Chapter 5

Modeling of Survival Data

Now we will explore the relationship between survival and explanatory variables by modeling. In this class, we consider two broad classes of regression models:

Proportional Hazards (PH) models

$$\lambda(t; \mathbf{Z}) = \lambda_{\mathbf{0}}(\mathbf{t})\mathbf{\Psi}(\mathbf{Z})$$

Most commonly, we write the second term as: $\Psi(\mathbf{Z}) = \mathbf{e}^{\beta \mathbf{Z}}$

Suppose $Z = 1$ for treated subjects and $Z = 0$ for untreated subjects. Then this model says that the hazard is increased by a factor of e^{β} for treated subjects versus untreated subjects (c^{β} might be < 1).

This is an example of a **semi-parametric** model.

Accelerated Failure Time (AFT) models

$$\log(T) = \mu + \beta \mathbf{Z} + \sigma \mathbf{w}$$

where w is an "error distribution". Typically, we place a **parametric** assumption on w:

- exponential, Weibull, Gamma
- lognormal

5.1 Covariates

In general, **Z** is a *vector* of covariates of interest.

Z may include:

- continuous factors (eg, age, blood pressure)
- discrete factors (gender, marital status)
- possible interactions (age by sex interaction)

5.1.1 Covariates

Just as in standard linear regression, if we have a discrete covariate A with a levels, then we will need to include $(a-1)$ dummy variables $(U_1, U_2, \ldots, U_a)$ such that $U_j = 1$ if $A = j$. Then

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_2 U_2 + \beta_3 U_3 + \cdots + \beta_a U_a)$$

(In the above model, the subgroup with $A = 1$ or $U_1 = 1$ is the reference group.)

5.1.2 Interactions

Two factors, A and B, interact if the hazard of death depends on the combination of levels of A and B.

We follow the principle of hierarchical models, and only include interactions if all of the associated main effects are also included.

The example I just gave was based on a proportional hazards model, but the description of the types of covariates we might want to include in our model applies to both the AFT and PH model.

We'll start out by focusing on the Cox PH model, and address some of the following questions:

- What does the term $\lambda_0(t)$ mean?
- What's "proportional" about the PH model?
- How do we estimate the parameters in the model?
- How do we interpret the estimated values?

- How can we construct tests of whether the covariates have a significant effect on the distribution of survival times?
- How do these tests compare to the logrank test or the Wilcoxon test?

5.2 The Cox Proportional Hazards model

$$\lambda(t; \mathbf{Z}) = \lambda_{\mathbf{0}}(\mathbf{t}) \exp(\beta \mathbf{Z})$$

This is the most common model used for survival data.
Why?

- flexible choice of covariates
- fairly easy to fit
- standard software exists

References: Collett, Chapter 3
Allison, Chapter 5
Cox and Oakes, Chapter 7
Kleinbaum, Chapter 3
Klein and Moeschberger, Chapters 8 & 9
Kalbfleisch and Prentice
Lee

Note Some books (like Collett) use $h(t; \mathbf{X})$ as their standard notation instead of $\lambda(t; \mathbf{Z})$. **Why do we call it proportional hazards?**

Think of the first example, where $Z = 1$ for treated and $Z = 0$ for control. Then if we think of $\lambda_1(t)$ as the hazard rate for the treated group, and $\lambda_0(t)$ as the hazard for control, then we can write:

$$\begin{aligned} \lambda_1(t) &= \lambda(t; Z = 1) = \lambda_0(t) \exp(\beta Z) \\ &= \lambda_0(t) \exp(\beta) \end{aligned}$$

This implies that the ratio of the two hazards is a constant, ϕ, which does NOT depend on time, t. In other words, the hazards of the two groups remain proportional over time.

$$\phi = \frac{\lambda_1(t)}{\lambda_0(t)} = e^{\beta}$$

ϕ is referred to as the **hazard ratio**. **What is the interpretation of β here?**

5.2.1 The Baseline Hazard Function

In the example of comparing two treatment groups, $\lambda_0(t)$ is the hazard rate for the control group.

In general, $\lambda_0(t)$ is called the **baseline hazard function**, and reflects the underlying hazard for subjects with all covariates $Z_1, ..., Z_p$ equal to 0 (i.e., the "reference group").

The general form is:

$$\lambda(t; \mathbf{Z}) = \lambda_0(t) \exp(\beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_p Z_p)$$

So when we substitute all of the Z_j's equal to 0, we get:

$$\begin{aligned} \lambda(t, \mathbf{Z} = \mathbf{0}) &= \lambda_0(t) \exp(\beta_1 * 0 + \beta_2 * 0 + \cdots + \beta_p * 0) \\ &= \lambda_0(t) \end{aligned}$$

In the general case, we think of the i-th individual having a set of covariates $\mathbf{Z_i} = (\mathbf{Z_{1i}}, \mathbf{Z_{2i}}, ..., \mathbf{Z_{pi}})$, and we model their hazard rate as some multiple of the baseline hazard rate:

$$\lambda_i(t, \mathbf{Z_i}) = \lambda_0(t) \exp(\beta_1 Z_{1i} + \cdots + \beta_p Z_{pi})$$

This means we can write the log of the hazard ratio for the i-th individual to the reference group as:

$$\log\left(\frac{\lambda_i(t)}{\lambda_0(t)}\right) = \beta_1 Z_{1i} + \beta_2 Z_{2i} + \cdots + \beta_p Z_{pi}$$

The Cox Proportional Hazards model is a linear model for the log of the hazard ratio

One of the biggest advantages of the framework of the Cox PH model is that we can estimate the parameters β which reflect the effects of treatment and other covariates without having to make any assumptions about the form of $\lambda_0(t)$.

In other words, we don't have to assume that $\lambda_0(t)$ follows an exponential model, or a Weibull model, or any other particular parametric model.

That's what makes the model *semi-parametric*.

Questions:

1. Why don't we just model the hazard ratio,
$\phi = \lambda_i(t)/\lambda_0(t)$**, directly as a linear function of the covariates Z?**

2. Why doesn't the model have an intercept?

5.2.2 Estimation of the model parameters

The basic idea is that under PH, information about β can be obtained from the relative orderings (i.e., ranks) of the survival times, rather than the actual values. Why?

Suppose T follows a PH model:

$$\lambda(t; \mathbf{Z}) = \lambda_{\mathbf{0}}(\mathbf{t})\mathbf{e}^{\beta \mathbf{Z}}$$

Now consider $T^* = g(T)$, where g is a monotonic increasing function. We can show that T^* also follows the PH model, with the same multiplier, $e^{\beta \mathbf{Z}}$. Therefore, when we consider likelihood methods for estimating the model parameters, we only have to worry about the ranks of the survival times.

5.2.3 Likelihood Estimation for the PH Model

Kalbfleisch and Prentice derive a likelihood involving only β and $\mathbf{Z}$ (not $\lambda_0(t)$) based on the marginal distribution of the ranks of the observed failure times (in the absence of censoring).

Cox (1972) derived the same likelihood, and generalized it for censoring, using the idea of a **partial likelihood**

Suppose we observe $(X_i, \delta_i, \mathbf{Z}_i)$ for individual i, where

- X_i is a censored failure time random variable
- δ_i is the failure/censoring indicator (1=fail,0=censor)
- $\mathbf{Z}_i$ represents a set of covariates

The covariates may be continuous, discrete, or time-varying.

Suppose there are K distinct failure (or death) times, and let $\tau_1, ...\tau_K$ represent the K ordered, distinct death times.

For now, assume there are no tied death times.

Let $\mathcal{R}(t) = \{i : x_i \geq t\}$ denote the set of individuals who are "at risk" for failure at time t.

More about risk sets:

- I will refer to $\mathcal{R}(\tau_j)$ as the risk set at the jth failure time
- I will refer to $\mathcal{R}(X_i)$ as the risk set at the failure time of individual i
- There will still be r_j individuals in $\mathcal{R}(\tau_j)$.
- r_j is a number, while $\mathcal{R}(\tau_j)$ identifies the actual subjects at risk

What is the partial likelihood?

Intuitively, it is a product over the set of observed death times of the conditional probabilities of seeing the observed deaths, given the set of individuals at risk at those times.

At each death time τ_j, the contribution to the likelihood is:

$$\begin{aligned} L_j(\beta) &= Pr(\text{individual j fails}|\text{1 failure from } \mathcal{R}(\tau_j)) \\ &= \frac{Pr(\text{individual } j \text{ fails}|\text{ at risk at } \tau_j)}{\sum_{\ell \in \mathcal{R}(\tau_j)} Pr(\text{individual } \ell \text{ fails}|\text{ at risk at } \tau_j)} \\ &= \frac{\lambda(\tau_j; \mathbf{Z}_j)}{\sum_{\ell \in \mathcal{R}(\tau_j)} \lambda(\tau_j; \mathbf{Z}_\ell)} \end{aligned}$$

Under the PH assumption, $\lambda(t; \mathbf{Z}) = \lambda_0(t)e^{\beta \mathbf{Z}}$, so we get:

$$\begin{aligned} L^{partial}(\beta) &= \prod_{j=1}^{K} \frac{\lambda_0(\tau_j)e^{\beta \mathbf{Z_j}}}{\sum_{\ell \in \mathcal{R}(\tau_j)} \lambda_0(\tau_j)e^{\beta \mathbf{Z}_\ell}} \\ &= \prod_{j=1}^{K} \frac{e^{\beta \mathbf{Z_j}}}{\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta \mathbf{Z}_\ell}} \end{aligned}$$

Another derivation:
In general, the likelihood contributions for censored data fall into two categories:

- **Individual is censored at X_i:**

$$L_i(\beta) = \mathbf{S(X_i)} = \exp[-\int_{\mathbf{0}}^{\mathbf{X_i}} \lambda_{\mathbf{i}}(\mathbf{u})\mathbf{du}]$$

- **Individual fails at X_i:**

$$L_i(\beta) = \mathbf{S(X_i)}\lambda_{\mathbf{i}}(\mathbf{X_i}) = \lambda_{\mathbf{i}}(\mathbf{X_i}) \exp[-\int_{\mathbf{0}}^{\mathbf{X_i}} \lambda_{\mathbf{i}}(\mathbf{u})\mathbf{du}]$$

Thus, everyone contributes $S(X_i)$ to the likelihood, and only those who fail contribute $\lambda_i(X_i)$.

This means we get a total likelihood of:

$$L(\beta) \quad = \quad \prod_{i=1}^{n} \lambda_i(X_i)^{\delta_i} \ \exp[-\int_0^{X_i} \lambda_i(u)du]$$

The above likelihood holds for all censored survival data, with general hazard function $\lambda(t)$. In other words, we haven't used the Cox PH assumption at all yet.

Now, let's multiply and divide by the term $\left[\sum_{j\in\mathcal{R}(X_i)} \lambda_i(X_i)\right]^{\delta_i}$:

$$L(\beta) = \prod_{\mathbf{i=1}}^{\mathbf{n}} \left[\frac{\lambda_{\mathbf{i}}(\mathbf{X_i})}{\sum_{\mathbf{j}\in\mathcal{R}(\mathbf{X_i})} \lambda_{\mathbf{i}}(\mathbf{X_i})}\right]^{\delta_{\mathbf{i}}} \left[\sum_{\mathbf{j}\in\mathcal{R}(\mathbf{X_i})} \lambda_{\mathbf{i}}(\mathbf{X_i})\right]^{\delta_{\mathbf{i}}} \exp[-\int_{\mathbf{0}}^{\mathbf{X_i}} \lambda_{\mathbf{i}}(\mathbf{u})\mathbf{du}]$$

Cox (1972) argued that the first term in this product contained almost all of the information about β, while the second two terms contained the information about $\lambda_0(t)$, i.e., the baseline hazard.

If we just focus on the first term, then under the Cox PH assumption:

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{n} \left[\frac{\lambda_i(X_i)}{\sum_{j \in \mathcal{R}(X_i)} \lambda_i(X_i)} \right]^{\delta_i} \\ &= \prod_{i=1}^{n} \left[\frac{\lambda_0(X_i) \exp(\beta \mathbf{z_i})}{\sum_{j \in \mathcal{R}(X_i)} \lambda_0(X_i) \exp(\beta \mathbf{z_j})} \right]^{\delta_i} \\ &= \prod_{i=1}^{n} \left[\frac{\exp(\beta \mathbf{z_i})}{\sum_{j \in \mathcal{R}(X_i)} \exp(\beta \mathbf{z_j})} \right]^{\delta_i} \end{aligned}$$

This is the partial likelihood defined by Cox. Note that it does not depend on the underlying hazard function $\lambda_0(\cdot)$. Cox recommends treating this as an ordinary likelihood for making inferences about β in the presence of the nuisance parameter $\lambda_0(\cdot)$.

A simple example:

individual	X_i	δ_i	Z_i
1	9	1	4
2	8	0	5
3	6	1	7
4	10	1	3

Now let's compile the pieces that go into the partial likelihood contributions at each failure time:

j	ordered failure time X_i	$\mathcal{R}(X_i)$	i_j	Likelihood contribution $\left[e^{\beta Z_i} / \sum_{j \in \mathcal{R}(X_i)} e^{\beta Z_j}\right]^{\delta_i}$
1	6	{1,2,3,4}	3	$e^{7\beta}/[e^{4\beta} + e^{5\beta} + e^{7\beta} + e^{3\beta}]$
2	8	{1,2,4}	2	1
3	9	{1,4}	1	$e^{4\beta}/[e^{4\beta} + e^{3\beta}]$
4	10	{4}	4	$e^{3\beta}/e^{3\beta} = 1$

The partial likelihood would be the product of these four terms.

5.2.4 Notes on the partial likelihood

$$\begin{aligned} L(\beta) &= \prod_{j=1}^{n} \left[\frac{e^{\beta \mathbf{z_j}}}{\sum_{\ell \in \mathcal{R}(X_j)} e^{\beta \mathbf{z}_\ell}} \right]^{\delta_j} \\ &= \prod_{j=1}^{K} \frac{e^{\beta \mathbf{Z_j}}}{\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta \mathbf{Z}_\ell}} \end{aligned}$$

where the product is over the K death (or failure) times.

- contributions only at the death times
- the partial likelihood is NOT a product of independent terms, but of conditional probabilities
- There are other choices besides $\Psi(\mathbf{z}) = \mathbf{e}^{\beta \mathbf{z}}$, but this is the most common and the one for which software is generally available.

5.2.5 Partial Likelihood inference

Inference can be conducted by treating the partial likelihood as though it satisfied all the regular likelihood properties (take the more advanced failure time course to see why!!) The **log-partial likelihood** is

$$\begin{aligned} \ell(\beta) &= \log \left\{ \prod_{j=1}^{n} \left[\frac{e^{\beta \mathbf{z_j}}}{\sum_{\ell \in \mathcal{R}(X_j)} e^{\beta \mathbf{z}_\ell}} \right]^{\delta_j} \right\} \\ &= \log \left[\prod_{j=1}^{K} \frac{e^{\beta \mathbf{z_j}}}{\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta \mathbf{z}_\ell}} \right] \\ &= \sum_{j=1}^{K} \left[\beta \mathbf{z_j} - \log \left(\sum_{\ell \in \mathcal{R}(\tau_\mathbf{j})} \mathbf{e}^{\beta \mathbf{z}_\ell} \right) \right] = \sum_{j=1}^{K} l_j(\beta) \end{aligned}$$

where l_j is the log-partial likelihood contribution at the j-th ordered death time.

Suppose there is only one covariate (β is one-dimensional).

The **partial likelihood score equations** are:

$$U(\beta) = \frac{\partial}{\partial \beta}\ell(\beta) = \sum_{j=1}^{n} \delta_j \left[Z_j - \frac{\sum_{\ell \in \mathcal{R}(X_j)} Z_\ell e^{\beta Z_\ell}}{\sum_{\ell \in \mathcal{R}(X_j)} e^{\beta Z_\ell}} \right]$$

We can express $U(\beta)$ intuitively as a sum of "observed" minus "expected" values:

$$U(\beta) = \frac{\partial}{\partial \beta}\ell(\beta) = \sum_{j=1}^{n} \delta_j (Z_j - \bar{Z}_j)$$

where $\bar{Z}_j$ is the "weighted average" of the covariate Z over all the individuals in the risk set at time τ_j. Note that β is involved through the term $\bar{Z}_j$.

The maximum partial likelihood estimators can be found by solving $U(\beta) = 0$.

Like standard likelihood theory, it can be shown (not easily) that

$$\frac{(\widehat{\beta} - \beta)}{se(\hat{\beta})} \sim N(0, 1)$$

The variance of $\hat{\beta}$ can be obtained by inverting the second derivative of the partial likelihood,

$$var(\hat{\beta}) \sim \left[-\frac{\partial^2}{\partial \beta^2}\ell(\beta) \right]^{-1}$$

From the above expression for $U(\beta)$, we have:

$$\frac{\partial^2}{\partial \beta^2}\ell(\beta) = \sum_{j=1}^{n} \delta_j \left[-\frac{\sum_{\ell \in \mathcal{R}(X_j)} (Z_j - \bar{Z}_j)^2 e^{\beta Z_\ell}}{\sum_{\ell \in \mathcal{R}(X_j)} e^{\beta Z_\ell}} \right]$$

Note: The true variance of $\hat{\beta}$ is a function of the unknown β. We calculate the "observed" information by substituting the partial likelihood estimate of β into the above variance formula.

Simple Example for 2-group comparison: (no ties)

Group 0: $4^+, 7, 8^+, 9, 10^+ \implies Z_i = 0$

Group 1: $3, 5, 5^+, 6, 8^+ \implies Z_i = 1$

j	ordered failure time X_i	Number at risk Group 0	Number at risk Group 1	Likelihood contribution $\left[e^{\beta Z_i} / \sum_{j \in \mathcal{R}(X_i)} e^{\beta Z_j}\right]^{\delta_i}$
1	3	5	5	$e^{\beta}/[5+5e^{\beta}]$
2	5	4	4	$e^{\beta}/[4+4e^{\beta}]$
3	6	4	2	$e^{\beta}/[4+2e^{\beta}]$
4	7	4	1	$e^{0}/[4+1e^{\beta}] = 1/[4+1e^{\beta}]$
5	9	2	0	$e^{0}/[2+0] = 1/2$

Again, we take the product over the likelihood contributions, then maximize to get the partial MLE for β.

What does β represent in this case?

Notes

- The "observed" information matrix is generally used because in practice, people find it has better properties. Also, the "expected" is very hard to calculate.
- There is a nice analogy with the score and information matrices from more standard regression problems, except that here we are summing over observed death times, rather than individuals.
- Newton Raphson is used by many of the computer packages to solve the partial likelihood equations.

5.2.6 Fitting Cox PH model with R

R uses the "`coxph`" command. First, try typing "`help(‘‘coxph"`"

```
coxph(formula, data=, weights, subset,
      na.action, init, control,
      ties=c("efron","breslow","exact"),
      singular.ok=TRUE, robust=FALSE,
      model=FALSE, x=FALSE, y=TRUE, tt, method, ...)
```

Exaxmple Leukemia Data

```
Call:
coxph(formula = Surv(weeks, remiss) ~ trt, data = leukemia, ties = "breslow")

  n= 42, number of events= 30

       coef exp(coef) se(coef)      z Pr(>|z|)
trt -1.5092    0.2211   0.4096 -3.685 0.000229 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

    exp(coef) exp(-coef) lower .95 upper .95
trt    0.2211      4.523   0.09907    0.4934

Concordance= 0.69  (se = 0.053 )
Rsquare= 0.304   (max possible= 0.989 )
Likelihood ratio test= 15.21  on 1 df,   p=9.615e-05
Wald test            = 13.58  on 1 df,   p=0.0002288
Score (logrank) test = 15.93  on 1 df,   p=6.571e-05
```

More Notes:

- The Cox Proportional hazards model has the advantage over a simple logrank test of giving us an estimate of the "risk ratio" (i.e., $\phi = \lambda_1(t)/\lambda_0(t)$). This is more informative than just a test statistic, and we can also form confidence intervals for the risk ratio.
- In this case, $\hat{\phi} = 0.221$, which can be interpreted to mean that the hazard for relapse among patients treated with 6-MP is less than 25% of that for placebo patients.
- From the STS LIST command in Stata or PROC LIFETEST in SAS, we were able to get estimates of the entire survival distribution $\hat{S}(t)$ for each treatment group; we can't immediately get this from our Cox model without further assumptions. **Why not?**

5.2.7 Adjustments for ties

The proportional hazards model assumes a continuous hazard – ties are not possible. There are four proposed modifications to the likelihood to adjust for ties.

(1) **Cox's (1972) modification:** "discrete" method

(2) **Peto-Breslow method**

(3) **Efron's (1977) method**

(4) **Exact method (Kalbfleisch and Prentice)**

(5) **Exact marginal method**

Some notation:

$\tau_1, ...\tau_K$	the K ordered, distinct death times
d_j	the number of failures at τ_j
H_j	the "history" of the entire data set, up to the j-th death or failure time, including the time of the failure, but not the identities of the d_j who fail there.
$i_{j1}, ...i_{jd_j}$	the identities of the d_j individuals who fail at τ_j

Cox's (1972) modification: "discrete" method

Cox's method assumes that if there are tied failure times, they truly happened at the same time. It is based on a discrete likelihood.

The **partial likelihood** is:

$$\begin{aligned} L(\beta) &= \prod_{j=1}^{K} Pr(i_{j1}, ...i_{jd_j} \text{ fail} \mid d_j \text{ fail at } \tau_j, \text{ from } \mathcal{R}) \\ &= \prod_{j=1}^{K} \frac{Pr(i_{j1}, ...i_{jd_j} \text{ fail} \mid \text{ in } \mathcal{R}(\tau_j))}{\sum_{\ell \in s(j,d_j)} Pr(\ell_1, ...\ell_{d_j} \text{ fail} \mid \text{ in } \mathcal{R}(\tau_j))} \\ &= \prod_{j=1}^{K} \frac{\exp(\beta \mathbf{z_{i_{j1}}}) \cdots \exp(\beta \mathbf{z_{i_{jd_j}}})}{\sum_{\ell \in s(j,d_j)} \exp(\beta \mathbf{z}_{\ell_1}) \cdots \exp(\beta \mathbf{z}_{\ell_{d_j}})} \\ &= \prod_{j=1}^{K} \frac{\exp(\beta \mathbf{S_j})}{\sum_{\ell \in s(j,d_j)} \exp(\beta \mathbf{S_{j\ell}})} \end{aligned}$$

In the previous formula

- $s(j, d_j)$ is the set of all possible sets of d_j individuals that can possibly be drawn from the risk set at time τ_j
- S_j is the sum of the Z's for all the d_j individuals who fail at τ_j
- $S_{j\ell}$ is the sum of the Z's for all the d_j individuals in the ℓ-th set drawn out of $s(j, d_j)$

What does this all mean??!!

Let's modify our previous simple example to include ties.

Simple Example (with ties)

Group 0: $4^+, 6, 8^+, 9, 10^+ \implies Z_i = 0$

Group 1: $3, 5, 5^+, 6, 8^+ \implies Z_i = 1$

j	Ordered failure time X_i	Number at risk Group 0	Number at risk Group 1	Lik. Contribution $e^{\beta S_j} / \sum_{\ell \in s(j,d_j)} e^{\beta S_{j\ell}}$
1	3	5	5	$e^\beta/[5+5e^\beta]$
2	5	4	4	$e^\beta/[4+4e^\beta]$
3	6	4	2	$e^\beta/[6+8e^\beta+e^{2\beta}]$
4	9	2	0	$e^0/2 = 1/2$

The tie occurs at $t = 6$, when $\mathcal{R}(\tau_j) = \{Z = 0 : (6, 8^+, 9, 10^+),$
$Z = 1 : (6, 8^+)\}$. Of the $\binom{6}{2} = 15$ possible pairs of subjects at risk at t=6, there are 6 pairs formed where both are from group 0 ($S_j = 0$), 8 pairs formed with one in each group ($S_j = 1$), and 1 pairs formed with both in group 1 ($S_j = 2$).

Problem: With numbers of ties, the denominator can have many many terms and be difficult to calculate.

Breslow method: (default)

Breslow and Peto suggested replacing the term $\sum_{\ell \in s(j,d_j)} e^{\beta S_{j\ell}}$ in the denominator by the term $\left(\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta Z_\ell}\right)^{d_j}$, so that the following modified partial likelihood would be used:

$$L(\beta) = \prod_{j=1}^{K} \frac{e^{\beta S_j}}{\sum_{\ell \in s(j,d_j)} e^{\beta S_{j\ell}}} \approx \prod_{j=1}^{K} \frac{e^{\beta S_j}}{\left(\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta Z_\ell}\right)^{d_j}}$$

Justification:

Suppose individuals 1 and 2 fail from $\{1, 2, 3, 4\}$ at time τ_j. Let $\phi(i)$ be the

hazard ratio for individual i (compared to baseline).

$$\begin{aligned}\frac{e^{\beta S_j}}{\sum_{\ell \in s(j,d_j)} e^{\beta S_{j\ell}}} &= \frac{\phi(1)}{\phi(1)+\phi(2)+\phi(3)+\phi(4)} \times \frac{\phi(2)}{\phi(2)+\phi(3)+\phi(4)} \\ &\quad + \frac{\phi(2)}{\phi(1)+\phi(2)+\phi(3)+\phi(4)} \times \frac{\phi(1)}{\phi(1)+\phi(3)+\phi(4)} \\ &\approx \frac{2\phi(1)\phi(2)}{[\phi(1)+\phi(2)+\phi(3)+\phi(4)]^2}\end{aligned}$$

The Peto (Breslow) approximation will break down when the number of ties are relative to the size of the risk sets, and then tends to yield estimates of β which are biased toward 0.

Efron's (1977) method:

Efron suggested an even closer approximation to the discrete likelihood:

$$L(\beta) = \prod_{j=1}^{K} \frac{e^{\beta S_j}}{\left(\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta Z_\ell} + \frac{j-1}{d_j} \sum_{\ell \in \mathcal{D}(\tau_j)} e^{\beta Z_\ell}\right)^{d_j}}$$

Like the Breslow approximation, Efron's method will yield estimates of β which are biased toward 0 when there are many ties.

However, (1995) Allison recommends the Efron approximation since it is much faster than the exact methods and tends to yield much closer estimates than the default Breslow approach.

Exact method (Kalbfleisch and Prentice)

The "discrete" option that we discussed in (1) is an exact method based on a discrete likelihood (assuming that tied events truly ARE tied).

This second exact method is based on the continuous likelihood, under the assumption that if there are tied events, that is due to the imprecise nature of our measurement, and that there must be some *true* ordering.

All possible orderings of the tied events are calculated, and the probabilites

of each are summed.

Example with 2 tied events (1,2) from riskset (1,2,3,4):

$$\frac{e^{\beta S_j}}{\sum_{\ell \in s(j,d_j)} e^{\beta S_{j\ell}}} = \frac{e^{\beta S_1}}{e^{\beta S_1} + e^{\beta S_2} + e^{\beta S_3} + e^{\beta S_4}} \times \frac{e^{\beta S_2}}{e^{\beta S_2} + e^{\beta S_3} + e^{\beta S_4}} + \frac{e^{\beta S_2}}{e^{\beta S_1} + e^{\beta S_2} + e^{\beta S_3} + e^{\beta S_4}} \times \frac{e^{\beta S_1}}{e^{\beta S_1} + e^{\beta S_3} + e^{\beta S_4}}$$

Bottom Line Implications of Ties (See Allison (1995), p.127-137)

(1) **When there are no ties**, all four options give *exactly* the same results.

(2) **When there are only a few ties**, it won't make much difference which method is used. However, since the exact methods won't take much extra computing time, you might as well use one of them.

(3) **When there are many ties** (relative to the number at risk), the Breslow option (default) performs poorly (Farewell & Prentice, 1980; Hsieh, 1995). Both of the approximate methods, Breslow and Efron, yield coefficients that are attenuated (biased toward 0).

(4) **The choice of which exact method to use** should be based on substantive grounds - are the tied event times truly tied? ...or are they the result of imprecise measurement?

(5) **Computing time of exact methods** is much longer than that of the approximate methods. However, in most cases it will still be less than 30 seconds even for the exact methods.

(6) **Best approximate method** - the Efron approximation nearly always works better than the Breslow method, with no increase in computing time, so use this option if exact methods are too computer-intensive.

R commands for PH Model with Ties:

Stata offers four options for adjustments with tied data:

- `breslow` (default)

- `efron`
- `exact` (same as the "discrete" option in SAS and the "exactp" option in Stata)

Fecundability Data Example:

```
coxph(formula = Surv(cycle, censor) ~ smoker, data = fecund)

  n= 586, number of events= 567

          coef exp(coef) se(coef)      z Pr(>|z|)
smoker -0.3878    0.6786   0.1140 -3.401 0.000671 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

       exp(coef) exp(-coef) lower .95 upper .95
smoker    0.6786      1.474    0.5427    0.8485

Concordance= 0.537  (se = 0.014 )
Rsquare= 0.021   (max possible= 1 )
Likelihood ratio test= 12.57  on 1 df,   p=0.000392
Wald test            = 11.57  on 1 df,   p=0.0006712
Score (logrank) test = 11.71  on 1 df,   p=0.0006218
```

A special case: the two-sample problem

Previously, we derived the logrank test from an intuitive perspective, assuming that we have $(X_{01}, \delta_{01}) \ldots (X_{0n_0}, \delta_{0n_0})$ from group 0 and $(X_{11}, \delta_{11}), \ldots, (X_{1n_1}, \delta_{1n_1})$ from group 1.

Just as a χ^2 test for binary data can be derived from a logistic model, we will see here that the logrank test can be derived as a special case of the Cox Proportional Hazards model.

First, let's re-define our notation in terms of (X_i, δ_i, Z_i):

$$\begin{aligned}(X_{01}, \delta_{01}), \ldots, (X_{0n_0}, \delta_{0n_0}) &\implies (X_1, \delta_1, 0), \ldots, (X_{n0}, \delta_{n0}, 0) \\ (X_{11}, \delta_{11}), \ldots, (X_{1n_1}, \delta_{1n_1}) &\implies (X_{n0+1}, \delta_{n0+1}, 1), \ldots, (X_{n0+n1}, \delta_{n0+n1}, 1)\end{aligned}$$

In other words, we have $n0$ rows of data $(X_i, \delta_i, 0)$ for the group 0 subjects, then $n1$ rows of data $(X_i, \delta_i, 1)$ for the group 1 subjects.

Using the proportional hazards formulation, we have

$$\lambda(t; Z) = \lambda_0(t)\, e^{\beta Z}$$

Group 0 hazard: $\lambda_0(t)$

Group 1 hazard: $\lambda_0(t)\, e^{\beta}$

The log-partial likelihood is:

$$\begin{aligned} logL(\beta) &= \log\left[\prod_{j=1}^{K} \frac{e^{\beta Z_j}}{\sum_{\ell\in\mathcal{R}(\tau_j)} e^{\beta Z_\ell}}\right] \\ &= \sum_{j=1}^{K}\left[\beta Z_j - \log[\sum_{\ell\in\mathcal{R}(\tau_j)} e^{\beta Z_\ell}]\right] \end{aligned}$$

Taking the derivative with respect to β, we get:

$$\begin{aligned} U(\beta) &= \frac{\partial}{\partial\beta}\ell(\beta) \\ &= \sum_{j=1}^{n}\delta_j\left[Z_j - \frac{\sum_{\ell\in\mathcal{R}(\tau_j)} Z_\ell e^{\beta Z_\ell}}{\sum_{\ell\in\mathcal{R}(\tau_j)} e^{\beta Z_\ell}}\right] \\ &= \sum_{j=1}^{n}\delta_j(Z_j - \bar{Z}_j) \end{aligned}$$

$$\text{where } \bar{Z}_j = \frac{\sum_{\ell\in\mathcal{R}(\tau_j)} Z_\ell e^{\beta Z_\ell}}{\sum_{\ell\in\mathcal{R}(\tau_j)} e^{\beta Z_\ell}}$$

$U(\beta)$ is called the **"score"**.

As we discussed earlier in the class, one useful form of a likelihood-based test is the **score test**. This is obtained by using the score $U(\beta)$ evaluated at H_o as a test statistic.

Let's look more closely at the form of the score:

$\delta_j Z_j$ **observed** number of deaths in group 1 at τ_j

$\delta_j \bar{Z}_j$ **expected** number of deaths in group 1 at τ_j

Why?

Under $H_0 : \beta = 0$, $\bar{Z}_j$ is simply the number of individuals from group 1 in the risk set at time τ_j (call this r_{1j}), divided by the total number in the risk

set at that time (call this r_j). Thus, $\bar{Z}_j$ approximates the probability that given there is a death at τ_j, it is from group 1.

Thus, the score statistic is of the form:

$$\sum_{j=1}^{n}(O_j - E_j)$$

When there are ties, the likelihood has to be replaced by one that allows for ties.

In R:

R produces the score test as follows (for the fecundability example above):

```
Likelihood ratio test= 12.68  on 1 df,   p=0.0003695
Wald test            = 12.12  on 1 df,   p=0.0004985
Score (logrank) test = 12.25  on 1 df,   p=0.0004642
```

... which is the same as what would have been produced by the logrank test:

```
Call:
survdiff(formula = Surv(cycle, censor) ~ smoker, data = fecund)

           N Observed Expected (O-E)^2/E (O-E)^2/V
smoker=0 486      474      446      1.76      12.3
smoker=1 100       93      121      6.50      12.3

 Chisq= 12.3  on 1 degrees of freedom, p= 0.000464
```

Chapter 10

Parametric Survival Analysis

So far, we have focused primarily on nonparametric and semi-parametric approaches to survival analysis, with heavy emphasis on the Cox proportional hazards model:

$$\lambda(t, \mathbf{Z}) \quad = \quad \lambda_0(t) \ \exp(\beta \mathbf{Z})$$

We used the following estimating approach:

- We estimated $\lambda_0(t)$ nonparametrically, using the Kaplan-Meier estimator, or using the Kalbfleisch/Prentice estimator under the PH assumption
- We estimated β by assuming a linear model between the log HR and covariates, under the PH model

Both estimates were based on maximum likelihood theory.

Reading: for parametric models see Collett.

There are several reasons why we should consider some alternative approaches based on parametric models:

- The assumption of proportional hazards might not be appropriate (based on major departures)
- If a parametric model actually holds, then we would probably gain efficiency

- We may want to handle non-standard situations like
  - interval censoring
  - incorporating population mortality
- We may want to make some connections with other familiar approaches (e.g. use of the Poisson likelihood)
- We may want to obtain some estimates for use in designing a future survival study.

10.1 Exponential Regression

Observed data: $(X_i, \delta_i, \mathbf{Z}_i)$ for individual i,

$\mathbf{Z}_i = (Z_{i1}, Z_{i2}, ..., Z_{ip})$ represents a set of p covariates.

Right censoring: Assume that $X_i = \min(T_i, U_i)$

Survival distribution: Assume T_i follows an exponential distribution with a parameter λ that depends on $\mathbf{Z}_i$, say $\lambda_i = \Psi(\mathbf{Z}_i)$.

Then we can write:

$$T_i \sim exponential(\Psi(\mathbf{Z}_i))$$

First, let's review some facts about the exponential distribution (from our first survival lecture):

$$\begin{aligned}
f(t) &= \lambda e^{-\lambda t} \qquad \text{for } t \geq 0 \\
S(t) &= P(T \geq t) = \int_t^\infty f(u)du = e^{-\lambda t} \\
F(t) &= P(T < t) = 1 - e^{-\lambda t} \\
\lambda(t) &= \frac{f(t)}{S(t)} = \lambda \qquad \text{constant hazard!} \\
\Lambda(t) &= \int_0^t \lambda(u)\, du = \int_0^t \lambda\, du = \lambda t
\end{aligned}$$

Now, we say that λ is a constant *over time* t, but we want to let it depend on the covariate values, so we are setting

$$\lambda_i = \Psi(\mathbf{Z}_i)$$

The hazard rate would therefore be the same for any two individuals with the same covariate values.

Although there are many possible choices for Ψ, one simple and natural choice is:

$$\Psi(\mathbf{Z}_i) = \exp[\beta_0 + Z_{i1}\beta_1 + Z_{i2}\beta_2 + ... + Z_{ip}\beta_p]$$

WHY?

- ensures a positive hazard
- for an individual with $\mathbf{Z} = \mathbf{0}$, the hazard is e^{β_0}.

The model is called **exponential regression** because of the natural generalization from regular linear regression

10.1.1 Exponential regression for the 2-sample case

Assume we have only a single covariate $\mathbf{Z} = Z$,
i.e., $(p = 1)$.

Hazard Rate:

$$\Psi(\mathbf{Z}_i) = \exp(\beta_0 + Z_i\beta_1)$$

Define: $Z_i = 0$ if individual i is in group 0
$Z_i = 1$ if individual i is in group 1

What is the hazard for group 0?

What is the hazard for group 1?

What is the hazard ratio of group 1 to group 0?

What is the interpretation of β_1?

10.1.2 Likelihood for Exponential Model

Under the assumption of right censored data, each person has one of two possible contributions to the likelihood:

(a) they have an **event** at X_i $(\delta_i = 1) \Rightarrow$ contribution is

$$L_i = \underbrace{S(X_i)}_{\text{survive to } X_i} \cdot \underbrace{\lambda(X_i)}_{\text{fail at } X_i} = e^{-\lambda X_i}\,\lambda$$

(b) they are **censored** at X_i ($\delta_i = 0$) $\Rightarrow$ contribution is

$$L_i = \underbrace{S(X_i)}_{\text{survive to } X_i} = e^{-\lambda X_i}$$

The **likelihood** is the product over all of the individuals:

$$\begin{aligned}
\mathcal{L} &= \prod_i L_i \\
&= \prod_i \underbrace{\left(\lambda e^{-\lambda X_i}\right)^{\delta_i}}_{\text{events}} \underbrace{\left(e^{-\lambda X_i}\right)^{(1-\delta_i)}}_{\text{censorings}} \\
&= \prod_i \lambda^{\delta_i} \left(e^{-\lambda X_i}\right)
\end{aligned}$$

10.1.3 Maximum Likelihood for Exponential

How do we use the likelihood?

- first take the log
- then take the partial derivative with respect to β
- then set to zero and solve for $\widehat{\beta}$
- this gives us the **maximum likelihood estimators**

The log-likelihood is:

$$\begin{aligned}
\log \mathcal{L} &= \log \left[\prod_i \lambda^{\delta_i} \left(e^{-\lambda X_i}\right) \right] \\
&= \sum_i \left[\delta_i \log(\lambda) - \lambda X_i\right] \\
&= \sum_i \left[\delta_i \log(\lambda)\right] - \sum_i \lambda X_i
\end{aligned}$$

For the case of exponential regression, we now substitute the hazard $\lambda = \Psi(\mathbf{Z}_i)$ in the above log-likelihood:

$$\log \mathcal{L} = \sum_i \left[\delta_i \log(\Psi(\mathbf{Z}_i))\right] - \sum_i \Psi(\mathbf{Z}_i) X_i \tag{10.1}$$

General Form of Log-likelihood for Right Censored Data

In general, whenever we have right censored data, the likelihood and corresponding log likelihood will have the following forms:

$$\begin{aligned} \mathcal{L} &= \prod_i [\lambda_i(X_i)]^{\delta_i} \; S_i(X_i) \\ \log \mathcal{L} &= \sum_i [\delta_i \log (\lambda_i(X_i))] - \sum_i \Lambda_i(X_i) \end{aligned}$$

where

- $\lambda_i(X_i)$ is the hazard for the individual i who fails at X_i
- $\Lambda_i(X_i)$ is the cumulative hazard for an individual at their failure or censoring time

For example, see the derivation of the likelihood for a Cox model on p.11-13 of Lecture 4 notes. We started with the likelihood above, then substituted the specific forms for $\lambda(X_i)$ under the PH assumption.

Consider our model for the hazard rate:

$$\lambda = \Psi(\mathbf{Z}_i) = \exp[\beta_0 + Z_{i1}\beta_1 + Z_{i2}\beta_2 + ... + Z_{ip}\beta_p]$$

We can write this using vector notation, as follows:

$$\begin{aligned} \text{Let} \quad \mathbf{Z}_i &= (1, Z_{i1}, ...Z_{ip})^T \\ \text{and} \quad \beta &= (\beta_0, \beta_1, ...\beta_p) \end{aligned}$$

(Since β_0 is the intercept (i.e., the log hazard rate for the baseline group), we put a "1" as the first term in the vector $\mathbf{Z}_i$.)

Then, we can write the hazard as:

$$\Psi(\mathbf{Z}_i) = \exp[\beta \mathbf{Z}_i]$$

Now we can substitute $\Psi(\mathbf{Z}_i) = \exp[\beta \mathbf{Z}_i]$ in the log-likelihood shown in (10.1):

$$\log \mathcal{L} = \sum_{i=1}^{n} \delta_i (\beta \mathbf{Z}_i) - \sum_{i=1}^{n} X_i \exp(\beta \mathbf{Z}_i)$$

10.1.4 Score Equations

Taking the derivative with respect to β_0, the score equation is:

$$\frac{\partial \log \mathcal{L}}{\partial \beta_0} = \sum_{i=1}^{n} [\delta_i - X_i \exp(\beta \mathbf{Z}_i)]$$

For β_k, $k = 1, ...p$, the equations are:

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \beta_k} &= \sum_{i=1}^{n} [\delta_i Z_{ik} - X_i Z_{ik} \exp(\beta \mathbf{Z}_i)] \\ &= \sum_{i=1}^{n} Z_{ik} [\delta_i - X_i \exp(\beta \mathbf{Z}_i)] \end{aligned}$$

To find the MLE's, we set the above equations to 0 and solve (simultaneously). The equations above imply that the MLE's are obtained by setting the weighted number of failures ($\sum_i Z_{ik}\delta_i$) equal to the weighted cumulative hazard ($\sum_i Z_{ik}\Lambda(X_i)$). To find the variance of the MLE's, we need to take the second derivatives:

$$-\frac{\partial^2 \log \mathcal{L}}{\partial \beta_k \partial \beta_j} = \sum_{i=1}^{n} Z_{ik} Z_{ij} X_i \exp(\beta \mathbf{Z}_i)$$

Some algebra (see Cox and Oakes section 6.2) reveals that

$$Var(\widehat{\beta}) = I(\beta)^{-1} = \left[\mathbf{Z}(\mathbf{I} - \Pi)\mathbf{Z}^T\right]^{-1}$$

where

- $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ is a $(p+1) \times n$ matrix
  (p covariates plus the "1" for the intercept β_0)

- $\Pi = diag(\pi_1, \ldots, \pi_n)$ (this means that Π is a diagonal matrix, with the terms $\pi_1, \ldots, \pi_n$ on the diagonal)

- π_i is the probability that the i-th person is censored, so $(1 - \pi_i)$ is the probability that they failed.

- **Note:** The information $I(\beta)$ (inverse of the variance) is proportional to the number of failures, not the sample size. This will be important when we talk about study design.

10.1.5 The Single Sample Problem ($Z_i = 1$ for everyone)

First, what is the MLE of β_0?

We set $\frac{\partial \log \mathcal{L}}{\partial \beta_0} = \sum_{i=1}^{n}[\delta_i - X_i \exp(\beta_0 Z_i)]$ equal to 0 and solve:

$$\begin{aligned} \Rightarrow \quad \sum_{i=1}^{n} \delta_i &= \sum_{i=1}^{n}[X_i \exp(\beta_0)] \\ d &= \exp(\beta_0) \sum_{i=1}^{n} X_i \\ \exp(\widehat{\beta}_0) &= \frac{d}{\sum_{i=1}^{n} X_i} \\ \hat{\lambda} &= \frac{d}{t} \end{aligned}$$

where d is the total number of deaths (or events), and $t = \sum X_i$ is the total person-time contributed by all individuals.

If d/t is the MLE for λ, what does this imply about the MLE of β_0?

Using the previous formula $Var(\hat{\beta}) = \left[\mathbf{Z}(\mathbf{I} - \Pi)\mathbf{Z}^T\right]^{-1}$,
what is the variance of $\widehat{\beta}_0$?:

With some matrix algebra, you can show that it is:

$$Var(\widehat{\beta}_0) = \frac{1}{\sum_{i=1}^{n}(1 - \pi_i)} = \frac{1}{d}$$

What about $\hat{\lambda} = e^{\hat{\beta}_0}$?

By the delta method,

$$\begin{aligned} Var(\hat{\lambda}) &= \hat{\lambda}^2 \, Var(\widehat{\beta}_0) \\ &= \quad ? \end{aligned}$$

10.1.6 The Two-Sample Problem

	Z_i	Subjects	Events	Follow-up
Group 0:	$Z_i = 0$	n_0	d_0	$t_0 = \sum_{i=1}^{n_0} X_i$
Group 1:	$Z_i = 1$	n_1	d_1	$t_1 = \sum_{i=1}^{n_1} X_i$

The log-likelihood

$$\begin{aligned}
\log \mathcal{L} &= \sum_{i=1}^{n} \delta_i(\beta_0 + \beta_1 Z_i) - \sum_{i=1}^{n} X_i \exp(\beta_0 + \beta_1 Z_i) \\
\text{so } \frac{\partial \log \mathcal{L}}{\partial \beta_0} &= \sum_{i=1}^{n} [\delta_i - X_i \exp(\beta_0 + \beta_1 Z_i)] \\
&= (d_0 + d_1) - (t_0 e^{\beta_0} + t_1 e^{\beta_0 + \beta_1}) \\
\frac{\partial \log \mathcal{L}}{\partial \beta_1} &= \sum_{i=1}^{n} Z_i [\delta_i - X_i \exp(\beta_0 + \beta_1 Z_i)] \\
&= d_1 - t_1 e^{\beta_0 + \beta_1} \\
\text{This implies: } \hat{\lambda}_1 &= e^{\hat{\beta}_0 + \hat{\beta}_1} = ? \\
\hat{\lambda}_0 &= e^{\hat{\beta}_0} = ? \\
\hat{\beta}_0 &= ? \\
\hat{\beta}_1 &= ?
\end{aligned}$$

Important Result:

> **The maximum likelihood estimates (MLE's) of the hazard rates under the exponential model are the number of events divided by the person-years of follow-up!**

(this result will be relied on heavily when we discuss study design)

10.1.7 Exponential Regression: Means and Medians

Mean Survival Time

For the exponential distribution, $E(T) = 1/\lambda$.

- **Control Group:**

$$\overline{T}_0 = 1/\hat{\lambda}_0 = 1/\exp(\hat{\beta}_0)$$

- **Treatment Group:**

$$\overline{T}_1 = 1/\hat{\lambda}_1 = 1/\exp(\hat{\beta}_0 + \hat{\beta}_1)$$

Median Survival Time

This is the value M at which $S(t) = e^{-\lambda t} = 0.5$, so $M = \text{median} = \frac{-\log(0.5)}{\lambda}$

- **Control Group:**

$$\hat{M}_0 = \frac{-\log(0.5)}{\hat{\lambda}_0} = \frac{-\log(0.5)}{\exp(\hat{\beta}_0)}$$

- **Treatment Group:**

$$\hat{M}_1 = \frac{-\log(0.5)}{\hat{\lambda}_1} = \frac{-\log(0.5)}{\exp(\hat{\beta}_0 + \hat{\beta}_1)}$$

10.1.8 Exponential Regression: Variance Estimates and Test Statistics

We can also calculate the variances of the MLE's as simple functions of the number of failures:

$$var(\hat{\beta}_0) = \frac{1}{d_0}$$

$$var(\hat{\beta}_1) = \frac{1}{d_0} + \frac{1}{d_1}$$

So our test statistics are formed as:

For testing $H_o : \beta_0 = 0$:

$$\chi^2_w = \frac{\left(\hat{\beta}_0\right)^2}{var(\hat{\beta}_0)} = \frac{[\log(d_0/t_0)]^2}{1/d_0}$$

For testing $H_o : \beta_1 = 0$:

$$\chi^2_w = \frac{\left(\hat{\beta}_1\right)^2}{var(\hat{\beta}_1)} = \frac{\left[\log(\frac{d_1/t_1}{d_0/t_0})\right]^2}{\frac{1}{d_0} + \frac{1}{d_1}}$$

How would we form confidence intervals for the hazard ratio?

10.1.9 The Likelihood Ratio Test Statistic: An alternative to the Wald test

A likelihood ratio test is based on 2 times the log of the ratio of the likelihoods under the null and alternative. We reject H_0 if $2\log(\text{LR}) > \chi^2_{1,0.05}$, where

$$LR = \frac{\mathcal{L}(H_1)}{\mathcal{L}(H_0)} = \frac{\mathcal{L}(\widehat{\lambda}_0, \widehat{\lambda}_1)}{\mathcal{L}(\widehat{\lambda})}$$

For a sample of n independent exponential random variables with parameter λ, the Likelihood is:

$$\begin{aligned} L &= \prod_{i=1}^{n} [\lambda^{\delta_i} \exp(-\lambda x_i)] \\ &= \lambda^d \exp(-\lambda \sum x_i) \\ &= \lambda^d \exp(-\lambda n \bar{x}) \end{aligned}$$

where d is the number of deaths or failures. The log-likelihood is

$$\ell = d\log(\lambda) - \lambda n \bar{x}$$

and the MLE is

$$\widehat{\lambda} = d/(n\bar{x})$$

Two-Sample Case: LR test calculations

Data:

Group 0:	d_0 failures among the n_0 females mean failure time is $\bar{x}_0 = (\sum_i^{n_0} X_i)/n_0$
Group 1:	d_1 failures among the n_1 males mean failure time is $\bar{x}_1 = (\sum_i^{n_1} X_i)/n_1$

Under the **alternative hypothesis**:

$$\begin{aligned} \mathcal{L} &= \lambda_1^{d_1} \exp(-\lambda_1 n_1 \bar{x}_1) \times \lambda_0^{d_0} \exp(-\lambda_0 n_0 \bar{x}_0) \\ \log(\mathcal{L}) &= d_1 \log(\lambda_1) - \lambda_1 n_1 \bar{x}_1 + d_0 \log(\lambda_0) - \lambda_0 n_0 \bar{x}_0 \end{aligned}$$

The MLE's are:

$$\begin{aligned} \widehat{\lambda}_1 &= d_1/(n_1\bar{x}_1) \qquad \text{for males} \\ \widehat{\lambda}_0 &= d_0/(n_0\bar{x}_0) \qquad \text{for females} \end{aligned}$$

Under the **null hypothesis**:

$$\begin{aligned} \mathcal{L} &= \lambda^{d_1+d_0}\exp[-\lambda(n_1\bar{x}_1 + n_0\bar{x}_0)] \\ \log(\mathcal{L}) &= (d_1+d_0)\log(\lambda) - \lambda[n_1\bar{x}_1 + n_0\bar{x}_0] \end{aligned}$$

The corresponding MLE is

$$\widehat{\lambda} = (d_1+d_0)/[n_1\bar{x}_1 + n_0\bar{x}_0]$$

A likelihood ratio test can be constructed by taking twice the difference of the log-likelihoods under the alternative and the null hypotheses:

$$-2\left[(d_0+d_1)\log\left(\frac{d_0+d_1}{t_0+t_1}\right) - d_1\log[d_1/t_1] - d_0\log[d_0/t_0]\right]$$

10.1.10 Nursing home example

For the females:

- $n_0 = 1173$
- $d_0 = 902$
- $t_0 = 310754$
- $\bar{x}_0 = 265$

For the males:

- $n_1 = 418$
- $d_1 = 367$
- $t_1 = 75457$
- $\bar{x}_1 = 181$

Plugging these values in, we get a LR test statistic of 64.20.

Hand Calculations using events and follow-up

By adding up "LOS" for males to get t_1 and for females to get t_0, I obtained:

- $d_0 = 902$ (females)
  $d_1 = 367$ (males)
- $t_0 = 310754$ (female follow-up)
  $t_1 = 75457$ (male follow-up)

This yields an estimated log HR:

$$\hat{\beta}_1 \;=\; \log\left[\frac{d1/t1}{d0/t0}\right] = \log\left[\frac{367/75457}{902/310754}\right] = \log(1.6756) = 0.5162$$

The estimated standard error is:

$$\sqrt{var(\hat{\beta}_1)} \;=\; \sqrt{\frac{1}{d_1}+\frac{1}{d_0}} \;=\; \sqrt{\frac{1}{902}+\frac{1}{367}} \;=\; 0.06192$$

So the Wald test becomes:

$$\chi^2_W \;=\; \frac{\hat{\beta}_1^2}{var(\hat{\beta}_1)} \;=\; \frac{(0.51619)^2}{0.061915} \;=\; 69.51$$

We can also calculate $\hat{\beta}_0 = log(d_0/t_0) = -5.842$,
along with its standard error $se(\hat{\beta}_0) = \sqrt{(1/d0)} = 0.0333$

10.1.11 Exponential Regression in R

We use the `survreg` command with the `dist="exp"` option:

```
Call:
survreg(formula = Surv(losyr, fail) ~ gender, data = nurshome,
    dist = "exp")
             Value Std. Error     z        p
(Intercept) -0.0578     0.0333 -1.73 8.28e-02
gender      -0.5162     0.0619 -8.34 7.62e-17

Scale fixed at 1

Exponential distribution
```

```
Loglik(model)= -1006.3   Loglik(intercept only)= -1038.4
Chisq= 64.2 on 1 degrees of freedom, p= 1.1e-15
Number of Newton-Raphson Iterations: 5
n= 1591
```

Since $Z = 8.337$, the chi-square test is $Z^2 = 69.51$.

10.2 The Weibull Regression Model

At the beginning of the course, we saw that the survivorship function for a Weibull random variable is:

$$S(t) \quad = \quad \exp[-\lambda(t^\kappa)]$$

and the hazard function is:

$$\lambda(t) \quad = \quad \kappa\,\lambda\,t^{(\kappa-1)}$$

The Weibull regression model assumes that for someone with covariates $\mathbf{Z}_i$, the survivorship function is

$$S(t;\mathbf{Z}_i) = \exp[-\Psi(\mathbf{Z}_i)(t^\kappa)]$$

where $\Psi(\mathbf{Z}_i)$ is defined as in exponential regression to be:

$$\Psi(\mathbf{Z}_i) = \exp[\beta_0 + Z_{i1}\beta_1 + Z_{i2}\beta_2 + ...Z_{ip}\beta_p]$$

For the 2-sample problem, we have:

$$\Psi(\mathbf{Z}_i) = \exp[\beta_0 + Z_{i1}\beta_1]$$

10.2.1 Weibull MLEs for the 2-sample problem

Log-likelihood:

$$\log\mathcal{L} \quad = \quad \sum_{i=1}^{n}\delta_i \log\left[\kappa\exp(\beta_0+\beta_1 Z_i)X_i^{\kappa-1}\right] - \sum_{i=1}^{n} X_i^\kappa \exp(\beta_0+\beta_1 Z_i)$$

$$\Rightarrow \qquad \exp(\hat\beta_0) = d_0/t_0\kappa \qquad \exp(\hat\beta_0+\hat\beta_1) = d_1/t_1\kappa$$

where

$$t_{j\kappa} = \sum_{i=1}^{n_j} X_i^{\hat\kappa} \text{ among } n_j \text{ subjects}$$

$$\hat\lambda_0(t) = \hat\kappa\ \exp(\hat\beta_0)\ t^{\hat\kappa-1}\ \ \hat\lambda_1(t) = \hat\kappa\ \exp(\hat\beta_0+\hat\beta_1)\ t^{\hat\kappa-1}$$

$$\begin{aligned}\widehat{HR} \quad &= \quad \hat\lambda_1(t)/\hat\lambda_0(t) = \exp(\hat\beta_1) \\ &= \quad \exp\left(\frac{d_1/t_1\kappa}{d_0/t_0\kappa}\right)\end{aligned}$$

10.2.2 Weibull Regression: Means and Medians

Mean Survival Time

For the Weibull distribution, $E(T) = \lambda^{(-1/\kappa)}\Gamma[(1/\kappa)+1]$.

- **Control Group:**

$$\overline{T}_0 = \hat{\lambda}_0^{(-1/\hat{\kappa})}\ \Gamma[(1/\hat{\kappa})+1]$$

- **Treatment Group:**

$$\overline{T}_1 = \hat{\lambda}_1^{(-1/\hat{\kappa})}\ \Gamma[(1/\hat{\kappa})+1]$$

Median Survival Time

For the Weibull distribution, $M = \text{median} = \left[\frac{-\log(0.5)}{\lambda}\right]^{1/\kappa}$

- **Control Group:**

$$\hat{M}_0 = \left[\frac{-\log(0.5)}{\hat{\lambda}_0}\right]^{1/\hat{\kappa}}$$

- **Treatment Group:**

$$\hat{M}_1 = \left[\frac{-\log(0.5)}{\hat{\lambda}_1}\right]^{1/\hat{\kappa}}$$

where $\hat{\lambda}_0 = \exp(\hat{\beta}_0)$ and $\hat{\lambda}_1 = \exp(\hat{\beta}_0 + \hat{\beta}_1)$.

Note: the symbol Γ is the "gamma" function. If x is an integer, then

$$\Gamma(x) = (x-1)!$$

In cases where x is not an integer, this function has to be evaluated numerically. In homework and labs, I will supply this value to you.

The Weibull regression model is very easy to fit:

- In STATA: Just specify `dist(weibull)` instead of `dist(exp)` within the `streg` command
- In SAS: use model option `dist=weibull` within the `proc lifereg` procedure
- In R: we use the `survreg` command with the `dist="weibull"` option.

Note: to get more information on these modeling procedures, use the online help facilities.

10.2.3 Fitting the Weibull model in R

The nursing home data example

```
Call:
survreg(formula = Surv(losyr, fail) ~ gender, data = nurshome,
    dist = "weibull")
             Value Std. Error     z        p
(Intercept) -0.143     0.0542 -2.65 8.13e-03
gender      -0.673     0.1011 -6.66 2.67e-11
Log(scale)   0.487     0.0232 20.99 8.94e-98

Scale= 1.63

Weibull distribution
Loglik(model)= -731.1   Loglik(intercept only)= -751.9
Chisq= 41.73 on 1 degrees of freedom, p= 1e-10
Number of Newton-Raphson Iterations: 5
n= 1591
```

10.3 Comparison of parametric and nonparametric models

10.3.1 Comparison of Exponential with Kaplan-Meier

We can see how well the Exponential model fits by comparing the survival estimates for males and females under the exponential model, i.e., $P(T \geq t) = e^{(-\hat{\lambda}_z t)}$, to the Kaplan-Meier survival estimates:

10.3.2 Comparison of Weibull with Kaplan-Meier

We can see how well the Weibull model fits by comparing the survival estimates, $P(T \geq t) = e^{(-\hat{\lambda}_z t^{\hat{\kappa}})}$, to the Kaplan-Meier survival estimates. **Which do you think fits best?**

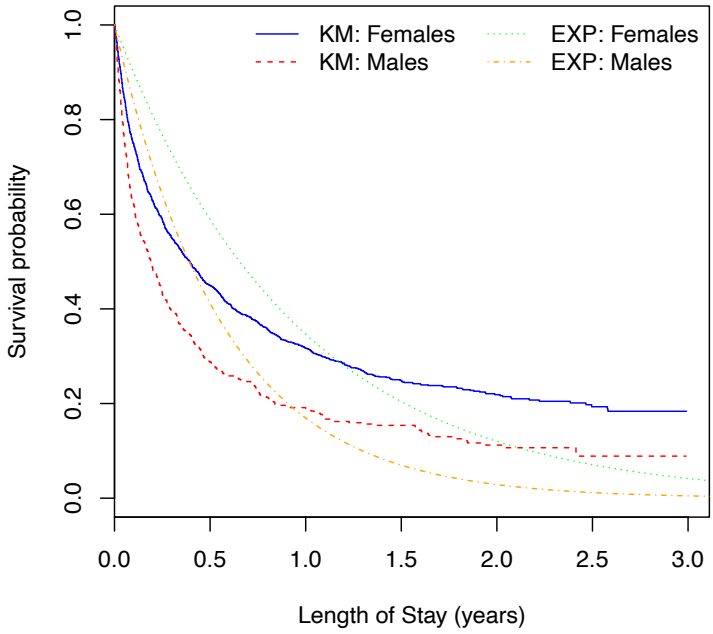
10.3.3 Other useful plots for evaluating goodness of fit

Other visual tests of goodness of fit are as follows:

- $-\log(\hat{S}(t))$ vs t
- $\log[-\log(\hat{S}(t))]$ vs $\log(t)$

Why are these useful?

Figure 10.1: Comparison of exponential with Kaplan Meier



If T is exponential, then $S(t) = \exp(-\lambda t))$

$$\begin{aligned} \text{so} \quad \log(S(t)) &= -\lambda t \\ \text{and} \quad \Lambda(t) &= \lambda\, t \end{aligned}$$

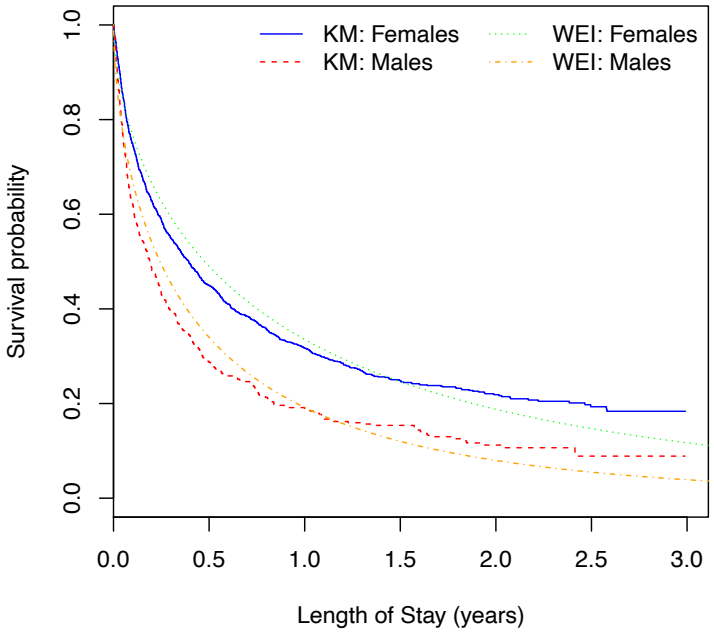
a straight line in t with slope λ and intercept=0

If T is Weibull, then $S(t) = \exp(-(\lambda t)^{\kappa})$

$$\begin{aligned} \text{so} \quad \log(S(t)) &= -\lambda t^{\kappa} \\ \text{then} \quad \Lambda(t) &= \lambda t^{\kappa} \\ \text{and} \quad \log(-\log(S(t))) &= \log(\lambda) + \kappa * \log(t) \end{aligned}$$

a straight line in $\log(t)$ with slope κ and intercept $\log(\lambda)$. So we can calculate our estimated $\Lambda(t)$ and plot it versus t, and if it seems to form a straight line, then the exponential distribution is probably appropriate for our dataset.

Figure 10.2: Comparison between Weibull and Kaplan Meier



Or we can plot $\log \hat{\Lambda}(t)$ versus $\log(t)$, and if it seems to form a straight line, then the Weibull distribution is probably appropriate for our dataset.

10.3.4 Comparison of Methods for the two-sample problem:

Data:

	Z_i	Subjects	Events	Follow-up
Group 0:	$Z_i = 0$	n_0	d_0	$t_0 = \sum_{i=1}^{n_0} X_i$
Group 1:	$Z_i = 1$	n_1	d_1	$t_1 = \sum_{i=1}^{n_1} X_i$

In General:

$$\lambda_z(t) \quad = \quad \lambda(t, Z = z) \qquad \text{for } z = 0 \text{ or } 1.$$

The hazard rate depends on the value of the covariate Z. In this case, we are assuming that we only have a single covariate, and it is binary ($Z = 1$ or $Z = 0$)

Figure 10.3: Log-log plot of the exponential model

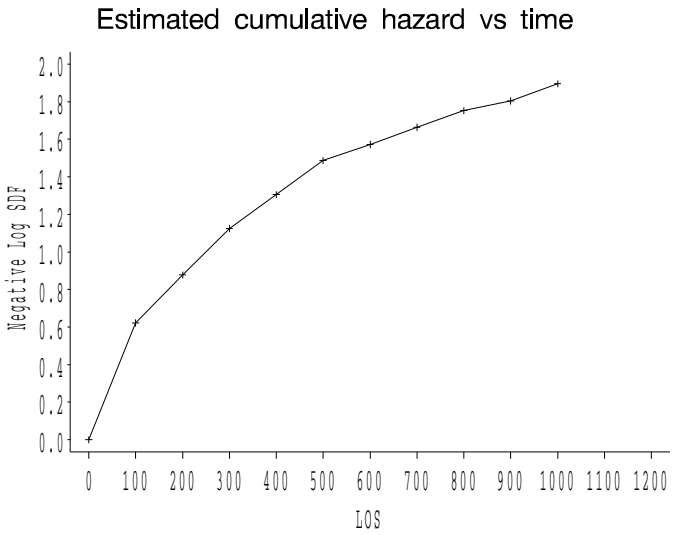
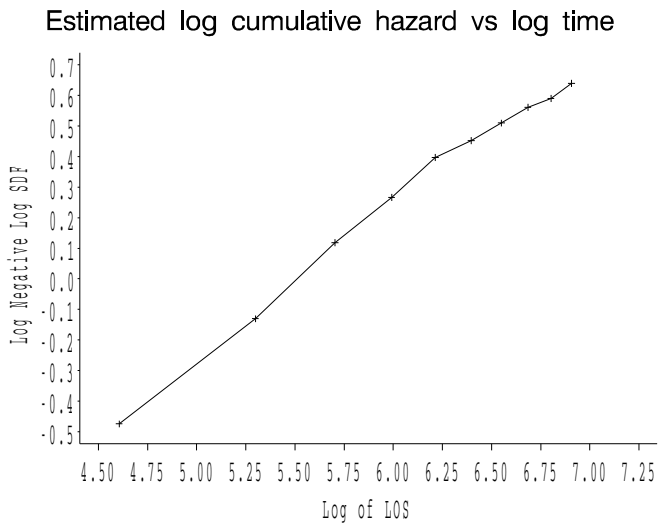


Figure 10.4: Log-log plot of the Weibull model



10.4 Models

10.4.1 Exponential Regression

$$\begin{aligned} \lambda_z(t) &= \exp(\beta_0 + \beta_1 Z) \\ \Rightarrow \lambda_0 &= \exp(\beta_0) \\ \lambda_1 &= \exp(\beta_0 + \beta_1) \\ HR &= \exp(\beta_1) \end{aligned}$$

10.4.2 Weibull Regression:

$$\begin{aligned} \lambda_z(t) &= \kappa\ \exp(\beta_0 + \beta_1 Z)\ t^{\kappa-1} \\ \Rightarrow \lambda_0 &= \kappa\ \exp(\beta_0)\ t^{\kappa-1} \\ \lambda_1 &= \kappa\ \exp(\beta_0 + \beta_1)\ t^{\kappa-1} \\ HR &= \exp(\beta_1) \end{aligned}$$

10.4.3 Proportional Hazards Model:

$$\begin{aligned} \lambda_z(t) &= \lambda_0(t) \exp(\beta_1) \\ \Rightarrow \lambda_0 &= \lambda_0(t) \qquad \text{KM?} \\ \lambda_1 &= \lambda_0(t) \exp(\beta_1) \\ HR &= \exp(\beta_1) \end{aligned}$$

10.4.4 Remarks

Exponential model is a special case of the Weibull model with $\kappa = 1$ (note: Collett uses γ instead of κ)

Exponential and Weibull models are both special cases of the Cox PH model.

How can you show this?

If either the exponential model or the Weibull model is valid, then these models will tend

to be more efficient than PH (smaller s.e.'s of estimates). This is because they assume a particular form for $\lambda_0(t)$, rather than estimating it at every death time.

For the Exponential model, the hazards are constant over time, given the value of the covariate Z_i:

$$\begin{aligned} Z_i = 0 \Rightarrow \quad \hat{\lambda}_0 &= \exp(\hat{\beta}_0) \\ Z_i = 1 \Rightarrow \quad \hat{\lambda}_0 &= \exp(\hat{\beta}_0 + \hat{\beta}_1) \end{aligned}$$

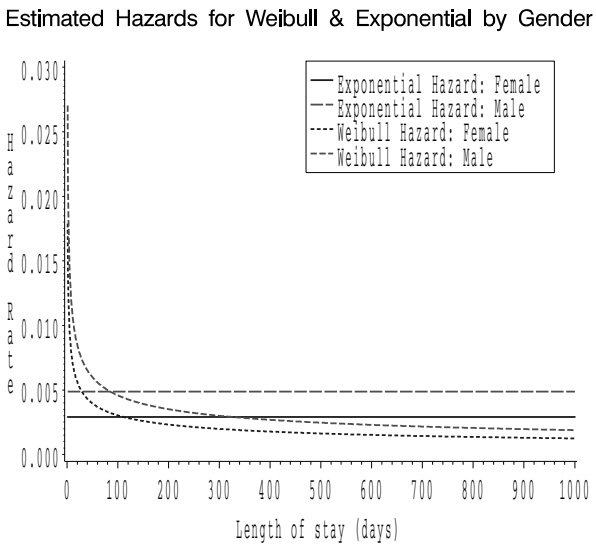
For the Weibull model, we have to estimate the hazard as a function of time, given the estimates of β_0, β_1 and κ:

$$\begin{aligned} Z_i = 0 \Rightarrow \quad \hat{\lambda}_0(t) &= \hat{\kappa}\ \exp(\hat{\beta}_0)\ t^{\hat{\kappa}-1} \\ Z_i = 1 \Rightarrow \quad \hat{\lambda}_1(t) &= \hat{\kappa}\ \exp(\hat{\beta}_0 + \hat{\beta}_1)\ t^{\hat{\kappa}-1} \end{aligned}$$

However, the ratio of the hazards is still just $\exp(\hat{\beta}_1)$, since the other terms cancel out.

Here's what the estimated hazards look like for the nursing home data:

Figure 10.5: Comparison of the Weibull and Exponential hazards



10.4.5 Proportional Hazards Model:

To get the MLE's for this model, we have to maximize the Cox partial likelihood iteratively. There are not closed form expressions like above.

$$L(\beta) = \prod_{i=1}^{n} \left[\frac{e^{\beta \mathbf{Z}_i}}{\sum_{\ell \in \mathcal{R}(X_i)} e^{\beta \mathbf{Z}_\ell}} \right]^{\delta_i}$$

$$= \prod_{i=1}^{n} \left[\frac{e^{\beta_0 + \beta_1 Z_i}}{\sum_{\ell \in \mathcal{R}(X_i)} e^{\beta_0 + \beta_1 Z_\ell}} \right]^{\delta_i}$$

Comparison with Proportional Hazards Model

For the PH model, $\hat{\beta}_1 = 0.394$ and $\widehat{HR} = e^{0.394} = 1.483$.

Comparison with the Logrank and Wilcoxon Tests

```
\begin{verbatim}
Call:
survdiff(formula = Surv(losyr, fail) ~ gender, data = nurshome)

            N Observed Expected (O-E)^2/E (O-E)^2/V
gender=0 1173      902      995      8.76      41.1
gender=1  418      367      274     31.88      41.1

 Chisq= 41.1  on 1 degrees of freedom, p= 1.46e-10
```

The Gehan-Wilcoxon test. Note that this is fit by adding `rho=1` in R:

```
Call:
survdiff(formula = Surv(losyr, fail) ~ gender, data = nurshome,
    rho = 1)

            N Observed Expected (O-E)^2/E (O-E)^2/V
gender=0 1173      529      592      6.66      41.8
gender=1  418      236      173     22.73      41.8

 Chisq= 41.8  on 1 degrees of freedom, p= 9.94e-11
```

10.4.6 Comparison of Hazard Ratios and Test Statistics

For effect of Gender

Model/Method	λ_0	λ_1	HR	log(HR)	se(log HR)	Wald Statistic
Exponential	0.0029	0.0049	1.676	0.5162	0.0619	69.507
Weibull						
$t = 50$	0.0040	0.0060	1.513	0.4138	0.0636	42.381
$t = 100$	0.0030	0.0046	1.513			
$t = 500$	0.0016	0.0025	1.513			
Logrank						41.085
Wilcoxon						41.468
Cox PH						
Ties=Breslow			1.483	0.3944	0.0621	40.327
Ties=Discrete			1.487	0.3969	0.0623	40.565
Ties=Efron			1.486	0.3958	0.0621	40.616
Ties=Exact			1.486	0.3958	0.0621	40.617
Score (Discrete)						41.085

Comparison of Mean and Median Survival Times by Gender

	Mean Survival		Median Survival	
Model/Method	Female	Male	Female	Male
Exponential	344.5	205.6	238.8	142.5
Weibull	461.6	235.4	174.2	88.8
Kaplan-Meier	318.6	200.7	144	70
Cox PH (Kalbfleisch/Prentice)			131	72