COXREG

Cox (1972) first suggested the models in which factors related to lifetime have a multiplicative effect on the hazard function. These models are called proportional hazards (PH) models.

Under the proportional hazards assumption, the hazard function *h* of *t* given *X* is of the form

$$h(t\,|\,\mathbf{x}) = h_0(t)\,e^{\mathbf{x}'\beta} \qquad \textbf{(1)}$$

where **x** is a known vector of regressor variables associated with the individual, β is a vector of unknown parameters, and $h_0(t)$ is the baseline hazard function for an individual with $\mathbf{x}=0$. Hence, for any two covariates sets $\mathbf{x}_1$ and $\mathbf{x}_2$, the log hazard functions $h(t\,|\,\mathbf{x}_1)$ and $h(t\,|\,\mathbf{x}_2)$ should be parallel across time.

When a factor does not affect the hazard function multiplicatively, stratification may be useful in model building. Suppose that individuals can be assigned to one of *m* different strata, defined by the levels of one or more factors. The hazard function for an individual in the *j*th stratum is defined as

$$h_j(t\,|\,\mathbf{x}) = h_{0j}(t)\,e^{\mathbf{x}'\beta} \qquad \textbf{(2)}$$

There are two unknown components in the model: the regression parameter β and the baseline hazard function $h_{0j}(t)$. The estimation for the parameters is described below.

Estimation

We begin by considering a nonnegative random variable *T* representing the lifetimes of individuals in some population. Let $f(t\,|\,\mathbf{x})$ denote the probability density function (pdf) of *T* given a regressor **x** and let $S(t\,|\,\mathbf{x})$ be the survivor function (the probability of an individual surviving until time *t*). Hence

$$S(t|\mathbf{x}) = \int_t^{\infty} f(u|\mathbf{x})\,du \qquad \textbf{(3)}$$

The hazard $h(t|\mathbf{x})$ is then defined by

$$h(t|\mathbf{x}) = \frac{f(t|\mathbf{x})}{S(t|\mathbf{x})} \qquad \textbf{(4)}$$

Another useful expression for $S(t|\mathbf{x})$ in terms of $h(t|\mathbf{x})$ derived from equations **(3)** and **(4)** is

$$S(t|\mathbf{x}) = \exp\left(-\int_0^t h(u|\mathbf{x})\,du\right) \qquad \textbf{(5)}$$

Thus,

$$\ln S(t|\mathbf{x}) = -\int_0^t h(u|\mathbf{x})\,du \qquad \textbf{(6)}$$

For some purposes, it is also useful to define the cumulative hazard function

$$H(t|\mathbf{x}) = \int_0^t h(u|\mathbf{x})\,du = -\ln S(t|\mathbf{x}) \qquad \textbf{(7)}$$

Assume that the hazard function has the form of equation **(1)**. The survivor function can be written as

$$S(t|\mathbf{x}) = \left[S_0(t)\right]^{\exp(\mathbf{x}'\beta)} \qquad \textbf{(8)}$$

where $S_0(t)$ is the baseline survivor function defined by

$$S_0(t) = \exp\left(-H_0(t)\right) \qquad \textbf{(9)}$$

and

$$H_0(t) = \int_0^t h_0(u)\,du$$

Some relationships between $S(t|\mathbf{x})$, $H(t|\mathbf{x})$ and $H_0(t)$, $S_0(t)$ and $h_0(t)$ which will be used later are

$$\ln S(t|\mathbf{x}) = -H(t|\mathbf{x}) = -\exp(\mathbf{x}'\beta)H_0(t) \quad \textbf{(10)}$$

$$\ln\left(-\ln S(t|\mathbf{x})\right) = \mathbf{x}'\beta + \ln H_0(t) \quad \textbf{(11)}$$

To estimate the survivor function $S(t|\mathbf{x})$, we can see from equation **(8)** that there are two components, β and $S_0(t)$, which need to be estimated. The approach we use here is to estimate β from the partial likelihood function and then to maximize the full likelihood for $S_0(t)$.

Estimation of Beta

Assume that

- There are *m* levels for the stratification variable.
- Individuals in the same stratum have proportional hazard functions.
- The relative effect of the regressor variables is the same in each stratum.

Let $t_{j1} < \cdots < t_{jk_j}$ be the observed uncensored failure time of the n_j individuals in the *j*th stratum and $x_{j1}, \ldots, x_{jk_j}$ be the corresponding covariates. Then the partial likelihood function is defined by

$$L(\beta) = \prod_{j=1}^{m} \prod_{i=1}^{k_j} \frac{e^{\mathbf{S}'_{ji}\beta}}{\left(\sum_{l \in R_{ji}} w_l e^{\mathbf{x}'_l \beta}\right)^{d_{ji}}} \quad \textbf{(12a)}$$

where d_{ji} is the sum of case weights of individuals whose lifetime is equal to t_{ji} and $\mathbf{S}_{ji}$ is the weighted sum of the regression vector **x** for those d_{ji} individuals, w_l is the case weight of individual *l*, and R_{ji} is the set of individuals alive and uncensored just prior to t_{ji} in the *j*th stratum. Thus the log-likelihood arising from equation **(12a)** is

$$l = \ln L(\beta) = \sum_{j=1}^{m}\sum_{i=1}^{k_j}\mathbf{S}'_{ji}\beta - \sum_{j=1}^{m}\sum_{i=1}^{k_j} d_{ji}\ln\left(\sum_{l\in R_{ji}} w_l e^{\mathbf{x}'_l\beta}\right) \qquad \textbf{(12b)}$$

and the first derivatives of *l* are

$$D_{\beta_r} = \frac{\partial l}{\partial \beta_r} = \sum_{j=1}^{m}\sum_{i=1}^{k_j}\left(S_{ji}^{(r)} - d_{ji}\frac{\sum_{l\in R_{ji}} w_l x_{lr} e^{\mathbf{x}'_l\beta}}{\sum_{l\in R_{ji}} w_l e^{\mathbf{x}'_l\beta}}\right), \quad r = 1,\dots,p \qquad \textbf{(13)}$$

In equation **(13)**, $S_{ji}^{(r)}$ is the *r*th component of $\mathbf{S}_{ji} = \left(S_{ji}^{(1)},\dots,S_{ji}^{(p)}\right)'$. The maximum partial likelihood estimate (MPLE) of β is obtained by setting $\frac{\partial l}{\partial \beta_r}$ equal to zero for $r = 1,\dots,p$, where *p* is the number of independent variables in the model. The equations $\frac{\partial l}{\partial \beta_r} = 0 \quad (r = 1,\dots,p)$ can usually be solved by using the Newton-Raphson method.

Note that from equation **(12a)** the partial likelihood function $L(\beta)$ is invariant under translation. All the covariates are centered by their corresponding overall mean. The overall mean of a covariate is defined as the sum of the product of weight and covariate for all the censored and uncensored cases in each stratum. For notational simplicity, $\mathbf{x}_l$ used in the Estimation Section denotes centered covariates.

Three convergence criteria for the Newton-Raphson method are available:

- Absolute value of the largest difference in parameter estimates between iterations (δ) divided by the value of the parameter estimate for the previous iteration; that is,

$$\text{BCON} = \left| \frac{\delta}{\text{parameter estimate for previous iteration}} \right|$$

- Absolute difference of the log-likelihood function between iterations divided by the log-likelihood function for previous iteration.

- Maximum number of iterations.

The asymptotic covariance matrix for the MPLE $\hat{\beta} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)'$ is estimated by $\mathbf{I}^{-1}$ where $\mathbf{I}$ is the information matrix containing minus the second partial derivatives of $\ln L$. The (*r*, *s*)-th element of **I** is defined by

$$\mathbf{I}_{rs} = -E \frac{\partial^2}{\partial \beta_r \partial \beta_s} \ln L$$

$$= \sum_{j=1}^{m} \sum_{i=1}^{k_j} d_{ji} \left[\frac{\sum_{l \in R_{ji}} w_l x_{ls} x_{lr} e^{\mathbf{x}_l' \beta}}{\sum_{l \in R_{ji}} w_l e^{\mathbf{x}_l' \beta}} - \frac{\left(\sum_{l \in R_{ji}} w_l x_{lr} e^{\mathbf{x}_l' \beta} \right) \left(\sum_{l \in R_{ji}} w_l x_{ls} e^{\mathbf{x}_l' \beta} \right)}{\left(\sum_{l \in R_{ji}} w_l e^{\mathbf{x}_l' \beta} \right)^2} \right] \quad \textbf{(14)}$$

We can also write **I** in a matrix form as

$$I_{rs} = \sum_{j=1}^{m} \sum_{i=1}^{k_j} d_{ji} \left(x'\left(t_{ji}\right) \right) V\left(t_{ji}\right) \left(x\left(t_{ji}\right) \right)$$

where $\mathbf{x}(t_{ji})$ is a $n_{ji} \times p$ matrix which represents the *p* covariate variables in the model evaluated at time t_{ji}, n_{ji} is the number of distinct individuals in R_{ji}, and $\mathbf{V}(t_{ji})$ is a $n_{ji} \times n_{ji}$ matrix with the *l*th diagonal element $v_{ll}(t_{ji})$ defined by

$$v_{ll}(t_{ji}) = p_l(t_{ji}) w_l - \left(w_l p_l(t_{ji})\right)^2$$

$$p_l(t_{ji}) = \frac{\exp(\mathbf{x}_l' \hat{\beta})}{\sum_{h \in R_{ji}} w_h \exp(\mathbf{x}_h' \hat{\beta})}$$

and the (*l*, *k*) element $v_{lk}(t_{ji})$ defined by

$$v_{lk}(t_{ji}) = w_l p_l(t_{ji}) \times w_k p_k(t_{ji})$$

Estimation of the Baseline Function

After the MPLE $\hat{\beta}$ of β is found, the baseline survivor function $S_{0j}(t)$ is estimated separately for each stratum. Assume that, for a stratum, $t_1 < \cdots < t_k$ are observed lifetimes in the sample. There are n_i at risk and d_i deaths at t_i, and in the interval $[t_{i-1}, t_i)$ there are λ_i censored times. Since $S_0(t)$ is a survivor function, it is non-increasing and left continuous, and thus $\hat{S}_0(t)$ must be constant except for jumps at the observed lifetimes $t_1, \ldots, t_k$.

Further, it follows that

$$\hat{S}_0(t_1) = 1$$

and

$$\hat{S}_0(t_i +) = \hat{S}_0(t_{i+1})$$

Writing $\hat{S}_0(t_i+) = p_i \; (i=1,\ldots,k)$, the observed likelihood function is of the form

$$L_1 = \prod_{i=1}^{k} \left\{ \prod_{l \in D_i} \left(p_{i-1}^{\exp(\mathbf{x}_l'\beta)} - p_i^{\exp(\mathbf{x}_l'\beta)} \right)^{w_l} \prod_{l \in C_i} \left(p_{i-1}^{\exp(\mathbf{x}_l'\beta)} \right)^{w_l} \right\} \prod_{l \in C_{k+1}} \left(p_k^{\exp(\mathbf{x}_l'\beta)} \right)^{w_l}$$

where D_i is the set of individuals dying at t_i and C_i is the set of individuals with censored times in $[t_{i-1}, t_i)$. (Note that if the last observation is uncensored, C_{k+1} is empty and $p_k = 0$.)

If we let $\alpha_i = p_i / p_{i-1} \; (i=1,\ldots,k)$, L_1 can be written as

$$L_1 = \prod_{i=1}^{k} \prod_{l \in D_i} \left(1 - \alpha_i^{\exp(\mathbf{x}_l'\beta)} \right)^{w_l} \prod_{l \in R_i - D_i} \alpha_i^{w_l \exp(\mathbf{x}_l'\beta)}$$

Differentiating $\ln L_1$ with respect to $\alpha_1,\ldots,\alpha_k$ and setting the equations equal to zero, we get

$$\sum_{l \in D_i} \frac{w_l \exp(\mathbf{x}_l'\beta)}{1 - \alpha_i^{\exp(\mathbf{x}_l'\beta)}} = \sum_{l \in R_i} w_l \exp(\mathbf{x}_l'\beta) \qquad i = 1,\ldots,k \tag{15}$$

We then plug the MPLE $\hat{\beta}$ of β into equation **(15)** and solve these *k* equations separately.

There are two things worth noting:

- If any $|D_i| = 1$, $\hat{\alpha}_i$ can be solved explicitly.

$$\hat{\alpha}_i = \left[1 - \frac{w_i \exp(\mathbf{x}_i'\hat{\beta})}{\sum_{l \in R_i} w_l \exp(\mathbf{x}_l'\hat{\beta})} \right]^{\exp(-\mathbf{x}_i'\hat{\beta})} \tag{16}$$

- If $|D_i| > 1$, equation **(7)** must be solve iteratively for $\hat{\alpha}_i$. A good initial value for $\hat{\alpha}_i$ is

$$\hat{\alpha}_i = \exp\left(\frac{-d_i}{\sum_{l \in R_i} w_l \exp\left(\mathbf{x}_l'\hat{\beta}\right)}\right) \tag{17}$$

where $d_i = \sum_{l \in D_i} w_l$ is the weight sum for set D_i. (See Lawless, 1982, p. 361.)

Once the $\hat{\alpha}_i$, $i = 1, \ldots, k$ are found, $S_0(t)$ is estimated by

$$\hat{S}_0(t) = \prod_{i:(t_i < t)} \hat{\alpha}_i \tag{18}$$

Since the above estimate of $S_0(t)$ requires some iterative calculations when ties exist, Breslow (1974) suggests using equation **(17)** as an estimate for α_i; however, we will use this as an initial estimate.

The asymptotic variance for $-\ln \hat{S}_0(t)$ can be found in Chapter 4 of Kalbfleisch and Prentice (1980). At a specified time *t*, it is consistently estimated by

$$\operatorname{var}\left(-\ln \hat{S}_0(t)\right) = \sum_{t_i < t} |D_i| \left(\sum_{l \in R_i} w_l \exp\left(\mathbf{x}_l'\hat{\beta}\right)\right)^{-2} + \mathbf{a}'\mathbf{I}^{-1}a \tag{19}$$

where *a* is a $p \times 1$ vector with the *j*th element defined by

$$\sum_{t_i < t} |D_i| \frac{\sum_{l \in R_i} w_l x_{lj} \exp\left(\mathbf{x}_l'\hat{\beta}\right)}{\left(\sum_{l \in R_i} w_l \exp\left(\mathbf{x}_l'\hat{\beta}\right)\right)^2}$$

and **I** is the information matrix. The asymptotic variance of $\hat{S}(t|x)$ is estimated by

$$e^{2\mathbf{x}'\hat{\beta}}\left(\hat{S}(t|\mathbf{x})\right)^2 \operatorname{var}\left(-\ln \hat{S}_0(t)\right) \quad \textbf{(20)}$$

Selection Statistics for Stepwise Methods

COX REGRESSION offers the same methods for variable selection as LOGISTIC REGRESSION. For the details of these methods, and stepwise algorithms, see the LOGISTIC REGRESSION chapter. Here we will only define the three removal statistics—Wald, LR, and Conditional—and the Score entry statistic.

Score Statistic

The score statistic is calculated for every variable not in the model to decide which variable should be added to the model. First we compute the information matrix **I** for all eligible variables based on the parameter estimates for the variables in the model and zero parameter estimates for the variables not in the model. Then we partition the resulting **I** into four submatrices as follows:

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \quad \textbf{(21)}$$

where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are square matrices for variables in the model and variables not in the model, respectively, and $\mathbf{A}_{12}$ is the cross-product matrix for variables in and out. The score statistic for variable $\mathbf{x}_i$ is defined by

$$\mathbf{D}'_{x_i}\mathbf{B}_{22,i}\mathbf{D}_{x_i}$$

where $\mathbf{D}_{x_i}$ is the first derivative of the log-likelihood with respect to all the parameters associated with x_i and $\mathbf{B}_{22,i}$ is equal to $\left(\mathbf{A}_{22,i} - \mathbf{A}_{21,i}\mathbf{A}_{11}^{-1}\mathbf{A}_{12,i}\right)^{-1}$, and $\mathbf{A}_{22,i}$ and $\mathbf{A}_{12,i}$ are the submatrices in $\mathbf{A}_{22}$ and $\mathbf{A}_{12}$ associated with variable x_i.

Wald Statistic

The Wald statistic is calculated for the variables in the model to select variables for removal. The Wald statistic for variable $\mathbf{x}_j$ is defined by

$$\hat{\beta}'_j \mathbf{B}_{11,j} \hat{\beta}_j$$

where $\hat{\beta}_j$ is the parameter estimate associated with $\mathbf{x}_j$ and $\mathbf{B}_{11,j}$ is the submatrix of $\mathbf{A}_{11}^{-1}$ associated with $\mathbf{x}_j$.

LR (Likelihood Ratio) Statistic

The LR statistic is defined as twice the log of the ratio of the likelihood functions of two models evaluated at their own MPLES. Assume that *r* variables are in the current model and let us call the current model the full model. Based on the MPLES of parameters for the full model, *l*(*full*) is defined in equation **(12b)**. For each of *r* variables deleted from the full model, MPLES are found and the reduced log-likelihood function, *l*(*reduced*), is calculated. Then LR statistic is defined as

$$-2(l(reduced) - l(full))$$

Conditional Statistic

The conditional statistic is also computed for every variable in the model. The formula for conditional statistic is the same as LR statistic except that the parameter estimates for each reduced model are conditional estimates, not MPLES. The conditional estimates are defined as follows. Let $\hat{\beta} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_r\right)'$ be the MPLES for the *r* variables (blocks) and *C* be the asymptotic covariance for the parameters left in the model given $\hat{\beta}_i$ is

$$\tilde{\beta}_{(i)} = \hat{\beta}_{(i)} - \mathbf{C}_{12}^{(i)} \left(\mathbf{C}_{22}^{(i)}\right)^{-1} \hat{\beta}_i$$

where $\hat{\beta}_i$ is the MPLE for the parameter(s) associated with $\mathbf{x}_i$ and $\hat{\beta}_{(i)}$ is $\hat{\beta}$ without $\hat{\beta}_i$, $\mathbf{C}_{12}^{(i)}$ is the covariance between the parameter estimates left in the model $\hat{\beta}_{(i)}$ and $\hat{\beta}_i$, and $\mathbf{C}_{22}^{(i)}$ is the covariance of $\hat{\beta}_i$. Then the conditional statistic for variable $\mathbf{x}_i$ is defined by

$$-2\left(l\left(\tilde{\beta}_{(i)}\right)-l(full)\right)$$

where $l\left(\tilde{\beta}_{(i)}\right)$ is the log-likelihood function evaluated at $\tilde{\beta}_{(i)}$.

Note that all these four statistics have a chi-square distribution with degrees of freedom equal to the number of parameters the corresponding model has.

Statistics

Initial Model Information

The initial model for the first method is for a model that does not include covariates. The log-likelihood function l is equal to

$$l(0)=-\sum_{j=1}^{m}\sum_{i=1}^{k_j} d_{ji}\ln\left(n_{ji}^*\right)$$

where n_{ji}^* is the sum of weights of individuals in set R_{ji}.

Model Information

When a stepwise method is requested, at each step, -2 log-likelihood function and three chi-square statistics (model chi-square, improvement chi-square, and overall chi-square) and their corresponding degrees of freedom and significance are printed.

–2 Log-Likelihood

$$-2\sum_{j=1}^{m}\sum_{i=1}^{k_j}\left(\mathbf{s}'_{ji}\hat{\beta}-d_{ji}\ln\left(\sum_{l\in R_{ji}}w_l\exp\left(\mathbf{x}'_l\hat{\beta}\right)\right)\right)$$

where $\hat{\beta}$ is the MPLE of β for the current model.

Improvement Chi-Square

(–2 log-likelihood function for previous model) – (–2 log-likelihood function for current model).

The previous model is the model from the last step. The degrees of freedom are equal to the absolute value of the difference between the number of parameters estimated in these two models.

Model Chi-Square

(–2 log-likelihood function for initial model) – (–2 log-likelihood function for current model).

The initial model is the final model from the previous method. The degrees of freedom are equal to the absolute value of the difference between the number of parameters estimated in these two model.

Note: The values of the model chi-square and improvement chi-square can be less than or equal to zero. If the degrees of freedom are equal to zero, the chi-square is not printed.

Overall Chi-Square

The overall chi-square statistic tests the hypothesis that all regression coefficients for the variables in the model are identically zero. This statistic is defined as

$$\mathbf{u}'(0)\mathbf{I}^{-1}\mathbf{u}(0)$$

where $\mathbf{u}(0)$ represents the vector of first derivatives of the partial log-likelihood function evaluated at $\beta=0$. The elements of **u** are defined in equation **(13)** and **I** is defined in equation **(14)**.

Information for Variables in the Equation

For each of the single variables in the equation, MPLE, SE for MPLE, Wald statistic, and its corresponding *df*, significance, and partial *R* are given. For a single variable, *R* is defined by

$$R = \left[\frac{\text{Wald} - 2}{-2 \text{ log - likelihood for the intial model}} \right]^{1/2} \times \text{sign of MPLE}$$

if $\text{Wald} > 2$. Otherwise *R* is set to zero.

For a multiple category variable, only the Wald statistic, *df*, significance, and partial *R* are printed, where *R* is defined by

$$R = \left[\frac{\text{Wald} - 2 * \text{df}}{-2 \text{ log - likelihood for the intial model}} \right]^{1/2}$$

if $\text{Wald} > 2\,\text{df}$. Otherwise *R* is set to zero.

Information for the Variables Not in the Equation

For each of the variables not in the equation, the Score statistic is calculated and its corresponding degrees of freedom, significance, and partial *R* are printed. The partial *R* for variables not in the equation is defined similarly to the *R* for the variables in the equation by changing the Wald statistic to the Score statistic.

There is one overall statistic called the residual chi-square. This statistic tests if all regression coefficients for the variables not in the equation are zero. It is defined by

$$\mathbf{u}'\left(\hat{\beta}\right)\mathbf{B}_{22}\,\mathbf{u}\left(\hat{\beta}\right)$$

where $\mathbf{u}\left(\hat{\beta}\right)$ is the vector of first derivatives of the partial log-likelihood function with respect to all the parameters not in the equation evaluated at MPLE $\hat{\beta}$ and $\mathbf{B}_{22}$ is equal to $\left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1}$ and **A** is defined in equation **(21)**.

Survival Table

For each stratum, the estimates of the baseline cumulative survival (S_0) and hazard (H_0) function and their standard errors are computed. The estimate $\hat{S}_0$ for S_0 has been discussed in equations **(15)** through **(18)**. It is easy to see from equation **(9)** that $H_0(t)$ is estimated by

$$\hat{H}_0(t) = -\ln \hat{S}_0(t)$$

and the asymptotic variance of $\hat{H}_0(t)$ is defined in equation **(19)**. Finally, the cumulative hazard function $H(t|\mathbf{x})$ and survival function $S(t|\mathbf{x})$ are estimated by

$$\hat{H}(t|\mathbf{x}) = \exp\left(\mathbf{x}'\hat{\beta}\right)\hat{H}_0(t)$$

and, for a given **x**,

$$\hat{S}(t|\mathbf{x}) = \left[\hat{S}_0(t)\right]^{\exp\left(\mathbf{x}'\hat{\beta}\right)} = \left[\hat{S}_0^*(t)\right]^{\exp\left(\mathbf{x}'\hat{\beta}-a\right)}$$

The asymptotic variances are

$$\operatorname{var}\left(\hat{H}(t|\mathbf{x})\right) = \exp\left(2\mathbf{x}'\hat{\beta}\right)\operatorname{var}\left(\hat{H}_0(t)\right)$$

and

$$\operatorname{var}\left(\hat{S}(t|\mathbf{x})\right) = \exp\left(2\mathbf{x}'\hat{\beta}\right)\left(\hat{S}(t|\mathbf{x})\right)^2 \operatorname{var}\left(\hat{H}_0(t)\right)$$

Diagnostic Statistics

Three casewise diagnostic statistics, Residual, Partial Residual, and DFBETAs, are produced. Both Residual and DFBETA are computed for all distinct individuals. Partial Residuals are calculated only for uncensored individuals.

Assume that there are n_j subjects in stratum j and k_j distinct observed events $t_1 < \cdots < t_{k_j}$. Define the selected probability for the lth individual at time t_i as

$$p_l(t_i) = \begin{cases} \dfrac{\exp\left(\mathbf{x}_l'(t_i)\hat{\beta}\right)}{\sum_{h \in R_i} w_h \exp\left(\mathbf{x}_h'(t_i)\hat{\beta}\right)} & \text{if } l\text{th individual is in } R_i \\ 0 & \text{otherwise} \end{cases}$$

and

$$u_l = \sum_{i=1}^{k_j} d_i \left[p_l(t_i) - p_l^2(t_i) \right]$$

$$y_l(t_i) = \begin{cases} 1 & \text{if } l\text{th individual is in } D_i \\ 0 & \text{otherwise} \end{cases}$$

$$r_l = \sum_{i=1}^{k_j} \left[y_l(t_i) - d_i p_l(t_i) \right]$$

DFBETA

The changes in the maximum partial likelihood estimate of beta due to the deletion of a single observation have been discussed in Cain and Lange (1984) and Storer and Crowley (1985). The estimate of DFBETA computed is derived from augmented regression models. The details can be found in Storer and Crowley (1985). When the lth individual in the jth stratum is deleted, the change $\Delta\beta_j$ is estimated by

$$\Delta\beta_l = -\frac{1}{\mathbf{m}}\mathbf{I}^{-1}v_l r_l$$

where

$$w = \text{diag}\left(w_1, \ldots, w_{n_{ji}}\right)$$

$$v_l = \sum_{i=1}^{k_j} d_i p_l(t_i)\left(\mathbf{x}_l(t_i) - \mathbf{x}(t_i)\mathbf{w}\,\mathbf{p}(t_i)\right)$$

$$\mathbf{p}(t_i) = \left(p_1(t_i), \ldots, p_{n_{ji}}(t_i)\right)'$$

$$m_l = u_l - v_l\mathbf{I}^{-1}v_l$$

and $\mathbf{x}'(t_i)$ is an $n_{ji} \times p$ matrix which represents the p covariate variables in the model evaluated at t_i, and n_{ji} is the number of individuals in R_{ji}.

Partial Residuals

Partial residuals can only be computed for the covariates which are not time dependent. At time t_i in stratum j, x_g is the $p \times 1$ observed covariate vector for any gth individual in set D_i, where D_i is the set of individuals dying at t_i. The partial residual γ_g is defined by

$$\gamma_g = \begin{pmatrix} \gamma_{g1} \\ \cdots \\ \gamma_{gp} \end{pmatrix} = \mathbf{x}_g - p(t_i)\mathbf{x}$$

Rewriting the above formula in a univariate form, we get

$$\gamma_{gh} = x_{gh} - \frac{\sum_{l \in R_i} w_l x_{lh} \exp\left(\mathbf{x}_l'\hat{\beta}\right)}{\sum_{l \in R_i} w_l \exp\left(\mathbf{x}_l'\hat{\beta}\right)}, \quad h = 1, \ldots, p, g \in D_i$$

where x_{gh} is the *h*th component for x_g. For every variable, the residuals can be plotted against times to test the proportional hazards assumption.

Residuals

The residuals e_i are computed by

$$e_i = \hat{H}(t_i \mid \mathbf{x}_i) = \exp\left(\mathbf{x}_i'\hat{\beta}\right)\left(\hat{H}_0(t_i)\right)$$

which is the same as the estimate of the cumulative hazard function.

Plots

For a specified pattern, the covariate values $\mathbf{x}_c$ are determined and $\mathbf{x}_c'\hat{\beta}$ is computed. There are three plots available in COXREG.

Survival Plot

For stratum *j*, $\left(t_i, \hat{S}_0(t_i \mid \mathbf{x}_c)\right)$, $i = 1, \ldots, k_j$ are plotted where

$$\hat{S}(t_i \mid \mathbf{x}_c) = \left(\hat{S}_0(t_i)\right)^{\exp\left(\mathbf{x}_c'\hat{\beta}\right)}$$

When PATTERN(ALL) is requested, for every uncensored time t_i in stratum *j* the survival function is estimated by

$$\hat{S}(t_i) = \frac{\sum_{l=1}^{k_j} w_l \hat{S}(t_i \mid \mathbf{x}_c)}{\sum_{l=1}^{k_j} w_l} = \frac{\sum_{l=1}^{k_j} w_l \left(\hat{S}_0(t_i)\right)^{\exp\left(\mathbf{x}_c'\hat{\beta}\right)}}{\sum_{l=1}^{k_j} w_l}$$

Then $\left(t_i, \hat{S}(t_i)\right)$, $i = 1, \ldots, k_j$ are plotted for stratum *j*.

Hazard Plot

For stratum *j*, $\left(t_i, \hat{H}(t_i \mid \mathbf{x}_c)\right)$, $i = 1, \ldots, k_j$ are plotted where

$$\hat{H}(t_i \mid \mathbf{x}_c) = \exp\left(\mathbf{x}_c'\hat{\beta}\right) \hat{H}_0(t_i)$$

LML Plot

The log-minus-log plot is used to see whether the stratification variable should be included as a covariate. For stratum *j*, $\left(t_i, \mathbf{x}_c'\hat{\beta} + \ln \hat{H}_0(t_i)\right)$, $i = 1, \ldots, k_j$ are plotted.

If the plot shows parallelism among strata, then the stratum variable should be a covariate.

References

Breslow, N. E. 1974. Covariance analysis of censored survival data. *Biometrics*, 30: 89–99.

Cain, K. C., and Lange, N. T. 1984. Approximate case influence for the proportional hazards regression model with censored data. *Biometrics*, 40: 493–499.

Cox, D. R. 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34: 187–220.

Kalbfleisch, J. D., and Prentice, R. L. 1980. *The statistical analysis of failure time data.* New York: John Wiley & Sons, Inc.

Lawless, J. F. 1982. *Statistical models and methods for lifetime data*. New York: John Wiley & Sons, Inc.

Storer, B. E., and Crowley, J. 1985. A diagnostic for Cox regression and general conditional likelihoods. *Journal of the American Statistical Association*, 80: 139–147.