^2 = 2\{LL(\mathbf{b}) - LL[\beta_{10}, \mathbf{b}_2(\beta_{10})]\} \quad (8.5.2)$$

which has a large sample chi-squared distribution with q degrees of freedom under H_0 .

To test $H_0 : \beta_1 = \beta_{10}$ using the score statistic, let $\mathbf{U}_1[\beta_{10}, \mathbf{b}_2(\beta_{10})]$ be the $q \times 1$ vector of scores for β_1 , evaluated at the hypothesized value of β_{10} and at the restricted partial maximum likelihood estimator for β_2 . Then,

$$X_{SC}^2 = \mathbf{U}_1[\beta_{10}, \mathbf{b}_2(\beta_{10})]' [\mathbf{I}^{11}(\beta_{10}, \mathbf{b}_2(\beta_{10}))] \mathbf{U}_1[\beta_{10}, \mathbf{b}_2(\beta_{10})] \quad (8.5.3)$$

which has a large sample chi-squared distribution with q degrees of freedom under H_0 . We shall illustrate these tests in the following example.

EXAMPLE 8.2

(continued) In section 8.4, a global test was performed on stage of cancer in a study of 90 males diagnosed with cancer of the larynx. Here we shall test the hypothesis that there is no difference in survival between patients with different stages of disease, adjusting for the age of the patient. Our test is based on the model with covariates Z_1 , Z_2 , and Z_3 , which, as in section 8.2, are the indicators of stage II, III, and IV disease, respectively, and a covariate Z_4 which is the patient's age at diagnosis. The local hypothesis of interest is $H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$ against the alternative hypothesis that, at least, one of these β 's is nonzero.

To apply the score test or the likelihood ratio test, we need to estimate the coefficient for age, β_4 , in the model, with $\beta_1 = \beta_2 = \beta_3 = 0$. This involves fitting a Cox model with only the single covariate, age. Fitting this model, we find $b_4 = 0.023$ with a log partial likelihood of -195.906 .

Using this value of b_4 , we find that the score is

$$\mathbf{U}(0, 0, 0, 0.023) = (-2.448, 3.0583, 7.4400, 0.000)'$$

and that

$$\mathbf{I}(0, 0, 0, 0.023) = \begin{pmatrix} 7.637 & -2.608 & -0.699 & -24.730 \\ -2.608 & 9.994 & -0.979 & -8.429 \\ -0.699 & -0.979 & 3.174 & 11.306 \\ -24.730 & -8.429 & 11.306 & 4775.716 \end{pmatrix}$$

The inverse of this information matrix is given by

$$\mathbf{I}^{-1}(0, 0, 0, 0.023) = \begin{pmatrix} 0.1529 & 0.0449 & 0.0448 & 0.0008 \\ 0.0449 & 0.1164 & 0.0446 & 0.0003 \\ 0.0448 & 0.0446 & 0.3404 & -0.0005 \\ 0.0008 & 0.0003 & -0.0005 & 0.0002 \end{pmatrix}$$

so the score statistic is given by

$$X_{SC}^2 = (-2.448, 3.0583, 7.4400) \begin{pmatrix} 0.1529 & 0.0449 & 0.0448 \\ 0.0449 & 0.1164 & 0.0446 \\ 0.0448 & 0.0446 & 0.3404 \end{pmatrix} \begin{pmatrix} -2.448 \\ 3.0583 \\ 7.4400 \end{pmatrix} = 20.577.$$

Comparing this quantity to a chi-squared distribution with three degrees of freedom, we find that the p -value of the test of no stage effect is 0.0001.

To perform the Wald and likelihood ratio tests, we need to fit the full model with all four covariates. Here, we find

$$\mathbf{b}' = (0.1386, 0.6383, 1.6931, 0.0189)$$

with a partial log likelihood of -188.179 . The likelihood ratio test of H_0 is

$$X_{LR}^2 = 2[-188.179 - (-195.906)] = 15.454.$$

The p -value of this test is 0.0015.

To perform the Wald test, we need the information matrix based on \mathbf{b} . This matrix is

$$\mathbf{I}(\mathbf{b}) = \begin{pmatrix} 5.9987 & -2.3913 & -1.4565 & -22.8634 \\ -2.3913 & 10.9917 & -3.3123 & -14.0650 \\ -1.4565 & -3.3123 & 7.4979 & 25.6149 \\ -22.8634 & -14.0650 & 25.6149 & 5088.5378 \end{pmatrix}$$

The inverse of this matrix is the covariance matrix of \mathbf{b} , given by

$$\mathbf{I}^{-1}(\mathbf{b}) = \begin{pmatrix} 0.2137 & 0.0683 & 0.0690 & 0.0008 \\ 0.0683 & 0.1268 & 0.0682 & 0.0003 \\ 0.0690 & 0.0682 & 0.1783 & -0.0004 \\ 0.0008 & 0.0003 & -0.0004 & 0.0002 \end{pmatrix} \quad (8.5.4)$$

The Wald chi-squared statistic is given by

$$X_W^2 = (0.1386, 0.6383, 1.6931) \begin{pmatrix} 0.2137 & 0.0683 & 0.0690 \\ 0.0683 & 0.1268 & 0.0682 \\ 0.0690 & 0.0682 & 0.1783 \end{pmatrix}^{-1} \begin{pmatrix} 0.1386 \\ 0.6383 \\ 1.6931 \end{pmatrix}$$

$$= 17.63$$

which has a p -value of 0.0005.

Often it is desirable to perform tests involving one (or more) linear combination(s) of the parameters. For the Wald test, one can form a matrix of q linear combinations of parameters to test such hypotheses. Here one forms a $q \times p$ matrix of full rank ($q \leq p$),

$$C = \begin{pmatrix} c_1' \\ c_2' \\ \vdots \\ c_q' \end{pmatrix} \quad (8.5.5)$$

where $c_k' = (c_{k1}, c_{k2}, \dots, c_{kp})$ is a vector of coefficients for the k th linear combination of the betas, and the hypothesis to be tested is

$$H_0 : C\beta = C\beta_0. \quad (8.5.6)$$

From large-sample theory,

$$(Cb - C\beta_0)'[CI^{-1}(b)C']^{-1}(Cb - C\beta_0) \quad (8.5.7)$$

will have an asymptotic chi-squared distribution with q degrees of freedom.

EXAMPLE 8.2

(continued) In the previous example, we fitted a model to data on patients with cancer of the larynx. In this example, we wish to test the hypothesis $H_0 : \beta_1 = 0$. Note that the upper 1×1 submatrix of $\hat{V}(b)$ is precisely the matrix $I^{11}(b)$ required in (8.5.1) and the Wald chi-squared test is calculated as $(0.1386)(0.2137)^{-1}(0.1386) = 0.0898$. If we choose the linear combination approach, $c = (1, 0, 0, 0)$ and $(Cb)'[CI^{-1}(b)C']^{-1}Cb = 0.0898$, the same result as above. Note that this statistic, which has a large-sample chi-squared distribution with one degree of freedom under H_0 , is testing for a difference in risk of death between stage I and stage II cancer patients, adjusting for age. Here the p -value of that test is 0.7644 which suggests no difference between stage I and II patients.

Most statistics packages will produce an "Analysis of Variance" (ANOVA) table describing all such univariate Wald tests along with the estimated standard error and relative risk of the effects. Note that, in such tables, the relative risk, $\exp(b)$, is the relative risk in a differ-

TABLE 8.1

Analysis of Variance Table for Stage of the Laryngeal Cancer Patients, Utilizing the "Breslow" Method of Handling Ties

Variables	Degrees of Freedom	Parameter Estimates	Standard Errors	Wald Chi Square	p-Value	Relative risk
Z ₁ : Stage II	1	0.1386	0.4623	0.09	0.7644	1.15
Z ₂ : Stage III	1	0.6383	0.3561	3.21	0.0730	1.89
Z ₃ : Stage IV	1	1.6931	0.4222	16.08	<0.0001	5.44
Z ₄ : Age	1	0.0189	0.0143	1.76	0.1847	1.02

ence of one unit in the covariate values. So the relative risk for age in Table 8.1 reflects the excess risk of dying for each additional year of age at diagnosis. Similarly, the risks of death for a patient in Stages II, III, and IV relative to a patient in Stage I are 1.15, 1.89, and 5.44, respectively. The corresponding confidence intervals for the β_i are $[b_i - z_{1-\alpha/2}SE(b_i), b_i + z_{1-\alpha/2}SE(b_i)]$ which may be obtained from the ANOVA table and tables of the unit normal distribution. Confidence intervals for the relative risk may be found by exponentiating the lower and upper limits, respectively. For example, a 95% confidence interval for the risk of death for patients in Stage IV relative to the risk of death for patients in Stage I would be $\{\exp[1.6931 - 1.96(0.4222)], \exp[1.6931 + 1.96(0.4222)]\} = (2.38, 12.44)$. This means that, with approximately 95% confidence, $\exp(\beta_3)$ will lie between 2.38 and 12.44, so that we will reject the hypothesis that $\beta_3 = 0$ when $\alpha = 0.05$, as indicated in the table (p -value < 0.0001).

Often, one is interested in relative risks that may not appear directly in the table. For example, the risk of death for patients in Stage III relative to the risk of death for patients in Stage II is found by taking $\exp(\beta_2)/\exp(\beta_1) = \exp(\beta_2 - \beta_1)$. The point estimate of this risk is $\exp(0.6383 - 0.1386) = 1.65$ which could also have been obtained directly from the table as 1.89/1.15 (aside from round-off error). The confidence interval for this relative risk cannot be obtained directly from the table. One needs the standard error of $b_2 - b_1$ which means we need the variance-covariance matrix of the b_i 's as given in (8.5.4). Calculating $\text{Var}(b_2 - b_1) = \text{Var}(b_2) + \text{Var}(b_1) - 2\text{Cov}(b_2, b_1) = 0.1268 + 0.2137 - 2(0.0683) = 0.2039$ we are led, by taking the square root, to the standard error of $(b_2 - b_1) = 0.4515$. Now, we can find a 95% confidence interval for $\beta_2 - \beta_1$ as $[b_2 - b_1 - 1.96 SE(b_2 - b_1), b_2 - b_1 + 1.96 SE(b_2 - b_1)] = [0.4997 - 1.96(0.4515), 0.4997 + 1.96(0.4515)] = (-0.3852, 1.3846)$. Exponentiating the lower and upper limit leads to the approximate 95% confidence interval for $\exp(\beta_2 - \beta_1)$ as (0.68, 3.99). Thus, this relative risk cannot be judged to differ from one.

Both the Wald and the likelihood ratio test can be used to test the hypothesis $H_0 : \beta_1 = \beta_2, \beta_2 = \beta_3$ or, equivalently, $H_0 : \beta_1 = \beta_2 = \beta_3$. This is a test of the hypothesis that, adjusted for age of the patient, survival is the same for stage II, III and IV patients, but not necessarily the same as for stage I patients. To perform the likelihood ratio test, we fit the full model with all four covariates which has a log partial likelihood of -188.179 , and we fit a model with two covariates $Z^* = Z_1 + Z_2 + Z_3$ and age (Z_4). Here Z^* is the indicator of stage II, III or IV disease. The log partial likelihood from this model is -193.137 . The likelihood ratio chi-squared is $2[-188.179 - (-193.137)] = 9.916$. For large samples under H_0 , this statistic has a chi-squared distribution with two degrees of freedom. (The degrees of freedom are the number of parameters in the full model minus the number of parameters in the reduced model.) The p -value of this test is 0.0070 which suggests that survival is different for at least one of the three advanced stages.

To perform the Wald test, we define the C matrix with two contrasts, namely,

$$C = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 \end{pmatrix},$$

and apply (8.5.7). The resulting statistic has a value of 10.7324 with two degrees of freedom for the large-sample chi square. The p -value of the test is 0.0047 , so the conclusion is the same as for the likelihood ratio test.

Now we turn our attention to a discussion of interaction. The first example is an example of an interaction between two categorical variables.

EXAMPLE 8.3

(continued) An alternative coding scheme for the data in section 1.7 discussed earlier is to code the variables as two main effect terms, race and gender, and an interaction term. For example

$$Z_1 = 1 \text{ if the subject is a female, } 0 \text{ otherwise,}$$

$$Z_2 = 1 \text{ if the subject is black, } 0 \text{ otherwise,}$$

and

$$Z_3 = Z_1 \times Z_2 \text{ if the subject is a black female, } 0 \text{ otherwise.}$$

Again the full model will be

$$b(t | Z) = b_0(t) \exp\{\beta'Z\} = b_0(t) \exp\{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3\}.$$

Note that the parameters β_i will have a different interpretation. The estimates of the parameters are obtained as

$$b_1 = -0.2484, b_2 = 0.0888, b_3 = 0.7455.$$

The complete analysis is in Table 8.2.

TABLE 8.2

Analysis of Variance Table for Race, Gender, and Interaction for the Kidney Transplant Patients Utilizing the "Breslow" Method of Handling Ties

Variables	Degrees of Freedom	Parameter Estimates	Standard Errors	Wald Chi-square	p-Values	Relative Risk
Z ₁ : Female	1	-0.2484	0.1985	1.57	0.2108	0.78
Z ₂ : Black	1	-0.0888	0.2918	0.09	0.7609	0.92
Z ₃ : Interaction	1	0.7455	0.4271	3.05	0.0809	2.11

Here the interest will center on the interaction term β_3 . However, it is instructive to see that the relative risks for black male, white male, and black female relative to white female are $\exp(-0.0888 - (-0.2484)) = 1.17$, $\exp(0 - (-0.2484)) = 1.28$, $\exp(-0.2484 - 0.0888 + 0.7455 - (-0.2484)) = 1.93$, respectively, just as we obtained for the earlier coding. These are two different coding schemes; the first treats the samples as four groups and the second treats the samples as a 2×2 factorial where interest may center on the interaction between gender and race. The interpretation of the two coding schemes are not inconsistent in that they lead to the same relative risks.

Next, we shall consider an example of an interaction between a continuous and a categorical variable.

EXAMPLE 8.2

(continued) The interaction between age and stage will involve three product terms, namely, $Z_5 = Z_1 Z_4$; $Z_6 = Z_2 Z_4$ and $Z_7 = Z_3 Z_4$, where $Z_i, i = 1, \dots, 4$ are defined as before. Thus, for a 50-year-old man with Stage II cancer, the three interaction variables will take on the following values: $Z_5 = Z_1 Z_4 = (1)(50) = 50$; $Z_6 = Z_2 Z_4 = (0)(50) = 0$ and $Z_7 = Z_3 Z_4 = (0)(50) = 0$. Other combinations of age and stage can be appropriately formed.

For this model, the estimates of the b 's are $b_1 = -7.9461, b_2 = -0.1225, b_3 = 0.8470, b_4 = -0.0026, b_5 = 0.1203, b_6 = 0.0114,$ and $b_7 = 0.0137$. The estimated variance-covariance matrix of the estimated parameters, obtained as the inverse of the Fisher information matrix, is

$$\hat{V}(b) = \begin{pmatrix} 13.529 & 2.932 & 2.956 & 0.045 & -0.191 & -0.044 & -0.045 \\ 2.932 & 6.093 & 2.957 & 0.044 & -0.044 & -0.091 & -0.044 \\ 2.956 & 2.957 & 5.884 & 0.044 & -0.044 & -0.044 & -0.086 \\ 0.045 & 0.044 & 0.044 & 0.001 & -0.001 & -0.001 & -0.001 \\ -0.191 & -0.044 & -0.044 & -0.001 & 0.003 & 0.001 & 0.001 \\ -0.044 & -0.091 & -0.044 & -0.001 & 0.001 & 0.001 & 0.001 \\ -0.045 & -0.044 & -0.086 & -0.001 & 0.001 & 0.001 & 0.001 \end{pmatrix}$$

Table 8.3 gives the analysis of variance table for this model.

TABLE 8.3
Analysis of Variance Table for Stage, Age, and the Interaction of Stage by Age for Laryngeal Cancer Patients, Utilizing the "Breslow" Method of Handling Ties

Variables	Degrees of Freedom	Parameter Estimates	Standard Errors	Wald Chi Square	p-Value
Z ₁ : Stage II	1	-7.946	3.6782	4.67	0.03
Z ₂ : Stage III	1	-0.1225	2.4683	0.003	0.96
Z ₃ : Stage IV	1	0.8470	2.4257	0.12	0.73
Z ₄ : Age	1	-0.0026	0.0261	0.01	0.92
Z ₅ : Z ₁ × Z ₄	1	0.1203	0.0523	5.29	0.02
Z ₆ : Z ₂ × Z ₄	1	0.0114	0.0375	0.09	0.76
Z ₇ : Z ₃ × Z ₄	1	0.0137	0.0360	0.14	0.70

Table 8.3 suggests that the effect of stage II on survival may be different for different ages because a local test of $\beta_5 = 0$ may be rejected (p -value = 0.02). Furthermore, it is suggested by the local tests of $\beta_6 = 0$ (p -value = 0.76) and $\beta_7 = 0$ (p -value = 0.70) that the effects of stages III and IV on survival may not be different for different ages.

To test the hypothesis that $\beta_6 = \beta_7 = 0$, we need the full $-2 \log$ likelihood for all seven parameters which is 370.155 and the reduced $-2 \log$ likelihood for the first five parameters which is 370.316. The local likelihood ratio chi-squared statistic for testing that there is no interaction between age and either stage III or IV disease ($H_0: \beta_6 = \beta_7 = 0$) is the difference between the reduced $-2 \log$ likelihood for the first five parameters minus the full $-2 \log$ likelihood for all seven parameters = $370.316 - 370.155 = 0.161$ with two degrees of freedom (p -value = 0.92). This provides strong confirmation that the latter two interaction terms may be dropped from the model and that the risks of dying for patients with Stages III and IV relative to the risk of dying for patients with Stage I does not depend on age.

In Table 8.4 the analysis of variance table for the reduced model with only an interaction between age and stage II disease is presented.

This table suggests that there is a significant interaction between age and stage II disease, that is, the relative risk of dying for a stage II

TABLE 8.4
Analysis of Variance Table for Stage, Age, and One Interaction Term (Stage II by Age) for Laryngeal Cancer Patients, Utilizing the "Breslow" Method of Handling Ties

Variables	Degrees of Freedom	Parameter Estimates	Standard Errors	Wald Chi Square	p-Value
Z ₁ : Stage II	1	-7.3820	3.4027	4.71	0.03
Z ₂ : Stage III	1	0.6218	0.3558	3.05	0.08
Z ₃ : Stage IV	1	1.7534	0.4240	17.11	<0.0001
Z ₄ : Age	1	0.0060	0.0149	0.16	0.69
Z ₅ : Z ₁ × Z ₄	1	0.1117	0.0477	5.49	0.02

patient of age Z₄ as compared to a stage I patient of the same age depends on that age. This relative risk is $\exp(\beta_1 + \beta_5 Z_4) = \exp(-7.382 + 0.1117 \text{ Age})$. For example, for a 76-year-old patient, this relative risk is 3.03 whereas for a 60-year-old it is 0.51. This linear combination of the estimated coefficients not only leads one to an estimated relative risk which depends on a patient's age at diagnosis, but also allows us to test the hypothesis that the risk of dying for stage I and II patients is the same for a given age, that is, we wish to test that the relative risk is one or, equivalently, that $\beta_1 + \beta_5(\text{age}) = 0$. To test the hypothesis that this linear combination of the parameters is zero, one forms the quadratic form based on $C = (1, 0, 0, 0, \text{age})'$. The resulting chi-squared statistic is

$$X_W^2 = \frac{(b_1 + b_5 \text{ age})^2}{V(b_1) + \text{age}^2 V(b_5) + 2 \text{ age Cov}(b_1, b_5)}$$

which has a large-sample chi-squared distribution with one degree of freedom. In this example, $V(b_1) = 11.5787$, $V(b_5) = 0.00227$ and $\text{Cov}(b_1, b_5) = -0.1607$, so for a 76-year-old person, X_W^2 equals $(1.1072)^2 / 0.2638 = 4.65$ (p -value = 0.03). For a 60-year-old we have a chi-square of 0.99 with a p -value of 0.32. This suggests that for "young" ages there is little difference in survival between stage I and II patients whereas, for older patients, those with stage II disease are more likely to die.

Practical Notes

1. A Monte Carlo study (Li et al., 1996) of the small-sample properties of the likelihood ratio, Wald, and scores tests was performed with respect to inference on a dichotomous covariate effect in a Cox proportional hazards model, as assessed by size and power considerations, under a variety of censoring fractions, sample sizes, and

hazard distributions. The general conclusion of this study was that the likelihood ratio test and Wald test performed similarly (although the likelihood test ratio test slightly outperformed the Wald test for smaller sample sizes). The score test was markedly inferior and is not recommended because it tends to inflate the size of the test. These conclusions held for a variety of censoring fractions.

2. Proc PHREG provides local tests based on the Wald statistic. Tests for contrasts are available.
3. S-Plus provides building blocks of the likelihood ratio test by running a series of models. Wald tests can be constructed using the estimated covariance matrix and parameter estimates.

8.6 Discretizing a Continuous Covariate

As we saw in the previous section the Cox model can be applied when the covariates are continuous or categorical. The interpretation of the model, however, is simplest when the covariate is a binary. Here the relative risk, $\exp(b)$, is the ratio of the risk of the event for a patient with the characteristic versus a patient without the characteristic. Often a medical investigator would like to treat a continuous covariate, X , as a binary covariate by assigning a score of 1 to subjects with large values of X and 0 to those with small values of X . This may be done to assign patients to poor- and high-risk groups based on the value of X , to aid in making graphical plots of patients with good or bad prognosis based on the binary covariate or simply to make the resulting relative risk calculations simpler for others to understand.

In most cases a major problem is determining the value of the cut point between high- and low-risk groups. In some cases this cut point can be based on biological reasoning and this is the optimal strategy for determination of the cut point. When no *a priori* information is available a "data-oriented" method is sometimes used to choose the cut point. These methods look at the distribution of the continuous covariate and divide subjects into groups based on some predetermined statistic on the covariate. For example, quite often subjects are divided into two equal groups based on whether they are larger or smaller than the sample median. These methods tend not to perform well.

In this section we will look at the "outcome-oriented" approach to this problem. Here we seek a cut point for the covariate which gives us the largest difference between individuals in the two data-defined groups. That is, for a continuous covariate, X , we seek a binary covariate Z defined by $Z = 1$ if $X \geq C$ and 0 if $X < C$, which makes the outcomes of the groups with $Z = 1$ as different from the group with $Z = 0$ as possible based on some statistic. We would also like to test the

hypothesis that this covariate in its discretized version has no effect on outcome. This test must be adjusted for the fact that we have biased the outcome of the test by considering the cut point which gives the maximum separation between the two groups.

The inference procedure we describe is due to Contal and O'Quigley (1999) and is based on the log rank test statistic discussed in section 7.3. This statistic is the score statistic from the Cox model. For the procedure we look at all possible cut points; and for each cut point, C_k , we compute the log rank statistic based on the groups defined by X being less than the cut point or greater than the cut point. That is, at each event time, t_i , we find the total number of deaths, d_i , and the total number at risk, r_i . We also find the total number of deaths with $X \geq C_k$, d_i^+ and the total number at risk with $X \geq C_k$, r_i^+ . We then compute the log rank statistic,

$$S_k = \sum_{i=1}^D \left[d_i^+ - d_i \frac{r_i^+}{r_i} \right], \quad (8.6.1)$$

where D is the total number of distinct death times.

The estimated cut point \hat{C} is the value of C_k which yields the maximum $|S_k|$. At this cut point the Cox regression model is

$$h(t|X) = h_0(t)\exp\{bZ\},$$

where $Z = 1$ if $X \geq \hat{C}$, 0 otherwise. The usual tests of $H_0: b = 0$ can not be used here since we picked the cut point \hat{C} , which is most favorable to rejecting H_0 . To compute the proper test statistic we need first to compute the quantity s^2 defined by

$$s^2 = \frac{1}{D-1} \sum_{i=1}^D \left\{ 1 - \sum_{j=1}^i \frac{1}{D-j+1} \right\} \quad (8.6.2)$$

The test statistic is then

$$Q = \frac{\max |S_k|}{s\sqrt{D-1}} \quad (8.6.3)$$

which under the null hypothesis has a limiting distribution of the supremum of the absolute value of a Brownian Bridge. For $Q > 1$ the p -value of the test is approximately equal to $2\exp\{-2Q^2\}$.

EXAMPLE 8.3

(continued) In section 1.7 we discussed a trial of 863 kidney transplant patients. We would like to examine categorizing the patients into high- or low-risk groups based on their age at transplant. We shall look at separate analyses by race and sex.

Consider first the sample of 92 black males. Here the transplants occurred at 43 distinct ages, which are potential candidates for a cut

point. There were 14 distinct death times, which gives $s^2 = 0.8268$. The maximum value of $|S_k|$ is at age 58 and $Q = 0.8029$, which gives a p -value of at least 0.30 (see Theoretical Note 3). This suggests that age is not related to outcome for black males.

The following table gives the results for other race and sex combinations. It also gives the estimated relative risk of the high-risk (Age \geq cut-point) group compared to the low-risk group. Also presented are the results of a model which treats age continuously. Figure 8.2 depicts the estimates $|S_k|$ for each of the four sex and race combinations that are used to find the estimated cut point. Here we find close agreement between the discrete model for age and the continuous model.

TABLE 8.5
Age Cut Points for Kidney Transplant Patients

Race/Sex	Discrete Model for Age				Continuous Model for Age	
	Cut Point	Q	p-value	RR(95%CI)	b(SE)	p
Black/male	58	0.8029	> 0.30	2.3(0.5 - 10.4)	0.036(0.024)	0.14
White/male	41	3.1232	< 0.001	2.6(1.6 - 4.1)	0.060(0.010)	< 0.001
Black/female	48	0.9445	> 0.30	2.6(0.8 - 8.4)	0.034(0.026)	< 0.20
White/female	36	1.9310	0.001	4.4(1.9 - 10.6)	0.042(0.012)	< 0.001

Theoretical Notes

1. Wu (2001) shows that if a test is based on the best cut point without some adjustment for multiple testing then this test rejects too often when the null hypothesis is true.
2. The method discussed here, based on the score statistic, is due to Contal and O'Quigley (1999). An alternative method, due to Jespersen (1986), is also based on the supremum of the absolute value of the log rank tests. His variance is slightly different than that presented here and in a Monte Carlo study. Wu (2001) shows that this statistic's performance is not quite as good as the Contal and O'Quigley statistic.
3. The limiting distribution of Q under the null hypothesis is the same as the supremum of the absolute value of a Brownian bridge. The p -value can be found by

$$P\{Q \geq q\} = 2 \sum_{j=1}^{\infty} (-1)^{j+1} \exp\{-2j^2 q^2\}, \quad (8.6.4)$$

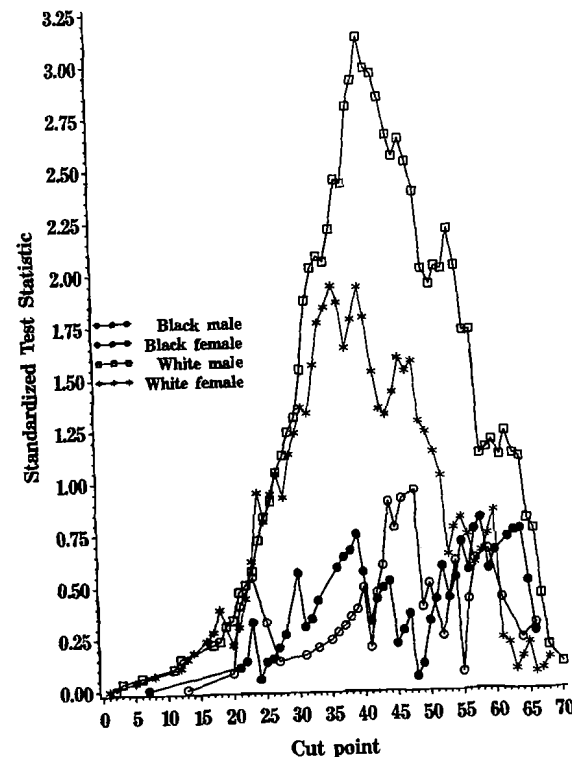


Figure 8.2

which is approximately equal to $2\exp\{-2q^2\}$ when $q > 1$. For $q \leq 1$ the p -value is at least 0.30.

Practical Notes

1. Estimation of the cut point can be performed by finding the cut point which maximizes any of the three statistics discussed in section 8.4, the Wald, Score, or likelihood ratio tests. All give approximately the correct answer.
2. Basing inference on an unadjusted Cox model with a binary covariate based on the cut-point model leads to tests which falsely reject the null hypothesis of treatment effect too often. Some correction must

be made to these tests to ensure that the overall level of inference is correct.

3. Estimation of the cut point can be performed by using existing software to identify the model which gives the maximum test statistic. The test discussed here requires additional calculations.
4. The Martingale residual plot discussed in section 11.3 can be used to check the appropriateness of discretizing a continuous covariate.

8.7 Model Building Using the Proportional Hazards Model

In an earlier example, we explored the modeling of age and stage on the survival of patients with laryngeal cancer. In many studies, a variety of explanatory factors are measured and a major question in analyzing such data sets is how to incorporate these factors in the modeling procedure.

The distinction between factors and variables is sometimes a bit vague although we shall refer to single-degree-of-freedom independent variables (such as age that is treated continuously) as either factors or variables, whereas we shall refer to multiple-degree-of-freedom independent variables (such as stage) as factors.

As mentioned at the beginning of this chapter, two distinctly different, yet important, problems in regression are i) to adjust for potential confounding (or explanatory) variables when one has a specific hypothesis in mind and the desire is to compare two or more groups with respect to survival times or ii) to predict the distribution of the time to some event from a list of explanatory variables with no particular prior hypothesis in mind. Utilizing the proportional hazards model introduced in section 2.6 and the testing procedures more fully explained in this chapter, we shall detail the approaches used for these two situations and illustrate them with two examples.

First, if one has a particular hypothesis in mind, then interest centers upon that particular hypothesis and any model building will be done to adjust that particular comparison (or comparisons) for other noncontrollable factors. Often, the other explanatory factors are simply viewed as adjusters or confounders and interest in them matters only insofar as they affect the assessment of the basic hypothesis. Examples of such possible confounders are demographic variables, such as age, gender, race, etc.; patient clinical variables at the onset of the clinical trial that may reflect the patient's condition, such as severity of disease, size of tumor, physiological variables, etc.; and, in the case of transplantation, characteristics of the donor.

The starting point of the model building process for this problem is to perform the global test of the primary hypothesis described in sections 8.3 and 8.4. This gives the investigator an impression of the simple, unadjusted relationship between the basic hypothesized factor and survival. In searching for possible confounders, it is useful to examine the simple relationship between the other explanatory factors and survival, adjusted for the factor of interest because, if there is obviously no relationship between a factor and survival, then, it is not likely to be a confounder. Thus, the next step is to consider the relationship between each of the other explanatory factors and survival, given that the factor stated in the basic hypothesis is already in the model. These local tests were described in detail in section 8.5. This process is continued by exploring the relationships between each of the remaining explanatory variables and survival, given that the factor stated in the basic hypothesis and the one next most related to survival (assuming that the basic variable is in the model) are in the model. If no significant confounders are found at any step in this process, then we stop and base our inference about the primary hypothesis on the last model. This approach is illustrated in the next example.

Another approach advocated as useful in model building, is one due to Akaike (1973) which examines the likelihood and the number of parameters included in the model. It attempts to balance the need for a model which fits the data very well to that of having a simple model with few parameters. More specifically, the Akaike information criterion (AIC), examines the statistic

$$\text{AIC} = -2 \log L + kp,$$

where p is the number of regression parameters in the model, k is some predetermined constant (which we shall take as 2), and L is the usual likelihood function. This criterion, will decrease as variables are added to the model. At some point, the criterion will increase which is a signal that the added variables are unnecessary. The AIC is reminiscent of the adjusted R^2 in least-squares regression, in that both are attempting to adjust the fit of the model by the number of parameters included. This criterion will also be recorded in the following example.

EXAMPLE 8.5

Continuing the discussion of the study of acute leukemia patients being given a bone marrow transplant, as introduced in section 1.3 and continued in Examples 4.2 and 7.5, we shall adjust the basic comparisons of the three risk groups, acute lymphoblastic leukemia (ALL), low-risk acute myelocytic leukemia (AML low-risk), and high-risk acute myelocytic leukemia (AML high-risk), so as to reduce the possible bias which may exist in making those comparisons (because this was not a randomized clinical trial). Because this chapter discusses only fixed-time covariates,

we will use only the fixed-time covariates as possible confounders in making the comparisons among risk groups.

The first step in the model-building process is the global test of the hypothesis of no difference in disease-free survival. As discussed in section 8.2, we define two binary covariates ($Z_1 = 1$ if AML low-risk; $Z_2 = 1$ if AML high-risk) for the factor of interest. The global Wald chi-squared (with two degrees of freedom) statistic is 13.01 (p -value = 0.001). The AIC for this model is 737.29.

In this example, there are two sets of factors. The first set of factors is measured only on the patient. These are Z_3 : waiting time from diagnosis to transplant, Z_4 : indicator of FAB (French-American-British) classification M4 or M5 for AML patients, and Z_5 : indicator of whether the patient was given a graft-versus-host prophylactic combining methotrexate (MTX) with cyclosporine and possibly methylprednisolone. Tests involving these factors will have one degree of freedom.

The second set of factors is based on combinations of patient and donor characteristics and cannot be described by a single covariate. These factors are sex ($Z_6 = 1$ if male donor, $Z_7 = 1$ if male recipient, and $Z_8 = Z_6 \times Z_7 = 1$ if donor and recipient are male), CMV status ($Z_9 = 1$ if donor is CMV positive, $Z_{10} = 1$ if recipient is CMV positive, and $Z_{11} = Z_9 \times Z_{10} = 1$ if donor and recipient are CMV positive), and age ($Z_{12} = \text{donor age} - 28$, $Z_{13} = \text{recipient age} - 28$, and $Z_{14} = Z_{12} \times Z_{13}$). Tests involving these factors will have three degrees of freedom.

Table 8.6 gives the local Wald tests for the six factors. Here, all models include the covariates Z_1 and Z_2 for the factor of primary interest. We find that the factor FAB classification (Z_4) has the smallest Akaike information criterion and the smallest p -value. This factor is added to the model. Table 8.7 gives the local Wald tests of all other factors not in the model with Z_1 , Z_2 , and Z_4 in the model. From this table, we see that the factor age (Z_{12} , Z_{13} , and Z_{14}) should be added to the model. Table 8.8 continues the model building by testing for factors not in the model, adjusted for risk group, FAB class, and age. In this table, we

TABLE 8.6
Local Tests for Possible Confounders, Adjusted for Risk Groups

Factor	Degrees of Freedom	Wald Chi-Square	p-Value	AIC
Waiting time (Z_3)	1	1.18	0.277	737.95
FAB class (Z_4)	1	8.08	0.004	731.02
MTX (Z_5)	1	2.03	0.155	737.35
Sex (Z_6, Z_7, Z_8)	3	1.91	0.591	741.44
CMV status (Z_9, Z_{10}, Z_{11})	3	0.19	0.980	743.10
Age (Z_{12}, Z_{13}, Z_{14})	3	11.98	0.007	733.18

see that all the local tests are nonsignificant and that the AIC is larger than that for the model with disease group, FAB class, and age alone. Thus, the model building process stops and the final model is given in Table 8.9. In this model, the local Wald test of the primary hypothesis of no difference between risk groups has a p -value of 0.003, which suggests that there is a difference in survival rates between at least two of the risk groups after adjustment for the patient's FAB class and for the donor and patient's age. Although we have used Wald tests in this

TABLE 8.7
Local Tests for Possible Confounders, Adjusted for Risk Groups and FAB Class

Factor	Degrees of Freedom	Wald Chi-Square	p-Value	AIC
Waiting time (Z_3)	1	1.18	0.277	731.68
MTX (Z_5)	1	2.05	0.152	731.06
Sex (Z_6, Z_7, Z_8)	3	0.92	0.820	736.11
CMV status (Z_9, Z_{10}, Z_{11})	3	0.02	0.999	737.00
Age (Z_{12}, Z_{13}, Z_{14})	3	13.05	0.004	725.98

TABLE 8.8
Local Tests for Possible Confounders, Adjusted for Risk Groups, FAB Class, and Age

Factor	Degrees of Freedom	Wald Chi-Square	p-Value	AIC
Waiting time (Z_3)	1	0.46	0.495	727.48
MTX (Z_5)	1	1.44	0.229	726.58
Sex (Z_6, Z_7, Z_8)	3	1.37	0.713	730.61
CMV status (Z_9, Z_{10}, Z_{11})	3	0.58	0.902	731.42

TABLE 8.9
Analysis of Variance Table for the Final Model for Bone Marrow Transplants

	Degrees of Freedom	b	SE(b)	Wald Chi-Square	p-Value
Z_1	1	-1.091	0.354	9.48	0.002
Z_2	1	-0.404	0.363	1.24	0.265
Z_4	1	0.837	0.279	9.03	0.003
Z_{12}	1	0.004	0.018	0.05	0.831
Z_{13}	1	0.007	0.020	0.12	0.728
Z_{14}	1	0.003	0.001	11.01	0.001

example, similar conclusions are obtained if the likelihood ratio statistic is used throughout.

The second situation, where regression techniques are useful, is in modeling the distribution of the time-to-some-event from a list of explanatory variables with no particular prior hypothesis in mind. Here, interest centers upon identifying a set of variables which will aid an investigator in modeling survival or identifying a set of variables which may be used in testing a hypothesis in some future study (hypothesis generating).

The starting point of model building for this problem is to perform separate global tests for each explanatory factor, as described in sections 8.3 and 8.4, so as to examine the simple relationship between the explanatory variables and survival. The purpose in this step is to ascertain which factor is most related to survival. The next step is to consider the relationship between each of the other explanatory factors (not the one identified as the most significant one) and survival, given that the factor identified as the most significant is already in the model. These local tests are also described in detail in section 8.5. This process is continued by exploring the relationship between each of the remaining explanatory factors and survival, assuming that the variable identified as the most significant one and the one next most related to survival (given the first variable is in the model) are already in the model. The p -value approach requires a significance level for entering variables into the model. This approach is illustrated in the next example. Furthermore, the Akaike information criterion may be used to assess the extent to which the investigator wishes to include variables into the model. This approach is especially useful for deciding how many variables to include.

EXAMPLE 8.6

In section 1.14 (see Example 5.4), a data set including times to weaning of breast-fed infants was described. In this example, we wish to find a model predictive of the distribution of time to weaning. Fixed-time factors measured by questionnaire include race of mother (black, white, or other), poverty status indicator, smoking status of mother at birth of child, alcohol drinking status of mother at birth of child, age of mother at child's birth, education of mother at birth of child (less than high school, high school graduate, some college), and lack of prenatal care indicator (mother sought prenatal care after third month or never sought prenatal care).

In building a model, we are mainly interested in finding factors which contribute to the distribution of the time to weaning. Because there are many ties in this data set, we shall use the "discrete" likelihood for handling ties. Table 8.10 contains the results of the single-factor

TABLE 8.10
Global Tests for Each Factor Potentially Related to Weaning Time

Factor	Degrees of Freedom	Wald Chi-Square	p-Value	AIC
Race of mother	2	8.03	0.018	5481.67
Poverty at birth	1	0.71	0.399	5486.69
Smoking	1	10.05	0.002	5477.61
Alcohol	1	2.01	0.157	5485.48
Age	1	0.15	0.698	5487.26
Education	2	6.95	0.031	5482.36
No prenatal care	1	0.16	0.687	5487.25

TABLE 8.11
Local Tests for Each Factor Potentially Related to Weaning Time, Adjusted for Mother's Smoking Status

Factor	Degrees of Freedom	Wald Chi-Square	p-Value	AIC
Race of mother	2	12.38	0.002	5469.71
Poverty at birth	1	1.42	0.234	5478.17
Alcohol	1	1.04	0.307	5478.59
Age	1	0.01	0.954	5479.61
Education	2	3.87	0.145	5477.71
No prenatal care	1	0.02	0.888	5479.59

Wald tests. Race of mother, mother's smoking status, and education of mother are all significantly related to the time to weaning in the simple regressions. The most significant factor, mother's smoking status, is added to the model, and local tests for the remaining factors are given in Table 8.11. From this table, we add the race factor to the model and perform the local tests for the remaining factors (Table 8.12). In Table 8.12, we see that all the remaining risk factors are not significant at a 5 percent significance level. If model selection criterion is based on the p -value (< 0.05) of the local tests, we would stop at this point and take, as our final model, one with two factors, smoking status and race. The ANOVA Table for this model is given in Table 8.13A. Model building based on the AIC, however, suggests adding the poverty factor to the model because the AIC with this factor is smaller than that without the factor. Proceeding based on the AIC, we find that the AIC is increased when any other factor is added to a model with race, smoking, and poverty included as factors (table not shown). The ANOVA table for the final model, based on the AIC, is in Table 8.13B.

TABLE 8.12

Local Tests for Each Factor Potentially Related to Weaning Time, Adjusted for Mother's Smoking Status and Race

Factor	Degrees of Freedom	Wald Chi-Square	p-Value	AIC
Poverty at birth	1	2.99	0.084	5468.65
Alcohol	1	1.16	0.281	5470.58
Age	1	0.19	0.660	5471.51
Education	2	2.08	0.353	5471.60
No prenatal care	1	0.03	0.854	5471.67

TABLE 8.13A

ANOVA Table for the Time to Weaning Based on the p-Value Approach

	Degrees of Freedom	b	SE(b)	Wald Chi-Square	p-Value
Smoking	1	0.308	0.081	14.34	<0.001
Race-Black	1	0.156	0.111	1.98	0.159
Race-Other	1	0.350	0.102	11.75	<0.001

TABLE 8.13B

ANOVA Table for the Time to Weaning, Based on the AIC Approach

	Degrees of Freedom	b	SE(b)	Wald Chi-Square	p-Value
Smoking	1	0.328	0.082	15.96	<0.001
Race-Black	1	0.184	0.112	2.70	0.100
Race-Other	1	0.374	0.103	13.18	<0.001
Poverty	1	-0.163	0.094	2.99	0.084

Practical Notes

1. In the example, the stepwise model building was based on the Wald statistic. The choice of this statistic is arbitrary and either the score or likelihood ratio statistic could be used. For data sets with a large number of covariates, the score statistic may be more efficient in early steps of this process because high-dimensional models need not be fit at each step. Automated procedures which can be used, when all factors are a single covariate, are available in SAS using either the score or Wald statistic.

2. The model selection procedure discussed here is a forward selection procedure. An alternative model building procedure is a backward selection procedure which starts with the model with all factors, and, at each step, removes the least significant factor from the model. A stepwise model selection procedure combines the forward and backward procedures. All three procedures for single covariate factors are available in SAS.
3. The choice of k in the AIC reflects how conservative one wishes to be in model building (larger values of k will include fewer variables).

8.8 Estimation of the Survival Function

Once we have obtained estimates of the risk coefficients β from a proportional hazards regression model, it may be of interest to estimate the survival probability for a new patient with a given set of covariates \mathbf{Z}_0 . The estimator of the survival function is based on Breslow's estimator of the baseline cumulative hazard rate derived in the Theoretical Notes of section 8.3.

To construct this estimator we, first, fit a proportional hazards model to the data and obtain the partial maximum likelihood estimators \mathbf{b} and the estimated covariance matrix $\hat{\mathbf{V}}(\mathbf{b})$ from the inverse of the information matrix. Let $t_1 < t_2 < \dots < t_D$ denote the distinct death times and d_i be the number of deaths at time t_i . Let

$$W(t_i; \mathbf{b}) = \sum_{j \in R(t_i)} \exp \left(\sum_{b=1}^p b_b Z_{jb} \right). \quad (8.8.1)$$

The estimator of the cumulative baseline hazard rate $H_0(t) = \int_0^t b_0(u) du$ is given by

$$\hat{H}_0(t) = \sum_{k \leq t} \frac{d_k}{W(t_k; \mathbf{b})}, \quad (8.8.2)$$

which is a step function with jumps at the observed death times. This estimator reduces to the Nelson-Aalen estimator of section 4.2, when there are no covariates present, and can be derived naturally using a profile likelihood construction (see Theoretical Note 2 of section 8.3). The estimator of the baseline survival function, $S_0(t) = \exp[-H_0(t)]$ is given by

$$\hat{S}_0(t) = \exp[-\hat{H}_0(t)]. \quad (8.8.3)$$

This is an estimator of the survival function of an individual with a baseline set of covariate values, $\mathbf{Z} = \mathbf{0}$. To estimate the survival function

for an individual with a covariate vector $\mathbf{Z} = \mathbf{Z}_0$, we use the estimator

$$\hat{S}(t | \mathbf{Z} = \mathbf{Z}_0) = \hat{S}_0(t) \exp(\hat{\beta}' \mathbf{z}_0). \quad (8.8.4)$$

Under rather mild regularity conditions this estimator, for fixed t , has an asymptotic normal distribution with mean $S(t | \mathbf{Z} = \mathbf{Z}_0)$ and a variance which can be estimated by

$$\hat{V}[\hat{S}(t | \mathbf{Z} = \mathbf{Z}_0)] = [\hat{S}(t | \mathbf{Z} = \mathbf{Z}_0)]^2 [Q_1(t) + Q_2(t; \mathbf{Z}_0)]. \quad (8.8.5)$$

Here,

$$Q_1(t) = \sum_{t_i \leq t} \frac{d_i}{W(t_i, \mathbf{b})^2} \quad (8.8.6)$$

is an estimator of the variance of $\hat{H}_0(t)$ if \mathbf{b} were the true value of β . Here

$$Q_2(t; \mathbf{Z}_0) = \mathbf{Q}_3(t; \mathbf{Z}_0)' \hat{V}(\mathbf{b}) \mathbf{Q}_3(t; \mathbf{Z}_0) \quad (8.8.7)$$

with \mathbf{Q}_3 the p -vector whose k th element is defined by

$$Q_{3k}(t; \mathbf{Z}_0) = \sum_{t_i \leq t} \left[\frac{W^{(k)}(t_i; \mathbf{b})}{W(t_i; \mathbf{b})} - Z_{0k} \right] \left[\frac{d_i}{W(t_i; \mathbf{b})} \right], \quad k = 1, \dots, p \quad (8.8.8)$$

where

$$W^{(k)}(t_i; \mathbf{b}) = \sum_{j \in R(t_i)} Z_{jk} \exp(\mathbf{b}' \mathbf{Z}_j)$$

Q_2 reflects the uncertainty in the estimation process added by estimating β . Here, $Q_3(t; \mathbf{Z}_0)$ is large when \mathbf{Z}_0 is far from the average covariate in the risk set. Using this variance estimate, pointwise confidence intervals for the survival function can be constructed for $S(t | \mathbf{Z} = \mathbf{Z}_0)$ using the techniques discussed in section 4.3. Again, the log-transformed or arcsine-square-root-transformed intervals perform better than the naive, linear, confidence interval.

EXAMPLE 8.2

(continued) We shall estimate the survival functions for survival after detection of laryngeal cancer based on the Cox regression model summarized in Table 8.1. Here, we wish to produce a curve for each of the four stages of disease. Because an adjustment for age is included in our model, we shall provide an estimate for a sixty-year-old male. The baseline survival function S_0 , is estimated directly from Eqs. (8.8.2) and (8.8.3). The estimate of survival for a stage I cancer patient (of age 60 at diagnosis) is $S_0(t) \exp(0.0189 \times 60)$; for a stage II patient $S_0(t) \exp(0.0189 \times 60 + 0.1386)$; for a stage III patient $S_0(t) \exp(0.0189 \times 60 + 0.6383)$; and for a stage IV patient $S_0(t) \exp(0.0189 \times 60 + 1.6931)$. Figure 8.3 shows the estimates of the four survival curves.

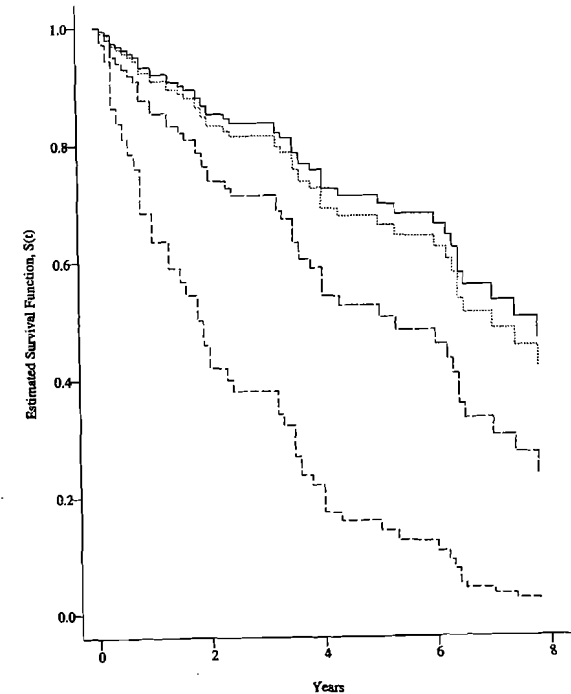


Figure 8.3 Estimated survival functions for a 60 year old larynx cancer patient. Stage I Cancer (——) Stage II Cancer (— — —) Stage II Cancer (— · — ·) Stage IV (— · — ·)

At five years, the estimated survival probabilities for a 60-year-old are 0.7031 for a Stage I patient, 0.6672 for a Stage II patient, 0.5132 for a Stage III patient, and 0.1473 for a Stage IV patient. Using Eqs. (8.8.5)–(8.8.8), we find that the standard errors of these estimators are 0.0737, 0.1059, 0.0949, and 0.0996, respectively. At 5 years, 95% confidence intervals for the survival function, based on the log transformation (4.3.2), are (0.5319, 0.8215), (0.4176, 0.8290), (0.3171, 0.6788), and (0.0218, 0.3834), for stages I, II, III, and IV, respectively.

Practical Notes

1. An alternative estimator of the baseline hazard rate has been proposed by Kalbfleisch and Prentice (1973). When there is at most a

single death at each time, this estimator is given by

$$\hat{H}_0(t) = \sum_{t_i \leq t} \left[1 - \left(1 - \frac{\delta_i \exp(\mathbf{b}'\mathbf{Z}_{t_i})}{W(t_i; \mathbf{b})} \right)^{\exp(-b'z_i)} \right]$$

When there are ties between the death times, then, the estimator is found numerically by solving a system of equations. This statistic is used in SAS.

- Alternative estimators of the baseline survival function are the product integral of the Breslow estimator of the cumulative hazard rate, which, in the case of no tied deaths, is given by

$$\hat{S}_2(t|0) = \prod_{t_i \leq t} \left(1 - \frac{\delta_i}{W(t_i; \mathbf{b})} \right),$$

or the product integral of the Kalbfleisch and Prentice estimator (see Note 1 above)

$$\hat{S}_3(t|0) = \prod_{t_i \leq t} \left[1 - \frac{\delta_i \exp(\mathbf{b}'\mathbf{Z}_{t_i})}{W(t_i; \mathbf{b})} \right]^{\exp(-b'z_i)}$$

Each of these can be used in Eq. (8.8.4) to obtain an estimator of the survival for an individual with a covariate vector \mathbf{Z}_0 . An alternative estimator to $S(t | \mathbf{Z}_0)$ is given by first adjusting $\hat{H}_0(t)$ by the factor $\exp(\mathbf{b}'\mathbf{Z}_0)$ and, then, constructing a product-limit estimator based on $\hat{H}(t | \mathbf{Z}_0) = \hat{H}(t) \exp(\mathbf{b}'\mathbf{Z}_0)$ given by

$$\hat{S}_i(t | \mathbf{Z}_0) = \prod_{t_i \leq t} \left[1 - \frac{\delta_i \exp(\mathbf{b}'\mathbf{Z}_0)}{W(t_i; \mathbf{b})} \right]$$

Under rather mild regularity conditions, each of the four estimators of $S(t | \mathbf{Z}_0)$ is asymptotically normal with the correct mean and a variance estimated by (8.8.5).

- The estimators S_2 and S_4 can take negative values. This is only a problem when the risks sets are small in the right-hand tail of the estimates. Typically, this happens when one is attempting to predict survival for a covariate value which is extreme as compared to the covariate values of those remaining at risk when the prediction is being made. The negative value is a signal to the investigator that predictions for this covariate value should not be made in this region.
- The SAS procedure PHREG uses the Kalbfleisch and Prentice estimator described in Note 1. A recent Monte Carlo study by Andersen and Klein (1996) shows that this estimator has a larger bias and mean-squared error than the other three estimators of survival. Breslow's estimator is also available in SAS.
 - S-Plus has both the Breslow and Kalbfleisch and Prentice estimator available in the function `surv.fit`.

- As was the case for confidence intervals for the survival function discussed in Chapter 4, Andersen and Klein (1996) show that the log-transformed confidence interval for $S(t | \mathbf{Z}_0)$ seems to work best, and the arcsine-square-root confidence interval is a close second. The routine use of the linear confidence interval is not recommended.
- Based on an extensive Monte Carlo study by Andersen and Klein (1996), it can be shown that the estimators S_2 and S_4 have the smallest bias and are recommended. The estimator S_3 , available in SAS, seems to perform quite poorly for continuous and mixed, continuous covariate models.

8.9 Exercises

8.1

In section 1.10, times to death or relapse (in days) are given for 23 non-Hodgkin's lymphoma (NHL) patients, 11 receiving an allogenic (Allo) transplant from an HLA-matched sibling donor and 12 patients receiving an autologous (Auto) transplant. Also, data on 20 Hodgkin's lymphoma (HOD) patients, 5 receiving an allogenic (Allo) transplant from an HLA-matched sibling donor and 15 patients receiving an autologous (Auto) transplant is given.

- Treating NHL Allo as the baseline hazard function, state the appropriate coding which would allow the investigator to test for any difference in survival functions for the four groups, treating them as four independent groups.
- Treating NHL Allo as the baseline hazard function, state the appropriate coding which would allow the investigator to test for an interaction between type of transplant and disease type using main effects and interaction terms.
- Suppose that we have the following model for the hazard rates in the four groups:

$$h(t | \text{NHL Allo}) = b_0(t)$$

$$h(t | \text{HOD Allo}) = b_0(t) \exp(2)$$

$$h(t | \text{NHL Auto}) = b_0(t) \exp(1.5)$$

$$h(t | \text{HOD Auto}) = b_0(t) \exp(.5)$$

What are the risk coefficients, β_i , $i = 1, 2, 3$, for the interaction model in part b?

8.2

In section 1.6 a study is described which evaluates a protocol change in disinfectant practices in a large midwestern university medical center. Of primary interest in the study is a comparison of two methods of

body cleansing. The first method, used exclusively from January 1983 to June 1984, consisted of a routine bathing care method (initial surface decontamination with 10% povidone-iodine followed with regular bathing with Dial soap). From June 1984 to the end of the study period in December 1985, body cleansing was initially performed using 4% chlorhexidine gluconate. Eighty-four patients were in the group who received the new bathing solution, chlorhexidine, and 70 patients served as the control group who received routine bathing care, povidone-iodine. Included in the data set is a covariate that measures the total surface area burned. The data is reported on our web site.

State the appropriate coding which would allow the investigator to test for:

- any difference in survival functions for the two groups.
- any difference in survival functions for the two groups adjusting for total area burned.

8.3 In section 1.11, a study was conducted on the effects of ploidy on the prognosis of patients with cancer of the tongue. Tissue samples were examined to determine if the tumor had a aneuploid or diploid DNA profile. Times to death for these two groups of patients are recorded in Table 1.6. To analyze this data create a single indicator variable, Z , which reflects the type of tumor.

- Find the p -value of a test of the hypothesis of no effect of ploidy on survival using the score test and the Breslow method of handling ties.
- Estimate β and its standard error using the Breslow method of handling ties. Find a 95% confidence interval for the relative risk of death of an individual with an aneuploid tumor as compared to an individual with a diploid tumor.
- Repeat (a) using the likelihood test. Compare your answer to that of part a.
- Repeat (a) using the Wald test. Compare your answer to those in parts a and c.

8.4 In Exercise 7 of Chapter 7, three different treatments were administered to rats who had F98 glioma cells implanted into their brains. The data for the three groups of rats lists the death times (in days) in that exercise. Create two dummy variables, $Z_1 = 1$ if animal is in the "radiation only" group, 0 otherwise; $Z_2 = 1$ if animal is in the "radiation plus BPA" group, 0 otherwise. Use the Breslow method of handling ties in the problems below.

- Estimate β_1 and β_2 and their respective standard errors. Find a 95% confidence interval for the relative risk of death of an animal radiated only compared to an untreated animal.

- Test the global hypothesis of no effect of either radiation or radiation plus BPA on survival. Perform the test using all the three tests (Wald, likelihood ratio, and score test).
- Test the hypothesis that the effect a radiated only animal has on survival is the same as the effect of radiation plus BPA (i.e., Test $H_0 : \beta_1 = \beta_2$).
- Find an estimate and a 95% confidence interval for the relative risk of death for a radiation plus BPA animal as compared to a radiated only animal.
- Test the hypothesis that any radiation given as a treatment (either radiation alone or with BPA) has a different effect on survival than no radiation. Use the likelihood ratio test.
- Repeat part (e) using a Wald test.

8.5 Using the data set in Exercise 1, using the Breslow method of handling ties,

- Analyze the data by performing a global test of no effect of group as defined in Exercise 8.1(a) on survival. Construct an ANOVA table to summarize estimates of the risk coefficients and the results of the one degree of freedom tests for each covariate in the model.
- Repeat part (a) using the coding as described in Exercise 8.1(b). Furthermore, test the hypothesis of disease type by transplant interaction using a likelihood ratio test based on this coding. Repeat using the Wald test.
- Find point estimates and 95% confidence intervals for the relative risk of death for an NHL Auto transplant patient as compared to an NHL Allo transplant patient.
- Find the p -value of a test of the hypothesis that the hazard rates are the same for HOD Allo transplants and NHL Allo patients, using the Wald test. Repeat a similar test for Auto patients.
- Test the hypothesis, using the Wald test, that the hazard rates for Auto transplant and Allo transplant patients are the same for each disease group against the alternative that the hazard rates for Auto transplant and Allo transplant patients for at least one group are different using a two-degree of freedom test of $H_0 : h(t | \text{NHL Allo}) = h(t | \text{NHL Auto})$ and $H_0 : h(t | \text{HOD Allo}) = h(t | \text{HOD Auto})$.

8.6 In section 1.13, data on the time to hospitalization of pneumonia in young children was discussed. The data is presented on our web site. In the sample there were 3,470 annual personal interviews. An investigator is interested in assessing race, poverty status, and their interaction on time to hospitalization of pneumonia. Use the discrete method for handling ties to answer the following questions.

- Estimate the parameters of your model and their standard errors. Construct and interpret an "ANOVA" table for this model.

- (b) Provide point estimates and 95% confidence intervals for the relative risk of hospitalization for pneumonia for a person raised in poverty relative to a person not raised in poverty for each race.
- (c) Test that blacks raised in poverty have a different hospitalization for pneumonia rate than whites not raised in poverty.
- 8.7** In section 1.6 a study is described which evaluates the relationship of various covariates to staphylococcus infection in a large midwestern university medical center (see Exercise 8.2). One of the covariates recorded in the data set is the total surface area burned. Use Breslow's method for handling ties to answer the following questions.
- (a) Find the optimal cutpoint to categorize patients into high- or low-risk groups for staphylococcus infection based on their total surface area burned for each disinfectant practice.
- (b) Test the hypothesis that there is a difference in times to infection for high- and low-risk groups using the cutpoints obtained in (a). Using the cut points obtained in (a) find the relative risk of the high-risk group compared to the low-risk group for each disinfectant practice.
- (c) Analyze the data using total surface area burned as a continuous variable. Give the parameter estimate, standard error, and relative risk for total surface area burned. Compare with the answer in (b).
- 8.8** In section 1.3, data gathered from a multicenter trial of patients in three groups (ALL, AML low-risk, and AML high-risk) was followed after transplantation until relapse, death, or end of study. One of the covariates recorded in the data set is the waiting time to transplant (in days). Use Breslow's method for handling ties in the following.
- (a) You are asked to categorize patients into high- or low-risk groups for disease-free survival based on the waiting time to transplant variable for the ALL group.
- (b) Analyze the data using waiting time to transplant as a categorized variable using the cut point obtained in (a). Give the parameter estimate, standard error, and relative risk of the high-risk group compared to the low-risk group for the ALL group.
- (c) Analyze the data using waiting time to transplant as a continuous variable. Give the parameter estimate, standard error, and relative risk for waiting time to transplant for the ALL group. Compare with answer in (b).
- 8.9** Use the Breslow method for handling ties and the Wald test in the following.
- (a) Using the data set in section 1.6, test the hypothesis that the distributions of the times to staphylococcus infection are the same in the two disinfectant groups.
- (b) Test the hypothesis that the distributions of the times to staphylococcus infection are the same in the two disinfectant groups adjusting for the total area burned, Z_4 . Compare your results to those in part a.
- (c) Also available in the data set is information on other factors that may be associated with the timing of staphylococcus infection. Some of these factors are gender, race, total surface area burned, and type of burn (chemical, scald, electrical, flame). For each factor create a set of fixed-time covariates. Test the hypothesis that the times to staphylococcus infection are the same for the two disinfectant groups using a model which adjusts for each of these factors.
- (d) Since one is primarily interested in comparing the two bathing solutions, interest will center upon building a model with the view of testing that particular comparison adjusting for the other non-controllable factors in part (c). Using a forward selection approach, build such a model using the p -value approach. Based on the final model, test the hypothesis of primary interest.
- 8.10** In section 1.3, several event times are described for patients receiving a bone marrow transplant for leukemia. Consider the time to development of acute graft-versus-host disease (AGVHD). As a prophylactic treatment, patients at two of the hospitals were given a treatment combining methotrexate (MTX) with cyclosporine and possibly methylprednisilone. Patients at the other hospitals were not given methotrexate but rather a combination of cyclosporine and methylprednisilone. Of primary interest in studying AGVHD is a test of the effectiveness of the MTX regime to prevent AGVHD. Use Breslow's method for handling ties to answer the following exercises.
- (a) Using an appropriate Cox model test the hypothesis of no difference in the rate of development of AGVHD between MTX and no MTX patients. Find a point estimate and a 95% confidence interval for the relative risk of AGVHD for patients on the MTX protocol as compared to those not given MTX.
- (b) Patients were also grouped into risk categories based on their status at the time of transplantation. These categories were as follows: acute lymphoblastic leukemia (ALL) with 38 patients and acute myeloid leukemia (AML). The latter category was further subdivided into low-risk—first remission (54 patients) and high-risk—second remission or untreated first relapse or second or greater relapse or never in remission (45 patients). Test the hypothesis of interest (no effect of MTX on development of AGVHD) adjusting for the three disease categories.
- (c) Test for the possibility of an interaction effect on AGVHD between the disease categories and the use MTX.
- (d) Using the factors of age, sex, CMV status, FAB class, waiting time to transplant, and disease category as defined in Example 8.5, find the best model to test the primary hypothesis of no MTX effect on

the occurrence of AGVHD. Test the primary hypothesis and find an estimate of the relative risk of occurrence of AGVHD for an MTX patient as compared to a non-MTX patient.

- 8.11** In section 1.13, data gathered from annual personal interviews conducted for the National Longitudinal Survey of Youth (NLSY) from 1979 through 1986 was presented. This data was used to study whether or not the mother's feeding choice protected the infant against hospitalized pneumonia in the first year of life. Ages of young children at the time they were hospitalized with pneumonia were recorded as well as the observed ages of those infants that were not hospitalized with pneumonia during the study period. The data is available from our web site, which can be reached via the authors' pages at <http://www.springer-ny.com>. Use the discrete method for handling ties in the following.
- (a) Consider the dummy variable $Z = 1$ if infants were breast fed at birth, 0 if infants were never breast fed, and test the hypothesis $H_0: \beta = 0$, i.e., the survival functions for the two types of breast feeding are equal, using the score, likelihood ratio, and Wald tests. Find the estimate of β , b , the standard error of b , and the relative risk using the Wald test.
- (b) Also available in the data set is information on other factors that may be associated with the timing of hospitalized pneumonia. These factors are age of the mother at the infant's birth, rural-urban environment of the mother, use of alcohol by the mother (no drinks, less than one drink, 1–2 drinks, 3–4 drinks, or more than 4 drinks per month), mother's cigarette use (none, less than 1 pack/day, 1 or more pack/day), region of country (northeast, north central, south, or west), birthweight of infant (less than 5.5 lbs or 5.5 lbs or more), poverty status of mother (yes/no), race of mother (white, black, or other), or number of siblings of infant. For each factor create a set of fixed-time covariates. Test the hypothesis that the times to hospitalized pneumonia are the same for the two feeding groups adjusting for each of these factors in a separate model using the Wald test.
- (c) Since one is primarily interested in comparing the two types of breast feeding, interest will center upon building a model with the view of testing the particular comparison of interest adjusting for the other noncontrollable fixed covariates in part b. Build such a model using the AIC approach and the Wald test.
- (d) Summarize your findings from this data set.
- 8.12** A major problem in certain sub-populations is the occurrence of sexually transmitted diseases (STD). Even if one ignores the lethal effects of the acquired immune deficiency syndrome, other STD's still have a significant impact on the morbidity of the community. Two of these STD's are the focus of this investigation—gonorrhea and chlamydia. Both of

these diseases can be prevented and effectively treated. The purpose of the study described in section 1.12 is to identify those factors which are related to time until reinfection by either gonorrhea or chlamydia given a patient with an initial infection of gonorrhea or chlamydia. The data for this study is available from our web site.

Possible factors related to reinfection are the individual's race (black/white), marital status (divorced/separated, married, single), age at time of initial infection, years of schooling, initial infection type (gonorrhea, chlamydia, both), number of partners within the last 30 days, oral sex within the last year, rectal sex within the past year, presence of symptoms (abdominal pain, discharge, dysuria, itch, lesion, rash, lymph node involvement), and condom use. If the factors that are related to a greater risk of reinfection can be identified, then interventions could be targeted to those individuals who are at greatest risk for reinfection. Use regression techniques to find those factors which are most predictive of the distribution of the time until reinfection from this list of fixed explanatory factors with no particular prior hypothesis in mind. Build such a model using the p -value approach. Use the Breslow method for handling ties and the Wald test in the model building.

- 8.13** Find 95% confidence intervals for the survival functions for the two bathing solutions at 20 days for a patient with 25% of total surface area of body burned, using data in Section 1.6.
- 8.14** (a) Estimate the survival functions of the time to AGVHD for the MTX and no MTX treatment groups discussed in Exercise 8.10, adjusted for disease category. Provide a separate estimate for each disease group.
- (b) Find 95% confidence intervals for the survival functions for the two patient treatment groups at 80 days for AML high-risk patients.