7. Cox Regression Models

(Part II)

Tied Data

In practice, it is quite common for our data to contain tied survival times. Therefore, we need a different technique to construct the partial likelihood in the presence of tied data.

A simple example:

Patient	x	δ	z
1	x_1	1	z_1
2	x_2	1	z_2
3	x_3	0	z_3
4	x_4	1	z_4
5	x_5	1	z_5

$$x_1 = x_2 < x_3 < x_4 < x_5$$

We assume the following proportional hazards model

$$\lambda(t|z_i) = \lambda_0(t)\exp(z_i\beta)$$

Since there are 3 distinct survival times $(i.e,\ x_1, x_4, x_5)$, intuitively, the partial likelihood function of β will take the following form

$$L(\beta) = L_1(\beta)L_2(\beta)L_3(\beta)$$

$L_2(\beta)$ and $L_3(\beta)$ can be constructed in the usual way. So we will focus on the construction of $\boldsymbol{L_1(\beta)}$.

$L_1(\beta)$: The Exact Method

This method assumes that the survival time has a continuous distribution and the true survival times of patients 1 and 2 are **different**. These two patients have the same survival times in our data because our measurement does not have enough accuracy or the original data was rounded for convenience and this information got lost, etc. For patients 1 and 2, there are 2! = 2 possible orderings.

Let A_1 denote the event that patient 1 died before patient 2, and A_2 denote the event that patient 2 died before patient 1. Then by the law of total probability, we have

$$L_1(\beta) = P[\text{observed two deaths at } x_1] = P[A_1 \cup A_2] = P[A_1] + P[A_2]$$

and

$$P[A_1] = \frac{e^{z_1\beta}}{e^{z_1\beta} + e^{z_2\beta} + e^{z_3\beta} + e^{z_4\beta} + e^{z_5\beta}} \times \frac{e^{z_2\beta}}{e^{z_2\beta} + e^{z_3\beta} + e^{z_4\beta} + e^{z_5\beta}}$$

$$P[A_2] = \frac{e^{z_2\beta}}{e^{z_1\beta} + e^{z_2\beta} + e^{z_3\beta} + e^{z_4\beta} + e^{z_5\beta}} \times \frac{e^{z_1\beta}}{e^{z_1\beta} + e^{z_3\beta} + e^{z_4\beta} + e^{z_5\beta}}$$

After the partial likelihood $L(\beta)$ is constructed, the inference of β is exactly the same as the case where there is no tied survival time (tied survival time and censoring time have no effect on the partial likelihood construction). That is, we maximize the new partial likelihood $L(\beta)$ to obtain MPLE of β, use inverse of minus second derivative of the log partial likelihood function to estimate the variability in the MPLE of β. We can also perform score test and likelihood ratio test.

The exact method is implemented in *Proc Phreg* in SAS. Suppose in our data set *mydata,* we use *time* to denote the (censored) survival times with *cens* the censoring indicator, and z the covariate, then the PH model can be fitted with the exact method using the following SAS code:

```
Proc Phreg data=mydata;
model time*cens(0) = z / ties=exact;
run;
```

(other choices for *ties*: breslow; efron; discrete)

The exact method will yield optimal estimate of β. However, this method can be potentially computationally intensive. For example, suppose there are d_j tied survival times at the jth distinct survival time, then $d_j!$ different orderings have to be considered and $L_j(\beta)$ is the sum of $d_j!$ different terms, each of which is the product of d_j terms (conditional probabilities). This number could be very large. For example, when $d_j = 5$, then $d_j * d_j! = 5 * 5! = 6000$ different terms have to be calculated to get $L_j(\beta)$. Because of this computational difficulties, another two methods have been proposed to approximate the exact partial likelihood.

$L_1(\beta)$: Breslow's Approximation

Notice that

$$\frac{e^{z_2\beta}}{e^{z_2\beta}+e^{z_3\beta}+e^{z_4\beta}+e^{z_5\beta}} \approx \frac{e^{z_2\beta}}{e^{z_1\beta}+e^{z_2\beta}+e^{z_3\beta}+e^{z_4\beta}+e^{z_5\beta}}$$

$$\frac{e^{z_1\beta}}{e^{z_1\beta}+e^{z_3\beta}+e^{z_4\beta}+e^{z_5\beta}} \approx \frac{e^{z_1\beta}}{e^{z_1\beta}+e^{z_2\beta}+e^{z_3\beta}+e^{z_4\beta}+e^{z_5\beta}}$$

So, $P[A_1]$, $P[A_2]$ and hence $L_1(\beta)$ can be approximated by

$$\frac{e^{z_1\beta}}{e^{z_1\beta}+e^{z_2\beta}+e^{z_3\beta}+e^{z_4\beta}+e^{z_5\beta}} \times \frac{e^{z_2\beta}}{e^{z_1\beta}+e^{z_2\beta}+e^{z_3\beta}+e^{z_4\beta}+e^{z_5\beta}} = \frac{e^{(z_1+z_2)\beta}}{\left[\sum_{l=1}^{5} e^{z_l\beta}\right]^2}$$

In general, if there are d_j tied survival times at the jth distinct survival time, then $L_j(\beta)$ is approximated by

$$L_j(\beta) = \frac{e^{\sum_{l\in D_j} z_l\beta}}{\left[\sum_{l\in R_j} e^{z_l\beta}\right]^{d_j}}$$

where R_j is the risk set at the jth survival time and D_j is the event (death) set at the jth distinct survival time. So the partial likelihood of β is

$$L(\beta) = \prod_{j=1}^{D} L_j(\beta) \approx \prod_{j=1}^{D} \frac{e^{\sum_{l\in D_j} z_l\beta}}{\left[\sum_{l\in R_j} e^{z_l\beta}\right]^{d_j}}$$

where D is the total distinct events.

(Default method in *Proc Phreg*; *ties=breslow*)

$L_1(\beta)$: Efron's Approximation

Notice that $\frac{bc}{a(a-b)} + \frac{bc}{a(a-c)} \approx \frac{2bc}{a\left[a - \frac{b+c}{2}\right]}$

Let $a = e^{z_1\beta} + e^{z_2\beta} + e^{z_3\beta} + e^{z_4\beta} + e^{z_5\beta}$; $b = e^{z_1\beta}$; $c = e^{z_2\beta}$

Then, $L_1(\beta) \approx \frac{2bc}{a\left[a - \frac{b+c}{2}\right]}$

In general, if there are d_j tied survival times at the jth distinct survival time, then $L_j(\beta)$ is approximated by

$$L_j(\beta) = \frac{e^{\sum_{l \in D_j} z_l\beta}}{\left[\sum_{l \in R_j} e^{z_l\beta}\right]^{d_j}}$$

where R_j is the risk set at the jth survival time and D_j is the event (death) set at the jth distinct survival time. So the partial likelihood of β is

$$L(\beta) = \prod_{j=1}^{D} L_j(\beta) \approx \prod_{j=1}^{D} \frac{e^{\sum_{l \in D_j} z_l\beta}}{\prod_{k=1}^{d_j}\left(\sum_{l \in R_j} e^{z_l\beta} - \frac{k-1}{d_j}\sum_{l \in D_j} e^{z_l\beta}\right)}$$

where D is the total distinct events.

(in *Proc Phreg; ties=efron*)

$L_1(\beta)$: Discrete Method

This method does NOT assume that there is underlying ordering of the tied survival times. Instead, the time is assumed to be discrete. In this case, we consider the following model: for any death time t, let

$$\pi_{it} = P[\text{subject } i \text{ will die at } t | \text{subject } i \text{ survive up to } t]$$

then assume the following proportional odds model (a logistic regression with time-varying intercepts)

$$\log\left(\frac{\pi_{it}}{1-\pi_{it}}\right) = \alpha_t + z_i\beta$$

Where α_t's are nuisance parameters and β is the parameter of interest (treatment effect, for example). In this case, $L_1(\beta)$ can be interpreted as

$$L_1(\beta) = P[\text{deaths occurred to subjects 1 and 2} | \text{there are 2 deaths out of 5 subjects}]$$

It can be shown that the above probability is equal to

$$L_1(\beta) = \frac{e^{\sum_{l\in D_j} z_l\beta}}{\sum_{all\ D_j} e^{s_j\beta}}$$

where D_js are $\binom{5}{2} = 10$ possible combinations, and $s_j = \sum_{l\in D_j} z_l$

Note:

(1) the model considered here is not a proportional hazards model. However, when there is no tied observation in the data set, the resulting likelihood is exactly the same as the Cox partial likelihood. This is the main reason that discrete method is included in Proc Phreg with *ties=discrete*.

(2) this method can be even more computationally intensive than the exact method.

Multiple Covariates

The real strength of the proportional hazards model is that it allows us to model the relationship of survival time, through its hazard function, to many covariates simultaneously:

$$\lambda(t|z) = \lambda_0(t)e^{\beta_1 z_1 + \cdots + \beta_p z_p} = \lambda_0(t)\mathrm{e}^{z^T\beta}$$

where z is a $p \times 1$ vector of covariates such as treatment indicators, prognostic factors, etc., and $\beta = (\beta_1, \ldots, \beta_p)^T$ is a $p \times 1$ vector of regression coefficients. Note that there is no intercept β_0 in the model. So $\lambda_0(t)$ is often called the baseline hazard function. It can be interpreted as the hazard function for the population of subjects with $z = 0$, i.e. $\lambda(t|z = 0) = \lambda_0(t)$.

Estimation of β is exactly similar to the case of one covariate. The partial likelihood of β is given by

$$PL(\beta) = \prod_{\{\text{all grid pt } u\}} \left[\frac{e^{z_{i(u)}^T \beta}}{\sum_{l=1}^{n} e^{z_l^T \beta} Y_l(t_j)} \right]^{dN(u)}$$

and the log partial likelihood of β is

$$l(\beta) = \sum_{\{\text{all grid pt } u\}} dN(u) \left[z_{I(u)}^T \beta - \log\left(\sum_{l=1}^{n} e^{z_l^T \beta} Y_l(u) \right) \right]$$

z_l is the covariate value for the l th individual; i.e. $z_l = (z_{l1}, \ldots, z_{lp})^T$

MPLE of β: $\hat{\beta}$

Maximize $l(\beta)$ by setting score vector to be zero:

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

where $\frac{\partial l(\beta)}{\partial \beta} = \left(\frac{\partial l(\beta)}{\partial \beta_1}, \dots, \frac{\partial l(\beta)}{\partial \beta_p}\right)^T$. Similarly, we have

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{\{\text{all grid pt } u\}} dN(u)\left[z_{I(u),j} - \bar{z}_j(u,\beta)\right]$$

where $z_{I(u),j}$ denotes the jth element of the covariate vector for the individual $I(u)$ who died at time u, and

$$\bar{z}_j(u,\beta) = \frac{\sum_{l=1}^n z_{lj} e^{z_l^T \beta} Y_l(u)}{\sum_{l=1}^n e^{z_l^T \beta} Y_l(u)} = \sum_{l=1}^n z_{lj} w_l \,, w_l = \frac{e^{z_l^T \beta} Y_l(u)}{\sum_{l=1}^n e^{z_l^T \beta} Y_l(u)}$$

If we denote $Z_{I(u)}^{p\times 1} = \begin{pmatrix} z_{I(u),1} \\ \vdots \\ \vdots \\ z_{I(u),p} \end{pmatrix}$, $Z^{p\times 1}(u,\beta) = \begin{pmatrix} \bar{z}_1(u,\beta) \\ \vdots \\ \vdots \\ \bar{z}_p(u,\beta) \end{pmatrix}$,

Then, $$U(\beta) = \sum_{\{\text{all grid pt } u\}} dN(u)\left[Z_{I(u)}^{p\times 1} - Z^{p\times 1}(u,\beta)\right] = 0^{p\times 1}$$

In order for the partial likelihood equation to have a unique solution, it is sufficient that the Hessian matrix H be negative definite

$$a^T H a < 0 \text{ for all } a_{p\times 1} \neq 0, \text{ and } H = \frac{\partial^2 l(\beta)}{\partial \beta^T \partial \beta} = \left[\frac{\partial^2 l(\beta)}{\partial \beta_j \partial \beta_{j'}}\right]_{p\times p}$$

Or, equivalently,

$$J(\beta) = -\frac{\partial^2 l(\beta)}{\partial \beta^T \partial \beta} = -\left[\frac{\partial^2 l(\beta)}{\partial \beta_j \partial \beta_{j'}}\right]_{p\times p} \text{ is positive definite.}$$

It can be easily shown that the (j, j')th element of $J(\beta)$ is

$$\begin{aligned} J_{j,j'} &= \sum_u dN(u)\left[\frac{\sum_{l=1}^n z_{lj} z_{lj'} e^{z_l^T \beta} Y_l(u)}{\sum_{l=1}^n e^{z_l^T \beta} Y_l(u)} - \bar{z}_j(u,\beta)\bar{z}_{j'}(u,\beta)\right] \\ &= \sum_u dN(u)\left[\frac{\sum_{l=1}^n \left(z_{lj} - \bar{z}_j(u,\beta)\right)\left(z_{lj'} - \bar{z}_{j'}(u,\beta)\right) e^{z_l^T \beta} Y_l(u)}{\sum_{l=1}^n e^{z_l^T \beta} Y_l(u)}\right] \\ &= \sum_u dN(u) V_{j,j'}(u,\beta) \end{aligned}$$

where $V_{j,j'}(u,\beta)$ is weighted sample covariance between the jth and j'th element of the covariate vector among individuals *at risk* at time *u* with the weight being w_l.

If the weighted $p \times p$ covariate matrix of the covariate vector among individuals at risk at time u is denoted as

$$V(u,\beta) = \begin{pmatrix} V_{11}(u,\beta) & \cdots & V_{1p}(u,\beta) \\ \vdots & \vdots & \vdots \\ V_{p1}(u,\beta) & \cdots & V_{pp}(u,\beta) \end{pmatrix}$$

then the information matrix is

$$J^{p\times p}(\beta) = \sum_{u} dN(u)V(u,\beta)$$

Note: In matrix notation, $V(u,\beta)$ can be expressed as

$$V(u,\beta) = \frac{\sum_{l=1}^{n}\big(z_l - \bar{z}(u,\beta)\big)\big(z_l - \bar{z}(u,\beta)\big)^T e^{z_l^T\beta}Y_l(u)}{\sum_{l=1}^{n} e^{z_l^T\beta}Y_l(u)}$$

$$= \sum_{l=1}^{n} w_l\big(z_l - \bar{z}(u,\beta)\big)\big(z_l - \bar{z}(u,\beta)\big)^T,$$

which is a weighted variance matrix of the covariate vectors among the individuals at risk at time u. Thus $V(u,\beta)$ is positive definite. Therefore the information matrix $J^{p\times p}(\beta)$ is also a positive definite matrix. So the Hessian matrix $\mathrm{H} = -J^{p\times p}(\beta)$ is negative definite. This implies that log partial likelihood is a concave function of β and hence it has a unique maximum, which can be obtained by setting the first derivative of the log partial likelihood, *i.e.*, score $U(\beta)$, to be zero.

Statistical Properties

1. The score vector $U(\beta_0)$ evaluated at the true value of β will be asymptotically distributed as a multivariate normal with mean vector zero and covariance matrix which can be estimated unbiasedly by $J(\beta_0)$.

$$U(\beta_0) \overset{a}{\sim} N\big(0, J(\beta_0)\big)$$

2. The MPLE $\hat{\beta}$ will also be asymptotically normal

$$\hat{\beta} \overset{a}{\sim} N(\beta_0, J^{-1}(\beta_0))$$

where $J^{-1}(\beta_0)$is the inverse of $J(\beta_0)$. Since $J(\beta_0)$is positive definite, so its unique inverse exists and is also positive definite.

Inference on Parameter of interest

Sometimes we may only be interested in making inferential statements about part of model parameters. Towards this end, let us partition the parameter vectorβinto two parts: $\beta = (\theta^T, \phi^T)^T$, where θ is a $g(\leq p)$ dimensional vector. Here, we refer θ to as the parameter of interest and call ϕ as the nuisance parameter. Of course, the parameter of interest θ can be the entire parameter vector β. Correspondingly, the score vector is partitioned as

$$U(\beta) = U(\theta, \phi) = \begin{pmatrix} U_\theta(\theta, \phi) \\ U_\phi(\theta, \phi) \end{pmatrix}, \text{ where } U_\theta(\theta, \phi) = \frac{\partial l(\theta, \phi)}{\partial \theta}, U_\phi(\theta, \phi) = \frac{\partial l(\theta, \phi)}{\partial \phi}$$

The partial likelihood information matrix can also be partitioned into

$$J(\beta) = \begin{pmatrix} J_{\theta\theta}(\theta, \phi) & J_{\theta\phi}(\theta, \phi) \\ J_{\phi\theta}(\theta, \phi) & J_{\phi\phi}(\theta, \phi) \end{pmatrix}$$

and its inverse into

$$J^{-1}(\beta) = \begin{pmatrix} J^{\theta\theta}(\theta, \phi) & J^{\theta\phi}(\theta, \phi) \\ J^{\phi\theta}(\theta, \phi) & J^{\phi\phi}(\theta, \phi) \end{pmatrix}$$

NOTE: The superscript notation is used to index the partition of an inverse matrix and subscript notation to index the original matrix.

With these notations, $\hat{\beta} \overset{a}{\sim} N\left(\beta_0, J^{-1}(\hat{\beta})\right)$ is equivalent to

$$\begin{pmatrix}\hat{\theta}\\ \hat{\phi}\end{pmatrix} \overset{a}{\sim} N\left(\begin{pmatrix}\theta_0\\ \phi_0\end{pmatrix}, \begin{pmatrix}J^{\theta\theta}(\hat{\theta},\hat{\phi}) & J^{\theta\phi}(\hat{\theta},\hat{\phi})\\ J^{\phi\theta}(\hat{\theta},\hat{\phi}) & J^{\phi\phi}(\hat{\theta},\hat{\phi})\end{pmatrix}\right)$$

Therefore, $\hat{\theta}$ has the asymptotic distribution

$$\hat{\theta} \overset{a}{\sim} N\left(\theta_0, J^{\theta\theta}(\hat{\theta},\hat{\phi})\right) \quad \text{or,} \quad (\hat{\theta}-\theta_0) \overset{a}{\sim} N\left(0, J^{\theta\theta}(\hat{\theta},\hat{\phi})\right)$$

Therefore

$$(\hat{\theta}-\theta_0)^T\left[J^{\theta\theta}(\hat{\theta},\hat{\phi})\right]^{-1}(\hat{\theta}-\theta_0) \overset{a}{\sim} \chi^2_g$$

$\left[J^{\theta\theta}(\hat{\beta},\hat{\phi})\right]^{-1}$ is the inverse of the partition of the inverse of the information matrix.

In general, $\left[J^{\theta\theta}(\hat{\beta},\hat{\phi})\right]^{-1} \neq J^{\theta\theta}(\hat{\theta},\hat{\phi})$

Using this notation, we can find a confidence region for the parameter of interest θ.

Let $\chi^2_{\alpha,g}$ be the $(1-\alpha)$ quantile of a χ^2 with g degrees of freedom, *i.e.*, $\mathrm{P}\left[\chi^2_g \geq \chi^2_{\alpha,g}\right] = \alpha$

Hence, $\mathrm{P}\left[(\theta-\hat{\theta})^T\left[J^{\theta\theta}(\hat{\theta},\hat{\phi})\right]^{-1}(\theta-\hat{\theta}) \geq \chi^2_{\alpha,g}\right] = \alpha$

For a given data set, the inequality $(\theta-\hat{\theta})^T\left[J^{\theta\theta}(\hat{\theta},\hat{\phi})\right]^{-1}(\theta-\hat{\theta}) \geq \chi^2_{\alpha,g}$ describes a g-dimensional ellipsoid centered at $\hat{\theta}$ and whose orientation is dictated by the eigenvalues and eigenvectors of $\left[J^{\theta\theta}(\hat{\theta},\hat{\phi})\right]^{-1}$

If θ is one-dimensional, then $J^{\theta\theta}(\hat{\theta}, \hat{\phi})$ is also one-dimensional.

If $\theta = \beta_j$, in this case, $\hat{\theta} = \hat{\beta}_j$, $\theta_0 = \beta_{j0}$, and $J^{\theta\theta}(\hat{\theta}, \hat{\phi}) = [se(\hat{\beta}_j)]^2$.

Also, the confidence region above simplifies to an interval, and the $(1-\alpha)$th confidence interval of θ or β_j would be

$$\hat{\beta} \pm z_{\alpha/2} se(\hat{\beta}_j)$$

Where $se(\hat{\beta}_j) = [J^{\theta\theta}(\hat{\theta}, \hat{\phi})]^{1/2}$

Hypothesis Testing

Now suppose we are interested in testing the null hypothesis

$$H_0: \theta = \theta_0$$

Three tests: Wald test, score test, Likelihood ratio test

Wald test:

Since under H_0, we have

$$(\hat{\theta} - \theta_0)^T [J^{\theta\theta}(\hat{\theta}, \hat{\phi})]^{-1} (\hat{\theta} - \theta_0) \overset{a}{\sim} \chi^2_g$$

If the null hypothesis H_0 were not true, we would expect the above quadratic form to get larger since $\hat{\theta}$ would not be close to θ_0. This suggests that we will reject $H_0: \theta = \theta_0$ at the α level of significance if

$$(\theta - \hat{\theta})^T [J^{\theta\theta}(\hat{\theta}, \hat{\phi})]^{-1} (\theta - \hat{\theta}) \geq \chi^2_{\alpha,g}$$

This is the Wald test.

Restricted MPLE & Score Test

H_0 only specifies a specific value of θ, however, nothing is assumed about the nuisance parameters ϕ. Therefore, even under H_0, an estimate of ϕ will be necessary in order to derive tests as a function of the data. Under$H_0: \theta = \theta_0$, an obvious estimator for ϕ is to maximize the log partial likelihood as a function of ϕ. This is referred to as a restricted MPLE and will be denoted by $\hat{\phi}(\theta_0)$. That is, $\hat{\phi}(\theta_0)$ is the value of ϕ which maximizes the function $l(\theta_0, \phi)$. This restricted MPLE can be obtained by solving the $(p-g)$ equations of $(p-g)$ unknowns

$$U_\phi\left(\theta_0, \hat{\phi}(\theta_0)\right) = 0,$$

using the $(p-g)$ dimensional subset of the score vector corresponding to the partial derivatives of the log partial likelihood with respect to the nuisance parameters.

The score test of the hypothesis $H_0: \theta = \theta_0$ is based on the score vector

$$U_\theta\left(\theta_0, \hat{\phi}(\theta_0)\right)$$

It can be shown that under$H_0: \theta = \theta_0$,

$$U_\theta\left(\theta_0, \hat{\phi}(\theta_0)\right) \overset{a}{\sim} N\left(0, \left[J^{\theta\theta}\left(\theta_0, \hat{\phi}(\theta_0)\right)\right]^{-1}\right)$$

Thus a level α score test of the hypothesis $H_0: \theta = \theta_0$ is to reject H_0 whenever

$$\left[U_\theta\left(\theta_0, \hat{\phi}(\theta_0)\right)\right]^T \left[J^{\theta\theta}\left(\theta_0, \hat{\phi}(\theta_0)\right)\right]\left[U_\theta\left(\theta_0, \hat{\phi}(\theta_0)\right)\right] \geq \chi^2_{\alpha,g}$$

Likelihood Ratio Test

We define the MPLE for β, or equivalently (θ, ϕ), as the value of (θ, ϕ)that maximizes the log partial likelihood $l(\theta, \phi)$. We denote this estimate as $\hat{\beta}$, or $\left(\hat{\theta}, \hat{\phi}\right)$. We also defined the restricted MPLE $\hat{\phi}(\theta_0)$ as the value of ϕ that maximizes the following function $l(\theta_0, \phi)$.

It must be the case that, for any set of data, $l\left(\hat{\theta}, \hat{\phi}\right)$ must be greater than or equal to $l\left(\theta_0, \hat{\phi}(\theta_0)\right)$, since $l\left(\hat{\theta}, \hat{\phi}\right)$ is maximized over a larger parameter space. We would expect, however, that if H_0 were true, $\hat{\theta}$ would be close to θ_0 and consequently $l\left(\hat{\theta}, \hat{\phi}\right)$ would be close to $l\left(\theta_0, \hat{\phi}(\theta_0)\right)$. It is therefore reasonable to expect that H_0 would not be true if the difference

$$l\left(\hat{\theta}, \hat{\phi}\right) - l\left(\theta_0, \hat{\phi}(\theta_0)\right)$$

is sufficiently large.

Since under H_0, we have $2\left[l\left(\hat{\theta}, \hat{\phi}\right) - l\left(\theta_0, \hat{\phi}(\theta_0)\right)\right] \overset{a}{\sim} \chi^2_g$

Therefore, the likelihood ratio test rejects $H_0: \theta = \theta_0$ at the α level of significance if

$$2\left[l\left(\hat{\theta}, \hat{\phi}\right) - l\left(\theta_0, \hat{\phi}(\theta_0)\right)\right] \geq \chi^2_{\alpha,g}$$

Model with multiple covariates

- The proportional hazards model with multiple covariates is ideal for studying the relationship that the variable of interest has on survival while adjusting for the effect of the other covariates, by including both the variable of interest as well as other variables (which may be confounders, or other variables we may wish to adjust for) in the model.
- In epidemiology, if our interest is the relationship of survival to drinking, we would say that smoking was a confounding variable. That is, smoking was a prognostic factor (*i.e.*, is related to survival) and smoking is correlated to drinking.
- Even in controlled studies, *i.e.*, randomized clinical trials, we may wish to adjust for other variables. Such adjusted analyses often lead to more precise estimate of the effect of interest and greater power to detect differences.
- In some cases, enforced balance of certain prognostic factors by treatment, necessitates the need for adjusted analyses.

Cautionary Remark: All of the above statements are based on the premise that the models being considered are adequate representations of the distribution of the data. So, for example, if proportional hazards is not a good model of the relationship of survival to the covariates, the results derived from such a model may be misleading.

A Hypothetic Example

Let S denote the smoking indicator (1 = smoker, 0 = nonsmoker), and D denote drinking indicator (1 = drinker, 0 = non-drinker). If we were to study the effect of drinking on survival, we may identify a cohort of individuals, say, individuals enrolling into a health insurance program or HMO. At the time of enrollment certain information may be gathered; including Age, Sex, Smoking and Drinking status, for example. Using either information from the insurance company or a death register, we identify who has died, when they died, as well as who is currently alive. That is, we obtain censored survival data. Suppose, we use the following proportional hazards model:

$$\lambda(t|D) = \lambda_0(t)e^{D\beta}$$

the parameter β is interpreted as the log hazard ratio between drinkers and non-drinkers (assumed constant over time t) and exp(β) as the hazard ratio.

Although this interpretation is correct, it may be causally misleading as it does not adjust for potential confounding factors. Consequently, we may use the following proportional hazards model with multiple covariates

$$\lambda(t|\cdot) = \lambda_0(t)e^{D\theta+S\phi_1+A\phi_2+Sx\phi_3}$$

where *S* = smoking status, *A* = age, *Sx* = sex.

Here the parameter θ corresponds to the log hazard ratio for a drinker compared to a non-drinker with the same smoking, age and sex variables; *i.e.*, adjusted for smoking, age and sex. And exp(θ) is the adjusted hazard ratio.

Note: Here θ is the parameter of interest and $\phi = (\phi_1, \phi_2, \phi_3)$ is the nuisance parameters.

Reminder: The hazard ratio is

$$\frac{\lambda(t|D=1,S=s,A=a,Sx=sx)}{\lambda(t|D=0,S=s,A=a,Sx=sx)}=\frac{\lambda_0(t)e^{\theta+s\phi_1+a\phi_2+sx\phi_3}}{\lambda_0(t)e^{0+s\phi_1+a\phi_2+sx\phi_3}}=e^{\theta}$$

The data collected necessary to fit this model would be at the form

$$(x_i,\delta_i,d_i,s_i,a_i,sx_i),\qquad i=1,2,\dots,n$$

The proportional hazards model

$$\lambda(t|\cdot)=\lambda_0(t)e^{D\theta+S\phi_1+A\phi_2+Sx\phi_3}$$

would be fit using *Proc Phreg* in SAS, using partial likelihood methods.
The output would yield the MPLE $\left(\hat{\theta},\hat{\phi}_1,\hat{\phi}_2,\hat{\phi}_3\right)$ as well as their estimated standard errors. From this we would construct a $(1-\alpha)$ confidence interval for

$$\hat{\theta}\pm z_{\alpha/2}se(\hat{\theta})$$

We could also test the null hypothesis $\mathrm{H}_0\colon\theta=0$ using a Wald test, score test, or partial likelihood ratio test, for θ with ϕ corresponding to the nuisance parameters.

A Real Example

This is one of the first successful trials of adjuvant chemotherapy for colon cancer. Levamisole is a low-toxicity compound previously used to treat worm infestations in animals; 5-FU is a moderately toxic (as these things go) chemotherapy agent. The data contains the following entries:

Time: days until event or censoring
Status: censoring status
Rx: Treatment - Lev(amisole), Lev(amisole)+5-FU
Sex (S): 1=male
Age (A): in years
Obstruct (Ob): obstruction of colon by tumour
Perfor (P): perforation of colon
Adhere (Ad): adherence to nearby organs
Nodes (Nd): number of lymph nodes with detectable cancer

The primary purpose is to evaluate the new treatment (Lev+5-FU) on colon cancer, adjusting for the above prognostic factors. The mode is

$$\lambda(t|\cdot) = \lambda_0(t)e^{Rx\theta+S\phi_1+A\phi_2+Ob\phi_3+P\phi_4+Ad\phi_5+Nd\phi_6}$$

From the SAS output:

$$\hat{\theta} = -0.3854, se(\hat{\theta}) = 0.0867$$

A 95% confidence interval for θ is

$$\hat{\theta} \pm z_{\alpha/2} se\left(\hat{\theta}\right) = -0.3854 \pm 1.96 * 0.0867 = [-0.5553, -0.2155\ ]$$

The estimate of the adjusted treatment hazard ratio is

$$e^{\hat{\theta}} = e^{-0.3854} = 0.6802$$

with a 95% CI of

$$[e^{-0.5553}, e^{-0.2155}] = [0.5739, 0.8962]$$

If we want to test the hypothesis H_0: $\theta = 0$; *i.e.*, no treatment effect adjusting for the other covariates, we can use

1. The Wald test: $\left[\frac{\hat{\theta}}{se\left(\hat{\theta}\right)}\right]^2 = \left[\frac{-0.3854}{0.0867}\right]^2 = 19.75$

   with p-value $= 8.8 \times 10^{-6}$.

2. Likelihood ratio test:

   $$2\left[l\left(\hat{\theta}, \hat{\phi}\right) - l\left(\theta = 0, \hat{\phi}(\theta = 0)\right)\right] = -7401.384 - (-7421.421) = 20.037$$

   with p-value $= 7.6 \times 10^{-6}$.

3. Score test: *Proc Phreg* will not automatically calculate the score test for H_0: $\theta = 0$ in the presence of nuisance parameters. (See the program for the score test.) The observed $\chi^2 = 19.99$, yielding the same p-value as other two tests.

Remark: In many clinical trials, not all the data are collected on all the individuals (such as nodes in this colon dataset). Consequently, one or more variables may be missing per individuals. In SAS the default for missing data is a ".". The way that SAS handles missing data is to delete an entire recorded if any of the variables being considered for a particular analysis is missing. Therefore, we must be careful when we are considering analysis with different sub-models. This is especially the case when we consider the likelihood ratio test for nested models. For example, fewer recorded may be missing when we consider one covariate as opposed to a model with that covariate and additional covariates. This might necessitate running a model on a subset of the data, where the subset corresponds to all data records with complete covariate information for the larger model (*i.e.*, the model with the most covariates)

The impact that missing data on the results of a study can be very complicated and only recently has been studied seriously. The strategy to eliminate entire record if any of the data are missing is very crude and can give biased results depending on the reasons for missingness. It may be useful to conduct some sensitivity analyses on different sets of data corresponding to different levels of missingness, *i.e.* looking at the effect of treatment (unadjusted) within the subset of non-missing patients to see if the results were comparable to the full data.

A serious study of identifying confounding factors, assuming we felt it was important to study, would take some work. It may start with the correlation analysis among the covariates.

Note:

The colon cancer study is originally described in Laurie (1989). The main report is found in Moertel (1990). This data set is closest to that of the final report in Moertel (1991). A version of the data with less follow-up time was used in the paper by Lin (1994).

References:

- JA Laurie, CG Moertel, TR Fleming, HS Wieand, JE Leigh, J Rubin, GW McCormack, JB Gerstner, JE Krook and J Malliard. Surgical adjuvant therapy of large-bowel carcinoma: An evaluation of levamisole and the combination of levamisole and fluorouracil: The North Central Cancer Treatment Group and the Mayo Clinic. J Clinical Oncology, 7:1447-1456, 1989.
- DY Lin. Cox regression analysis of multivariate failure time data: the marginal approach. Statisticsin Medicine, 13:2233-2247, 1994.
- CG Moertel, TR Fleming, JS MacDonald, DG Haller, JA Laurie, PJ Goodman, JS Ungerleider, WA Emerson, DC Tormey, JH Glick, MH Veeder and JA Maillard. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. New England J of Medicine, 332:352-358, 1990.
- CG Moertel, TR Fleming, JS MacDonald, DG Haller, JA Laurie, CM Tangen, JS Ungerleider, WA Emerson, DC Tormey, JH Glick, MH Veeder and JA Maillard, Fluorouracil plus Levamisole as and effective adjuvant therapy after resection of stage II colon carcinoma: a final report. Annals of Internal Med, 122:321-326, 1991.