6. Modeling Survival Data with Cox Regression Models

6.1 The Proportional Hazards Model

A proportional hazards model proposed by D.R. Cox (1972) assumes that

$$\lambda(t|z) = \lambda_0(t)\mathrm{e}^{z_1\beta_1+\cdots+z_p\beta_p} = \lambda_0(t)\mathrm{e}^{z^T\beta}, \tag{1}$$

where z is a $p \times 1$ vector of covariates such as treatment indicators, prognositc factors, etc., and β is a $p \times 1$ vector of regression coefficients. Note that there is no intercept β_0 in model (1).

Obviously,

$$\lambda(t|z = 0) = \lambda_0(t).$$

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 $z = 0$.

The baseline hazard function $\lambda_0(t)$ in model (1) can take any shape as a function of t. The only requirement is that $\lambda_0(t) > 0$. This is the nonparametric part of the model and $z^T\beta$ is the parametric part of the model. So Cox's proportional hazards model is a semiparametric model.

Interpretation of a proportional hazards model

1. It is easy to show that under model (1)

$$S(t|z) = [S_0(t)]^{\exp(z^T\beta)},$$

where $S(t|z)$ is the survival function of the subpopulation with covariate z and $S_0(t)$ is the survival function of baseline population ($z = 0$). That is

$$S_0(t) = \mathrm{e}^{-\int_0^t \lambda_0(u)du}.$$

2. For any two sets of covariates z_0 and z_1,

$$\frac{\lambda(t|z_1)}{\lambda(t|z_0)} = \frac{\lambda_0(t)\mathrm{e}^{z_1^T\beta}}{\lambda_0(t)\mathrm{e}^{z_0^T\beta}} = \mathrm{e}^{(z_1-z_0)^T\beta}, \quad \text{for all } t \geq 0,$$

which is a constant over time (so the name of proportional hazards model). Equivalently,

$$\log\left[\frac{\lambda(t|z_1)}{\lambda(t|z_0)}\right] = (z_1 - z_0)^T\beta, \quad \text{for all } t \geq 0.$$

3. With one unit increase in z_k while other covariate values being held fixed, then

$$\log\left[\frac{\lambda(t|z_k+1)}{\lambda(t|z_k)}\right] = \log(\lambda(t|z_k+1)) - \log(\lambda(t|z_k)) = \beta_k.$$

Therefore, β_k is the increase in log hazard (*i.e.*, log hazard-ratio) at **any** time with unit increase in the kth covariate z_k. Equivalently,

$$\frac{\lambda(t|z_k+1)}{\lambda(t|z_k)} = \mathrm{e}^{\beta_k}, \quad \text{for all } t \geq 0.$$

So $\exp(\beta_k)$ is the hazard ratio associated with one unit increase in z_k. Furthermore, since $P[t \leq T < t + \Delta t | T \geq t, z] \approx \lambda(t|z)\Delta t$, we have

$$\frac{P[t \leq T < t + \Delta t | T \geq t, z_k + 1]}{P[t \leq T < t + \Delta t | T \geq t, z_k]} \approx \mathrm{e}^{\beta_k}, \quad \text{for all } t \geq 0.$$

so $\exp(\beta_k)$ can be loosely interpreted as the ratio of two conditional probabilities of dying in the near future given a subject is alive at any time t. Since

$$\frac{\lambda(t|z_k+1) - \lambda(t|z_k)}{\lambda(t|z_k)} = \mathrm{e}^{\beta_k} - 1.$$

So $\mathrm{e}^{\beta_k} - 1$ can be interpreted as the percentage change (increase or decrease) in hazard with one unit increase in z_k while adjusting for other covariates.

Inferential Problems

From the interpretation of the model, it is obvious that β characterizes the "effect" of z. So β should be the focus of our inference while $\lambda_0(t)$ is a nuisance "parameter". Given a sample of censored survival data, our inferential problems include:

1. Estimate β; derive its statistical properties.

2. Testing hypothesis $H_0 : \beta = 0$ or for part of β.

3. Diagnostics.

Estimation

Since the baseline hazard $\lambda_0(t)$ is left completely unspecified (infinite dimensional), ordinary likelihood methods can't be used to estimate β. Cox conceived of the idea of a partial likelihood to remove the nuisance parameter $\lambda_0(t)$ from the proposed estimating equation.

Historical Note: Cox described the proportional hazards model in JRSSB (1972), in what is now the most quoted statistical papers in history. He also outlined in this paper the method for estimation which he referred to as using conditional likelihood. It was pointed out to him in the literature that what he proposed was not a conditional likelihood and that there may be some flaws in his logic. Cox (1975) was able to recast his method of estimation through what he called "partial likelihood" and published this in *Biometrika*. This approach seemed to be based on sound inferential principles. Rigorous proofs showing the consistency and asymptotic normality were not published until 1981 when Tsiatis (*Annals of Statistics*) demonstrated these large sample properties. In 1982, Anderson and Gill (*Annals of Statistics*) simplified and generalized these results through the use of counting processes.

6.2 Estimation Using Partial Likelihood

Data and Model

1. Data: (X_i, Δ_i, z_i), $i = 1, \cdots, n$, where for the ith individual

   $X_i = \min(T_i, C_i)$.

   $\Delta_i = I(T_i \leq C_i)$.

   $z_i = (z_{i1}, z_{i2}, \cdots, z_{ip})^T$ is a vector of covariates.

2. Model: Proportional hazards model

$$\lambda(t|z_i) = \lambda_0(t)\mathrm{e}^{z_i^T\beta},$$

where

$$\lambda(t|z_i) = \lim_{h\to 0^+}\left\{\frac{P[t \le T_i < t+h | T_i \ge t, z_i]}{h}\right\}.$$

Assume that C_i and T_i are conditionally independent given z_i. Then the cause-specific hazard can be used to represent the hazard of interest. That is (in terms of conditional probabilities)

$$P[x \le X_i < x+\Delta x, \Delta_i = 1 | X_i \ge x, z_i] = P[x \le T_i < x+\Delta x | T_i \ge x, z_i] \approx \lambda_{T_i}(x|z_i)\Delta x.$$

Similar to the case of log rank test, we need to define some notation. Let us break the time axis (patient time) into a grid of points. Assume the survival time is continuous. We hence can take the grid points dense enough so that at most one death can occur within any interval.

Let $dN_i(u)$ denote the indicator for the ith individual being observed to die in $[u, u+\Delta u)$. Namely,

$$dN_i(u) = I(X_i \in [u, u+\Delta u), \Delta_i = 1).$$

Let $Y_i(u)$ denote the indicator for whether or not the ith individual is at risk at time u. Namely,

$$Y_i(u) = I(X_i \ge u).$$

Let $dN(u) = \sum_{i=1}^n dN_i(u)$ denote the number of deaths for the whole sample occurring in $[u, u+\Delta u)$. Since we are assuming Δu is sufficiently small, so $dN(u)$ is either 1 or 0 at any time u.

Let $Y(u) = \sum_{i=1}^n Y_i(u)$ be the total number from the entire sample who are at risk at time u.

Let $\mathcal{F}(x)$ denote the information up to time x (one of the grid points)

$$\mathcal{F}(x) = \{(dN_i(u), Y_i(u), z_i), i = 1, \cdots, n; \text{for grid points } u < x \text{ and } dN(x)\}.$$

Note: Conditional on $\mathcal{F}(x)$, we know who has died or was censored prior to x, when they died or were censored, together with their covariate values. We know the individuals at risk at time x and their corresponding covariate value. In addition, we also know if a death occurs at interval $[x, x + \Delta x)$.

What we don't know is the individual who was observed to die among those at risk at time x if $dN(x) = 1$.

Let $I(x)$ denote the individual in the sample who died at time x if someone died. If no one dies at time x, then $I(x) = 0$

For example, if $I(x) = j$, then this means that the jth individual in the sample with covariate vector z_j died in $[x, x + \Delta x)$.

Let $\mathcal{F}(\infty)$ denote all the data in the sample. Namely

$$\mathcal{F}(\infty) = \{(X_i, \Delta_i, z_i), i = 1, \cdots, n\}.$$

If we let $u_1 < u_2 < \cdots$ denote the value of the grid points along the time axis, then the data (with redundancy) can be expressed as

$$(\mathcal{F}(u_1), I(u_1), \mathcal{F}(u_2), I(u_2), \cdots, \mathcal{F}(\infty)).$$

Denote the observed values of the above random variables by lower cases. Then the likelihood of the parameter $\lambda_0(t)$ and β can be written as

$$\begin{aligned}
&P[\mathcal{F}(u_1) = f(u_1); \lambda_0(\cdot), \beta] \times P[I(u_1) = i(u_1) | \mathcal{F}(u_1) = f(u_1); \lambda_0(\cdot), \beta] \\
&\times P[\mathcal{F}(u_2) = f(u_2) | \mathcal{F}(u_1) = f(u_1), I(u_1) = i(u_1); \lambda_0(\cdot), \beta] \\
&\times P[I(u_2) = i(u_2) | \mathcal{F}(u_1) = f(u_1), I(u_1) = i(u_1), \mathcal{F}(u_2) = f(u_2); \lambda_0(\cdot), \beta] \\
&\times \cdots
\end{aligned}$$

and the last term can be simplified as

$$\begin{aligned} & P[I(u_2) = i(u_2)|\mathcal{F}(u_1) = f(u_1), I(u_1) = i(u_1), \mathcal{F}(u_2) = f(u_2); \lambda_0(\cdot), \beta] \\ = \; & P[I(u_2) = i(u_2)|\mathcal{F}(u_2) = f(u_2); \lambda_0(\cdot), \beta]. \end{aligned}$$

That is, the full likelihood can be written as the product of a series of conditional likelihoods.

The partial likelihood (as defined by D.R. Cox) consists of the product of every other conditional probabilities in the above presentation. That is

$$PL = \prod_{\{\text{all grid pt } u\}} P[I(u) = i(u)|\mathcal{F}(u) = f(u); \lambda_0(\cdot), \beta].$$

Suppose we have the following small data set, we will try to find find out this partial likelihood:

Patient ID	x	δ	z
1	2	1	2
2	2	0	2
3	3	1	1
4	4	1	3

It turns out that the partial likelihood is

$$PL(\beta) = \frac{\mathrm{e}^{2\beta}}{\mathrm{e}^{2\beta} + \mathrm{e}^{2\beta} + \mathrm{e}^{\beta} + \mathrm{e}^{3\beta}} \times \frac{\mathrm{e}^{\beta}}{\mathrm{e}^{\beta} + \mathrm{e}^{3\beta}} \times \frac{\mathrm{e}^{3\beta}}{\mathrm{e}^{3\beta}}. \tag{2}$$

In general, we have to consider two cases in calculating the above partial likelihood.

Case 1: Suppose conditional on $\mathcal{F}(u)$ we have $dN(u) = 0$. That is, no death is observed at time u. In such a case, $I(u) = 0$ with probability 1.

Hence for any grid point u where $dN(u) = 0$, we have

$$P[I(u) = 0|\mathcal{F}(u) = f(u)] = 1.$$

Therefore, the partial likelihood is not affected at any point u such that $dN(u) = 0$.

Case 2: $dN(u) = 1$. Conditional on $\mathcal{F}(u)$, if we know that one individual dies at time u, then it must be one of the individuals still at risk (alive and not censored) at time u; *i.e.*, among the following individuals

$$\{i : Y_i(u) = 1\}.$$

Also conditional on $\mathcal{F}(u)$, we know the covariate vector z_i associated to each individual i such that $Y_i(u) = 1$. Therefore, we ask the following question:

> *Among* $Y(u) = \sum_{i=1}^{n} Y_i(u)$ *individuals, what is the probability that the observed death happened to the ith subject (who is actually observed to die at u) rather than to the other patients?*

Unlike the null hypothesis case for the two-sample problem, the probabilities of choosing these subjects are not equally likely, but rather, they are **proportional to** their cause-specific hazard of dying at time u, which can be derived as follows:

Let A_i = the event that subject i is going to die in $[u, u + \Delta u)$ given that he/she is still alive at u. If a patient is not at risk at u (*i.e.*, $Y_i(u) = 0$), then $A_i = \phi$. Since we chose Δu to be so small that there is at most one death in $[u, u + \Delta u)$, so we know

$$A_1, A_2, \cdots, A_n \text{ are mutually exclusive.}$$

Because of the independence of survival times and censoring times, those $Y(u)$ patients who are at risk at u (not censored and still alive at u) make up a random sample of the subpopulation consisting of the patients who will survive up to u (and with the same covariate value). Under independent censoring assumption, we already showed in Chapter 3 that the cause-specific hazard is the same as the hazard of interest; *i.e.*,

$$\lambda(u, \delta_i = 1|z_i) = \lambda(u, |z_i).$$

Since Δu is chosen to be very small, so

$$\begin{aligned} P[A_i] &\approx Y_i(u)\lambda(u, \delta_i = 1|z_i)\Delta u \\ &= Y_i(u)\lambda(u, |z_i)\Delta u. \\ &= Y_i(u)\lambda_0(u)\exp(z_i^T\beta)\Delta u, \end{aligned}$$

where the last equation is due to the assumption of the cox model. Therefore

$$\begin{aligned} & P[I(u) = i(u)|\mathcal{F}(u) = f(u); \lambda_0(\cdot), \beta] \\ = \; & P[A_{i(u)}|A_1 \cup \cdots \cup A_n] \\ = \; & \frac{P[A_{i(u)}]}{\sum_{l=1}^n P[A_l]} \\ \approx \; & \frac{\lambda_0(u) \exp(z_{i(u)}^T \beta) \Delta u}{\sum_{l=1}^n \lambda_0(u) \exp(z_l^T \beta) Y_l(u) \Delta u} \\ = \; & \frac{\exp(z_{i(u)}^T \beta)}{\sum_{l=1}^n \exp(z_l^T \beta) Y_l(u)}. \end{aligned}$$

Here $Y_{i(u)}(u) = 1$ since we know this patient had to be at risk at u (since we know that this patient died in $[u, u + \Delta u)$).

Combining these cases, the partial likelihood can be written as

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

Remark: To be formal, we need to define z_0 even though it is never used. We can, for example, take $z_0 = 0$.

Other equivalent ways of writing the partial likelihood include: Let $t_1, \cdots, t_d$ define the distinct death times, then

$$\begin{aligned} PL(\beta) &= \prod_{j=1}^d \left[\frac{\exp(z_{i(t_j)}^T \beta)}{\sum_{l=1}^n \exp(z_l^T \beta) Y_l(t_j)} \right]; \\ PL(\beta) &= \prod_{i=1}^n \prod_{\{\text{all grid pt } u\}} \left[\frac{\exp(z_i^T \beta)}{\sum_{l=1}^n \exp(z_l^T \beta) Y_l(u)} \right]^{dN_i(u)}; \\ PL(\beta) &= \prod_{i=1}^n \left[\frac{\exp(z_i^T \beta)}{\sum_{l=1}^n \exp(z_l^T \beta) Y_l(x_i)} \right]^{\delta_i}. \end{aligned}$$

Remark: Stare at these different representations for a while, you will convince yourself that they are all equivalent.

The importance of using the partial likelihood is that this function depends **only** on β,

the parameter of interest, and is free of the baseline hazard $\lambda_0(t)$, which is infinite dimensional nuisance function.

Cox suggested treating PL as a regular likelihood function and making inference on β accordingly. For example, we maximize the partial likelihood to get the estimate of β, often called MPLE (maximum partial likelihood estimate), and use the minus of the second derivative of the log partial likelihood as the information matrix, etc.

Properties of the score of the partial likelihood

For ease of presentation, let us focus on one covariate case. The extension is straightforward.

Obviously, the log partial likelihood function of β is

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

The score function is

$$U(\beta) = \frac{\partial \ell(\beta)}{\partial \beta} = \sum_{\{\text{all grid pts } u\}} dN(u) \left[z_{I(u)} - \frac{\sum_{l=1}^{n} z_l \exp(z_l\beta) Y_l(u)}{\sum_{l=1}^{n} \exp(z_l\beta) Y_l(u)} \right],$$

and the second derivative is

$$\frac{\partial^2 \ell(\beta)}{\partial \beta^2} = -\sum_u dN(u) \left[\frac{\sum_{l=1}^{n} z_l^2 \exp(z_l\beta) Y_l(u)}{\sum_{l=1}^{n} \exp(z_l\beta) Y_l(u)} - \left(\frac{\sum_{l=1}^{n} z_l \exp(z_l\beta) Y_l(u)}{\sum_{l=1}^{n} \exp(z_l\beta) Y_l(u)} \right)^2 \right].$$

Define

$$\bar{z}(u,\beta) = \frac{\sum_{l=1}^{n} z_l \exp(z_l\beta) Y_l(u)}{\sum_{l=1}^{n} \exp(z_l\beta) Y_l(u)} = \sum_{l=1}^{n} z_l w_l,$$

where

$$w_l = \frac{\exp(z_l\beta) Y_l(u)}{\sum_{l=1}^{n} \exp(z_l\beta) Y_l(u)}$$

is the weight that is proportional to the hazard of the individual failing. So $\bar{z}(u,\beta)$ can be interpreted as the weighted average of the covariate z among those individuals still at risk at time u with weights w_l.

Define

$$\begin{aligned} V_z(u,\beta) &= \left[\frac{\sum_{l=1}^n z_l^2 \exp(z_l\beta)Y_l(u)}{\sum_{l=1}^n \exp(z_l\beta)Y_l(u)} - \left(\frac{\sum_{l=1}^n z_l \exp(z_l\beta)Y_l(u)}{\sum_{l=1}^n \exp(z_l\beta)Y_l(u)}\right)^2\right] \\ &= \left[\frac{\sum_{l=1}^n z_l^2 \exp(z_l\beta)Y_l(u)}{\sum_{l=1}^n \exp(z_l\beta)Y_l(u)} - \left(\bar{z}(u,\beta)\right)^2\right] \\ &= \sum_{l=1}^n z_l^2 w_l - \left(\bar{z}(u,\beta)\right)^2. \end{aligned}$$

This can be shown to be equal to

$$V_z(u,\beta) = \sum_{l=1}^n \left[\frac{(z_l - \bar{z}(u,\beta))^2 \exp(z_l\beta)Y_l(u)}{\sum_{l=1}^n \exp(z_l\beta)Y_l(u)}\right] = \sum_{l=1}^n \left(z_l - \bar{z}(u,\beta)\right)^2 w_l.$$

This last representation says that $V_z(u,\beta)$ can be interpreted as the weighted variance of the covariates among those individuals still at risk at u and hence $V_z(u,\beta) > 0$. Consequently,

$$\frac{\partial^2 \ell(\beta)}{\partial \beta^2} = -\sum_u dN(u) V_z(u,\beta) < 0.$$

The above property can also be displayed graphically. For example, the partial likelihood function (2) looks like:

Therefore $\ell(\beta)$ has a unique maximizer and can be obtained uniquely by solving the following partial likelihood equation:

$$U(\beta) = \frac{\partial \ell(\beta)}{\partial \beta} = \sum_{\{\text{all grid pts } u\}} dN(u)\left[z_{I(u)} - \frac{\sum_{l=1}^n z_l \exp(z_l\beta)Y_l(u)}{\sum_{l=1}^n \exp(z_l\beta)Y_l(u)}\right] = 0.$$

This maximizer $\hat{\beta}$ defines the MPLE of β.

Terminology: The quantity

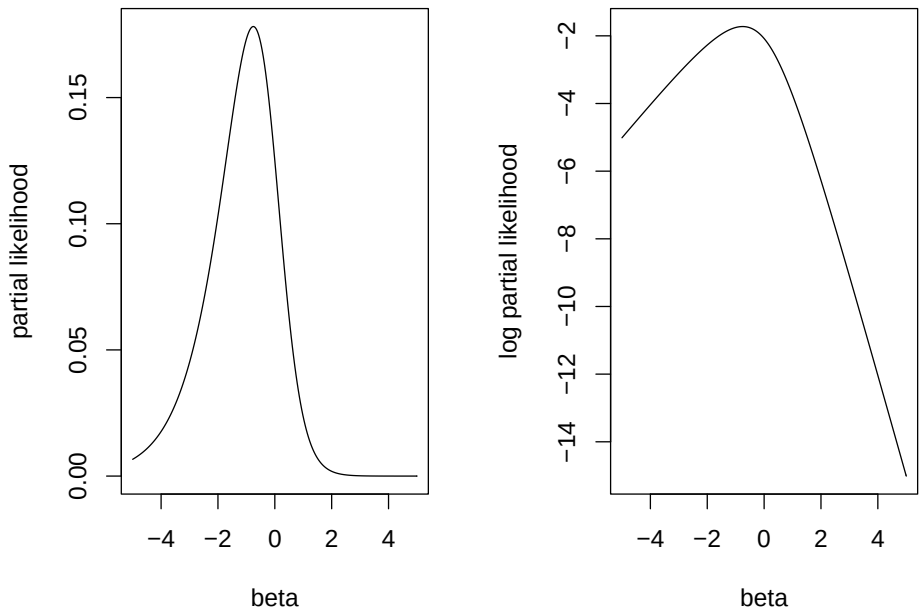
$$-\frac{\partial^2 \ell(\beta)}{\partial \beta^2} = \sum_u dN(u) V_z(u,\beta)$$

is defined as the partial likelihood observed information and is denoted by $J(\beta)$.

Ultimately, we want to show that the MPLE $\hat{\beta}$ has nice statistical properties. These include:

- Consistency: That is, $\hat{\beta}$ will converge to the true value of β which generated the data as the sample size gets larger. We call this true value β_0.

Figure 1: *The partial likelihood (2)*



- Asymptotic Normality: $\hat{\beta}$ will be approximately normally distributed with mean β_0 and a variance which can be estimated from the data. This approximation will be better as the sample size gets larger. This result is useful in making inference for the true β.

- Efficiency: Among all other competing estimators for β, the MPLE has the smallest variance, at least, when the sample size gets larger.

In order to show the properties for $\hat{\beta}$, we expand $U(\hat{\beta})$ at the true value β_0 using Taylor expansion:

$$0 = U(\hat{\beta}) \approx U(\beta_0) + \frac{\partial U(\beta_0)}{\partial \beta}(\hat{\beta} - \beta_0).$$

Since

$$\frac{\partial U(\beta_0)}{\partial \beta} = \frac{\partial^2 \ell(\beta_0)}{\partial \beta^2} = -J(\beta_0),$$

therefore

$$(\hat{\beta} - \beta_0) \approx [J(\beta_0)]^{-1} U(\beta_0)$$

This expression indicates that we need to investigate the properties of the score function $U(\beta_0)$

$$U(\beta_0) = \sum_u dN(u) \left[z_{I(u)} - \bar{z}(u, \beta_0) \right].$$

Properties of the score:

(1) $\mathrm{E}[U(\beta_0)] = 0$.

Since

$$\begin{aligned} \mathrm{E}[U(\beta_0)] &= \mathrm{E}\left[\sum_u dN(u) \left(z_{I(u)} - \bar{z}(u, \beta_0) \right) \right] \\ &= \sum_u \mathrm{E}\left[dN(u) \left(z_{I(u)} - \bar{z}(u, \beta_0) \right) \right], \end{aligned}$$

and

$$\begin{aligned} & \mathrm{E}\left[dN(u) \left(z_{I(u)} - \bar{z}(u, \beta_0) \right) \right] \\ = \; & \mathrm{E}\left[\mathrm{E}\left[dN(u) \left(z_{I(u)} - \bar{z}(u, \beta_0) \right) \middle| \mathcal{F}(u) \right] \right] \end{aligned}$$

Conditional on $\mathcal{F}(u)$, $dN(u)$ and $\bar{z}(u, \beta_0)$ are both known. Consequently the inner expectation can be written as

$$dN(u) \left[\mathrm{E}[z_{I(u)} | \mathcal{F}(u)] - \bar{z}(u, \beta_0) \right].$$

Remember that $I(u)$ is the patient identifier for the individual that dies at time u and is set to zero if no one dies at u. If no one dies at u, then $dN(u) = 0$, and hence the above quantity is zero. If someone dies at u, then $dN(u) = 1$, and conditional on $\mathcal{F}(u)$, we know it has to be one of the $Y(u)$ people at risk at time u; *i.e.*, $I(u)$ must be one of the values $\{i : Y_i = 1\}$.

The conditional distribution of $z_{I(u)}$ given $\mathcal{F}(u)$ can be derived through the conditional distribution of $I(u)$ given $\mathcal{F}(u)$ as shown in Table 1.

Table 1: *Conditional distribution of $z_{I(u)}$ given $\mathcal{F}(u)$*

Values of $I(u)$	Values of $z_{I(u)}$	Probability
1	z_1	$\exp(z_1\beta_0)Y_1(u)/\sum_{l=1}^n \exp(z_l\beta_0)Y_l(u) = w_1$
2	z_2	$\exp(z_2\beta_0)Y_2(u)/\sum_{l=1}^n \exp(z_l\beta_0)Y_l(u) = w_2$
$\vdots$	$\vdots$	$\vdots$
n	z_n	$\exp(z_n\beta_0)Y_n(u)/\sum_{l=1}^n \exp(z_l\beta_0)Y_l(u) = w_n$

Therefore

$$\mathrm{E}[z_{I(u)}|\mathcal{F}(u)] = \sum_{l=1}^n z_l w_l = \frac{\sum_{l=1}^n z_l \exp(z_l\beta_0)Y_l(u)}{\sum_{l=1}^n \exp(z_l\beta_0)Y_l(u)} = \bar{z}(u,\beta_0).$$

From this, we immediately get

$$\mathrm{E}[U(\beta_0)] = 0.$$

Note: From the conditional distribution of $z_{I(u)}$ given $\mathcal{F}(u)$, it is easy to see the conditional variance of $z_{I(u)}$

$$\begin{aligned}
\mathrm{Var}[z_{I(u)}|\mathcal{F}(u)] &= \sum_{l=1}^n \left(z_l - \mathrm{E}[z_{I(u)}|\mathcal{F}(u)]\right)^2 w_l \\
&= \frac{\sum_{l=1}^n (z_l - \bar{z}(u,\beta_0))^2 \exp(z_l\beta_0)Y_l(u)}{\sum_{l=1}^n \exp(z_l\beta_0)Y_l(u)} \\
&= V_z(u,\beta_0).
\end{aligned}$$

(2) Finding an unbiased estimate for the variance of $U(\beta_0)$

Since $\mathrm{E}[U(\beta_0)] = 0$, so

$$\begin{aligned}
\mathrm{Var}[U(\beta_0)] &= \mathrm{E}[U(\beta_0)]^2 \\
&= \mathrm{E}\left[\sum_u dN(u)\left[z_{I(u)} - \bar{z}(u,\beta_0)\right]\right]^2 \\
&= \mathrm{E}\left[\sum_u \left\{dN(u)\left[z_{I(u)} - \bar{z}(u,\beta_0)\right]\right\}^2\right] \\
&+ \mathrm{E}\left[\sum_{u\neq u'} \left\{dN(u)\left[z_{I(u)} - \bar{z}(u,\beta_0)\right]\right\}\left\{dN(u')\left[z_{I(u')} - \bar{z}(u',\beta_0)\right]\right\}\right]
\end{aligned}$$

As usual, we will take an arbitrary cross-product and show it has zero expectation. Assume $u' > u$ and denote

$$A(u) = dN(u)\left[z_{I(u)} - \bar{z}(u,\beta_0)\right], \quad A(u') = dN(u')\left[z_{I(u)} - \bar{z}(u',\beta_0)\right].$$

Then the expectation of the cross-product is

$$\begin{aligned}
&\mathrm{E}\left[A(u)A(u')\right] \\
= \;& \mathrm{E}\left[\mathrm{E}\left[A(u)A(u')\middle|\,\mathcal{F}(u')\right]\right].
\end{aligned}$$

Since $u' > u$, conditional on $\mathcal{F}(u')$, $A(u)$ is known. So

$$\mathrm{E}\left[A(u)A(u')\middle|\,\mathcal{F}(u')\right] = A(u)\mathrm{E}\left[A(u')\middle|\,\mathcal{F}(u')\right] = 0.$$

Therefore

$$\begin{aligned}
\mathrm{Var}[U(\beta_0)] &= \mathrm{E}\sum_u \left[A^2(u)\right] \\
&= \sum_u \mathrm{E}\left[A^2(u)\right] \\
&= \sum_u \mathrm{E}\left[\mathrm{E}\left[A^2(u)\middle|\,\mathcal{F}(u)\right]\right]
\end{aligned}$$

The inner conditional expectation is

$$\mathrm{E}\left[A^2(u)\middle|\,\mathcal{F}(u)\right] = \mathrm{E}\left[\left\{dN(u)\left[z_{I(u)} - \bar{z}(u,\beta_0)\right]\right\}^2\middle|\,\mathcal{F}(u)\right].$$

Since we pick the grid points in our partition of time fine enough so that $dN(u)$ is either 0 or 1, so $dN^2(u) = dN(u)$. Hence

$$\mathrm{E}\left[A^2(u)\middle|\,\mathcal{F}(u)\right] = \mathrm{E}\left[dN(u)\left[z_{I(u)} - \bar{z}(u,\beta_0)\right]^2\middle|\,\mathcal{F}(u)\right].$$

Conditional on $\mathcal{F}(u)$, $dN(u)$ is known, $\bar{z}(u,\beta_0)$ is also known and from Table 1

$$\bar{z}(u,\beta_0) = \mathrm{E}[z_{I(u)}|\mathcal{F}(u)].$$

Therefore

$$\begin{aligned}
\mathrm{E}\left[A^2(u)\middle|\,\mathcal{F}(u)\right] &= dN(u)\mathrm{E}\left[\left[z_{I(u)} - \bar{z}(u,\beta_0)\right]^2\middle|\,\mathcal{F}(u)\right] \\
&= dN(u)\mathrm{Var}[z_{I(u)}|\mathcal{F}(u)] \\
&= dN(u)V_z(u,\beta_0).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathrm{Var}\left[U(\beta_0)\right] &= \sum_u \mathrm{E}\left[dN(u)V_z(u,\beta_0)\right] \\
&= \mathrm{E}\left[\sum_u dN(u)V_z(u,\beta_0)\right].
\end{aligned}$$

Note that the quantity $\sum_u dN(u)V_z(u,\beta_0)$ is a statistic (can be calculated from the observed data), so $\sum_u dN(u)V_z(u,\beta_0)$ is an unbiased estimate of $\mathrm{Var}\left[U(\beta_0)\right]$. In fact, $\sum_u dN(u)V_z(u,\beta_0)$ is the partial likelihood observed information $J(\beta_0)$ we defined before.

Conclusion

The score $U(\beta_0) = \sum_u A(u)$ is a sum of conditionally uncorrelated mean zero random variables and its variance can be unbiasedly estimated by

$$J(\beta_0) = \sum_u dN(u)V_z(u,\beta_0).$$

By the martingale CLT, we have:

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

Previously, we have shown that

$$(\hat{\beta} - \beta_0) \approx [J(\beta_0)]^{-1} U(\beta_0).$$

Treating $J(\beta_0)$ as a constant, we get the approximate distribution of $(\hat{\beta} - \beta_0)$

$$(\hat{\beta} - \beta_0) \overset{a}{\sim} \mathrm{N}(0, J^{-1}(\beta_0)).$$

Of course, in practice, β_0 is unknown. But we can substitute $\hat{\beta}$ for β_0 and use $J^{-1}(\hat{\beta})$ as the estimated variance of $\hat{\beta}$. That is, we use the following approximate distribution for $(\hat{\beta} - \beta_0)$

$$(\hat{\beta} - \beta_0) \overset{a}{\sim} \mathrm{N}(0, J^{-1}(\hat{\beta})),$$

where

$$J(\hat{\beta}) = \sum_u dN(u) \left[V_z(u, \hat{\beta}) \right],$$

and $\hat{\beta}$ is the MPLE of β solving the following equation

$$U(\hat{\beta}) = \sum_u dN(u) \left[z_{I(u)} - \bar{z}(u, \hat{\beta}) \right] = 0.$$

Inference with a Single Covariate

Assume a proportional hazards model with a single covariate z

$$\lambda(t) = \lambda_0(t) \mathrm{e}^{z\beta}.$$

After we get our data (x_i, δ_i, z_i), we can obtain the MPLE $\hat{\beta}$ by solving the partial likelihood equation; *i.e.*, setting the partial score to zero. Then asymptotically,

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

We can use this fact to construct confidence interval for β and test the hypothesis $H_0 : \beta = \beta_0$, etc. For example, a $(1 - \alpha)$ CI of β is

$$\hat{\beta} \pm z_{\alpha/2} [J^{-1}(\hat{\beta})]^{1/2}.$$

Myelomatosis data revisited: We analyzed myelomatosis data and did not find statistically significant difference between treatments 1 and 2. We want to quantify the difference by assuming the hazards of these two treatments are proportional to each other. Define a treatment indicator `trt1` which takes value 0 for treatment 1 and takes value 1 for treatment 2. Then we can use `Proc Phreg` for this purpose.

```
proc phreg data=myel;
  model dur*status(0)=trt1;
run;
```

Part of the output is given as follows:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

                                                    16:43 Thursday, March 2, 2000  15

                                The PHREG Procedure

     Data Set: WORK.MYEL
     Dependent Variable: DUR
     Censoring Variable: STATUS
     Censoring Value(s): 0
     Ties Handling: BRESLOW

                              Summary of the Number of
                              Event and Censored Values

                                                                 Percent
                          Total       Event    Censored        Censored

                             25          17           8           32.00


                          Testing Global Null Hypothesis: BETA=0

                        Without        With
    Criterion         Covariates    Covariates    Model Chi-Square

    -2 LOG L             94.084        92.765       1.319 with 1 DF (p=0.2508)
    Score                   .             .         1.297 with 1 DF (p=0.2547)
    Wald                    .             .         1.263 with 1 DF (p=0.2610)


                          Analysis of Maximum Likelihood Estimates

                          Parameter     Standard      Wald         Pr >          Risk
Variable    DF             Estimate       Error    Chi-Square   Chi-Square       Ratio

TRT1         1             0.572807      0.50960      1.26344       0.2610       1.773
```

So $\hat{\beta} = 0.5728$ with standard error 0.5096. This means that compared to treatment 1, treatment 2 will increase the hazard of dying at *any* time by 77% ($\exp(\hat{\beta}) - 1$). A 95% CI of β is

$$\hat{\beta} \pm 1.96 * \text{se}[\hat{\beta}] = 0.5728 \pm 1.96 * 0.5096 = [-0.426, 1.572].$$

And a 95% CI for the hazard ratio $\exp(\beta)$ is

$$[\text{e}^{-0.426}, \text{e}^{1.572}] = [0.653, 4.816].$$

Note: The output also gives three tests for $H_0 : \beta = 0$: likelihood ratio, score and Wald tests.

Comparison of score test and two-sample log rank test

Assume z is the dichotomous indicator for treatment; *i.e.*,

$$z = \begin{cases} 1 & \text{for treatment 1} \\ 0 & \text{for treatment 0} \end{cases},$$

and the proportional hazards model:

$$\lambda(t) = \lambda_0(t)\text{e}^{z\beta}.$$

Score test: Under $H_0 : \beta = 0$, the score $U(0)$ (evaluated under H_0) has the distribution

$$U(0) \overset{a}{\sim} \text{N}(0, J(0)).$$

Or equivalently,

$$\left[\frac{U(0)}{J^{1/2}(0)}\right]^2 \overset{a}{\sim} \chi^2_1.$$

Since the score $U(0)$ has the expression

$$U(0) = \sum_u dN(u) \left[z_{I(u)} - \bar{z}(u, 0)\right].$$

Then

1. If a death occurs at time u, then $dN(u) = 1$, in which case there will a contribution to $U(0)$ by adding $[z_{I(u)} - \bar{z}(u, 0)]$. Otherwise no contribution.

2. Since $z = 1$ for treatment 1 and $z = 0$ for treatment 0, $z_{I(u)}$ will then the number of deaths at time u from treatment 1.

3. Under $H_0 : \beta = 0$, $\bar{z}(u, 0)$ is simplified to be

$$\bar{z}(u, 0) = \frac{\sum_{l=1}^{n} z_l Y_l(u)}{\sum_{l=1}^{n} Y_l(u)},$$

which is the proportion of individuals in group 1 among those at risk at time u. Since we only assume one death at time u, this proportion is the expected number of death for treatment 1 among those at risk at time u, under the null hypothesis of no treatment difference.

4. Therefore, $U(0)$ is the sum over the death times of the observed number of deaths from treatment 1 minus the expected number of deaths under the null hypothesis. This was the numerator of the two-sample log rank test:

$$\sum_u \left[dN_1(u) - \frac{Y_1(u)}{Y(u)} dN(u) \right]$$

where $dN_1(u) = \#$ of observed deaths from treatment 1, $Y_1(u) = \#$ at risk at time u from treatment 1, $Y(u) =$ total $\#$ at risk at time u from 2 treatments, $dN(u) =$ total $\#$ of deaths from 2 treatments.

5. The denominator of the score test was computed as

$$J^{1/2}(0) = \left[\sum_u dN(u) V_z(u, 0) \right]^{1/2},$$

where

$$V_z(u, 0) = \frac{\sum_l [z_l - \bar{z}(u, 0)]^2 Y_l(u)}{\sum_l Y_l(u)}.$$

Note: Among the $Y(u)$ individuals at risk at time u, there are $Y_1(u)$ individuals whose z_l value of $z_l = 1$ and $Y_0(u)$ individuals whose z_l value of $z_l = 0$. We already argued that

$$\bar{z}(u, 0) = \frac{Y_1(u)}{Y(u)}.$$

Therefore,

$$\begin{aligned} V_z(u,0) &= \frac{\sum_l [z_l - \bar{z}(u,0)]^2 Y_l(u)}{\sum_l Y_l(u)} \\ &= \frac{\left[1 - \frac{Y_1(u)}{Y(u)}\right]^2 Y_1(u) + \left[0 - \frac{Y_1(u)}{Y(u)}\right]^2 Y_0(u)}{Y(u)} \quad (z_l(u) \text{ takes 1 or 0}) \\ &= \frac{\frac{Y_0^2(u)Y_1(u)}{Y^2(u)} + \frac{Y_1^2(u)Y_0(u)}{Y^2(u)}}{Y(u)} \qquad (Y_1(u) + Y_0(u) = Y(u)) \\ &= \frac{Y_0(u)Y_1(u)Y(u)}{Y^3(u)} \\ &= \frac{Y_0(u)Y_1(u)}{Y^2(u)}. \end{aligned}$$

Therefore,

$$J(0) = \sum_u dN(u) \frac{Y_0(u)Y_1(u)}{Y^2(u)}.$$

Let us contrast this with the variance used to compute the logrank test statistic:

$$\sum_u \left[\frac{Y_1(u)Y_0(u)dN(u)[Y(u) - dN(u)]}{Y^2(u)[Y(u) - 1]} \right].$$

Note: In the special case where $dN(u)$ can only be one or zero, then above expression reduces to

$$\sum_u \left[\frac{Y_1(u)Y_0(u)dN(u)[Y(u) - 1]}{Y^2(u)[Y(u) - 1]} \right] = \sum_u \left[\frac{Y_1(u)Y_0(u)dN(u)}{Y^2(u)} \right],$$

which is exactly equal to $J(0)$.

Therefore, we have demonstrated with continuous survival time data with no ties, the score test of the hypothesis $H_0 : \beta = 0$ in the proportional hazards model is exactly the same as the logrank test for dichotomous covariate z.

The score test

$$\left[\frac{U(0)}{J^{1/2}(0)} \right]^2$$

can be used to test the hypothesis $H_0 : \beta = 0$ for the model

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

for any covariate value z, whether or not z is discrete or continuous. The null hypothesis $H_0 : \beta = 0$ implies that the hazard rate at any time t is unaffected by the covariate z. This also implies that the survival distribution does not depend on z. The alternative hypothesis $H_A : \beta \neq 0$ implies that hazard rate increases or decreases (depending on the sign of β) as z increases throughout all time. Therefore, belief in this alternative hypothesis would mean that individuals with a higher value of z would have stochastically larger (or smaller depending on the sign of β) survival distribution than those individuals with a smaller values of z. The `test` command in `Proc Lifetest` computes the score test of the hypothesis $H_0 : \beta = 0$ for the proportional hazards model. Consequently, when using the `test` command, the covariate z is not limited to being dichotomous, nor discrete.

For example, we can test the treatment difference between treatments 1 and 2 for myelomatosis data using the following `SAS` command:

```
proc lifetest data=myel;
  time dur*status(0);
  test trt;
run;
```

and part of the output is presented in the following:

Univariate Chi-Squares for the LOG RANK Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
TRT	-2.3376	2.0522	1.2975	0.2547

Covariance Matrix for the LOG RANK Statistics

Variable	TRT
TRT	4.21151

Forward Stepwise Sequence of Chi-Squares for the LOG RANK Test

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment

TRT	1	1.2975	0.2547	1.2975	0.2547

Likelihood Ratio Test

As in the ordinary likelihood theory, the (partial) likelihood ratio test can also be used to test the null hypothesis:

$$H_0 : \beta = \beta_0.$$

Recall that $\ell(\beta)$ is the log partial likelihood. Intuitively, if H_0 is true, then $\hat{\beta}$, the MPLE of β, should be close to β_0. Hence $\ell(\hat{\beta})$ should be close to $\ell(\beta_0)$. Since $\ell(\hat{\beta}) - \ell(\beta_0)$ is always non-negative, so we should reject H_0 when this difference is large.

The likelihood ratio test uses the fact that

$$2\left[\ell(\hat{\beta}) - \ell(\beta_0)\right] \overset{a}{\sim} \chi^2_1, \quad \text{under } H_0 : \beta = 0.$$

Therefore, for a given level of significance α, we reject $H_0 : \beta = \beta_0$ if

$$2\left[\ell(\hat{\beta}) - \ell(\beta_0)\right] \geq \chi^2_{1,\alpha}$$

where $\chi^2_{1,\alpha}$ is the value such that $P[\chi^2_1 > \chi^2_{1,\alpha}] = \alpha$.

Expanding $\ell(\beta_0)$ at the MPLE $\hat{\beta}$, we get

$$\ell(\beta_0) \approx \ell(\hat{\beta}) + \frac{d\ell(\hat{\beta})}{d\beta}(\beta_0 - \hat{\beta}) + \frac{1}{2!}\frac{d^2\ell(\hat{\beta})}{d^2\beta}(\beta_0 - \hat{\beta})^2.$$

Since MPLE $\hat{\beta}$ maximizes $\ell(\beta)$, *i.e.*,

$$U(\hat{\beta}) = \frac{d\ell(\hat{\beta})}{d\beta} = 0,$$

and

$$\frac{d^2\ell(\hat{\beta})}{d^2\beta} = -J(\hat{\beta}),$$

so

$$2\left[\ell(\hat{\beta}) - \ell(\beta_0)\right] \approx J(\hat{\beta})(\hat{\beta} - \beta_0)^2.$$

We already derived that

$$(\hat{\beta} - \beta_0) \overset{a}{\sim} \mathrm{N}(0, J^{-1}(\hat{\beta})).$$

Therefore,

$$\begin{aligned} 2\left[\ell(\hat{\beta}) - \ell(\beta_0)\right] &\approx J(\hat{\beta})(\hat{\beta} - \beta_0)^2 \\ &= \left[\frac{\hat{\beta} - \beta_0}{J^{-1/2}(\hat{\beta})}\right]^2 \overset{a}{\sim} \chi^2_1 \text{ under } H_0 : \beta = \beta_0. \end{aligned}$$

Note: The `SAS` procedure `Phreg` can ONLY handle right censored data.

6.3 Handling Tied Data in Proportional Hazards Models

So far we have assumed that there is no tied observed survival time in our data when we construct the partial likelihood function for the proportional hazards model. However, in practice, it is quite common for our data to contain tied survival times due to obvious reasons. Therefore, we need a different technique to construct the partial likelihood in the presence of tied data. Throughout this subsection, we will work with the following super 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

where $x_1 = x_2 < x_3 < x_4 < x_5$. So the first two patients have tied survival times. 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) in this data set, intuitively, the partial likelihood function of β will take the following form

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

where $L_j(\beta)$ is the component in the partial likelihood corresponding to the jth distinct survival time. Since the second and third survival times x_4 and x_5 are distinct, $L_2(\beta)$ and $L_3(\beta)$ can be constructed in the usual way. So we will focus on the construction of $L_1(\beta)$. In fact,

$$L_2(\beta) = \frac{e^{z_4\beta}}{e^{z_4\beta} + e^{z_5\beta}}, \quad \text{and} \quad L_3(\beta) = 1.$$

We will discuss 4 methods that are implemented in `SAS`.

1. 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.

Without any knowledge of the true ordering of the survival times of patients 1 and 2, we have to consider all possible orderings. 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

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

and $P[A_1], P[A_2]$ are given in the usual way:

$$\begin{aligned} P[A_1] &= \frac{e^{z_1\beta}}{e^{z_1\beta} + e^{z_2\beta} + e^{z_3\beta} + e^{z_4\beta} + e^{z_5\beta}} \times \frac{e^{z_2\beta}}{e^{z_2\beta} + e^{z_3\beta} + e^{z_4\beta} + e^{z_5\beta}} \\ P[A_2] &= \frac{e^{z_2\beta}}{e^{z_2\beta} + e^{z_1\beta} + e^{z_3\beta} + e^{z_4\beta} + e^{z_5\beta}} \times \frac{e^{z_1\beta}}{e^{z_1\beta} + e^{z_3\beta} + e^{z_4\beta} + e^{z_5\beta}} \end{aligned}$$

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). Specifically, 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 fit with the exact method using the following `SAS` code:

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

Of course, 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 $d_j \times d_j! = 5 \times 5! = 6000$ different terms have to be calculated to get $L_j(\beta)$. Because of this computational difficulties, two methods have been proposed to approximate the exact partial likelihood.

2. Breslow's Approximation (default in `Proc Phreg`): Obviously, we can have the following approximation for our example:

$$\begin{aligned}
\frac{e^{\beta z_2}}{e^{z_2\beta}+e^{z_3\beta}+e^{z_4\beta}+e^{z_5\beta}} &\approx \frac{e^{z_2\beta}}{e^{z_1\beta}+e^{z_2\beta}+e^{z_3\beta}+e^{z_4\beta}+e^{z_5\beta}} \\
\frac{e^{z_1\beta}}{e^{z_1\beta}+e^{z_3\beta}+e^{z_4\beta}+e^{z_5\beta}} &\approx \frac{e^{z_1\beta}}{e^{z_1\beta}+e^{z_2\beta}+e^{z_3\beta}+e^{z_4\beta}+e^{z_5\beta}}
\end{aligned}$$

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

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

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

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

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

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

where D is the total distinct events. This approximation was proposed by Breslow (1974) and is the default in `Proc Phreg` of `SAS`.

Obviously, if at each distinct survival time the number of events (failures) d_j is small or/and the number of patients at risk n_j is large (so the ratio d_j/n_j is small), then Breslow's approximation should work well (the approximated partial likelihood should be very close to the exact partial likelihood) However, if these conditions do not satisfy, the approximation can be poor. Therefore Efron (1977) suggested another approximation.

3. Efron's Approximation: For our example, $L_1(\beta)$ in the exact partial likelihood using the exact method can be written as

$$L_1(\beta) = \frac{bc}{a(a-b)} + \frac{bc}{a(a-c)},$$

which can be approximated by

$$L_1(\beta) = \frac{2bc}{a(a-(b+c)/2)}.$$

This motivates the general approximation:

$$L_1(\beta) = \frac{\exp^{\sum_{l\in D_1} z_l\beta}}{\prod_{j=1}^{d_1}\left(\sum_{l\in R_1}\exp^{z_l\beta} - \frac{j-1}{d_1}\sum_{l\in D_1}\exp^{z_l\beta}\right)}.$$

We can specify the option `ties=efron` in `Proc Phreg` for this approximation.

4. 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, which may arise in some applications. For example, suppose we are interested in studying the number of times we drop a dish before it breaks. 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

$$L_1(\beta) = \frac{\mathrm{e}^{(z_1+z_2)\beta}}{\sum_{\text{all } D_j} \mathrm{e}^{s_j\beta}},$$

where D_j are $\begin{pmatrix} 5 \\ 2 \end{pmatrix} = 10$ possible combinations.

Obviously, 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`.

Note that conditional logistic model is a special case of this model. So `Proc Phreg` can be used to fit conditional logistic model. Also note that this method can be even more computationally intensive than, say, the exact method.

6.4 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:

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

where z is a $(p \times 1)$ vector and $\beta = (\beta_1, \cdots, \beta_p)^T$ is a $(p \times 1)$ vector of regression coefficients.

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

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

and the log partial likelihood of β is

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

Note: z_l is the covariate value for the lth individual; *i.e.*, $z_l = (z_{l1}, \cdots, z_{lp})^T$.

The maximum partial likelihood estimate $\hat{\beta}$ (MPLE) of β is obtained by maximizing $\ell(\beta)$, *i.e.*, by setting the score vector to be zero

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

where

$$\frac{\partial \ell(\beta)}{\partial \beta} = \left(\frac{\partial \ell(\beta)}{\partial \beta_1}, \cdots, \frac{\partial \ell(\beta)}{\partial \beta_p} \right)^T.$$

Similar to the previous chapter, we have

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_u dN(u) \left[z_{I(u)j} - \bar{z}_j(u, \beta) \right],$$

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

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

is the weighted average of the jth element of the covariate vector for the individuals at risk at time u.

If we denote

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

then the partial likelihood equation can be expressed as

$$U(\beta) = \sum_u dN(u) \left[Z_{I(u)}^{p\times 1} - \bar{Z}^{p\times 1}(u, \beta) \right] = 0^{p\times 1}.$$

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

$$a^T H a < 0 \text{ for all } a_{p\times 1} \neq 0,$$

where

$$H = \frac{\partial^2 \ell(\beta)}{\partial \beta^T \partial \beta} = \left[\frac{\partial^2 \ell(\beta)}{\partial \beta_j \partial \beta_{j'}} \right]_{p \times p}.$$

Equivalently,

$$J(\beta) = -\frac{\partial^2 \ell(\beta)}{\partial \beta^T \partial \beta} = -\left[\frac{\partial^2 \ell(\beta)}{\partial \beta_j \partial \beta_{j'}} \right]$$

is positive definite.

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

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

where $V_{j,j'}(u, \beta)$ is the 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

$$w_l = \frac{\exp(z_l^T \beta) Y_l(u)}{\sum_{l=1}^n \exp(z_l^T \beta) Y_l(u)}.$$

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

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

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

$$\begin{aligned} V(u, \beta) &= \frac{\sum_{l=1}^n (z_l - \bar{z}(u, \beta))(z_l - \bar{z}(u, \beta))^T \exp(z_l^T \beta) Y_l(u)}{\sum_{l=1}^n \exp(z_l^T \beta) Y_l(u)} \\ &= \sum_{l=1}^n w_l (z_l - \bar{z}(u, \beta))(z_l - \bar{z}(u, \beta))^T, \end{aligned}$$

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

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

is also a positive definite matrix. So the Hessian matrix $H = -J^{p\times p}(\beta)$ 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 associated with the partial likelihood, the score vector, and the MPLE for multi-parameter problems (*i.e*, a vector of covariates) can also be generalized from the one parameter case.

Namely, 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 $J(\beta_0)$. Write this fact as

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

The MPLE $\hat{\beta}$ will also be asymptoticly normal

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

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.

When we use a model with a vector of parameters, we are often interested in making inferential statements about the entire vector simultaneously or part of the vector. Towards this end, let us partition the parameter vector β into two parts: $\beta = (\theta^T, \phi^T)^T$, where θ is a $g(\leq p)$ dimensional vector.

We should 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

$$U(\theta,\phi) = \begin{pmatrix} U_\theta(\theta,\phi) \\ U_\phi(\theta,\phi) \end{pmatrix},$$

where

$$U_\theta(\theta,\phi) = \frac{\partial \ell(\theta,\phi)}{\partial \theta}, \quad U_\phi(\theta,\phi) = \frac{\partial \ell(\theta,\phi)}{\partial \phi}.$$

The partial likelihood information matrix can also be partitioned into

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

and its inverse into

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

Note: Here we use superscript notation to index the partition of an inverse matrix and subscript notation to index the original matrix.

With this notation, the following distributional statement

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

is equivalent to

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

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

$$\hat{\theta} \overset{a}{\sim} \mathrm{N}(\theta_0, J^{\theta\theta}(\hat{\theta},\hat{\phi})).$$

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

$$\hat{\theta} = \hat{\beta}_j, \quad \theta_0 = \beta_{j0},$$

and

$$J^{\theta\theta}(\hat{\theta},\hat{\phi}) = \left[\text{se}(\hat{\beta}_j)\right]^2.$$

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

Since

$$\hat{\theta} \overset{a}{\sim} \text{N}(\theta_0, J^{\theta\theta}(\hat{\theta},\hat{\phi})),$$

which is equivalent to

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

Therefore

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

i.e., the quadratic form is distributed as a χ^2 with g degrees of freedom.

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

$$\left[J^{\theta\theta}(\hat{\theta},\hat{\phi})\right]^{-1} \neq J_{\theta\theta}(\hat{\theta},\hat{\phi}).$$

Let $\chi^2_{\alpha;g}$ be the $(1-\alpha)$ quantile of a χ^2 with g degrees of freedom, *i.e.*,

$$P[\chi^2_g \geq \chi^2_{\alpha,g}] = \alpha.$$

Then

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

or equivalently,

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

For a given data set, the following inequality

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

describes a g-dimensional ellipsoid centered at $\hat{\theta}$ and whose orientation is dictated by the eigenvalues and eigenvectors of $\left[J^{\theta\theta}(\hat{\theta}, \hat{\phi})\right]^{-1}$. The interior of such an ellipsoid is the $(1-\alpha)$th confidence region for θ.

Note: If θ is one-dimensional, then this confidence region simplifies to an interval. In fact, if $\theta = \beta_j$ (one of the element of β), then the $(1-\alpha)$th confidence interval of θ or β_j would be

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

where

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

Generalization of Wald, Score and Likelihood ratio tests

Wald Test: We are interested in testing the null hypothesis

$$H_0 : \theta = \theta_0.$$

Since under H_0, we have

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

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

$$(\hat{\theta} - \theta_0)^T \left[J^{\theta\theta}(\hat{\theta}, \hat{\phi})\right]^{-1} (\hat{\theta} - \theta_0) \geq \chi^2_{\alpha;g}.$$

This is the Wald test.

Score Test: Before we can describe the score test and likelihood ratio test for the hypothesis $H_0 : \theta = \theta_0$, we must first define the notion of a restricted maximum partial likelihood estimator. Since the interest is focused on the parameter of interest θ, our null hypothesis specifies a specific value of θ that we wish to entertain. Nothing, however, is assumed about the nuisance parameters ϕ. Therefore, even under the null hypothesis, an estimate of ϕ will be necessary in order to derive tests as a function of the data.

An obvious estimator for ϕ, if we assume the null hypothesis to be true, is to maximize the log partial likelihood as a function of ϕ, keeping θ fixed at the hypothesized value of θ_0. 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 $\ell(\theta_0, \phi)$. This restricted MPLE can be obtained by solving the $(p - g)$ equations of $(p - g)$ unknowns

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

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

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

It can be shown (the proof is omitted here) that if $H_0 : \theta = \theta_0$ is true then this score vector with respect to the *parameters of interest* would be multivariate normal with mean zero and covariance matrix that can be estimated by

$$\left[J^{\theta\theta}(\theta_0, \hat{\phi}(\theta_0))\right]^{-1}.$$

That is,

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

If the null hypothesis were not true, we would expect the score vector above (evaluated at

θ_0) to have mean different from zero. This suggests rejecting H_0 whenever the quadratic form

$$\left[U_\theta(\theta_0, \hat{\phi}(\theta_0))\right]^T \left[J^{\theta\theta}(\theta_0, \hat{\phi}(\theta_0))\right] \left[U_\theta(\theta_0, \hat{\phi}(\theta_0))\right]$$

is sufficiently large.

This quadratic form was computed with respect to the inverse of the covariance matrix. Therefore, under H_0, the distribution of the quadratic form is a chi-square with g degrees of freedom.

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

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

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 $\ell(\theta, \phi)$. We denote this estimate as $\hat{\beta}$, or $(\hat{\theta}, \hat{\phi})$. We also defined the restricted MPLE $\hat{\phi}(\theta_0)$ as the value of ϕ that maximizes the following function

$$\ell(\theta_0, \phi).$$

It must be the case that, for any set of data, $\ell(\hat{\theta}, \hat{\phi})$ must be greater than or equal to $\ell(\theta_0, \hat{\phi}(\theta_0))$, since $\ell(\hat{\theta}, \hat{\phi})$ 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 $\ell(\hat{\theta}, \hat{\phi})$ would be close to $\ell(\theta_0, \hat{\phi}(\theta_0))$. It is therefore reasonable to expect that H_0 would not be true if the difference

$$\ell(\hat{\theta}, \hat{\phi}) - \ell(\theta_0, \hat{\phi}(\theta_0))$$

is sufficiently large.

Under H_0, the distribution of

$$2\left[\ell(\hat{\theta}, \hat{\phi}) - \ell(\theta_0, \hat{\phi}(\theta_0))\right] \overset{H_0}{\sim} \chi^2_g.$$

Therefore, the likelihood ratio test rejects H_0 at level α whenever

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

Models with multiple covariates: When studying the relationship of survival to a potential factor, we may wish to adjust for the effect of other variables. For example, if we wish to study the relationship of alcohol drinking to survival, in an *observational study*, we may be concerned that alcohol drinking is correlated with smoking. Thus, if we don't adjust for the effect of smoking, then what may seem as an apparent relationship between survival and drinking may really be an artifact of the effect of smoking on survival which is being confounded with drinking.

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.

The proportional hazards model with multiple covariates is ideal for such purposes. 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), we obtain the relationship that the variable of interest has on survival while adjusting for the effect of the other covariates.

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.

Example: Let S denote the smoking indicator (1 = smoker, 0 = nonsmoker), and D denote drinking indicator (1 = drinker, 0 = nondrinker). 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

$$\lambda(t|D) = \lambda_0(t)\exp(D\beta).$$

As we know, the parameter β is interpreted as the log hazard ratio between drinkers and nondrinkers (assumed constant over time t) and $\exp(\beta)$ 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

$$\lambda(t|\cdot) = \lambda_0(t)\exp(D\theta + S\phi_1 + A\phi_2 + Sx\phi_3),$$

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

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

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

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

$$(x_i, \delta_i, d_i, s_i, a_i, sx_i), \quad i = 1, 2, \cdots, n.$$

The proportional hazards model

$$\lambda(t|\cdot) = \lambda_0(t)\exp(D\theta + S\phi_1 + A\phi_2 + Sx\phi_3),$$

would be fit using `Proc Phreg` in `SAS`, using partial likelihood methods.

The output would yield the MPLE $(\hat{\theta}, \hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3)$ as well as their estimated standard errors.

From this we would construct a $(1-\alpha)$ confidence interval for θ

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

We could also test the null hypothesis $H_0 : \theta = 0$ using a Wald test, score test, or partial likelihood ratio test, for θ with ϕ corresponding to the nuisance parameters.

A Real Example: We will discuss a dataset on breast cancer (CALGB 8082). The data set has the following variables:

Menopausal status (0 = pre menopausal, 1 = post menopausal)

Tumor size (largest dimension of tumor in cm)

number of positive nodes

Estrogen receptor status (0 = negative, 1 = positive)

The primary purpose of this study is to evaluate certain treatment on breast cancer, adjusting for the above prognostic factors.

Note: After adjusting for the other covariates, the estimate of treatment effect yielded a parameter estimate of 0.021 with a estimated standard error 0.101.

Let Rx denote treatment, MS denote menopausal status, TS denote tumor size, NN number of positive nodes and ER estrogen receptor status.

If our interest is the effect of treatment on survival adjusting for the other covariates, we

write our model as

$$\lambda(t|\cdot) = \lambda_0(t)\exp(Rx\theta + MS\phi_1 + TS\phi_2 + NN\phi_3 + ER\phi_4)$$
$$\hat{\theta} = 0.021, \text{se}(\hat{\theta}) = 0.101,$$

and a 95% confidence interval for θ is

$$\hat{\theta} \pm 1.96 * \text{se}(\hat{\theta}) = 0.021 \pm 1.96 * 0.101 = [-0.177, 0.219].$$

The estimate of the adjusted treatment hazard ratio is

$$\exp(\hat{\theta}) = \exp(0.021) = 1.021,$$

with a 95% CI of

$$[\exp(-0.177), \exp(0.219)] = [0.838, 1.245].$$

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:
$$\left[\frac{\hat{\theta}}{\text{se}(\hat{\theta})}\right]^2 = \left[\frac{0.021}{0.101}\right]^2 = 0.042,$$
with p-value = 0.838.

2. Likelihood ratio test:
$$\begin{aligned} & 2[\ell(\hat{\theta}, \hat{\phi}) - \ell(\theta = 0, \hat{\phi}(\theta = 0))] \\ = \; & -4739.685 - (-4739.727) = 0.042, \end{aligned}$$
with p-value=0.838.

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 $\chi^2 = 0.042$, yielding the same p-value as other two tests.

Now that we feel fairly confident that there is not treatment effect. Suppose we decide to use these data to study the relationship of tumor size to survival. With respect to this question, these data can be viewed as an observational dataset. Let us consider the model

$$\lambda(t|\cdot) = \lambda_0(t)\exp(TS\theta).$$

The result of this model gives an estimate $\hat{\theta} = 0.042$, $\text{se}(\hat{\theta}) = 0.019$. The Wald test for $H_0 : \theta = 0$ is

$$\left[\frac{0.042}{0.019}\right]^2 = 4.75, \quad \text{p-value} = 0.029.$$

The likelihood ratio test and score test yield similar conclusions; namely, there may be some prognostic effect of tumor size on survival.

Remark: A typical larger tumor size is about $7cm$ ($\approx$ 2 standard deviation above the mean for this sample of patients. A typical smaller tumor size is about $1cm$ (the smallest tumor size is $0.1cm$). Hence the relative risk (or hazard ratio) for a woman with tumor size $7cm$ as compared to a woman with tumor size $1cm$ is

$$\frac{\lambda_0(t)\exp(7\theta)}{\lambda_0(t)\exp(\theta)} = \exp(6\theta),$$

which is estimated to be

$$\exp(6\hat{\theta}) = \exp(6 * 0.042) = 1.28.$$

A 95% CI for θ is

$$\hat{\theta} \pm 1.96 * \text{se}(\hat{\theta}) = 0.042 \pm 1.96 * 0.019 = [0.0048, 0.079].$$

Consequently, a 95% CI for relative risk $\exp(6\theta)$ is

$$\exp(6 * 0.0048), \exp(6 * 0.079) = [1.029, 1.606].$$

It may be however that the effect of tumor size may be confounded with other covariates. To study this, we consider the model

$$\lambda(t|\cdot) = \lambda_0(t)\exp(TS\theta + MS\phi_1 + NN\phi_2 + ER\phi_3).$$

From this model, we get $\hat{\theta} = 0.02$, and $\text{se}(\hat{\theta}) = 0.019$.

The corresponding estimate for the relative risk $\exp(6\theta)$ is now

$$RR = \exp(6 * \hat{\theta}) = 1.128,$$

and its 95% CI (adjusted for the other covariates) is

$$[0.902, 1.41].$$

Summary result for $\exp(6\theta)$

	Unadjusted (All available data)	Adjusted (All available data)
# of patients	$n = 817$	$n = 723$
RR	1.28	1.13
95% CI	[1.029, 1.606]	[0.902, 1.41]
Wald test	4.75 (p-val = 0.03)	1.14 (p-val = 0.29)
LR test	4.02	1.03
Score test	4.65	1.14

Remark: Unfortunately, in many clinical trials, not all the data are collected on all the individuals. 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 recored 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. For example, fewer recored may be missing when we consider one covariate as opposed to a model with that covariate and additional covariates.

This is especially the case when we consider the likelihood ratio test for nested models. We must make sure that the nested models being compared are on the same set of individuals. 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 may have 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. For example, in our analysis for CALGB 8082, we note that nobody had missing treatment information. Therefore, the effect of treatment could be analyzed using all 905 women randomized to this study. However, only 723 women had all the covariate information we ultimately considered. We therefore also looked at the effect of treatment (unadjusted) within this subset of 723 patients to see if the results were comparable to the full data.

	All patients	Patients with complete covariates
	$n = 905$	$n = 723$
RR	1.061	1.075
95% CI	[0.890, 1.265]	[0.882, 1.331]

Similarly, when we consider the effect of tumor size on survival (unadjusted), we used 817 women for which tumor size was collected. However, for the adjusted analysis we could only use 723 women with complete data on all covariates.

Previously, we contrasted the relationship of tumor size to survival; unadjusted versus adjusted. However, this was done on different data sets, one with 817 women having tumor size information and the other with 723 women with all covariates. In order to make sure that the

differences seen between these two analyses is not due to the different datasets being considered, we also look at the unadjusted effect of tumor size on survival using the data set with 723 women.

The estimate of relative risk (hazard ratio between tumor size of $7cm$ vs. $1cm$) and 95% CI are

$$n = 723, \quad RR = 1.307, \quad 95\%CI = [1.036, 1.649].$$

These results are similar to the unadjusted results obtained on the 817 patients.

In order to compare the likelihood ratio test for $H_0 : \theta = 0$ (no effect of tumor size on survival) adjusted for the other covariates, we need to compute

$$2[\ell(\hat{\theta}, \hat{\phi}) - \ell(\theta = 0, \hat{\phi}(\theta = 0))]$$

or

$$[-2\ell(\theta = 0, \hat{\phi}(\theta = 0))] - [-2\ell(\hat{\theta}, \hat{\phi})].$$

In order to compute $\ell(\theta = 0, \hat{\phi}(\theta = 0))$, we must consider the model when $\theta = 0$; *i.e.*,

$$\lambda(t|\cdot) = \lambda_0(t)\exp(0 + MS\phi_1 + NN\phi_2 + ER\phi_3)$$

and find the maximized log likelihood for this sub-model. We must make sure however that this sub-model is run on the same set of data as the full model; *i.e.*, on 723 women.

This is how we get the value for the likelihood ratio test:

$$4740.759 - 4739.727 = 1.032.$$

Remark on confounding: Previously, we noted that the unadjusted effect of tumor size on survival was significant (p-value = 0.03, Wald test), whereas the adjusted effect was not significant (p-value = 0.29, Wald test). This suggests that at least one of the variables we adjusted for confounds the relationship of tumor size to survival.

A serious study of this issue, assuming we felt it was important to study, would take some work. However, at first glance, we note that the "number of nodes" was a highly significant prognostic factor (Wald chi-square > 65, adjusted or unadjusted) and that there was substantial and significant correlation between "number of nodes" and tumor size. I suspect that this is the primary confounding relationship that weakened the effect of "tumor size" as an independent prognostic factor of survival.

Appendix: SAS Program and output

The following is the program and output related to the breast cancer data set from CALGB 8082:

```
options ps=62 ls=72;

data bcancer;
  infile "cal8082.dat";
  input days cens trt meno tsize nodes er;
  trt1 = trt - 1;
  label days="(censored) survival time in days"
    cens="censoring indicator"
    trt="treatment"
    meno="menopausal status"
    tsize="size of largest tumor in cm"
    nodes="number of positive nodes"
    er="estrogen receptor status"
    trt1="treatment indicator";
run;

data bcancer1; set bcancer;
  if meno = . or tsize = . or nodes = . or er = . then delete;
run;

title "Univariate analysis of treatment effect";
proc phreg data=bcancer;
  model days*cens(0) = trt1;
run;
```

The output of the above univariate program is

```
                Univariate analysis of treatment effect                1
                                        09:37 Tuesday, April 2, 2002

                          The PHREG Procedure

                           Model Information

Data Set                  WORK.BCANCER
Dependent Variable        days              (censored) survival time in days
Censoring Variable        cens              censoring indicator
Censoring Value(s)        0
Ties Handling             BRESLOW
```

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
905	497	408	45.08

Convergence Status

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

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	6362.858	6362.421
AIC	6362.858	6364.421
SBC	6362.858	6368.629

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	0.4375	1	0.5083
Score	0.4375	1	0.5083
Wald	0.4374	1	0.5084

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
trt1	1	0.05935	0.08973	0.4374	0.5084

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	Variable Label
trt1	1.061	treatment indicator

Program 2: adjusting for `meno tsize nodes er`:

```
title "Analysis of treatment effect adjusting for meno tsize nodes er";
proc phreg data=bcancer;
  model days*cens(0) = trt1 meno tsize nodes er;
run;
```

The output of program 2:

Analysis of treatment effect adjusting for meno tsize nodes er 2
09:37 Tuesday, April 2, 2002

The PHREG Procedure

Model Information

Data Set	WORK.BCANCER	
Dependent Variable	days	(censored) survival time in days
Censoring Variable	cens	censoring indicator
Censoring Value(s)	0	
Ties Handling	BRESLOW	

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
723	391	332	45.92

Convergence Status

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

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	4833.945	4739.685
AIC	4833.945	4749.685
SBC	4833.945	4769.528

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	94.2607	5	<.0001
Score	113.4441	5	<.0001
Wald	111.1227	5	<.0001

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
trt1	1	0.02080	0.10147	0.0420	0.8376
meno	1	0.39108	0.10797	13.1198	0.0003
tsize	1	0.01992	0.01875	1.1289	0.2880
nodes	1	0.05252	0.00652	64.8325	<.0001
er	1	-0.52723	0.10485	25.2862	<.0001

Analysis of treatment effect adjusting for meno tsize nodes er 3
09:37 Tuesday, April 2, 2002

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	Variable Label
trt1	1.021	treatment indicator
meno	1.479	menopausal status
tsize	1.020	size of largest tumor in cm
nodes	1.054	number of positive nodes
er	0.590	estrogen receptor status

Program 3: a model without treatment indicator:

```
title "Model without treatment";
proc phreg data=bcancer;
  model days*cens(0) = meno tsize nodes er;
run;
```

Output of program 3:

```
                        Model without treatment                         4
                                        09:37 Tuesday, April 2, 2002

                          The PHREG Procedure

                           Model Information

Data Set                  WORK.BCANCER
Dependent Variable        days              (censored) survival time in days
Censoring Variable        cens              censoring indicator
Censoring Value(s)        0
Ties Handling             BRESLOW


           Summary of the Number of Event and Censored Values

                                                  Percent
         Total       Event    Censored          Censored

           723         391         332             45.92


                        Convergence Status

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


                        Model Fit Statistics

                                   Without            With
                Criterion       Covariates       Covariates

                -2 LOG L          4833.945         4739.727
                AIC               4833.945         4747.727
                SBC               4833.945         4763.601


                Testing Global Null Hypothesis: BETA=0

        Test                 Chi-Square       DF     Pr > ChiSq
```

```
       Likelihood Ratio        94.2187        4         <.0001
       Score                  113.4346        4         <.0001
       Wald                   111.2321        4         <.0001


               Analysis of Maximum Likelihood Estimates

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

 meno         1       0.39180       0.10791       13.1828        0.0003
 tsize        1       0.02006       0.01876        1.1426        0.2851
 nodes        1       0.05257       0.00651       65.1841        <.0001
 er           1      -0.52691       0.10483       25.2652        <.0001

                        Model without treatment                        5
                                              09:37 Tuesday, April 2, 2002

                          The PHREG Procedure

               Analysis of Maximum Likelihood Estimates

                    Hazard
         Variable    Ratio    Variable Label

         meno         1.480    menopausal status
         tsize        1.020    size of largest tumor in cm
         nodes        1.054    number of positive nodes
         er           0.590    estrogen receptor status
```

Program 4: Univariate analysis of treatment effect using the subsample:

```
title "Univariate analysis of treatment effect using subsample";
proc phreg data=bcancer1;
  model days*cens(0) = trt1;
run;
```

Output of program 4:

```
      Univariate analysis of treatment effect using subsample        6
                                              09:37 Tuesday, April 2, 2002

                          The PHREG Procedure

                           Model Information

Data Set                  WORK.BCANCER1
Dependent Variable        days            (censored) survival time in days
Censoring Variable        cens            censoring indicator
Censoring Value(s)        0
Ties Handling             BRESLOW


           Summary of the Number of Event and Censored Values

                                                 Percent
              Total       Event    Censored    Censored
```

```
                723         391         332       45.92

                        Convergence Status

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

                        Model Fit Statistics

                                   Without           With
                   Criterion     Covariates     Covariates

                   -2 LOG L        4833.945       4833.430
                   AIC             4833.945       4835.430
                   SBC             4833.945       4839.398


                Testing Global Null Hypothesis: BETA=0

          Test                 Chi-Square       DF     Pr > ChiSq

          Likelihood Ratio         0.5156        1         0.4727
          Score                    0.5155        1         0.4728
          Wald                     0.5149        1         0.4730


                Analysis of Maximum Likelihood Estimates

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

 trt1         1        0.07263       0.10121        0.5149        0.4730

                Analysis of Maximum Likelihood Estimates

                                  Hazard
                 Variable          Ratio    Variable Label

                 trt1              1.075    treatment indicator
```

Program 5: Univariate analysis of tumor size effect using the whole sample.

```
title "Univariate analysis of tumor size effect using whole sample";
proc phreg data=bcancer;
  model days*cens(0) = tsize;
run;
```

Output of program 5:

```
      Univariate analysis of tumor size effect using whole sample      7
                                          09:37 Tuesday, April 2, 2002

                          The PHREG Procedure

                           Model Information

Data Set                  WORK.BCANCER
Dependent Variable        days            (censored) survival time in days
```

```
Censoring Variable      cens            censoring indicator
Censoring Value(s)      0
Ties Handling           BRESLOW

           Summary of the Number of Event and Censored Values

                                                  Percent
        Total       Event    Censored    Censored

          817         451         366       44.80

                        Convergence Status

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

                        Model Fit Statistics

                                   Without           With
                   Criterion     Covariates     Covariates

                   -2 LOG L        5681.392       5677.370
                   AIC             5681.392       5679.370
                   SBC             5681.392       5683.481

                Testing Global Null Hypothesis: BETA=0

        Test                 Chi-Square       DF     Pr > ChiSq

        Likelihood Ratio         4.0225        1         0.0449
        Score                    4.6533        1         0.0310
        Wald                     4.7476        1         0.0293

                Analysis of Maximum Likelihood Estimates

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

 tsize        1        0.04153       0.01906        4.7476        0.0293

              Analysis of Maximum Likelihood Estimates

                        Hazard
          Variable       Ratio    Variable Label

          tsize          1.042    size of largest tumor in cm
```

Program 6: Univariate analysis of tumor size effect using the subsample:

```
title "Univariate analysis of tumor size effect using subsample";
proc phreg data=bcancer1;
  model days*cens(0) = tsize;
run;
```

Output of program 6:

Univariate analysis of tumor size effect using subsample 8
09:37 Tuesday, April 2, 2002

The PHREG Procedure

Model Information

Data Set	WORK.BCANCER1	
Dependent Variable	days	(censored) survival time in days
Censoring Variable	cens	censoring indicator
Censoring Value(s)	0	
Ties Handling	BRESLOW	

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
723	391	332	45.92

Convergence Status

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

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	4833.945	4829.744
AIC	4833.945	4831.744
SBC	4833.945	4835.712

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.2016	1	0.0404
Score	5.0066	1	0.0253
Wald	5.1128	1	0.0238

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
tsize	1	0.04465	0.01975	5.1128	0.0238

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	Variable Label
tsize	1.046	size of largest tumor in cm

Program 7: Reduced model with `meno nodes er`:

```
title "Reduced model with meno nodes er";
proc phreg data=bcancer1;
  model days*cens(0) = meno nodes er;
run;
```

Output of program 7:

```
                  Reduced model with meno nodes er                    9
                                        09:37 Tuesday, April 2, 2002

                        The PHREG Procedure

                         Model Information

Data Set                  WORK.BCANCER1
Dependent Variable        days            (censored) survival time in days
Censoring Variable        cens            censoring indicator
Censoring Value(s)        0
Ties Handling             BRESLOW

          Summary of the Number of Event and Censored Values

                                                  Percent
       Total       Event    Censored    Censored

         723         391         332       45.92

                        Convergence Status

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

                     Model Fit Statistics

                                Without           With
          Criterion     Covariates     Covariates

          -2 LOG L        4833.945       4740.759
          AIC             4833.945       4746.759
          SBC             4833.945       4758.666

          Testing Global Null Hypothesis: BETA=0

     Test                 Chi-Square       DF     Pr > ChiSq

     Likelihood Ratio        93.1858        3         <.0001
     Score                  112.3495        3         <.0001
     Wald                   110.3494        3         <.0001

          Analysis of Maximum Likelihood Estimates

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

meno         1        0.38742       0.10786       12.9016        0.0003
nodes        1        0.05379       0.00636       71.5972        <.0001
er           1       -0.51916       0.10452       24.6744        <.0001
```

```
            Analysis of Maximum Likelihood Estimates

                        Hazard
          Variable       Ratio    Variable Label

          meno           1.473    menopausal status
          nodes          1.055    number of positive nodes

                 Reduced model with meno nodes er                  10
                                          09:37 Tuesday, April 2, 2002

                        The PHREG Procedure

            Analysis of Maximum Likelihood Estimates

                        Hazard
          Variable       Ratio    Variable Label

          er             0.595    estrogen receptor status
```

Program 8: look at the correlation among covariates in the whole sample and the subsample:

```
title "Correlation of covariates using whole sample";
proc corr data=bcancer;
  var meno tsize nodes er;
run;

title "Correlation of covariates using subsample";
proc corr data=bcancer1;
  var meno tsize nodes er;
run;
```

Output of program 8:

```
              Correlation of covariates using whole sample           11
                                          09:37 Tuesday, April 2, 2002

                           The CORR Procedure

          4  Variables:    meno     tsize    nodes    er


                           Simple Statistics

     Variable           N          Mean       Std Dev           Sum

     meno             891       0.58810       0.49245     524.00000
     tsize            817       3.21603       1.98253          2627
     nodes            896       6.53125       6.65252          5852
     er               791       0.64855       0.47773     513.00000

                           Simple Statistics

  Variable       Minimum       Maximum    Label

  meno                 0       1.00000    menopausal status
  tsize          0.10000      30.00000    size of largest tumor in cm
  nodes                0      57.00000    number of positive nodes
```

```
  er                    0       1.00000    estrogen receptor status


                  Pearson Correlation Coefficients
                     Prob > |r| under H0: Rho=0
                       Number of Observations

                                    meno      tsize      nodes         er

meno                             1.00000   -0.05815    0.05115    0.10469
menopausal status                            0.0973     0.1275     0.0033
                                     891        814        889        786

tsize                           -0.05815    1.00000    0.16787   -0.02528
size of largest tumor in cm      0.0973                 <.0001     0.4967
                                     814        817        817        725

nodes                            0.05115    0.16787    1.00000   -0.09113
number of positive nodes         0.1275     <.0001                0.0106
                                     889        817        896        786

er                               0.10469   -0.02528   -0.09113    1.00000
estrogen receptor status         0.0033     0.4967     0.0106
                                     786        725        786        791

               Correlation of covariates using subsample              12
                                          09:37 Tuesday, April 2, 2002

                           The CORR Procedure

          4  Variables:    meno     tsize    nodes    er


                           Simple Statistics

     Variable           N          Mean       Std Dev           Sum

     meno             723       0.59474       0.49128     430.00000
     tsize            723       3.21646       1.97440          2325
     nodes            723       6.38036       6.48484          4613
     er               723       0.65560       0.47550     474.00000

                           Simple Statistics

  Variable       Minimum       Maximum    Label

  meno                 0       1.00000    menopausal status
  tsize          0.10000      30.00000    size of largest tumor in cm
  nodes          1.00000      43.00000    number of positive nodes
  er                   0       1.00000    estrogen receptor status


              Pearson Correlation Coefficients, N = 723
                     Prob > |r| under H0: Rho=0

                                    meno      tsize      nodes         er

meno                             1.00000   -0.07193    0.02758    0.10133
menopausal status                            0.0532     0.4590     0.0064

tsize                           -0.07193    1.00000    0.18031   -0.02508
size of largest tumor in cm      0.0532                 <.0001     0.5007
```

```
nodes                          0.02758     0.18031     1.00000    -0.08592
number of positive nodes        0.4590      <.0001                  0.0209

er                             0.10133    -0.02508    -0.08592     1.00000
estrogen receptor status        0.0064      0.5007      0.0209
```

Program 9: score test for treatment effect adjusting for other covariates:

```
title "Score test for treatment effect adjusting for other covariates";
proc phreg data=bcancer1;
  model days*cens(0) = tsize meno nodes er trt1
    / selection=forward include=4 details slentry=1.0;
run;
```

Output of program 9:

```
     Score test for treatment effect adjusting for other covariates   13
                                          09:37 Tuesday, April 2, 2002

                         The PHREG Procedure

                          Model Information

Data Set                  WORK.BCANCER1
Dependent Variable        days            (censored) survival time in days
Censoring Variable        cens            censoring indicator
Censoring Value(s)        0
Ties Handling             BRESLOW


           Summary of the Number of Event and Censored Values

                                                    Percent
                Total       Event    Censored      Censored

                  723         391         332         45.92

The following variable(s) will be included in each model:

         tsize  meno  nodes  er


                         Convergence Status

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


                      Model Fit Statistics

                                 Without           With
                Criterion     Covariates     Covariates

                -2 LOG L        4833.945       4739.727
                AIC             4833.945       4747.727
                SBC             4833.945       4763.601


                Testing Global Null Hypothesis: BETA=0
```

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	94.2187	4	<.0001
Score	113.4346	4	<.0001
Wald	111.2321	4	<.0001

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
tsize	1	0.02006	0.01876	1.1426	0.2851
meno	1	0.39180	0.10791	13.1828	0.0003
nodes	1	0.05257	0.00651	65.1841	<.0001
er	1	-0.52691	0.10483	25.2652	<.0001

Score test for treatment effect adjusting for other covariates 14
09:37 Tuesday, April 2, 2002

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	Variable Label
tsize	1.020	size of largest tumor in cm
meno	1.480	menopausal status
nodes	1.054	number of positive nodes
er	0.590	estrogen receptor status

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > ChiSq	Label
trt1	0.0420	0.8376	treatment indicator

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
0.0420	1	0.8376

Step 1. Variable trt1 is entered. The model contains the following explanatory variables:

tsize meno nodes er trt1

Convergence Status

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

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	4833.945	4739.685
AIC	4833.945	4749.685
SBC	4833.945	4769.528

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	94.2607	5	<.0001
Score	113.4441	5	<.0001
Wald	111.1227	5	<.0001

Score test for treatment effect adjusting for other covariates 15
09:37 Tuesday, April 2, 2002

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
tsize	1	0.01992	0.01875	1.1289	0.2880
meno	1	0.39108	0.10797	13.1198	0.0003
nodes	1	0.05252	0.00652	64.8325	<.0001
er	1	-0.52723	0.10485	25.2862	<.0001
trt1	1	0.02080	0.10147	0.0420	0.8376

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	Variable Label
tsize	1.020	size of largest tumor in cm
meno	1.479	menopausal status
nodes	1.054	number of positive nodes
er	0.590	estrogen receptor status
trt1	1.021	treatment indicator

NOTE: All variables have been entered into the model.

Summary of Forward Selection

Step	Variable Entered	Number In	Score Chi-Square	Pr > ChiSq	Variable Label
1	trt1	5	0.0420	0.8376	treatment indicator

Program 10: Score test of tumor size effect adjusting for other covariates:

```
title "Score test of tumor size effect adjusting for other covariates";
proc phreg data=bcancer1;
  model days*cens(0) = meno nodes er tsize
    / selection=forward include=3 details slentry=1.0;
```

```
run;
```

Ouput of program 10:

```
     Score test of tumor size effect adjusting for other covariates   16
                                          09:37 Tuesday, April 2, 2002

                         The PHREG Procedure

                          Model Information

Data Set                  WORK.BCANCER1
Dependent Variable        days            (censored) survival time in days
Censoring Variable        cens            censoring indicator
Censoring Value(s)        0
Ties Handling             BRESLOW


           Summary of the Number of Event and Censored Values

                                                    Percent
                 Total       Event    Censored    Censored

                   723         391         332       45.92

The following variable(s) will be included in each model:

         meno  nodes  er


                          Convergence Status

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


                       Model Fit Statistics

                                  Without           With
                  Criterion     Covariates     Covariates

                  -2 LOG L        4833.945       4740.759
                  AIC             4833.945       4746.759
                  SBC             4833.945       4758.666


                Testing Global Null Hypothesis: BETA=0

        Test                 Chi-Square       DF     Pr > ChiSq

        Likelihood Ratio        93.1858        3         <.0001
        Score                  112.3495        3         <.0001
        Wald                   110.3494        3         <.0001


                Analysis of Maximum Likelihood Estimates

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

 meno         1        0.38742       0.10786       12.9016        0.0003
 nodes        1        0.05379       0.00636       71.5972        <.0001
```

er	1	-0.51916	0.10452	24.6744	<.0001

Score test of tumor size effect adjusting for other covariates 17
09:37 Tuesday, April 2, 2002

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	Variable Label
meno	1.473	menopausal status
nodes	1.055	number of positive nodes
er	0.595	estrogen receptor status

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > ChiSq	Label
tsize	1.1448	0.2846	size of largest tumor in cm

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
1.1448	1	0.2846

Step 1. Variable tsize is entered. The model contains the following explanatory variables:

meno nodes er tsize

Convergence Status

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

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	4833.945	4739.727
AIC	4833.945	4747.727
SBC	4833.945	4763.601

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	94.2187	4	<.0001
Score	113.4346	4	<.0001
Wald	111.2321	4	<.0001

Score test of tumor size effect adjusting for other covariates 18

09:37 Tuesday, April 2, 2002

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
meno	1	0.39180	0.10791	13.1828	0.0003
nodes	1	0.05257	0.00651	65.1841	<.0001
er	1	-0.52691	0.10483	25.2652	<.0001
tsize	1	0.02006	0.01876	1.1426	0.2851

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	Variable Label
meno	1.480	menopausal status
nodes	1.054	number of positive nodes
er	0.590	estrogen receptor status
tsize	1.020	size of largest tumor in cm

NOTE: All variables have been entered into the model.

Summary of Forward Selection

Step	Variable Entered	Number In	Score Chi-Square	Pr > ChiSq	Variable Label
1	tsize	4	1.1448	0.2846	size of largest tumor in cm