

Q1 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

$$\lambda(t;Z) = \lambda_0(t)\Psi(Z)$$

Most commonly, we write the second term as:

$$\Psi(Z) = e^{\beta Z}$$

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 $e\beta$ for treated subjects versus untreated subjects ($e\beta$ might be < 1).

This is an example of a semi-parametric model.

- Accelerated Failure Time (AFT) models

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

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

- exponential, Weibull, Gamma
- lognormal

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)

Discrete 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, \dots, U_a) such that $U_j = 1$ if $A = j$. Then

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

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

Interactions:

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

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

$$\begin{aligned} \varphi &= \frac{\lambda_1(t)}{\lambda_0(t)} \\ &= e^\beta \end{aligned}$$

φ is referred to as the hazard ratio.

What is the interpretation of β here?

5

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, \dots, Z_p equal to 0 (i.e., the "reference group").

The general form is:

$$\lambda(t; Z) = \lambda_0(t) \exp(\beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_p Z_p)$$

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

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

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

$\lambda_i(t, Z_i) = \lambda_0(t) \exp(\beta_1 Z_{1i} + \dots + \beta_p Z_{pi})$ This means we can write the log of the hazard ratio for the
 i -th individual to the reference group as:

log

$\frac{\lambda_i(t, Z_i)}{\lambda_0(t)}$

$= \log \left(\frac{\lambda_i(t, Z_i)}{\lambda_0(t)} \right)$

$\lambda_i(t)$

$\lambda_0(t)$

$\frac{\lambda_i(t, Z_i)}{\lambda_0(t)}$

$\frac{\lambda_i(t, Z_i)}{\lambda_0(t)} =$

$\exp(\beta_1 Z_{1i} +$

$\beta_2 Z_{2i} +$

\dots

$\beta_p Z_{pi})$

\cdot

$+$

$\beta_p Z_{pi}$

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*.

Q2

Kaplan-Meier estimate and pointwise bounds:

```
survfit(formula, conf.int = 0.95, conf.type = "log")
```

The Kaplan-Meier estimate is a nonparametric maximum likelihood estimate (MLE) of the survival function, $S(t)$. This estimate is a step function with jumps at observed event times, t_i . In the mathematics below, it is assumed the t_i are ordered: $0 < t_1 < t_2 < \dots < t_D$. If the number of individuals with an observed event time t_i is d_i , and the value Y_i represents the number of individuals at risk at time t_i (where at risk means individuals who die at time t_i or later), then the Kaplan-Meier estimate of the survival function and its estimated variance are given by

$$\begin{aligned} \hat{S}(t) &= \\ & \left(\prod_{t_i \leq t} \frac{Y_i - d_i}{Y_i} \right) \\ & \text{if } t_1 \leq t < t_2 \\ & \text{if } t_1 \leq t < t_2 \end{aligned}$$
$$\text{bV}[\hat{S}(t)] =$$

h
 $\hat{S}(t)$
 i_2
 \hat{S}_2
 $(t) =$
 h
 $\hat{S}(t)$
 i_2X
 t_i
 d_i
 $Y_i(Y_i \neq d_i)$

The pointwise confidence bounds for the "plain" (linear) and "log-log" options provided in R are given by

$$-\hat{S} \pm Z_{1-\alpha/2} \sqrt{\hat{S}(t) \hat{S}(t)}; \hat{S} \pm Z_{1-\alpha/2} \sqrt{\hat{S}(t) \hat{S}(t)}$$

$$-\hat{S}_1 \pm \sqrt{\hat{S}_1(t) \hat{S}_1(t)}$$

, where $\hat{S}_1 = \exp$

$$\left(\frac{\hat{S}_1(t)}{\log \hat{S}(t)} \right)$$

The Kaplan-Meier estimate is \hat{S}_t in R using the function `survfit()`. The simplest \hat{S}_t takes as input

Q3

(A) The log-normal distribution is another commonly used parametric distribution for characterizing the survival time.

$$LN(\mu, \sigma^2) = \exp fN(\mu, \sigma^2)g$$

$$E(T) = e_{-2} = 2$$

$$\text{Var}(T) = e_{-2} = 2$$

$$(e_{-2} \approx 1)$$

(B) Weibull distribution is also a generalization of the simple exponential distribution.

Be careful about the parametrization $W(\lambda, p)$, $\lambda > 0$ (scale parameter) and $p > 0$ (shape parameter):

$$1. S(t) = e^{-(\lambda t)^p}$$

$$2. f(t) = p\lambda(\lambda t)^{p-1} e^{-(\lambda t)^p} / t^{p-1} e^{-(\lambda t)^p}$$

.

$$3. h(t) = p\lambda(\lambda t)^{p-1} / t^{p-1}$$

$$4. H(t) = (\lambda t)^p.$$

$$E(T) = \lambda^{-1} \Gamma(1 + 1/p).$$

$$\text{Var}(T) = \lambda^{-2}$$

[

$$\Gamma(1 + 2/p) \approx \Gamma(1 + 1/p)^2$$

]

$$W(\lambda, 1) \sim \text{EXP}(\lambda).$$

$$W(\lambda, p) \sim \text{fEXP}(\lambda p) g_1 = p$$

Q4

Menu location: Analysis_Survival_Kaplan-Meier.

This function estimates survival rates and hazard from data that may be incomplete.

The survival rate is expressed as the survivor function (S):

- where t is a time period known as the survival time, time to failure or time to event (such as death); e.g. 5 years in the context of 5 year survival rates. Some texts present S as the estimated probability of surviving to time t for those alive just before t multiplied by the proportion of subjects surviving to t . Thus it reflects the probability of no event before t . At $t=0$ $S(t) = 1$ and decreases toward 0 as t increases toward infinity.

The product limit (PL) method of [Kaplan and Meier \(1958\)](#) is used to estimate S :

- where t_i is duration of study at point i , d_i is number of deaths up to point i and n_i is number of individuals at risk just prior to t_i . S is based upon the probability that an individual survives at the end of a time interval, on the condition that the individual was present at the start of the time interval. S is the product (P) of these conditional probabilities.

If a subject is last followed up at time t_i and then leaves the study for any reason (e.g. lost to follow up) t_i is counted as their [censorship](#) time.

Assumptions:

- [Censored](#) individuals have the same prospect of survival as those who continue to be followed. This can not be tested for and can lead to a [bias](#) that artificially reduces S .
- Survival prospects are the same for early as for late recruits to the study (can be tested for).
- The event studied (e.g. death) happens at the specified time. Late recording of the event studied will cause artificial inflation of S .

The instantaneous hazard function $h(t)$ [also known as the hazard rate, conditional failure rate or force of mortality] is defined as the event rate at time t conditional on surviving up to or beyond time t . As $h(t)$ is a rate, not a probability, it has units of $1/t$. The cumulative hazard function $H_{\text{hat}}(t)$ is the integral of the hazard rates from time 0 to t , which represents the accumulation of the hazard over time - mathematically this quantifies the number of times you would expect to see the failure event in a given time period, if the event was repeatable. So it is more accurate to think of hazards in terms of rates than probabilities. The cumulative hazard is estimated by the method of [Peterson \(1977\)](#) as:

S and H with their standard errors and confidence intervals can be saved to a workbook for further analysis

Q5

$$S(t) = \Pr(T \geq t)$$

It follows from the definition of the survivor function that

$$S(t) = 1 - F(t)$$

=

$$\int_0^t f(u) du$$

t

f(u)du:

Conditional distribution:

Given an individual lives up to time t_0 , the distribution of the future survival time is given by:

$$\Pr(\text{dies in } (t_0; t_0 + t) \mid \text{alive at } t_0) = \Pr(T < t_0 + t \mid T \geq t_0)$$

=

$$F(t_0 + t) - F(t_0)$$

$$S(t_0)$$

:

The probability density of future lifetime is the derivative of this:

d

dt

$$F(t_0 + t) - F(t_0)$$

$$S(t_0)$$

=

$$f(t + t_0)$$

$$S(t_0)$$

:

6

Hazard function:

The hazard function at t_0 , denoted $h(t_0)$, is the instantaneous rate of death, i.e. the following

limit:

$$h(t_0) = \lim_{t \rightarrow 0} \frac{F(t_0 + t) - F(t_0)}{t S(t_0)}$$

$\Delta t \rightarrow 0$

1

Δt

Pr(die in interval $[t_0; t_0 + \Delta t]$ | alive at time t_0)

The hazard function is related to the survivor function in the following way:

$h(t_0) = \lim$

$\Delta t \rightarrow 0$

1

Δt

Pr(die in interval $[t_0; t_0 + \Delta t]$ | alive at time t_0)

$= \lim$

$\Delta t \rightarrow 0$

1

Δt

$F(t_0 + \Delta t) - F(t_0)$

$S(t_0)$

$=$

1

$S(t_0)$

\lim

$\Delta t \rightarrow 0$

$F(t_0 + \Delta t) - F(t_0)$

Δt

$=$

1

$S(t_0)$

d

dt

$F(t_0)$

$$= \frac{f(t_0)}{S(t_0)}$$

So we have $h(t) = f(t)/S(t)$.

The hazard function gives the following linear approximation

$\Pr(\text{dies in } (t_0; t_0 + \Delta t) \mid \text{alive at } t_0) \approx \Delta t \cdot h(t_0)$:

7

Cumulative hazard:

$$H(t) =$$

\int_0^t

$$h(u) du$$

We have written the hazard function in terms of the survivor function, so now we do the converse

{ and the cumulative hazard comes in handy. Since $S(t) = 1 - F(t)$ it follows that:

$\frac{d}{dt}$

$\log[S(t)] =$

$$-\int_0^t h(u) du$$

1

$$S(t)$$

—

$\frac{d}{dt}$

$\log S(t)$

$$= -\int_0^t h(u) du$$

$f(t)$

$$S(t)$$

$$= -\int_0^t h(u) du$$

$= -\int_0^t h(u) du$:

Integrating gives

$$\log S(t) = -H(t)$$

so

$$S(t) = \exp[-H(t)]$$

Sometimes it's very useful to assume the hazard (or survival) functions have specific forms. Of course, this is equivalent to assuming a specific form for the underlying probability distribution F of the survival time T . We look next at the simplest assumptions we might make.

Exponential Distribution

Suppose that the hazard function is constant: $h(t) = \lambda$. It follows that:

$$H(t) = \lambda t$$

$$S(t) = \exp(-\lambda t)$$

$$f(t) = \lambda \exp(-\lambda t):$$

The probability density function is that for an exponential random variable. It has the 'lack of memory property':

$$\Pr(T > t_1 + t_2 \mid T > t_1) =$$

$$\exp(-\lambda(t_1 + t_2))$$

$$\exp(-\lambda t_1)$$

$$= \exp(-\lambda t_2) = \Pr(T > t_2):$$

Weibull Distribution

More usefully, we would like the hazard function to vary with time: the Weibull distribution is the simplest such example. Suppose that $h(t) = ctk$ for some constants c and $k > 1$.

$$H(t) = c$$

$$tk^{k+1}$$

$$k + 1$$

;

$$S(t) = \exp$$

—

$$tk^{k+1}$$

$$c$$

$$k + 1$$

$$tk^{k+1}$$

$$\begin{aligned}
 & - \\
 & ; \\
 & \text{so} \\
 & f(t) = \\
 & d \\
 & dt \\
 & F(t) = \int_0^t f(u) du \\
 & d \\
 & dt \\
 & S(t) \\
 & = c t^k \exp(-c t^{k+1}) \\
 & - \\
 & \int_0^t \\
 & c \\
 & k + 1 \\
 & t^{k+1} \\
 & - \\
 & :
 \end{aligned}$$

This is the Weibull distribution. It's more usually written with a different parameterization in the following way. Let $k = k + 1$ and $c = c/(k + 1)$, ($c > 0$; $k > 0$). Then:

$$\begin{aligned}
 h(t) &= c t^{k-1} \\
 H(t) &= c t^k \\
 S(t) &= \exp(-c t^k) \\
 f(t) &= c t^{k-1} \exp(-c t^k)
 \end{aligned}$$

Suppose T is a Weibull random variable with parameters c, k . It can be shown that the expectation of T is

$$E(T) = \frac{1}{c} \Gamma(1 + \frac{1}{k})$$

Here Γ is the gamma function:

$$\Gamma(x) =$$

z 1

0

ux 1e 2udu:

Since the distribution is skewed, the median is different from the mean. The median t_m is computed as follows:

$$F(t_m) = 1/2$$

$$\exp(-t_m^\alpha)$$

$$m) = 1/2$$

$$t_m =$$

—

$$\log 2$$

—

$$_1 =$$

:

The 100p-th percentile can be computed in a similar way.

Warning! R uses different parameterizations of the Weibull distribution. If you ever fit a Weibull model in R, make sure you know which parameterization is being used. In the R parameterisation the pdf is

$$f(t) =$$

—

—

t

—

$$_1$$

$$\exp$$

—

?

—

t

—

—

so

$\mu = \mu_0$ and $\mu = \mu_1$;

the mean is

$E(T) = \mu_0(1 + \mu_1)$

and the median is

$t_m = \mu_0(\log 2)$;

Q6

Mortality from prostate cancer increased with tumor stage and grade. For men with stage T1 or T2, low or moderate grade tumors (59.1% of all cases), mortality from prostate cancer was 2.12%, versus a 6.40% mortality rate from heart disease and 3.83% mortality from other cancers.

Q7

—

$z >$

$c - m$

$\sqrt{s_2 + .2}$

—

$\rightarrow pr$

—

$z > -m$

s

(5)

Hence,

equation (5) = $1 - \{\text{one-sided Phase II p-value}\}$. Consequently, if the Phase III is very large, the chance of success by assurance is maximally 1 minus the one-sided

p-value from Phase II. For example, a Phase II with $p = .0020$ one-sided means that assurance in Phase III cannot exceed 80% even with

infinitely many patients entered. More generally, as $.2 \rightarrow 0$ then $c \rightarrow 0$ and the conventional

power curve approaches a step function. Then any observed difference > 0 rejects the null with 100% probability, and any difference ≤ 0 rejects with probability zero. Therefore, for any prior f , assurance

$R_4 \text{ power } .5f_4.5d. \rightarrow$

+

$of_4.5d.$, which is the fraction of the prior density that lies to the right of zero.

(iv) If Phase III is k Times Larger than Phase II More commonly, $.2 > 0$

Commented [M1]:

Commented [M2R1]:

Commented [M3R1]:

and $s_2 > 0$ with $.2 \leq s_2$. In this circumstance, if the size of the Phase III is k times

(A) exponential approximation, using the survival probability at 36 months. First find the hazard estimate $\hat{\lambda}_0$.

(B) We observe that the vector $(B(t_1), \dots, B(t_n))$ has a multivariate normal distribution because

the event

$$\{B(t_1) = x_1, \dots, B(t_n) = x_n\}$$

can be re-written in terms of independent increment events

$$(t_1) = x_1, B(t_2) - B(t_1) = x_2 - x_1, \dots, B(t_n) - B(t_{n-1}) = x_n - x_{n-1},$$

yielding the joint density of $(B(t_1), \dots, B(t_n))$

$$f(x_1, \dots, x_n) = f_{t_1}(x_1) f_{t_2-t_1}(x_2 - x_1) \cdots f_{t_n-t_{n-1}}(x_n),$$

(C) is the risk ratio that we need to detect with 90% power with a two-sided 5% level test, find the needed number of deaths d . (That is, in the formula, use $\alpha/2 = 0.025$.)

(D) In order to construct an error bar for $\hat{S}(t)$ we need to make some sort of distributional assumption.

The simplest is to assume that $\hat{S}(t)$ is normally distributed. Let $z_{\alpha/2}$ be such that

$\Pr(Z > z_{\alpha/2}) = \alpha/2$; where $Z \sim N(0, 1)$; then an approximate $100(1 - \alpha)\%$ interval is given by

$$\hat{S}(t) \pm z_{\alpha/2} s[\hat{S}(t)]$$

where $s[\hat{S}(t)]$ is the standard error.

A snag with this, of course is that the distribution of $\hat{S}(t)$ is not really normal. Problems arise when $\hat{S}(t)$ is close to 0 or 1. One possible solution is to transform $\hat{S}(t)$ onto a $(0, 1)$ scale.

(E) statistical considerations we have r groups of individuals, with $r > 2$. We're interested in testing whether all the

groups are the same ie. the null hypothesis is that all the groups have the same survival distribution.

Our approach and notation is the same as previously: we pool all the death times together to define

intervals $[0; t_1)$, $[t_1; t_2)$, etc. We then have d_{kj} deaths in group k on interval j and n_{kj} individuals alive and uncensored from group k at the start of interval j .

We can extend the simplified form of the log-rank test quite readily.