7 Parametric Survival Models

D.G. Kleinbaum and M. Klein, *Survival Analysis: A Self-Learning Text, Third Edition,*
Statistics for Biology and Health, DOI 10.1007/978-1-4419-6646-9_7,

Introduction

The Cox model is the most widely used survival model in the health sciences, but it is not the only model available. In this chapter we present a class of survival models, called parametric models, in which the distribution of the outcome (i.e., the time to event) is specified in terms of unknown parameters. Many parametric models are acceleration failure time models in which survival time is modeled as a function of predictor variables. We examine the assumptions that underlie accelerated failure time models and compare the acceleration factor as an alternative measure of association to the hazard ratio. We present examples of the exponential, Weibull, and log-logistic models and give a brief description of other parametric approaches. The parametric likelihood is constructed and described in relation to left, right, and interval-censored data. Binary regression is presented as an alternative approach for modeling interval-censored outcomes. The chapter concludes with a discussion of frailty models.

Abbreviated Outline

The outline below gives the user a preview of the material covered by the presentation. A detailed outline for review purposes follows the presentation.

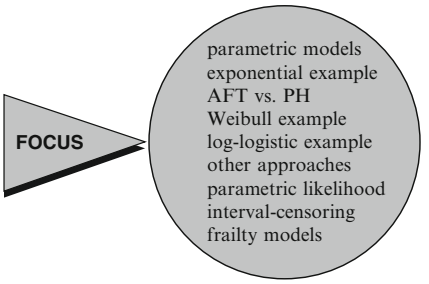
Objectives

Upon completing this chapter, the learner should be able to:

1. State or recognize the form of a parametric survival model and contrast it with a Cox model.
2. State common distributions used for parametric survival models.
3. Contrast an AFT model with a PH model.
4. Interpret output from an exponential survival model.
5. Interpret output from a Weibull survival model.
6. Interpret output from a log-logistic survival model.
7. State or recognize the formulation of a parametric likelihood.
8. State or recognize right-censored, left-censored, and interval-censored data.
9. State or recognize the form of a frailty model and the purpose of including a frailty component.
10. Interpret the output obtained from a frailty model.

Presentation

I. Overview



In this chapter we present parametric survival models and the assumptions that underlie these models. Specifically we examine the accelerated failure time (AFT) assumption and contrast it to the proportional hazards (PH) assumption. We present examples of several parametric models, including the exponential model, the Weibull model, and the log-logistic model. The parametric likelihood is discussed and how it accommodates left-, right-, and interval-censored data. We also consider models that include a frailty component to account for unobserved heterogeneity.

Parametric Modeling

- Outcome assumed to follow some family of distributions
- Exact distribution is unknown if parameters are unknown
- Data used to estimate parameters
- Examples of parametric models:
  - Linear regression
  - Logistic regression
  - Poisson regression

Linear regression, logistic regression, and Poisson regression are examples of parametric models that are commonly used in the health sciences. With these models, the **outcome is assumed to follow some distribution** such as the normal, binomial, or Poisson distribution. Typically, what is actually meant is that the outcome follows some family of distributions of similar form with unknown parameters. It is only when the value of the parameter(s) is known that the exact distribution is fully specified. For example, if one distribution is normal with a mean of 3 and another distribution is normal with a mean of 7, the distributions are of the same family (i.e., normal) but they are not exactly the same distribution. For parametric regression models, the data are typically used to estimate the values of the parameters that fully specify that distribution.

Distributions commonly used for parametric survival models:

- Weibull
- Exponential
- Log-logistic
- Lognormal
- Generalized gamma

A **parametric survival model** is one in which survival time (the outcome) is assumed to follow a known distribution. Examples of distributions that are commonly used for survival time are: the **Weibull**, the **exponential** (a special case of the Weibull), the **log-logistic**, the **lognormal**, and the **generalized gamma**, which are supported by SAS, Stata, and R software.

Parametric survival models

↘

Distribution specified for time

Cox model is semiparametric:

↘

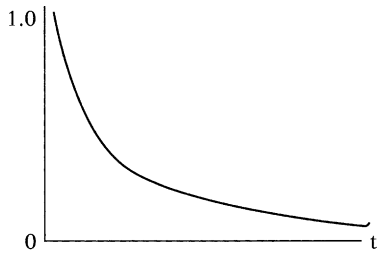
Baseline survival not specified

The Cox proportional hazards model, by contrast, is not a fully parametric model. Rather it is a **semi-parametric model** because even if the regression parameters (the betas) are known, the distribution of the outcome remains unknown. The baseline survival (or hazard) function is not specified in a Cox model.

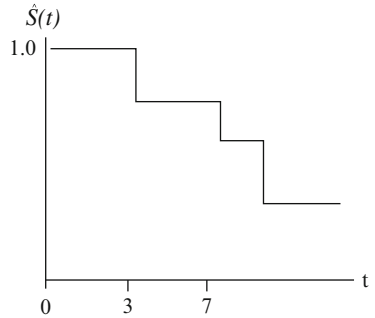
Cox model widely popular:

- No reliance on assumed distribution
- Computer packages can output Cox-adjusted survival estimates using algorithm that generalizes KM
- Baseline not necessary for estimation of hazard ratio

A key reason why the Cox model is widely popular is that it does not rely on distributional assumptions for the outcome. Although the baseline survival function is not estimated with a Cox model, computer packages such as SAS, Stata, SPSS, and R can output Cox-adjusted survival estimates (see Computer Appendix) by using a complicated algorithm that generalizes the Kaplan–Meier (KM) approach while making use of estimated regression coefficients obtained from a Cox model (Kalbfleisch and Prentice, 1980). Also, an estimation of the baseline hazard is not necessary for the estimation of a hazard ratio because the baseline hazard cancels in the calculation.



Theoretical $S(t)$



Step function (nondistributional estimates)

In theory, as time ranges from 0 to infinity, the survival function can be graphed as a smooth curve from $S(0) = 1$ to $S(\infty) = 0$ (see Chapter 1). Kaplan–Meier and Cox-adjusted survival estimates use empirical nondistributional methods that typically graph as step functions, particularly if the sample size is small. If in the data, for example, an event occurred at 3 weeks and the next event occurred at 7 weeks, then the estimated survival curve would be flat between 3 and 7 weeks using these nondistributional approaches. Moreover, if the study ends with subjects still remaining at risk, then the estimated survival function would not go all the way down to zero.

Appeal of Parametric Survival Models

- More consistent with theoretical S(t) than nondistributional approaches
- Simplicity
- Completeness — h(t) and S(t) specified

Survival estimates obtained from parametric survival models typically yield plots more consistent with a theoretical survival curve. If the investigator is comfortable with the underlying distributional assumption, then parameters can be estimated that completely specify the survival and hazard functions. This simplicity and completeness are the main appeals of using a parametric approach.

II. Probability Density Function in Relation to the Hazard and Survival Function

Probability density function known
⇓
Survival and hazard functions

For parametric survival models, time is assumed to follow some distribution whose probability density function f(t) can be expressed in terms of unknown parameters. Once a probability density function is specified for survival time, the corresponding survival and hazard functions can be determined. The survival function $S(t) = P(T > t)$ can be ascertained from the probability density function by integrating over the probability density function from time t to infinity. The hazard can then be found by dividing the negative derivative of the survival function by the survival function (see left).

$f(t) = dF(t)/dt$ where
$F(t) = Pr(T \leq t)$

$$S(t) = P(T > t) = \int_t^\infty f(u)du$$

$$h(t) = \frac{-d[S(t)]/dt}{S(t)}$$

Survival in terms of hazard

$$S(t) = \exp\left(-\int_0^t h(u)du\right)$$

The survival function can also be expressed in terms of the hazard function (see Chapter 1) by exponentiating the negative of the cumulative hazard function. The cumulative hazard function is the integral of the hazard function between integration limits of 0 and t.

Cumulative hazard: $\int_0^t h(u)du$

$f(t) = h(t)S(t)$

Finally, the probability density function can be expressed as the product of the hazard and the survival functions, $f(t) = h(t)S(t)$.

Key Point

Specifying one of f(t), S(t), or h(t) specifies all three functions

The key point is that specifying **any one of the probability density function, survival function, or hazard function allows the other two functions to be ascertained** by using the formulas shown on the left.

Survival and Hazard Functions for Selected Distributions

Distribution	$S(t)$	$h(t)$
Exponential	$\exp(-\lambda t)$	λ
Weibull	$\exp(-\lambda t^p)$	$\lambda p t^{p-1}$
Log-logistic	$\frac{1}{1+\lambda t^p}$	$\frac{\lambda p t^{p-1}}{1+\lambda t^p}$

On the left is a table containing the survival and hazard functions for three of the more commonly used distributions for survival models: the exponential, Weibull, and log-logistic distributions.

The exponential is a one-parameter distribution with a constant hazard λ. The Weibull and log-logistic distributions have two parameters λ and p. Notice that the Weibull distribution reduces to the exponential if $p = 1$. The probability density function for these distributions can be found by multiplying h(t) and S(t). As an example, the Weibull probability density function is shown on the left.

$f(t) = h(t)S(t)$

For example, Weibull:
$f(t) = \lambda p t^{p-1} \exp(-\lambda t^p)$
because $h(t) = \lambda p t^{p-1}$ and
$S(t) = \exp(-\lambda t^p)$

Typically in parametric models:

- λ reparameterized for regression
- p held fixed

Typically for parametric survival models, the parameter λ is reparameterized in terms of predictor variables and regression parameters and the parameter p (sometimes called the shape parameter) is held fixed. This is illustrated in the examples to come.

III. Exponential Example

Simplest parametric survival model:
Hazard function: $h(t) = \lambda$
(where λ is a constant)

EXAMPLE

Remission data (n = 42)

21 patients given treatment (TRT = 1)
21 patients given placebo (TRT = 0)

The first example we consider is the exponential model, which is the simplest parametric survival model in that the hazard is constant over time (i.e., $h(t) = \lambda$). The model is applied to the remission data (Freireich et al., 1963), in which 42 leukemia patients were followed until remission or censorship. Twenty-one patients received an experimental treatment (coded TRT = 1) and the other 21 received a placebo (coded TRT = 0). The data are listed in Chapter 1. The variable TRT is just a reverse coding of the variable RX presented in Chapter 3.

$h(t) = \lambda = \exp(\beta_0 + \beta_1 \text{ TRT})$

$\text{TRT} = 1$: $h(t) = \exp(\beta_0 + \beta_1)$
$\text{TRT} = 0$: $h(t) = \exp(\beta_0)$

$\text{HR}(\text{TRT} = 1 \text{ vs. } \text{TRT} = 0)$

$$= \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1)$$

For simplicity, we demonstrate an exponential model that has TRT as the only predictor. We state the model in terms of the hazard by reparameterizing λ as $\exp(\beta_0 + \beta_1 \text{TRT})$. With this model, the hazard for subjects in the treated group is $\exp(\beta_0 + \beta_1)$ and the hazard for the placebo group is $\exp(\beta_0)$. The hazard ratio comparing the treatment and placebo (see left side) is the ratio of the hazards $\exp(\beta_1)$. The exponential model is a proportional hazards model.

Constant Hazards
$\Rightarrow$ Proportional Hazards

Proportional Hazards
$\not\Rightarrow$ Constant Hazards

Exponential Model — Hazards are constant

Cox PH Model — Hazards are proportional not necessarily constant

The assumption that the hazard is constant for each pattern of covariates is a **much stronger assumption** than the PH assumption. If the hazards are constant, then of course the ratio of the hazards is constant. However, the **hazard ratio being constant does not necessarily mean that each hazard is constant**. In a Cox PH model the baseline hazard is not assumed constant. In fact, the form of the baseline hazard is not even specified.

Remission Data

Exponential regression
log hazard form

_t	Coef.	Std. Err.	z	P>|z|
trt	−1.527	.398	−3.83	0.00
_cons	−2.159	.218	−9.90	0.00

Output from running the exponential model is shown on the left. The model was run using Stata software. The parameter estimates are listed under the column called Coef. The parameter estimate for the coefficient of TRT (β_1) is −1.527. The estimate of the intercept (called cons in the output) is −2.159. The standard errors (Std. Err.), Wald test statistics (z), and p-values for the Wald test are also provided. The output indicates that the z test statistic for TRT is statistically significant with a p-value <0.005 (rounds to 0.00 in the output).

Coefficient estimates obtained by MLE

asymptotically normal

The regression coefficients are estimated **using maximum likelihood estimation (MLE)**, and are **asymptotically normally** distributed.

$\text{TRT} = 1$: $\hat{h}(t) = \exp(-2.159 + (-1.527)) = 0.025$
$\text{TRT} = 0$: $\hat{h}(t) = \exp(-2.159) = 0.115$

$\widehat{HR}\,(\text{TRT} = 1 \text{ vs } .0) = \exp(-1.527) = 0.22$
$95\%\text{CI} = \exp[-1.527 \pm 1.96(0.398)] = (0.10, 0.47)$

Results: suggest treatment lowers hazard

The estimated hazards for TRT = 1 and TRT = 0 are shown on the left. The estimated hazard ratio of 0.22 is obtained by exponentiating the estimated coefficient (−1.527) of the TRT variable. A 95% confidence interval can be calculated exp[−1.527 ± 1.96(0.398)] yielding a CI of (0.10, 0.47). These results suggest that the experimental treatment delays remission.

Parametric models

- Need not be PH models
- Many are AFT models

Exponential and Weibull

- Accommodate PH and AFT assumptions

Up to this point in the book, the key assumption for survival models has been the proportional hazard assumption. However, parametric survival models need not be PH models. **Many parametric models are acceleration failure time (AFT) models rather than PH models**. The exponential and Weibull distributions can accommodate both the PH and AFT assumptions.

Remission Data

Exponential regression accelerated failure-time form

_t	Coef.	Std. Err.	z	p > |z|
trt	1.527	.398	3.83	0.00
_cons	2.159	.218	9.90	0.00

On the left is Stata output from the AFT form of the exponential model with TRT as the only predictor. Stata can output both the PH or AFT form of an exponential or Weibull model (see Computer Appendix). SAS only runs the AFT form of parametric models and SPSS does not yet provide commands to run parametric models.

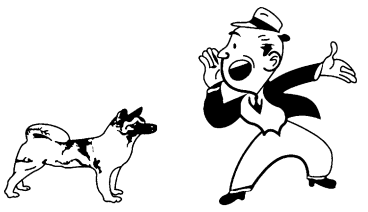
AFT vs. *PH*

- Different interpretation of parameters
- AFT applies to comparison of survival times
- PH applies to comparison of hazards

The interpretation of parameters differs for AFT and PH models. The AFT assumption is applicable for a comparison of survival times whereas the PH assumption is applicable for a comparison of hazards. In the following sections, we discuss the AFT assumption and then revisit this example and discuss the AFT form of this model.

IV. Accelerated Failure Time Assumption

AFT — Multiplicative effect with survival time
PH — Multiplicative effect with hazard



AFT models:
Describe "stretching out" or contraction of survival time

The underlying assumption for AFT models is that the effect of covariates is **multiplicative (proportional) with respect to survival time**, whereas for PH models the underlying assumption is that the effect of covariates is **multiplicative with respect to the hazard**.

To illustrate the idea underlying the AFT assumption, consider the lifespan of dogs. It is often said that dogs grow older seven times faster than humans. So a 10-year-old dog is in some way equivalent to a 70-year-old human. In AFT terminology we might say the probability of a dog surviving past 10 years equals the probability of a human surviving past 70 years. Similarly, we might say the probability of a dog surviving past 6 years equals the probability of a human surviving past 42 years because 42 equals 6 times 7. More generally we can say SD(t) = SH(7t), where SD(t) and SH(t) are the survival functions for dogs and humans, respectively. In this framework dogs can be viewed, on average, as accelerating through life 7 times faster than humans. Or from the other perspective, the lifespan of humans, on average, is stretched out 7 times longer than the lifespan of dogs. **AFT models describe this "stretching out" or contraction of survival time as a function of predictor variables**.

Second Illustration

S_1 (t) — Survival function for smokers
S_2 (t) — Survival function for nonsmokers

AFT assumption:
$S_2(t) = S_1(\gamma t)$ for $t \geq 0$
γ is the acceleration factor

If $\gamma = \exp(\beta)$
$S_2(t) = S_1([\exp(\alpha)]t)$
or
$S_2([\exp(-\alpha)]t) = S_1(t)$

For a second illustration of the accelerated failure time assumption consider a comparison of survival functions among smokers $S_1(t)$ and non-smokers $S_2(t)$. The AFT assumption can be expressed as $S_2(t) = S_1(\gamma t)$ for $t \geq 0$, where γ is a constant called the **acceleration factor** comparing smokers to nonsmokers. In a regression framework the acceleration factor γ could be parameterized as $\exp(\alpha)$ where α is a parameter to be estimated from the data. With this parameterization, the AFT assumption can be expressed as $S_2(t) = S_1(\exp(\alpha)t)$ or equivalently: $S_2(\exp(-\alpha)t) = S_1(t)$ for $t \geq 0$.

Suppose $\exp(\alpha) = 0.75$
then
$S_2(80) = S_1(60)$
$S_2(40) = S_1(30)$

More generally
$S_2(t) = S_1(0.75t)$

Suppose $\exp(\alpha) = 0.75$; then the probability of a nonsmoker surviving 80 years equals the probability of a smoker surviving 80(0.75) or 60 years. Similarly, the probability of a nonsmoker surviving 40 years equals the probability of a smoker surviving 30 years. More generally, the probability of a nonsmoker surviving t years equals the probability of a smoker surviving 0.75 times t years (i.e., $S_2(t) = S_1(0.75t)$).

T_1 — Survival time for smokers
T_2 — Survival time for nonsmokers

AFT assumption in terms of random variables:
$T_1 = \gamma T_2$

The AFT assumption can also be expressed in terms of random variables for survival time rather than the survival function. If T_2 is a random variable (following some distribution) representing the survival time for nonsmokers and T_1 is a random variable representing the survival time for smokers, then the AFT assumption can be expressed as $T_1 = \gamma\, T_2$.

Acceleration factor
Measure of association on survival time

Hazard ratio
Measure of association on the hazard

The **acceleration factor** is the key measure of association obtained in an AFT model. It allows the investigator to evaluate the effect of predictor variables on survival time just as the hazard ratio allows the evaluation of predictor variables on the hazard.

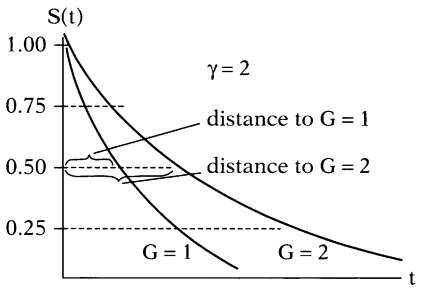
Acceleration factor (γ)

- Describes stretching or contraction of S(t)
- Ratio of times to any fixed value of S(t)

Suppose $\gamma = 2.0$
(Group 2 vs. Group 1)

- Time to $S(t) = 0.50$ (median) is double for Group 2
- Time to $S(t) = 0.20$ is double for Group 2
- Time to $S(t) = 0.83$ is double for Group 2
- Time to $S(t) = 0.98$ is double for Group 2
- Time to $S(t) =$ **q** is double for Group 2 (**generalization**)

The acceleration factor describes the "stretching out" or contraction of survival functions when comparing one group to another. More precisely, the **acceleration factor is a ratio of survival times corresponding to any fixed value of S(t)**. For example, if the acceleration factor comparing subjects in Group 2 vs. Group 1 is $\gamma = 2.0$, then the median survival time (value of t when $S(t) = 0.5$) for Group 2 is double the median survival time for Group 1. Moreover, the time it takes for S(t) to equal 0.2 or 0.83 or 0.98 is double for Group 2 compared to Group 1 for the same value of S(t). In general, the acceleration factor is a ratio of survival times corresponding to any quantile of survival time ($S(t) = q$).



Survival curves for Group 1 (G = 1) and Group 2 (G = 2)

Horizontal lines are twice as long to G = 2 compared to G = 1 because $\gamma = 2$

This idea is graphically illustrated by examining the survival curves for Group 1 (G = 1) and Group 2 (G = 2) shown on the left. For any fixed value of S(t), the distance of the horizontal line from the S(t) axis to the survival curve for G = 2 is double the distance to the survival curve for G = 1. Notice the median survival time (as well as the 25th and 75th percentiles) is double for G = 2. **For AFT models, this ratio of survival times is assumed constant for all fixed values of S(t)**.

V. Exponential Example Revisited

Remission data (n = 42)

21 patients given treatment (TRT = 1)
21 patients given placebo (TRT = 0)

We return to the exponential example applied to the remission data with treatment status (TRT) as the only predictor. In Section III, results from the PH form of the exponential model were discussed. In this section we discuss the AFT form of the model.

Previously discussed PH form of model
Now discuss AFT form of model

Exponential survival and hazard functions:

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

The exponential survival and hazard functions are shown on the left. Recall that the exponential hazard is constant and can be reparameterized as a PH model, $h(t) = \lambda = \exp(\beta_0 + \beta_1 \text{TRT})$. In this section we show how S(t) can be reparameterized as an AFT model.

Recall for PH model:

$h(t) = \lambda = \exp(\beta_0 + \beta_1 \text{TRT})$

AFT assumption
(comparing 2 levels of TRT)

- Ratio of times is constant for all fixed S(t)

Strategy for developing the model:

- Solve for t in terms of S(t)
- Scale t in terms of the predictors

The underlying AFT assumption, for comparing the two levels of the TRT covariate, is that the ratio of times to any fixed value of S(t) = q is constant for any probability q. We develop the AFT model with the survival function by solving for t in terms of S(t). We then scale t in terms of the predictors.

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

$$t = [-\ln(S(t)] \times \frac{1}{\lambda}$$

$$\text{let } \frac{1}{\lambda} = \exp(\alpha_0 + \alpha_1 TRT)$$

$$t = [-\ln(S(t)] \times \mathbf{exp}(\alpha_0 + \alpha_1 \mathbf{TRT})$$

Scaling of t

Median survival time, S(t) = 0.5:

$$t_m = [-\ln(0.5)] \times \exp(\alpha_0 + \alpha_1 TRT)$$

The exponential survival function is $S(t) = \exp(-\lambda t)$. By solving for t, we can obtain a formula for t in terms of S(t). Taking the natural log, multiplying by negative 1, and then multiplying by the reciprocal of λ, yields the expression for t shown on the left. By reparameterizing $1/\lambda = \exp(\alpha_0 + \alpha_1 TRT)$, or equivalently $\lambda = \exp[-(\alpha_0 + \alpha_1 TRT)]$, it can be seen how the predictor variable TRT is used to scale the time to any fixed value of S(t) (see left). For example, to find an expression for the median survival time t_m, substitute S(t) = 0.5 (see left).

Let S(t) = q

$$t = [-\ln(q)] \times \exp(\alpha_0 + \alpha_1 TRT)$$

Acceleration Factor:
γ(TRT = 1 vs. TRT = 0)

$$\gamma = \frac{[-\ln(q)]\exp(\alpha_0 + \alpha_1)}{[-\ln(q)]\exp(\alpha_0)}$$

$$= \exp(\alpha_1)$$

$$S_{TRT=0}(t) = S_{TRT=1}(\gamma t)$$

where $\gamma = \exp(\alpha_1)$

$$\text{AFT} : \frac{1}{\lambda} = \exp(\alpha_0 + \alpha_1 TRT)$$

$$\text{PH} : \lambda = \exp(\beta_0 + \beta_1 TRT)$$

The expression for t is restated on the left in terms of any fixed probability S(t) = q. The acceleration factor γ is found by taking the ratio of the times to S(t) = q for TRT = 1 and TRT = 0. After canceling, γ reduces to $\exp(\alpha_1)$.

Thus, the exponential hazard function for the remission data involving a single (0,1) TRT predictor satisfies the AFT model assumption shown at the left.

The formula for λ as the AFT model form is compared at the left with the (different) formula for λ in the PH model form.

Remission Data

Exponential regression accelerated failure-time form

_t	Coef.	Std. Err.	z	P>|z|
trt	1.527	.398	3.83	0.00
_cons	2.159	.218	9.90	0.00

$\hat{\gamma} = \exp(1.527) = 4.60$

95% CI: $\exp[1.527 \pm 1.96(0.398)] = (2.11, 10.05)$

On the left is Stata output from the AFT form of the exponential model with TRT as the only predictor. The estimate of the coefficient for TRT is 1.527 with a standard error of 0.398. An estimate of the acceleration factor for treatment is $\hat{\gamma} = \exp(1.527) = 4.60$. A 95% confidence interval for γ is calculated as $\exp[1.527 \pm 1.96(0.398)]$ yielding a CI of (2.11, 10.05).

$t = [-\ln(q)] \times \exp(\alpha_0 + \alpha_1 TRT)$

$\hat{t} = [-\ln(q)] \times \exp(2.159 + 1.527(TRT))$

Estimated Survival Times by S(t) Quartiles for TRT = 1 and TRT = 0 (Exponential Model)

S(t) = q	$\hat{t}_{TRT=0}$	$\hat{t}_{TRT=1}$
0.25	12.0	55.3
0.50	6.0	27.6
0.75	2.5	11.5

The parameter estimates can be used to estimate the time $\hat{t}$ to any value of S(t) = q. The table on the left lists the estimated time (in weeks) for the first, second (median), and third quartiles of S(t) using the expression for $\hat{t}$ shown above for both the treated and placebo groups. In this example, survival time is the time to remission for leukemia patients.

$\hat{\gamma} = 4.60$ (for TRT = 1 vs. TRT = 0)

Ratio of survival times:

$$\frac{55.3}{12.0} = \frac{27.6}{6.0} = \frac{11.5}{2.5} = 4.60$$

The ratio of survival times for each row in the table comparing TRT = 1 vs. TRT = 0 is 4.60, which not coincidently is the estimate of the acceleration factor (see left). The estimated acceleration factor suggests that the experimental treatment is effective for delaying remission by stretching survival time by a factor of 4.60. Although the hazard ratio is a more familiar measure of association for health scientists, the acceleration factor has an intuitive appeal, particularly for describing the efficacy of a treatment on survival.

Effect of treatment:

- Stretches survival by a factor of 4.6
- Interpretation of γ has intuitive appeal

HR and γ are reciprocals in exponential models:

$$\widehat{HR}(\text{TRT} = 1 \text{ vs. } 0) = \exp(-1.527) = 0.22$$

$$\hat{\gamma}(\text{TRT} = 1 \text{ vs. } 0) = \exp(1.527) = 4.60$$

In general

$\gamma > 1 \Rightarrow$ exposure benefits survival
HR $> 1 \Rightarrow$ exposure harmful to survival

$\gamma > 1 \Rightarrow$ exposure harmful to survival
HR $< 1 \Rightarrow$ exposure benefits survival

$\gamma = \text{HR} = 1 \Rightarrow$ no effect from exposure

Recall from Section III that the hazard ratio for the effect of treatment was estimated at $\exp(-1.527) = 0.22$ using the PH form of the exponential model. This result illustrates a key property of the exponential model: the corresponding acceleration factor and hazards ratio (e.g., TRT = 1 vs. TRT = 0) are reciprocals of each other. This property is unique to the exponential model. What can be generalized, however, is that **an acceleration factor greater than one for the effect of an exposure implies that being exposed (i.e., TRT = 1) is beneficial to survival whereas a hazard ratio greater than one implies being exposed is harmful to survival (and vice versa)**.

Exponential PH and AFT models:

- Same model
- Different parameterization
- Same estimates for
  - Survival function
  - Hazard function
  - Median survival

Although the exponential PH and AFT models focus on different underlying assumptions, they are in fact **the same model**. The only difference is in their parameterization. The resulting estimates for the survival function, hazard function, and median survival do not differ between these models (see Practice Exercises 6 and 7).

For those experienced with Poisson regression:

Exponential and Poisson models

- Assume a constant rate
- Different data structure
  - Poisson — aggregate counts
  - Exponential — individual level
- Use different outcomes
  - Poisson — number of cases
  - Exponential — time to event
- Yield equivalent parameter estimates
  - With same data and same covariates in the model

For those who have experience with Poisson regression, there is a close connection between the exponential and Poisson models. Both distributions assume an underlying constant rate. In fact, if the data are structured such that all the cases and the total time at risk are aggregated for each pattern of covariates (e.g., TRT = 1 and TRT = 0) and the log of the corresponding person-time at risk is used as an offset, then a Poisson model will yield equivalent parameter estimates as the exponential PH model. The difference is that the random outcome for the Poisson model is the count of events over a fixed amount of time at risk and the random outcome for the exponential model is the time (at risk) to event.

Exponential model is special case of Weibull model

We continue with the remission data example and present the more general Weibull model, which includes the exponential model as a special case. In the next section we also show a graphical approach for evaluating the appropriateness of the Weibull (and thus also the exponential) model.

VI. Weibull Example

Weibull Model:

Hazard function: $h(t) = \lambda p t^{p-1}$ (where $p > 0$ and $\lambda > 0$)

p is a shape parameter

- $p > 1$ hazard increases over time
- $p = 1$ constant hazard (exponential model)
- $p < 1$ hazard decreases over time

Additional shape parameter offers greater flexibility

The **Weibull model** is the most widely used parametric survival model. Its hazard function is $h(t) = \lambda p t^{p-1}$, where p and $\lambda > 0$. As with the exponential model, λ will be reparameterized with regression coefficients. The additional parameter p is called a **shape parameter** and determines the shape of the hazard function. If $p > 1$ then the hazard increases as time increases. If $p = 1$ then the hazard is constant and the Weibull model reduces to the exponential model ($h(t) = \lambda$). If $p < 1$ then the hazard decreases over time. The addition of this shape parameter gives the Weibull model greater flexibility than the exponential model, yet the hazard function remains relatively simple (basically a scaling of t raised to some fixed power).

Unique property for Weibull model
AFT ⇒ PH and PH ⇒ AFT
Holds if p is fixed

The Weibull model has the property that **if the AFT assumption holds then the PH assumption also holds** (and vice versa). This **property is unique** to the Weibull model (Cox and Oakes, 1984) and holds if p does not vary over different levels of covariates. The PH assumption allows for the estimation of a hazard ratio enabling a comparison of rates among different populations. The AFT assumption allows for the estimation of an acceleration factor, which can describe the direct effect of an exposure on survival time.

HR vs. AFT

Hazard ratio ⇒ Comparison of rates

Acceleration factor ⇒ Effect on survival

Useful Weibull property:

- ln[−ln S(t)] is linear with ln(t)
- Enables graphical evaluation using KM survival estimates

The Weibull model also has another key property: **the log(–log) of S(t) is linear with the log of time**. This allows a graphical evaluation of the appropriateness of a Weibull model by **plotting the log negative log of the Kaplan–Meier survival estimates against the log of time.**

Linearity of ln(t)

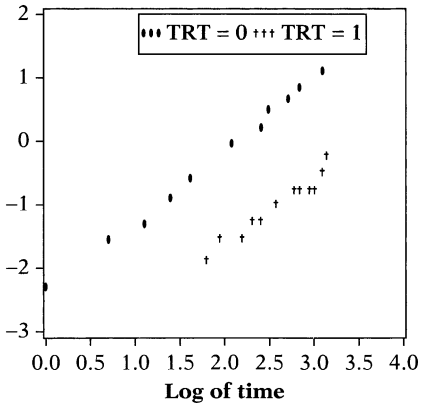
$$S(t) = \exp(-\lambda t^p)$$
$$\Rightarrow \ln[-\ln S(t)] = \ln(\lambda) + p \ln(t)$$

Intercept $= \ln(\lambda)$, Slope $= p$

To see this linear relationship: start with the Weibull survival function $S(t) = \exp(-\lambda t^p)$, take the log of S(t), multiply by negative one, and take the log again (see left). For the Weibull distribution, the $\ln[-\ln(S(t))]$ is a linear function of ln(t) with slope p and intercept $p \ln(\lambda)$. If the slope equals one then t follows an exponential distribution.

Remission data: evaluate Weibull assumption for TRT = 1 and TRT = 0

$\ln[-\ln \hat{S}(t)]$ plotted against ln(t)



We again return to the remission data and evaluate the appropriateness of the Weibull assumption for the treated (TRT = 1) and placebo (TRT = 0) groups. On the left is the plot of the log negative log Kaplan-Meier survival estimates against the log of time for TRT = 1 and TRT = 0. Both plots look reasonably straight suggesting that the Weibull assumption is reasonable. Furthermore, the lines appear to have the same slope (i.e., are parallel, same p) suggesting that the PH (and thus the AFT) assumptions hold. If this common slope equals one (i.e., $p = 1$), then survival time follows an exponential distribution. The Weibull model output containing the parameter estimates includes a statistical test for the hypothesis $p = 1$ or equivalently for $\ln(p) = 0$ (for testing the exponential assumption). This is examined later in this section.

Summary of possible results for plot of $\ln[-\ln \hat{S}(t)]$ against ln(t)

1. Parallel straight lines ⇒ Weibull, PH, and AFT assumptions hold
2. Parallel straight lines with slope of 1 ⇒ Exponential. PH and AFT
3. Parallel but not straight lines ⇒ PH but not Weibull, not AFT (can use Cox model)
4. Not parallel and not straight ⇒ Not Weibull, PH violated
5. Not parallel but straight lines ⇒ Weibull holds, but PH and AFT violated, different p

On the left is a summary of five possible results from an examination of the log negative log Kaplan-Meier survival estimates plotted against the log of time for two or more levels of covariates. The key points are that **straight lines support the Weibull assumption** and **parallel curves support the PH assumption**. If the plots are parallel but not straight then the PH assumption holds but not the Weibull. Assessing whether the curves are parallel is a familiar approach for evaluating the PH assumption in a Cox model (see Chapter 4 and Computer Appendix). An interesting scenario occurs if the lines are straight but not parallel. In this situation the Weibull assumption is supported but the PH and AFT assumptions are violated. If the lines are not parallel, then p is not constant across levels of covariates. In Section IX of this chapter, we present a method for modeling the shape parameter p as a function of predictor variables, but typically p is assumed fixed.

Previous plot suggests Weibull and PH assumption reasonable for TRT

An examination of the plot on the previous page suggests that the Weibull and PH assumptions are reasonable for treatment (TRT). First the PH form of the model is presented and then the AFT form.

Weibull PH model:

$h(t) = \lambda p t^{p-1}$
where $\lambda = \exp(\beta_0 + \beta_1 \text{ TRT})$.

The Weibull hazard function is $h(t) = \lambda p t^{p-1}$. A Weibull PH model is defined by reparameterizing lambda λ as $\exp(\beta_0 + \beta_1 \text{ TRT})$. The hazard ratio is obtained by substituting TRT = 1 and TRT = 0 into the hazard functions (see left). After canceling we obtain the familiar result $\exp(\beta_1)$. Note that this result depends on p having the same value for TRT = 1 and TRT = 0, otherwise time (t) would not cancel in the expression for the HR (i.e., PH assumption not satisfied).

Hazard ratio (TRT = 1 vs. TRT = 0)

$$\text{HR} = \frac{\exp(\beta_0 + \beta_1) p t^{p-1}}{\exp(\beta_0) p t^{p-1}} = \exp(\beta_1)$$

Remission Data

Weibull regression log relative-hazard form

_t	Coef.	Std. Err.	z	p > |z|
trt	−1.731	.413	−4.19	0.000
_cons	−3.071	.558	−5.50	0.000
/ln_p	.312	.147	2.12	0.034
p	1.366	.201		
1/p	.732	.109		

On the left is stata output from running the PH form of the Weibull model. There are parameter estimates for the coefficient of TRT, the intercept (called _cons), and for three forms of the shape parameter: p, $1/p$, and $\log(p)$. The estimate for p is 1.366 suggesting an increase in the hazard as survival time increases (because $\hat{p} > 1$). A statistical test for H_0: $\log(p) = 0$ yields a p-value of 0.034. At a significance level of 0.05 we would reject the null and decide p is not equal to 1, suggesting that the exponential model is not appropriate.

Weibull PH

$\widehat{HR}(\text{TRT} = 1 \text{ vs. } 0) = \exp(-1.731) = 0.18$

$95\%\ \text{CI} = \exp[-1.731 \pm 1.96(0.413)] = (0.08, 0.40)$

An estimated hazard ratio of 0.18 is obtained by exponentiating the estimated coefficient (−1.731) of the TRT variable. The 95% confidence interval for this HR is calculated to be (0.08, 0.40) indicating a significant preventive effect of treatment. These results are similar to those obtained from the exponential model in which the estimated hazard ratio was 0.22.

Weibull: $\widehat{HR} = 0.18$
Exponential: $\widehat{HR} = 0.22$
Suggests preventive effect of TRT

Comparing Cox and Weibull PH models

Cox: estimate β_1

$h(t) = h_0(t)\exp(\beta_1 \text{TRT})$

baseline hazard unspecified

It can be instructive to compare the Cox and Weibull PH models. The Cox PH model with treatment as the only predictor is stated as $h_0(t)\exp(\beta_1 \text{TRT})$. There is one parameter to estimate (β_1) and the distribution of the baseline hazard ($h_0(t)$) remains unspecified.

Weibull: estimate β_0, β_1, p

$h(t) = \lambda p t^{p-1}$ where

$\lambda = \exp(\beta_0 + \beta_1 \text{TRT})$.

$h(t) = [\exp(\beta_0) p t^{p-1}]\exp(\beta_1 \text{TRT})$.

baseline hazard specified parametrically

With some manipulation, the Weibull PH model can also be expressed as a product of a baseline hazard and $\exp(\beta_1$ TRT) (see left). There are three parameters to estimate β_0, β_1, and p that fully specify the hazard.

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

solve for t

$$t = [-\ln S(t)]^{1/p} \times \frac{1}{\lambda^{1/p}}$$

$$\text{let } \frac{1}{\lambda^{1/p}} = \exp(\alpha_0 + \alpha_1 RX)$$

$$t = [-\ln S(t)]^{1/p} \times \mathbf{exp}(\alpha_0 + \alpha_1 \mathbf{TRT})$$

Scaling of t

An AFT model can also be formulated with the Weibull distribution. We develop the AFT parameterization similarly to that done with the exponential model, by solving for t in terms of a fixed S(t). The Weibull survival function is $S(t) = \exp(-\lambda t^p)$. Taking the natural log, multiplying by negative 1, raising to the power $1/p$, and then multiplying by the reciprocal of $\lambda^{1/p}$, yields the expression for t shown on the left. By reparameterizing $1/\lambda^{1/p} = \exp(\alpha_0 + \alpha_1 TRT)$, it can be shown that α_1, the coefficient of the predictor variable TRT, is used to scale the time to any fixed value of S(t) (see left).

Let $S(t) = q$

$$t = [-\ln(q)]^{1/p} \times \exp(\alpha_0 + \alpha_1 TRT)$$

The expression for t is restated on the left in terms of any fixed probability $S(t) = q$. For example, to find an expression for the median survival time t_m, substitute $q = 0.5$ (see left).

Median survival time ($q = 0.5$)

$$t_m = [-\ln(0.5)]^{1/p} \times \exp(\alpha_0 + \alpha_1 TRT)$$

Acceleration factor, γ (TRT = 1 vs. TRT = 0)

$$\gamma = \frac{[-\ln(q)]^{1/p} \exp(\alpha_0 + \alpha_1)}{[-\ln(q)]^{1/p} \exp(\alpha_0)} = \exp(\alpha_1)$$

The acceleration factor γ is obtained as the ratio of the times to $S(t) = q$ for TRT = 1 and for TRT = 0. After canceling, γ reduces to $\exp(\alpha_1)$, where α_1 is the coefficient of the TRT variable. As with the PH form of the model, this result depends on p not varying by treatment status; otherwise γ would depend on q.

Remission Data

Weibull regression accelerated failure-time form

_t	Coef.	Std. Err.	z	P>|z|
trt	1.267	.311	4.08	0.000
_cons	2.248	.166	13.55	0.000
/ln_p	.312	.147	2.12	0.034
p	1.366	.201		
1/p	.732	.109		

Output from running a Weibull AFT model is shown on the left. The estimates for each form of the shape parameter (p, $1/p$, and $\ln(p)$) are the same as obtained from the previously shown PH form of the model.

The estimated acceleration factor of 3.55 is obtained by exponentiating the estimated coefficient (1.267) of the TRT variable. The 95% confidence interval for γ is calculated to be (1.93, 6.53). There results are shown at the top of the following page.

Weibull AFT:

$\hat{\gamma}(\text{TRT} = 1 \text{ vs. } 0) = \exp(1.267) = 3.55$

95% CI $= \exp[1.267 \pm 1.96(0.311)] = (1.93, 6.53)$

These results suggest that the median (or any other quantile of) survival time is increased by a factor of 3.55 for those receiving the treatment compared to the placebo. Recall that the acceleration factor was estimated at 4.60 using the exponential model. However, the exponential model uses a much stronger assumption: that the hazards are constant.

Weibull: $\hat{\gamma} = 3.55$
Exponential: $\hat{\gamma} = 4.60$ (assumes $h(t) = \lambda$)

Relating Weibull AFT and PH coefficients

Corresponding coefficients obtained from the PH and AFT forms of the Weibull models are related as follows: $\beta_j = -\alpha_j p$ for the jth covariate. This can most easily be seen by formulating the parameterization equivalently in terms of $\ln(\lambda)$ for both the PH and AFT forms of the model as shown on the left.

AFT: $\lambda^{1/p} = \exp[-(\alpha_0 + \alpha_1 \text{TRT})]$
$(1/p)\ln \lambda = -(\alpha_0 + \alpha_1 \text{TRT})$
$\ln \lambda = -p\,(\alpha_0 + \alpha_1 \text{TRT})$

PH: $\lambda = \exp(\beta_0 + \beta_1 \text{TRT})$
$\ln \lambda = \beta_0 + \beta_1 \text{TRT}$

Relationship of coefficients:

$\beta_j = -\alpha_j p$ so that
$\beta = -\alpha$ for exponential ($p = 1$)

This relationship is illustrated utilizing the coefficient estimates we obtained for TRT: $-1.731 = (-1.267)(1.366)$. Note for the exponential model in which $p = 1$, the PH and AFT coefficients are related, $\beta = -\alpha$.

Relating estimates for TRT (PH vs. AFT)

$-1.731 = (-1.267)(1.366)$

Next: log-logistic model

- Hazard may be nonmonotonic

Weibull model

- Hazard does not change direction

In the next example the log-logistic model is presented. In contrast to the Weibull, the hazard function for the log-logistic distribution allows for some nonmonotonic behavior in the hazard function.

VII. Log-Logistic Example

The log-logistic distribution accommodates an AFT model but not a PH model. Its hazard function is shown on the left. The shape parameter is $p(>0)$.

Log-logistic hazard: $h(t) = \dfrac{\lambda p t^{p-1}}{1 + \lambda t^p}$
(where $p > 0$ and $\lambda > 0$)

Shape of hazard function:

$p \leq 1$ hazard decreases over time
$p > 1$ hazard first increases and then decreases over time (unimodal)

If $p \leq 1$ then the hazard decreases over time. If $p > 1$, however, the hazard increases to a maximum point and then decreases over time. In this case ($p > 1$), the hazard function is said to be **unimodal**.

Log-logistic modeling assumptions:

AFT $\not\Longleftrightarrow$ PH

AFT $\Longleftrightarrow$ PO

PO: Survival Odds ratio constant over time

Unlike the Weibull model, a log-logistic AFT model is not a PH model. However, the log-logistic AFT model is a proportional odds (PO) model. **A proportional odds survival model is a model in which the survival odds ratio is assumed to remain constant over time**. This is analogous to a proportional hazard model where the hazard ratio is assumed constant over time.

Survival odds

$$\frac{S(t)}{(1-S(t))} = \frac{P(T > t)}{P(T \leq t)}$$

The survival odds is the odds of surviving beyond time t (i.e., S(t)/(1 − S(t)). This is the probability of not getting the event by time t divided by the probability of getting the event by time t.

Failure odds by time t

$$\frac{(1-S(t))}{S(t)} = \frac{P(T \leq t)}{P(T > t)}$$

The failure odds is the odds of getting the event by time t (i.e., (1 − S(t))/S(t)), which is the reciprocal of the survival odds (see left).

Log-logistic survival and failure functions

$$\mathrm{S(t)} = \frac{1}{1+\lambda t^p} \quad 1 - \mathrm{S(t)} = \frac{\lambda t^p}{1+\lambda t^p}$$

The log-logistic survival function (S(t)) and failure function (1 − S(t)) are shown on the left.

Failure odds

$$\frac{1-S(t)}{S(t)} = \frac{\left(\frac{\lambda t^p}{1+\lambda t^p}\right)}{\left(\frac{1}{1+\lambda t^p}\right)} = \lambda t^p$$

The failure odds simplifies in a log-logistic model to λt^p (see left).

Log-logistic PO model:

- Reparameterize λ in terms of Xs and βs

  $$\mathrm{SOR} = \frac{\mathrm{S_1(t)}/(1-\mathrm{S_1(t)})}{\mathrm{S_2(t)}/(1-\mathrm{S_2(t)})}$$

- SOR satisfies PO if SOR constant over time
- SOR constant

  ⇕

FOR = 1/SOR constant

A log-logistic proportional odds model can be formulated by reparameterizing λ in terms of predictor variables and regression parameters. We come back to this point later in this section.

A survival odds ratio (SOR) is defined as the ratio of survival odds for two groups of subjects, as shown on the left.

An SOR satisfies a PO assumption provided the SOR does not depend on time.

Also, if the SOR does not depend on time, then the failure odds ratio (FOR = 1/SOR) also does not depend on time.

Log Odds Is Linear with ln(t)

$$\log(\text{failure odds}) = \ln(\lambda t^p) = \ln(\lambda) + p[\ln(t)]$$

Intercept $= \ln(\lambda)$ slope $= p$

The log of the failure odds is $\ln(\lambda t^p)$, which can be rewritten as $\ln(\lambda) + p[\ln(t)]$. In other words, **the log odds of failure is a linear function of the log of time** with slope p and intercept $\ln(\lambda)$. This is a useful result enabling a graphical evaluation for the appropriateness of the log-logistic distribution.

Evaluate log-logistic assumption graphically

- Plot $\ln\left[\frac{(1-\hat{S}(t))}{\hat{S}(t)}\right]$ against ln(t)
- If log-logistic, then plot is linear with slope $= p$

The log-logistic assumption can be graphically evaluated by **plotting** $\ln(1 - \hat{S}(t))/(\hat{S}(t))$ **against ln(t) where** $\hat{S}(t)$ **are the Kaplan–Meier survival estimates**. If survival time follows a log-logistic distribution, then the resulting plots should be a straight line of slope p.

Alternatively

- Plot $\ln\left(\frac{((\hat{S}(t))}{(1-\hat{S}(t))}\right)$ against ln(t)
- If log-logistic, then plot is linear with slope $= -p$

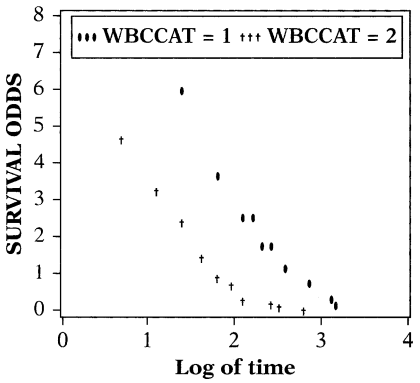
We could alternatively plot the log of the survival odds, $\ln(\hat{S}(t))/(1 - \hat{S}(t))$, against ln(t). If the log-logistic assumption is correct the resulting plots should be a straight line of slope $-p$.

Remission Data

WBCCAT: white blood cell count variable medium = 1 vs. high = 2

$\ln\left[\frac{\hat{S}(t)}{(1-\hat{S}(t))}\right]$ plotted against ln(t).

We next consider a different variable from the remission data: a dichotomous variable for white blood cell count (WBCCAT) coded medium = 1 and high = 2.



On the left is the plot of the log odds of survival (obtained from the Kaplan–Meier survival estimates) against the log of time comparing **medium** (WBCCAT = 1) and **high** (WBCCAT = 2) blood cell counts. The points for WBCCAT = 1 lie above the points for WBCCAT = 2 indicating that the survival odds are higher for those with a medium white blood cell count compared to high. The lines look reasonably straight and parallel, at least until the estimated odds of survival approaches zero.

If we accept the proposition that the lines look straight, then the log-logistic assumption is reasonable. Because the lines look parallel, the proportional odds (PO) assumption is also reasonable. If the PO assumption holds in a log-logistic model then the AFT assumption also holds.

Straight lines $\Longrightarrow$ Log-logistic

Parallel plots $\Longrightarrow$ PO

Log-logistic and PO $\Longrightarrow$ AFT

Log-logistic and Weibull graphical approach analogous

- Check PH for Weibull
- Check PO for log-logistic

AFT log-logistic model

$$S(t) = \frac{1}{1+\lambda t^p} = \frac{1}{1+\left(\lambda^{1/p} t\right)p}$$

solve for t to obtain

$$t = \left[\frac{1}{S(t)} - 1\right]^{1/p} \times \frac{1}{\lambda^{1/p}}$$

$$\text{let } \frac{1}{\lambda^{1/p}} = \exp(\alpha_0 + \alpha_1 \text{WBCCAT})$$

$$t = \left[\frac{1}{S(t)} - 1\right]^{1/p} \times \exp(\alpha_0 + \alpha_1 \mathbf{WBCCAT})$$

Scaling of t

Let S(t) − q

$$t = [q^{-1} - 1]^{1/p} \times \exp(\alpha_0 + \alpha_1 \text{ WBCCAT})$$

Median survival time (q = 0.5):

$$t_m = [2 - 1]^{1/p} \times \exp(\alpha_0 + \alpha_1 \text{ WBCCAT})$$

The key points from above are:

a. **straight lines support the log-logistic assumption**,
b. **parallel curves support the PO assumption**, and
c. **If the log-logistic and PO assumptions hold, then the AFT assumption also holds**.

The graphical evaluation for the log-logistic assumption is analogous to the graphical analysis of the Weibull assumption presented in the last section, except here the PO assumption rather than the PH assumption is evaluated by checking for parallel lines.

Next we consider an AFT log-logistic model with white blood cell count as the only predictor comparing WBCCAT = 2 (high count) and WBCCAT = 1 (medium count).

We develop the AFT parameterization by solving for t in terms of a fixed S(t). Starting with the expression for S(t), taking reciprocals, subtracting 1, raising to the power $1/p$, and then multiplying by the reciprocal of $\lambda^{1/p}$, yields the expression for t shown on the left. By reparameterizing $1/\lambda^{1/p} = \exp(\alpha_0 + \alpha_1 \text{ WBCCAT})$, we allow α_1, the coefficient of the predictor variable WBCCAT, to be used for the multiplicative scaling of time to any fixed value of S(t) (see left).

The expression for t is restated on the left in terms of any fixed probability S(t) = q. For example, to find an expression for the median survival time t_m, substitute q = 0.5 (see left).

Acceleration factor,
γ (WBCCAT = 2 vs. WBCCAT = 1)

$$= \frac{[q^{-1} - 1]^{1/p} \exp(\alpha_0 + 2\alpha_1)}{[q^{-1} - 1]^{1/p} \exp(\alpha_0 + 1\alpha_1)}$$
$$= \exp(\alpha_1)$$

The acceleration factor γ is found by taking the ratio of the times to S(t) = q for WBCCAT = 2 and for WBCCAT = 1. After canceling, γ reduces to $\exp(\alpha_1)$, where α_1 is the coefficient of the WBCCAT variable in the AFT model.

Log-logistic regression accelerated failure-time form

_t	Coef.	Std. Err.	z	P>|z|
wbccat	−.871	.296	−2.94	0.003
_cons	3.495	.498	7.09	0.000
ln_gam	−.779	.164	−4.73	0.000
gamma	.459	0.756		

$$p = 1/(0.459) = 2.18$$

The output from running the AFT log-logistic model is shown on the left. The coefficient estimate for WBCCAT is −0.871, which is statistically significant with a p-value of 0.003 (far right column of output).

Stata provides estimates of the reciprocal of p (gamma = $1/p$) rather than for p. The estimate for gamma is 0.459. Therefore, the estimate for p is 1/(0.459) = 2.18.

WBCCAT = 2 vs. WBCCAT = 1 (log-logistic):

$$\hat{\gamma} = \exp(-0.871) = 0.42$$
$$95\%\ \text{CI for}\ \gamma = \exp[-0.871 \pm 1.96(0.296)]$$
$$= (0.23,\ 0.75)$$

An estimate of the acceleration factor $\hat{\gamma}$ comparing WBCCAT = 2 to WBCCAT = 1 is found by exponentiating the estimate −0.871 of α_1 to obtain 0.42. The 95% confidence interval for γ is calculated to be (0.23, 0.75).

Comparing estimated survival

$$\hat{S}_1(t) = \hat{S}_2(0.42t)$$

Survival function for WBCCAT = 1

Survival function for WBCCAT = 2

These results suggest that the time for going out of remission is "accelerated" for patients with a **high** white blood cell count compared to those with a **medium** count by an estimated factor of 0.42. In terms of the survival functions estimated from this model, $\hat{S}_1(t) = \hat{S}_2(0.42t)$ where $\hat{S}_1(t)$ and $\hat{S}_2(t)$ are the respective survival functions for patients with medium and high blood cell counts.

Failure odds

$$\frac{1 - S(t)}{S(t)} = \frac{\left(\frac{\lambda t^p}{1+\lambda t^p}\right)}{\left(\frac{1}{1+\lambda t^p}\right)} = \lambda t^p$$

where $\lambda = \exp(\beta_0 + \beta_1\ \text{WBCCAT})$

The proportional odds form of the log-logistic model can also be formulated by reparameterizing λ. Recall that the log-logistic failure odds is λt^p.

By setting $\lambda = \exp(\beta_0 + \beta_1\ \text{WBCCAT})$, a (failure) odds ratio comparing WBCCAT = 2 to WBCCAT = 1 can be calculated (see left). After canceling, the odds ratio reduces to $\exp(\beta_1)$.

OR (WBCCAT = 2 vs. WBCCAT = 1)

$$= \frac{t^p \exp(\beta_0 + 2\beta_1)}{t^p \exp(\beta_0 + 1\beta_1)} = \exp(\beta_1)$$

Comparing AFT and PO (log-logistic)
Relationship of coefficients:

$\beta_j = -\alpha_j p$

Since $\hat{\alpha} = -0.871$ and $\hat{p} = 2.18$
Then,

$\hat{\beta}_1 = -(-0.871)(2.18) = 1.90$

and

$\widehat{OR} = \exp(1.90) = 6.69$

The corresponding coefficients for log-logistic PO and AFT models are related by $\beta_j = -\alpha_j p$ for the jth covariate. This result is obtained using a similar argument to that presented for the Weibull example in the previous section.

The estimate for α_1 in the AFT model is -0.871 and the estimate for p is 2.18. Therefore, an estimate for β_1 can be found by multiplying $-(-0.871)$ times 2.18 yielding 1.90. An estimate of the odds ratio is found by exponentiating this estimate, $\exp(1.90) = 6.69$. (Unfortunately, neither Stata nor SAS estimates the proportional odds form of the model.)

VIII. A More General Form of the AFT Model

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

- AFT Form: $\frac{1}{\lambda} = \exp(\alpha_0 + \alpha_1 \text{ TRT})$
- PH Form: $\lambda = \exp(\beta_0 + \beta_1 \text{ TRT})$

Weibull: $S(t) = \exp(-\lambda t^p)$

- AFT Form: $\frac{1}{\lambda^{1/p}} = \exp(\alpha_0 + \alpha_1 \text{ TRT})$
- PH Form: $\lambda = \exp(\beta_0 + \beta_1 \text{ TRT})$

Log-logistic: $S(t) = \frac{1}{1 + \lambda t^p}$

- AFT Form: $\frac{1}{\lambda^{1/p}} = \exp(\alpha_0 + \alpha_1 \text{ WBCCAT})$
- PO Form: $\lambda = \exp(\beta_0 + \beta_1 \text{ WBCCAT})$

On the left is a summary of the models discussed in the previous sections. These models were formulated by reparameterizing the survival (and hazard) functions in terms of regression parameters and predictor variables.

An advantage for stating the models in this form is that the interpretation and relationships between parameters are specific to each distribution.

However, there are more general ways these models could be stated. The Cox PH model is a more general way of stating the proportional hazards model. In this section we discuss a more general formulation of the AFT model.

General Form of AFT Model
(*One Predictor*)

$$\ln(T) = \alpha_0 + \alpha_1 TRT + \epsilon$$

ϵ: random error

Consider an AFT model with one predictor (TRT) in which T represents a random variable for survival time. The model can be expressed on the log scale as shown on the left, where ϵ is random error following some distribution.

With additional parameter

$$\ln(T) = \alpha_0 + \alpha_1 TRT + \sigma\epsilon$$

σ scales the error

Some distributions have an additional parameter (σ) scaling ϵ. The model including this additional parameter can be expressed as shown on the left, where the random error ϵ is multiplied by a scale parameter σ.

If $\epsilon \sim N(0, 1)$, then

$\ln(T) \sim N(\mu = \alpha_0 + \alpha_1 TRT, sd = \sigma)$

Similar to linear regression (except for inclusion of censorships)

If ϵ follows a standard normal distribution and $\ln(T) = \alpha_0 + \alpha_1 TRT + \sigma\epsilon$, then ln(T) would follow a normal distribution with mean $\mu = \alpha_0 + \alpha_1 TRT$ and standard deviation σ. For this situation, the model would look like a standard linear regression. The **key difference** between fitting this survival model and a standard linear regression **is the inclusion of censored observations** in the data.

In general,

$\mu_{\ln(T)} \neq (\alpha_0 + \alpha_1 TRT)$, $sd \neq \sigma$

Interpretation of parameters depends on distribution

In general, for other then the normal distribution, the mean of ln(T) is not necessarily $\alpha_0 + \alpha_1 TRT$ and its standard deviation is not σ. In other words, it should not be assumed that the mean of ϵ is 0 and the standard deviation is 1. The interpretation of the parameters depends on the underlying distribution.

Let $\sigma = \frac{1}{p}$, then

$$\ln(T) = \alpha_0 + \alpha_1 TRT + \frac{1}{p}\epsilon$$

Sometimes the model is parameterized using $\sigma = 1/p$. The model can then be restated by replacing $\sigma\epsilon$ with $(1/p)\,\epsilon$.

Additive model in terms of ln(T)

but

multiplicative model in terms of T

The AFT model is **additive on the log scale but a multiplicative model with respect to T**.

$$T = \exp\left(\alpha_0 + \alpha_1 TRT + \frac{1}{p}\epsilon\right)$$

$$= \exp[(\alpha_0 + \alpha_1 TRT)] \times \exp\left(\frac{1}{p}\epsilon\right)$$

In particular, the model can be expressed in terms of T by exponentiating ln(T), as shown on the left.

Collapse α_0 into baseline term

$$T_0 = \exp(\alpha_0)\exp\left(\frac{1}{p}\epsilon\right)$$

The model may also be expressed by collapsing the intercept into a baseline random term T_0 (see left). In this setting, T_0 is a random variable representing the survival time of the placebo group (TRT = 0).

so that $T = T_0 \exp(\alpha_1 \text{ TRT})$ where T_0 is a random variable for TRT = 0

Thus, the AFT model has a form analogous to the a Cox PH model, although the baseline term T_0 is a random variable rather than a constant.

AFT model may be expressed in terms of T or ln(T)

Comparing Distributions: T and ln(T)

T	ln(T)
Exponential	Extreme minimum value
Weibull	Extreme minimum value
Log-logistic	Logistic
Lognormal	Normal

In summary, an AFT model may be expressed by reparameterizing a specific distribution, or may be more generally expressed either in terms of a random variable T (for survival time), or ln (T). If T follows a Weibull distribution then ln (T) follows a distribution called the extreme minimum value distribution (see table on left). Similarly, if T follows a log-logistic or lognormal distribution then ln(T) follows a logistic or normal distribution, respectively. The logistic and normal distributions are similarly shaped, and are both symmetric about their mean.

IX. Other Parametric Models

In the previous sections we presented examples of the exponential, Weibull, and log-logistic models. In this section we briefly discuss some other parametric survival models.

Generalized Gamma Model

- Supported by SAS and Stata
- S(t), h(t) expressed in terms of integrals
- Contains three parameters
- Weibull, lognormal are special cases

The **generalized gamma model** is a parametric survival model that is supported by both SAS and Stata software. The hazard and survival function for this model is complicated and can only be expressed in terms of integrals. The generalized gamma distribution has three parameters allowing for great flexibility in its shape. The Weibull and lognormal distributions are special cases of the generalized gamma distribution (see Practice Exercises 12 to 14).

Lognormal Model

Similar to log-logistic
Difference:
Log-logistic: AFT and PO
Lognormal: AFT but not PO

The **lognormal model** also has a relatively complicated hazard and survival function that can only be expressed in terms of integrals. The shape of the lognormal distribution is very similar to the log-logistic distribution and yields similar model results. A difference is that although the lognormal model accommodates an accelerated failure time model, it is not a proportional odds model.

Gompertz Model

- PH model but not AFT
- One predictor (TRT) in model:

$h(t) = [\exp(\gamma t)] \times \exp(\beta_0 + \beta_1 \text{ TRT})$

↗ parametrically specified

$h_0(t) = \exp(\gamma t)$

$\gamma > 0$ hazard exponentially increases with t
$\gamma < 0$ hazard exponentially decreases with t
$\gamma = 0$ constant hazard (exponential model)

Parametric models need not be AFT models. The **Gompertz** model is a parametric proportional hazards model but not an AFT model. The model can be expressed in a form similar to that of a Cox PH model except that the baseline hazard is specified as the hazard of a Gompertz distribution containing a shape parameter γ (see left).

If $\gamma > 0$ then the hazard exponentially increases over time. If $\gamma < 0$ then the hazard exponentially decreases over time. If $\gamma = 0$ then the hazard is constant and reduces to the exponential model.

AFT model: multiplicative

$T = \exp(\alpha_0 + \alpha_1 \text{ TRT} + \epsilon)$
$= \exp(\alpha_0) \times \exp(\alpha_1 \text{ TRT}) \times \exp(\epsilon)$

but
additive on log scale:
$\ln(T) = \alpha_0 + \alpha_1 \text{ TRT} + \epsilon$

The AFT model is a **multiplicative model** (i.e., a multiplicative scaling of failure time). It becomes an additive model on the log scale (see left side).

Additive failure time model

$T = \alpha_0 + \alpha_1 \text{TRT} + \epsilon$

↗ T rather than log(T) is linear with TRT

An alternative parametric model is to define an **additive failure time model** in terms of T, rather than ln(T). Consider the model: $T = \alpha_0 + \alpha_1 \text{ TRT} + \epsilon$. Now T, rather than ln(T), is expressed as a linear function of the regression parameters. SAS supports such an additive failure time model (see Computer Appendix).

Modeling the Shape Parameter (e.g., Weibull and log-logistic)

Typical Weibull model

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

where $\lambda = \exp(\beta_0 + \beta_1 \text{ TRT})$
p unaffected by predictors

Many parametric models contain an extra shape (or ancillary) parameter beyond the regression parameters. For example, the Weibull and log-logistic models contain a shape parameter p. Typically, this parameter is considered fixed, unaffected by changes in the values of predictor variables.

Alternative Weibull model
models the ancillary parameter p

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

where $\lambda = \exp(\beta_0 + \beta_1 \text{ TRT})$
$p = \exp(\delta_0 + \delta_1 \text{ TRT})$

Not a PH or AFT model if $\delta_1 \neq 0$ but still a Weibull model

An alternative approach is to model the shape parameter in terms of predictor variables and regression coefficients. In the Weibull model shown on the left, both λ and p are modeled as functions of treatment status (TRT). If δ_1 is not equal to zero, then the value of p differs by TRT. For that situation, the PH (and thus the AFT) assumption is violated because t^{p-1} will not cancel in the hazard ratio for TRT (see Practice Exercises 15 to 17).

Choosing appropriate model

- Evaluate graphically
  - Exponential
  - Weibull
  - Log-logistic
- Akaike's information criterion
  - Compares model fit
  - Uses −2 log likelihood

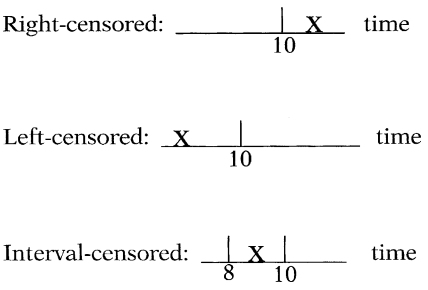
Choosing the most appropriate parametric model can be difficult. We have provided graphical approaches for evaluating the appropriateness of the exponential, Weibull, and log-logistic models. **Akaike's information criterion (AIC)** provides an approach for comparing the fit of models with different underlying distributions, making use of the −2 log likelihood statistic (described in Practice Exercises 11 and 14).

X. The Parametric Likelihood

- Function of observed data and unknown parameters
- Based on outcome distribution f(t)
- Censoring complicates survival data
  - Right-censored
  - Left-censored
  - Interval-censored

The likelihood for any parametric model is a function of the observed data and the model's unknown parameters. The form of the likelihood is based on the probability density function f(t) of the outcome variable. A complication of survival data is the possible inclusion of censored observations (i.e., observations in which the exact time of the outcome is unobserved). We consider three types of censored observations: right-censored, left-censored, and interval-censored.

Examples of Censored Subjects



Right-censored. Suppose a subject is lost to follow-up after 10 years of observation. The time of event is not observed because it happened after the 10th year. This subject is right-censored at 10 years because the event happened to the *right* of 10 on the time line (i.e., $t > 10$).

Left-censored. Suppose a subject had an event before the 10th year but the exact time of the event is unknown. This subject is left-censored at 10 years (i.e., $t < 10$).

Interval-censored. Suppose a subject had an event between the 8th and 10th year (exact time unknown). This subject is interval-censored (i.e., $8 < t < 10$).

Formulating the Likelihood

Barry, Gary, Larry, . . ., Outcome Distribution $f(t)$

Subject	Event Time	Likelihood Contribution
Barry	t = 2	$f(2)$
Gary	$t > 8$ (right-censored)	$\int_8^\infty f(t)dt$
Harry	t = 6	$f(6)$
Carrie	$t < 2$ (left-censored)	$\int_0^2 f(t)dt$
Larry	$4 < t < 8$ (interval-censored)	$\int_4^8 f(t)dt$

The table on the left illustrates how the likelihood is formulated for data on five subjects. We assume a probability density function f(t) for the outcome. Barry gets the event at time t = 2. His contribution to the likelihood is f(2). Gary is right-censored at t = 8. The probability that Gary gets the event after t = 8 is found by integrating f(t) from 8 to infinity. This is Gary's contribution to the likelihood. Harry gets the event at time t = 6. His contribution to the likelihood is f(6). Carrie is left-censored at t = 2. Her contribution to the likelihood is obtained by integrating f(t) from zero to 2. Finally, Larry is interval-censored between t = 4 and t = 8. His contribution to the likelihood is found by integrating f(t) from 4 to 8.

Likelihood (L)

Product of individual contributions

$$L = f(2)x\int_8^\infty f(t)dt \times f(6) \times \int_0^2 f(t)dt \times \int_0^8 f(t)dt$$

(Barry × Gary × Harry × Carrie × Larry)

The full likelihood (L) is found by taking the product of each subject's independent contribution to the likelihood. The likelihood for this example is shown on the left.

Assumptions for formulating L

- No competing risks
  - Competing event does not prohibit event of interest
  - Death of all causes is classic example of no competing risk

The formulation of this likelihood uses the assumption of **no competing risks**. In other words, we assume that no competing event will prohibit any subject from eventually getting the event of interest (see Chapter 9). Death from all causes is the classic example of an outcome that in reality has no competing risk. For other outcomes, the no competing risk assumption is more of a theoretical construct.

- Subjects independent
  - Allows L to be formulated as product of subjects' contributions

Another assumption is that individual contributions to the likelihood are independent. This assumption allows the full likelihood to be formulated as the product of each individual's contribution.

- Follow-up time continuous
  - No gaps in follow-up

A third assumption is that each subjects follow-up time is continuous without gaps (i.e., once subjects are out of the study, they do not return). If gaps are allowed, the likelihood can be modified to accommodate such a scenario.

Revisit example with Barry, Gary, Larry, ...
f(t) is Weibull
SMOKE is only predictor

1 = Smoker
0 = Nonsmoker

In the last example, we did not specify the probability density f(t), nor did we specify any covariates. We revisit this example, assuming f(t) is Weibull with one predictor SMOKE in the model (coded 1 for smokers and 0 for nonsmokers).

Weibull: $h(t) = \lambda pt^{p-1}$,
$S(t) = \exp(-\lambda t^p)$

$f(t) = h(t)S(t)$
$f(t) = \lambda pt\ p-1\ exp(-\lambda t\ p)$

where $\lambda = \exp(\beta 0 + \beta 1\ \text{SMOKE})$
(PH form of the model)

The Weibull hazard and survival functions are shown on the left. The probability density function f(t) is the product of the hazard and survival functions. The parameterization will use the proportional hazards (PH) form of the Weibull model: $\lambda = \beta_0 + \beta_1$ SMOKE.

Data Layout for Right-, Left-, and Interval-Censoring Using SAS

ID	LOWER	UPPER	SMOKE
Barry	2	2	1
Gary	8	-	0
Harry	6	6	0
Carrie	-	2	0
Larry	4	8	1

Right-censored: UPPER missing
Left-censored: LOWER missing
Interval-censored: LOWER < UPPER
Not censored: LOWER = UPPER

On the left is the data layout for running parametric models containing right-, left-, and interval censored data in a form suitable for using the SAS procedure PROC LIFETEST (version 8.2). There are two time variables LOWERand UPPER. Barry got the event at t = 2, so both LOWER and UPPER get the value 2. Gary was right-censored at 8 (t < 8) so LOWER gets the value 8 and UPPER is set to missing. Carrie is left-censored at 2 (t > 2) so LOWER is set to missing and UPPER gets the value 2. Larry was interval-censored with LOWER = 4 and UPPER = 8. Barry and Larry are smokers whereas Gary, Harry, and Carrie are nonsmokers.

Weibull Likelihood (L)

Product of individual contributions

$$L = f(2) \times \int_8^{\infty} f(t)dt \times f(6) \times \int_0^2 f(t)dt \times \int_4^8 f(t)dt$$

$$\begin{aligned} L = {} & \exp(\beta_0 + \beta_1)p(2)^{p-1} \exp(-\exp(\beta_0 + \beta_1)2^p) \\ & \times \int_8^{\infty} \exp(\beta_0)p(t)^{p-1} \exp(-\exp(\beta_0)t^p)dt \\ & \times \exp(\beta_0)p(6)^{p-1} \exp(-\exp(\beta_0)6^p) \\ & \times \int_0^2 \exp(\beta_0)p(t)^{p-1} \exp(-\exp(\beta_0)t^p)dt \\ & \times \int_4^8 \exp(\beta_0 + \beta_1)p(t)^{p-1} \\ & \quad \times \exp(-\exp(\beta_0 + \beta_1)t^p)dt \end{aligned}$$

The full likelihood using the Weibull distribution can now be formulated as a product of each individual's contribution (shown on the left). We have used a small dataset (five subjects) for ease of illustration but the process can be generalized for any number of subjects.

Obtaining maximum likelihood estimates

Solve system of equations:

$$\frac{\partial Ln(L)}{\partial \beta_j} = 0 \quad \text{j} = 1, 2, \ldots, \text{N}$$

where N = # of parameters

Once the likelihood is formulated, the question becomes: *which values of the regression parameters would maximize L*? The process of maximizing the likelihood is typically carried out by setting the partial derivative of the natural log of L to zero and then solving the system of equations (called the score equations). The parameter estimates (e.g., $\hat{p}$, $\hat{\beta}_0$, $\hat{\beta}_1$) that maximize L are the maximum likelihood estimates.

XI. Interval-Censored Data

Parametric likelihood

- Handles right-, left-, or interval-censored data

Cox likelihood

- Designed to handle right-censored data.

One advantage of a parametric model compared to a Cox model is that the parametric likelihood easily accommodates right-, left-, or interval-censored data. The Cox likelihood, by contrast, easily handles right-censored data but does not directly accommodate left- or interval-censored data.

Interval-censored study design

- Check for nonsymptomatic outcome once a year
- If outcome newly detected, exact time occurred during previous year
- Left-censoring special case of interval-censoring
  - Zero the lower boundary of the interval

Sometimes the design of a study is such that all the data are interval-censored. For example, consider a study in which healthcare workers examine subjects once a year, checking for a nonsymptomatic outcome. If an event was first detected in the beginning of the third year, then the exact time of the outcome occurred sometime between the second and third years. In this framework left-censoring can be considered a special case of interval-censoring with zero as the lower boundary of the interval.

Parametric model can be fitted

- f(t) specified
- Contribution to likelihood for each subject
  - Integrate f(t) over event interval

A parametric model can easily be fitted using the methods described in the previous section. Once a distribution for the outcome, f(t), is specified, each subject's contribution to the likelihood is obtained by integrating f(t) over the interval in which he or she had the event.

Binary regression

- Alternative approach for interval-censored data
- Outcome coded
  - 0 if subject survives interval
  - 1 if subject gets event during interval
- Useful approach if
  - Ample number of events in each interval
  - Prefer not to specify f(t)

A binary regression (e.g., logistic regression) is an alternative approach that may be considered if all the data are interval-censored. With this method, the outcome variable can be coded zero if the subject survives the interval and coded one if the subject gets the event during the interval. This approach is particularly useful if there are an ample number of events in each interval and the analyst prefers not to specify a distribution f(t) for continuous survival time.

Information on Three Subjects

Subject 1: Gets event in first interval
Subject 2: Survives first interval Survives second interval Gets event in third interval
Subject 3: Survives first interval Gets event in second interval

For illustration, consider a small dataset containing three subjects. Subject 1 gets the event in the first interval of follow-up, subject 2 gets the event in the third interval, and subject 3 gets the event in the second interval of follow-up.

Data Layout for Binary Regression

SUBJECT	EVENT	D_1	D_2	D_3	TRT
1	1	1	0	0	1
2	0	1	0	0	0
2	0	0	1	0	0
2	1	0	0	1	0
3	0	1	0	0	1
3	1	0	1	0	1

The data layout is shown on the left. Each observation represents one interval of follow-up time allowing multiple observations per subject. EVENT is the dichotomous outcome variable. Subject 1 had the event in the first interval (EVENT = 1) and thus has one observation. Subject 2 has three observations because she survived the first two intervals (EVENT = 0) but got the event in the third interval. D_1 is a dummy variable coded 1 if the observation represents the first interval and 0 otherwise. Similarly, D_2 is coded 1 for the second interval and D_3 is coded 1 for the third interval.

EVENT: dichotomous outcome coded 1 if event, 0 for no event during the interval

D_1, D_2, D_3: dummy variables for intervals 1, 2, and 3 coded 1 if in the corresponding interval, 0 otherwise

TRT: Treatment coded 1 for new treatment, 0 for placebo

TRT is the predictor of interest, coded 1 for the new treatment and 0 for the placebo. TRT could be coded as a time-independent or time-dependent variable. In this example, TRT is time-independent because TRT does not change values over different intervals corresponding to the same subject.

Logistic Model

$$\text{Logit } P(Y = 1) = \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 \text{TRT}$$

where $P(Y = 1)$ is the probability of event for a given interval conditioned on survival of previous intervals

A logistic model (shown at left) containing the three dummy variables and TRT can be formulated with the data in this form.

Interpretation of Parameters

β_1: Log odds of event in 1st interval among TRT = 0

β_2: Log odds of event in 2nd interval given survival of 1st interval among TRT = 0

β_3: Log odds of event in 3rd interval given survival of first two intervals among TRT = 0

β_4: Log odds ratio for TRT

Care must be taken with the interpretation of the parameters: β_1 is the log odds of the event occurring in the first interval among the placebo group; β_2 is the log odds of the event occurring in the second interval conditioned on survival of the first interval among the placebo group; β_3 is the log odds of the event occurring in the third interval conditioned on survival of the first and second intervals among the placebo group; and β_4 is the log odds ratio for TRT.

D_1, D_2, D_3 play similar role as intercept

- Baseline measure when covariates are zero
- 3 parameters rather than 1 intercept
  - Baseline measure may differ for each interval

The dummy variables play a similar role to that of the intercept in a conventional regression, providing a baseline outcome measure for the case in which all predictors are zero (e.g., TRT = 0). In general, the baseline measure may differ for each interval, which is the reason that the model contains 3 dummy variables rather than 1 intercept.

Odds Ratio (TRT = 1 vs. TRT = 0) = $\exp(\beta_4)$

Model uses PO assumption

- OR constant over time
- PO assumption can be tested
  - Include interaction terms with TRT and dummy variables
  - Significant interaction suggests PO violation
  - Need ample data to practically carry out test

The odds ratio comparing TRT = 1 to TRT = 0 is obtained by exponentiating β_4. This model uses the **proportional odds (PO) assumption** in that the odds ratio is assumed constant over time (or at least constant at the end of each interval). This assumption can be tested by including interaction (product) terms with TRT and two of the dummy variables in the model. A statistically significant product term would suggest a violation of the PO assumption. However, if there are sparse data corresponding to particular intervals, it will not be practical to carry out such a test on those intervals.

Alternative Binary Model

$$\log(-\log(1 - P(Y = 1))) = \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 TRT$$

where $1 - P(Y = 1)$ is the probability of surviving a given interval conditioned on survival of previous intervals

Logistic regression is not the only type of binary regression that may be considered for interval-censored data. An alternative binary model (shown on the left) uses the **complementary log–log link** function rather than the **logit link** function that is used for the more familiar logistic regression.

Complementary log–log link

- Log–log survival modeled as linear function of regression parameters

Logit link

- Log odds of failure modeled as linear function of regression parameters

A model using a complementary log–log link function expresses the log negative log survival probability as a linear function of regression parameters. By contrast, a model using a logit link function expresses the log odds of failure (i.e., getting the event) as a linear function of regression parameters.

Complementary log–log model is PH model

- HR (TRT = 1 vs. TRT = 0) = $\exp(\beta_4)$
- HR constant over time

The complementary log–log binary model is a proportional hazards model. The hazard ratio comparing TRT = 1 to TRT = 0 is obtained by exponentiating β_4.

Log–log survival curves:
parallel $\Rightarrow$ additive effects
$\Rightarrow$ PH

Complementary log–log link:
additive effects on log–log scale
$\Rightarrow$PH

Recall we can use log–log survival curves to evaluate the PH assumption for a Cox model. If the effects are additive (e.g., parallel for TRT = 1 and TRT = 0) then the PH assumption is assumed to hold. The underlying idea is similar for the complementary log–log link function in that additive effects are assumed on the log–log scale (e.g., comparing TRT = 1 to TRT = 0).

In theory
- Survival time is continuous

In practice
- Survival time measured in intervals
  - If event occurred in month 7 then event occurred in an interval of time

In theory, the time-to-event variable in survival analyses is thought of as a continuous variable. In practice, however, the time variable is typically an interval of time. For example, if time is measured in months and an event occurs in month 7 then the event is recorded as having occurred in a specific *interval* lasting a month.

Discrete survival analysis

- Discrete time
- for example, number of menstrual cycles to pregnancy rather than time to pregnancy
  - Fraction of cycle does not make sense

Discrete survival analysis is a survival analysis in which the outcome variable is discrete, both in theory and in practice. For example, consider a study in which women who stop using oral contraception are followed until pregnancy. The outcome is defined as the number of menstrual cycles until pregnancy. The number of cycles rather than the time to pregnancy is used because the cycle length varies among women and a woman ovulates only once per menstrual cycle (i.e., one opportunity per cycle to become pregnant). The number of cycles is a discrete outcome. A fraction of a cycle does not make sense.

Analyzing discrete survival data

- Can use binary regression
- Analogous to interval-censored data
  - Discrete outcome — subjects survive discrete units of time
  - Interval outcomes — subjects survive intervals of time

Binary regression, as described in this section, can be applied for discrete survival outcomes in a similar manner to that described for interval-censored outcomes. With this method, subjects can be conceptualized as surviving discrete units of time analogously as subjects surviving continuous intervals of time.

XII. Frailty Models

Frailty

- Random component
- Accounts for extra variability from unobserved factors

In this section we consider the inclusion of frailty to a survival model. Frailty is a random component designed to account for variability due to unobserved individual-level factors that is otherwise unaccounted for by the other predictors in the model.

Conceptualize S(t) two ways:

- For an individual
- Averaging over a theoretical large population

Consider a survival model with a continuous age variable and dichotomous smoking status variable as the only predictors. Under this model the survival function for a 33-year-old smoker might be conceptualized in different ways. One way is as the survival function for an individual 33-year-old smoker. The second way is as some kind of averaging over a theoretical large population of 33-year-old smokers.

With Frailty Component

Jake and Blake
1. May have different S(t) due to unobserved factors
2. Extra source of variability in outcome (e.g., more variation than expected under Weibull)

Now suppose a "frailty" component is included in the model. Under this model, we can conceptualize survival functions specific to each individual. If Jake and Blake are both 33-year-old smokers, not only might their observed failure times be different, but under this model **their individual survival functions could also be different**. Jake may be more "frail" than Blake due to unobserved factors accounting for individual level differences in his hazard and survival functions. These unobserved factors may contribute an extra layer of heterogeneity, leading to greater variability in survival times than might be expected under the model (e.g., Weibull) without the frailty component.

Without Frailty Component

Jake and Blake
1. Have same S(t)
2. May have different event times because event time is random, following some distribution (e.g., Weibull)

The frailty component α ($\alpha > 0$)

- Unobserved multiplicative effect on hazard
- Follows distribution $g(\alpha)$ with $\mu = E(\alpha) = 1$
- $Var(\alpha) = \theta$, parameter to be estimated

The frailty α is an unobserved multiplicative effect on the hazard function assumed to follow some distribution $g(\alpha)$ with $\alpha > 0$ and the mean of α equal to 1. The variance of α is a parameter θ (theta) that is typically estimated from the data.

Hazard and survival conditioned on frailty

$h(t|\alpha) = \alpha h(t)$

$S(t|\alpha) = S(t)^{\alpha}$

An individual's hazard function conditional on the frailty can be expressed as α **multiplied by h(t)**. Using the relationship between the survival and hazard functions, the corresponding conditional survival function can be expressed as **S(t) raised to the** α **power**.

$\alpha > 1$

- Increased hazard: $\alpha h(t) > h(t)$
- Decreased survival: $S(t)^{\alpha} < S(t)$

$\alpha < 1$

- Decreased hazard: $\alpha h(t) < h(t)$
- Increases survival: $S(t)^{\alpha} > S(t)$

$\alpha = 1$ (average frailty): $\alpha h(t) = h(t)$

Individuals with $\alpha > 1$ have an increased hazard and decreased probability of survival compared to those of average frailty ($\alpha = 1$). Similarly, individuals with $\alpha < 1$ have a decreased hazard and increased probability of survival compared to those of average frailty.

Survival functions (with frailty models)

1. Conditional, $S(t|\alpha)$, individual level
2. Unconditional, $S_U(t)$, population level

Unconditional survival function $S_U(t)$

$$S_U(t) = \int_0^{\infty} S(t|\alpha) g\{\alpha\} d\alpha$$

$$h_U(t) = \frac{-d[S_U(t)]/dt}{S_U(t)}$$

With frailty models, we distinguish the individual level or **conditional survival function** $S(t|\alpha)$ discussed above, from the population level or **unconditional survival function** $S_U(t)$, which represents a population average. Once the frailty distribution $g(\alpha)$ is chosen, the unconditional survival function is found by integrating over the conditional survival function $S(t|\alpha)$ times $g(\alpha)$, with respect to α. The corresponding unconditional hazard $h_U(t)$ can then be found using the relationship between the survival and hazard functions (see left).

Frailty distribution g(α), α > 0, E(α) = 1

Stata offers choices for g(α)
1. Gamma
2. Inverse-Gaussian

Both distributions parameterized in terms of θ, where Var(α) = θ

Any distribution for α > 0 with a mean of 1 can theoretically be used for the distribution of the frailty. Stata supports two distributions: the **gamma distribution** and the **inverse-Gaussian distribution** for the frailty. With the mean fixed at 1, both these distributions are parameterized in terms of the variance θ and typically yield similar results.

EXAMPLE

Vet Lung Cancer Trial
Predictors:
TX (dichotomous: 1 = standard, 2 = test)
PERF (continuous: 0 = worst, 100 = best)
DD (disease duration in months)
AGE (in years)
PRIORTX (dichotomous: 0 = none, 10 = some)

Model 1. No Frailty

Weibull regression (PH form)
Log likelihood = −206.20418

_t	Coef.	Std. Err.	z	p > |z|
tx	.137	.181	0.76	0.450
perf	−.034	.005	−6.43	0.000
dd	.003	.007	0.32	0.746
age	−.001	.009	−0.09	0.927
priortx	−.013	.022	−0.57	0.566
_cons	−2.758	.742	−3.72	0.000
/ln_p	−.018	.065	−0.27	0.786
p	.982	.064		
1/p	1.02	.066		

To illustrate the use of a frailty model, we apply the data from the Veteran's Administration Lung Cancer Trial described in Chapter 5. The exposure of interest is treatment status TX (standard = 1, test = 2). The control variables are performance status (PERF), disease duration (DD), AGE, and prior therapy (PRIORTX), whose coding is shown on the left. The outcome is time to death (in days).

Output from running a Weibull PH model without frailty using Stata software is shown on the left (Model 1). The model can be expressed: $h(t) = \lambda p t^{p-1}$ where

$$\lambda = \exp(\beta_0 + \beta_1 \text{TX} + \beta_2 \text{PERF} + \beta_3 \text{DD} + \beta_4 \text{AGE} + \beta_5 \text{PRIORTX}).$$

The estimate of the hazard ratio comparing TX = 2 vs. TX = 1 is exp(0.137) = 1.15 controlling for performance status, disease duration, age, and prior therapy. The estimate for the shape parameter is 0.982 suggesting a slightly decreasing hazard over time.

EXAMPLE: (continued)

Model 2. With Frailty

Weibull regression (PH form)
Gamma frailty
Log likelihood = −200.11338

_t	Coef.	Std. Err.	z	P>|z|
tx	.105	.291	0.36	0.719
perf	−.061	.012	−5.00	0.000
dd	−.006	.017	−0.44	0.663
age	−.013	.015	−0.87	0.385
priortx	−.006	.035	−0.18	0.859
_cons	−2.256	1.100	−2.05	0.040
/ln_p	.435	.141	3.09	0.002
/ln_the	−.150	.382	−0.39	0.695
p	1.54	.217		
1/p	.647	.091		
theta	.861	.329		

Likelihood ratio test of theta = 0:
chibar2(01) = 12.18
Prob>=chibar2=0.000

Model 2 (output on left) is the same Weibull model as Model 1 except that a frailty component has been included. The frailty in Model 2 is assumed to follow a gamma distribution with mean 1 and variance equal to theta (θ). The estimate of theta is 0.861 (bottom row of output). A variance of zero (theta = 0) would indicate that the frailty component does not contribute to the model. A likelihood ratio test for the hypothesis theta = 0 is shown directly below the parameter estimates and indicates a chi-square value of 12.18 with 1 degree of freedom yielding a highly significant p-value of 0.000 (rounded to three decimals).

Notice how all the **parameter estimates are altered with the inclusion of the frailty**. The estimate for the shape parameter is now 1.54, quite different from the estimate 0.982 obtained from Model 1. The inclusion of frailty not only has an impact on the parameter estimates but also complicates their interpretation.

Comparing Model 2 with Model 1

- There is one additional parameter to estimate in Model 2
- The actual values of individuals' frailty are not estimated in Model 2
- The coefficients for the predictor variables in Models 1 and 2 have different estimates and interpretation
- The estimate of the shape parameter is <1.0 for Model 1 and >1.0 for Model 2

Before discussing in detail how the inclusion of frailty influences the interpretation of the parameters, we overview some of the key points (listed on the left) that differentiate Model 2 (containing the frailty) and Model 1.

Model 2 contains one additional parameter, the variance of the frailty. However, the actual values of each subject's frailty are not estimated. The regression coefficients and Weibull shape parameter also differ in their interpretations for Model 2 compared to Model 1. We now elaborate on these points.

Model 2

Hazard for jth individual:

$h_j(t|\alpha_j) = \alpha_j h(t)$ j = 1, 2,. . ., n

where $h(t) = \lambda p t^{p-1}$
with $\lambda = \exp(\beta_0 + \beta_1 \text{TX} + \beta_2 \text{PERF} + \beta_3 \text{DD} + \beta_4 \text{AGE} + \beta_5 \text{PRIORTX})$
and where $\alpha \sim$ gamma ($\mu = 1$, variance = θ)

For Model 2 we can express the Weibull model with a gamma distributed frailty in terms of the individual level hazard for the jth subject.

If α_j denotes the frailty for the jth subject, then that subject's hazard $\mathbf{h_j}$ **(t|α_j)** can be expressed as α_j **multiplied by h(t)**, where h(t) is the Weibull hazard function parameterized in terms of the predictor variables and their regression coefficients (see left).

α_j not estimable
- An α_j associated with each subject
- Too many parameters

The values for each α_j are not estimable because there is a level of frailty associated with each data point. If we tried to estimate each subject's frailty, then there would be more parameters to estimate than observations in the dataset and the model would be overparameterized.

Rather, var[g(α)] is estimated
- Gamma is 2-parameter distribution
  - Mean set at 1.0
  - $\theta = \text{Var}(\alpha)$ is estimated

Rather, the variance of the frailty is estimated. The gamma distribution is a two-parameter distribution. Because the mean is set at 1, we need only estimate its variance to fully specify the frailty distribution.

Interpreting coefficients in Model 2

$$\widehat{HR} = \exp\left(\hat{\beta}_1\right) = 1.11$$

Estimates HR comparing two individuals
- With same α
- One with TX = 2, other with TX = 1
- With same levels of other predictors

The estimated coefficient for TX using Model 2 is 0.105. By exponentiating, we obtain exp (0.105) = 1.11. This is the **estimated hazard ratio for two individuals having the same frailty** in which one takes the test treatment and the other takes the standard treatment controlling for the other co-variates in the model. Thus, for two individuals with the same frailty, we can use the coefficient estimates from Model 2 to estimate the ratio of conditional hazards.

Recall: $h(t|\alpha) = \alpha h(t)$

$TX = 1$: $h_1(t|\alpha_1) = \alpha_1 h_1(t)$
$TX = 2$: $h_2(t|\alpha_2) = \alpha_2 h_1(t)$

If $\frac{h_2(t)}{h_1(t)} = \exp(\beta_1)$

then $\frac{\alpha_1 h_1(t)}{\alpha_2 h_2(t)} = \exp(\beta_1)$

only if $\alpha_1 = \alpha_2$

To clarify, recall that the individual level or conditional hazard function can be expressed as a multiplied by h(t). Suppose $h_1(t|\alpha_1)$ and $h_2(t|\alpha_2)$ are the conditional hazard functions for individuals who use the standard and test treatments, respectively, at the mean levels of the other covariates. If the ratio of $h_2(t)$ and $h_1(t)$ equals $\exp(\beta_1)$, then the ratio of $h_2(t|\alpha_2)$ and $h_1(t|\alpha_1)$ equals $\exp(\beta_1)$ only if the individuals have the same level of frailty (i.e., $\alpha_1 = \alpha_2$; see left).

Another interpretation for $\exp(\beta_1)$

- Ratio of conditional hazards from the same individual
- Effect for individual taking test rather than standard treatment

Another way to interpret the exponentiated coefficient for TRT, $\exp(\beta_1)$, is as a ratio of conditional hazards from the **same** individual. This measure can be used to estimate an effect for an individual taking the test treatment instead of the standard treatment.

Model 1 ($\hat{p} = 0.982$)
Decreasing hazard for individual and population because ($\hat{p} < 1$)

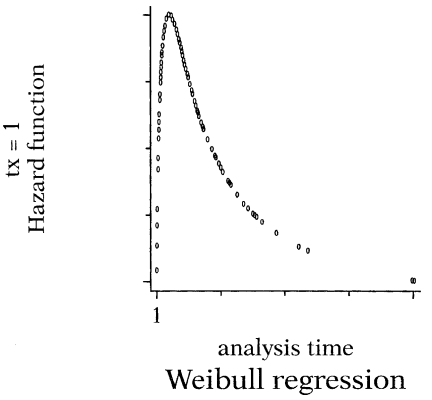
Model 2 ($\hat{p} = 1.54$)
Complication:
Individual level hazard
vs
Population level hazard

A somewhat striking difference in the output from Model 1 and Model 2 is the estimate of the shape parameter. The hazard estimated from Model 1 (without the frailty) is estimated to decrease over time because $\hat{p} < 1$. By contrast, the estimated individual level hazard from Model 2 is estimated to increase over time because $\hat{p} > 1$. However, the interpretation of the shape parameter in Model 2 has an additional complication that should be considered before making direct comparisons with Model 1. For frailty models, we have to distinguish between the **individual level** and **population level** hazards.

For Model 2
Conditional hazard increases
but
unconditional hazard unimodal

Although the estimated individual level or **conditional** hazard from Model 2 is estimated to increase over time, the estimated population level or **unconditional** hazard does not strictly increase. The unconditional hazard first increases but then decreases to zero, resulting in a **unimodal** shape due to the effect of the frailty, as we will now explain.

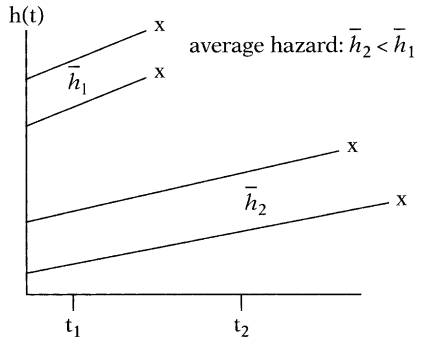
Estimated unconditional hazard Model 2 (TX = 1, mean level for other covariates, $\hat{p} = 1.54$)



On the left is a plot (from Model 2) of the estimated unconditional hazard for those on standard treatment (TX = 1) with mean values for the other covariates. The graph is unimodal, with the hazard first increasing and then decreasing over time. So each individual has an estimated increasing hazard ($\hat{p} = 1.54$), yet the hazard averaged over the population is unimodal, rather than increasing. How can this be?

The answer is that the population is comprised of individuals with different levels of frailty. The more frail individuals ($\alpha > 1$) have a greater hazard and are more likely to get the event earlier. Consequently, over time, the "at risk group" has an increasing proportion of less frail individuals ($\alpha > 1$), decreasing the population average, or unconditional, hazard.

Four increasing individual level hazards, but average hazard decreases from t_1 to t_2



To clarify the above explanation, consider the graph on the left in which the **hazards for four individuals increase** linearly over time until their event occurs. The two individuals with the highest hazards failed between times t_1 and t_2 and the other two failed after t_2. Consequently, the average hazard ($\bar{h}_2$) of the two individuals still at risk at t_2 is less than the average hazard ($\bar{h}_1$) of the four individuals at risk at t_1. Thus the average hazard of the "at risk" population **decreased** from t_1 to t_2 (i.e., $\bar{h}_2 < \bar{h}_1$) because the individuals surviving past t_2 were less frail than the two individuals who failed earlier.

Frailty Effect

$h_U(t)$ eventually decreases
because
"at risk group" becoming less frail over time

This property, in which the unconditional hazard eventually decreases over time because the "at risk group" has an increasing proportion of less frail individuals, is called the **frailty effect**.

Unconditional hazard $h_U(t)$ with gamma frailty

$$h_U(t) = \frac{h(t)}{1 - \theta \ln[S(t)]}$$

The unconditional hazard function $h_U(t)$, **with gamma frailty** is shown on the left.

If $\theta = 0$ then $h_U(t) = h(t)$ (no frailty)

If $\theta = 0$, then $h_U(t)$ reduces to $h(t)$ indicating that there is no frailty.

For Model 2:

- $h(t)$ and $S(t)$ are Weibull
- At $t = 0$
  - $h_U(t) = h(t)$ (increasing)
- As t gets large
  - If $\theta > 0$ then $h_U(t) \to 0$
- So $h_U(t)$ increases and then decreases (unimodal)

An examination of the expression for $h_U(t)$ gives further insight into how we obtained an estimated unconditional hazard of unimodal shape. $S(t)$ and $h(t)$ represent the survival and hazard functions ignoring the frailty, which for Model 2 corresponds to a Weibull distribution. If $t = 0$ then $h_U(t) = h(t)$, which for Model 2 yields an estimated increasing hazard. As t gets larger, and if $\theta > 0$, the denominator gets larger (because $\ln[S(t)]$ is negative) until eventually $h_U(t)$ approaches zero. So $h_U(t)$ is increasing at $t = 0$ but eventually decreases to zero, which means at some point in time, $h_U(t)$ changes direction.

Population level hazards (with gamma frailty)

$$h_{U1}(t) = \frac{h_1(t)}{1 - \theta \ln[S_1(t)]} \quad \text{for TX} = 1$$

$$h_{U2}(t) = \frac{h_2(t)}{1 - \theta \ln[S_2(t)]} \quad \text{for TX} = 2$$

Ratio of unconditional hazards (not PH)

$$\frac{h_{U2}(t)}{h_{U1}(t)} = \frac{h_2(t)}{h_1(t)} \times \frac{1 - \theta \ln[S_1(t)]}{1 - \theta \ln[S_2(t)]}$$

A consequence of the frailty effect is the need to distinguish between the ratio of individual level hazards and the ratio of population level hazards. For the population level hazards, the **PH assumption is violated** when a gamma (or inverse-Gaussian) distributed frailty is added to a PH model. To see this for gamma frailty, let $h_{U1}(t)$ and $h_{U2}(t)$ be the unconditional hazard functions representing the standard and test treatments, respectively, at the mean levels of the other covariates. The ratio of these hazards is shown on the left.

If

$$\frac{h_2(t)}{h_1(t)} = \exp(\beta_1)$$

then

$$\frac{h_{U2}(t)}{h_{U1}(t)} = \exp(\beta_1) \times \frac{1 - \theta \ln[S_1(t)]}{1 - \theta \ln[S_2(t)]}$$

not constant over time, PH violated

If the ratio of $h_2(t)$ and $h_1(t)$ equals $\exp(\beta_1)$, then the ratio of the unconditional hazards equals $\exp(\beta_1)$ times the ratio of $1 - \theta \ln[S_1(t)]$ and $1 - \theta \ln[S_2(t)]$. This latter ratio is a function of time and only cancels when t equals zero. Therefore the ratio of the unconditional hazards is not constant over time, thus violating the PH assumption.

Plots of $\hat{S}(t)$

- Generally averaged over population
  - An important consideration for frailty models

Generally, survival plots are estimated over a population average (e.g., Kaplan–Meier). When considering PH models without frailty, we do not need to distinguish between the conditional and unconditional survival functions. However, this distinction needs to be considered with frailty models.

Suppose $\ln[-\ln \hat{S}(t)]$ curves for TX start parallel but then converge over time:

1. It may be effect of TX weakens over time ⇓ PH model not appropriate
2. It may be effect of TX is constant over time but unobserved heterogeneity is in population ⇓ PH model with frailty is appropriate

Suppose we plot Kaplan–Meier log–log survival estimates evaluating the PH assumption for treatment (TX = 2 vs. TX = 1), and the plots start out parallel but then begin to converge over time. One interpretation is that the effect of the treatment weakens over time. For this interpretation, a PH model is not appropriate. Another interpretation is that the effect of the treatment remains constant over time but the plots converge due to unobserved heterogeneity in the population. For this interpretation, a PH model with frailty would be appropriate.

Model 2 (Weibull with frailty)

- Used PH parameterization
- Can equivalently use AFT parameterization

Recall, from Section VI of this chapter, that a Weibull PH model is also an AFT model. The only difference is in the way the model is parameterized. We next present the AFT form of Model 2.

Unconditional survival function $S_U(t)$ with gamma frailty $g(\alpha)$

$$S_U(t) = \int_0^\infty S(t|\alpha) g(\alpha) d\alpha$$
$$= [[1 - \theta \ln S(t)]^{-1/\theta}$$

Before stating the model, we show the unconditional survival function using gamma frailty. Recall that the unconditional survival function is obtained by integrating over the frailty, as shown on the left.

Model 3 (Weibull AFT with gamma frailty)

$$S_U(t) = [1 - \theta \ln S(t)]^{-1/\theta}$$

where $S(t) = \exp(-\lambda t^p)$ (Weibull) and

$$\frac{1}{\lambda^{1/p}} = \exp(\alpha_0 + a_1 TX + \alpha_2 PERF + \alpha_3 DD + \alpha_4 AGE + \alpha_5 PRIORTX)$$

Model 3 (the AFT form of Model 2) is presented in terms of the unconditional survival function $S_U(t)$. The unconditional survival function is a function of S(t), which represents the Weibull survival function. The Weibull survival function, in turn, is parameterized in terms of the shape parameter p and regression coefficients using AFT parameterization (see left).

Model 3 Output

Weibull regression (**AFT form**)
Gamma frailty
Log likelihood = −200.11338

_t	Coef.	Std. Err.	z	P > |z|
tx	−.068	.190	−0.36	0.721
perf	.040	.005	8.37	0.000
dd	.004	.009	0.44	0.661
age	.008	.009	0.89	0.376
priortx	.004	.023	0.18	0.860
_cons	1.460	.752	1.94	0.052
/ln_p	.435	.141	3.09	0.002
/ln_the	−.150	.382	−0.39	0.695
p	1.54	.217		
1/p	.647	.091		
theta	.861	.329		

Likelihood ratio test of theta = 0:
chibar2(01) = 12.18
Prob>=chibar2 = 0.000

The output for Model 3, shown on the left, is similar to that obtained from Model 2. The estimates for theta and p are identical to those obtained from Model 2. The difference is that the regression coefficients obtained with Model 3 use AFT parameterization, i.e., multiply by $-p = -1.54$ to get the PH coefficient estimates in Model 2.

$$\hat{\gamma}(\text{TX = 2 vs. 1}) = \exp(-0.068) = 0.93$$

An estimated acceleration factor of 0.93 **comparing two individuals with the same level of frailty**, for the effect of treatment (TX = 2 vs. TX = 1) and controlling for the other covariates, is obtained by exponentiating the estimated coefficient (−0.068) of the TX variable.

Comparing individuals with same α

Interpreting $\hat{\gamma}$

- Taking test treatment reduces individual's median survival time by factor of 0.93
- Suggests slightly harmful effect
- $\hat{\alpha}_1$ is not significant (p = 0.721)

Another interpretation for this estimate is that an individual taking the test treatment instead of the standard treatment reduces her median survival time (i.e., contracts her individual level survival function) by an estimated factor of 0.93. This estimate suggests a slight harmful effect from the test treatment compared to the standard treatment. However, the estimated coefficient for TX is not significant, with a p-value of 0.721.

PH assumption
Individual level PH $\nRightarrow$ Population level PH

AFT assumption
Individual level AFT $\Rightarrow$ Population level AFT

A key difference between the PH and AFT formulations of this model is that if the AFT assumption holds at the individual level, then it will also hold at the population level using the gamma (or inverse-Gaussian) distributed frailty.

Population level survival (with gamma frailty)

$S_{U1}(t) = [[1 - \theta \text{ In } S_1(t)]^{-1/\theta}$
$S_{U2}(t) = [[1 - \theta \text{ In } S_2(t)]^{-1/\theta}$

To see this for gamma frailty, let $S_{U1}(t)$ and $S_{U2}(t)$ be the unconditional survival functions representing the standard and test treatments respectively, at the mean levels of the other covariates.

Also let γ represent the individual level acceleration factor for treatment; that is, $S_1(t) = S_2(\gamma t)$. Then $S_{U1}(t) = S_{U2}(\gamma t)$ (see left).

If $S_1(t) = S_2(\gamma t)$
then

$$\begin{aligned} S_{U1}(t) &= [[1 - \theta In S_1(t)]^{-1/\theta} \\ &= [[1 - \theta In S_2(\gamma t)]^{-1/\theta} \\ &= S_{U2}(\gamma t) \end{aligned}$$

Thus,
Individual level AFT
$\Rightarrow$ Population level AFT

Thus, for models with gamma frailty, if the AFT assumption holds at the individual level then it also holds at the population level.

Coefficient estimates from Model 3

- Applies to individual or population
- Interpretation of $\exp(\hat{\alpha}_1) = 0.93$
  - Median survival time for individual reduced by factor of 0.93
  - Median survival time reduced in population by factor of 0.93

The coefficient estimates obtained from Model 3 can therefore be used at the population level as well as the individual level. So another interpretation for the estimated acceleration factor for treatment is that the test treatment reduces the median survival time in the population by an estimated factor of 0.93.

Models 2 and 3:
Same model, different parameterization
Same estimates for
$S(t)$, $S_U(t)$, $h(t)$, $h_U(t)$

Model 2 and Model 3 are the same model but use different parameterizations. The models provide identical estimates for the hazard and survival functions.

Models 2 and 3: Weibull with gamma frailty

- Unimodal unconditional hazard

Log-logistic model

- Accommodates unimodal hazard without a frailty component

Recall that the estimated unconditional hazard function obtained from this frailty model is of unimodal shape. Alternatively, a log-logistic (or lognormal) model, which accommodates a unimodal-shaped hazard function, could have been run without the frailty (see Practice Exercises 8 to 11 for comparison).

Parametric likelihood with frailty

- Uses $f_U(t)$, where $f_U(t) = h_U(t) S_U(t)$
- Formulated similarly to that described in Section X with $f_U(t)$ replacing $f(t)$
- Additional parameter θ

The likelihood for Model 3 can be formulated using the unconditional probability density function $f_U(t)$ which is the product of the unconditional hazard and survival functions. The likelihood is constructed in a similar manner to that described previously in this chapter except that $f_U(t)$ is used for the likelihood rather than $f(t)$ (see Section X). The main difference is that there is one additional parameter to estimate, the variance of the frailty.

Shared Frailty

- Clusters share same frailty
- For example, subjects from same family may share unobserved factors
  - Shared frailty designed to account for such similarities

Another type of frailty model is the **shared frailty model**. With this model, clusters of subjects are assumed to share the same frailty. For example, subjects from the same family may be similar with respect to some unobserved genetic or environmental factors. Allowing family members to share the same frailty is designed to account for such similarities.

Unshared Frailty

- The type of frailty we have described previous to this point
- Frailty distributed independently among subjects

By contrast, the frailty described previous to this point (**unshared frailty**) has been assumed to be distributed independently among subjects.

Shared Frailty Models

- Similar to random effect regression models
- Accounts for within-cluster correlation
- θ is a measure of the degree of correlation

Adding shared frailty to a survival model plays an analogous role to that of adding a random effect to a linear regression as a way to account for correlation between clusters of observations (Kleinbaum and Klein 2010). The estimate for the variance parameter θ in a shared frailty model can be thought of as a measure of the degree of correlation, where $\theta = 0$ indicates no within-cluster correlation.

Hazard conditional on shared frailty (for jth subject in kth cluster)

$h_{jk}(t|\alpha_k) = \alpha_k h_{jk}(t)$
where
$$h_{jk}(t) = h(t|\mathbf{X}_{jk})$$
for $j = 1, 2, \ldots, n_k$
and total n_k subjects in k^{th} cluster

If family is the cluster variable, then
subjects of same family have same α_k

For a shared frailty model, the conditional hazard function for the jth subject from the kth cluster can be expressed as α_k **multiplied by $\mathbf{h_{jk}(t)}$** where $h_{jk}(t)$ depends on the subject's covariates $\mathbf{X}_{jk}$. Notice that the frailty α_k is subscripted by k, but not by j. This indicates that subjects from the same cluster share the same frailty. If, for example, subjects are clustered by family, then subjects from the same family are assumed to have the same frailty.

Shared and unshared frailty

- Fundamentally the same
  - Accounts for variation due to unobservable factors
- Difference in data to which they are applied
  - Affects interpretation and methods of estimation

The frailty in a shared frailty model or unshared frailty model **is fundamentally the same**, a random effect to account for a source of variation due to unobservable, or latent, factors. **However, the data to which the shared and unshared frailty is applied are different, affecting differences in interpretation and methods of estimation**.

Unshared frailty models

- Subjects assumed independent

Shared frailty models

- Accounts for dependence among subjects who share frailty

For unshared frailty models, a subject's survival is assumed to be independent of the survival of other subjects in the study population. For shared frailty models, however, the frailty accounts for dependence among subjects who share the same frailty. Shared frailty provides an approach to account for correlation in the data due to unobservable factors common within clusters of subjects.

Likelihood for shared frailty models

- More complicated than for unshared frailty models
- Unconditional contribution of each cluster formulated separately by integrating out $g(\alpha)$
- Full likelihood formed as product of unconditional contribution from each cluster

The formulation of the likelihood is more complicated for shared frailty models than it is for unshared frailty models. To construct the shared frailty likelihood, the unconditional contribution for each cluster of subjects is formulated separately by integrating out the frailty from the product of each subject's conditional contribution. The full likelihood is then formulated as the product of the contributions from each cluster (see Gutierrez 2002 for details).

Shared frailty in Cox model

- Provided by Stata
  - Only gamma distributed shared frailty available
- Accounts for within-group correlation

Up to this point we have discussed frailty in terms of parametric models. Stata (version 8) allows shared frailty to be included in a Cox model in order to account for within-group correlation. The conditional hazard function for the jth subject from the kth cluster can be expressed as α_k multiplied by the baseline hazard $h_0(t)$ multiplied by $\exp(\beta \mathbf{X}_{jk})$. The frailty component is assumed to follow some distribution even though the distribution is unspecified for the rest of the model. Stata only allows a gamma distribution for the frailty to be included with a Cox model.

Cox shared frailty model

$h_{ij}(t|\alpha_j) = \alpha_k h_0(t) \exp(\beta X_{jk})$
for $j = 1, 2, \ldots, n_k$
total of n_k subjects in kth cluster

PH violation of $h_U(t)$ in Cox model

- if gamma-distributed frailty included
- Interpreting coefficient estimates
  - Only used for HR estimates among those who share same α

If a gamma-distributed frailty component is added to the Cox model, then the PH assumption is not satisfied for the unconditional hazards. In this framework, the frailty in a Cox model can be thought of as a source of random error that causes violation of the PH assumption at the population level. Consequently, care must be taken in the interpretation of the coefficient estimates. They can only be used to obtain estimates for hazard ratios conditioned on the same level of frailty.

Recurrent events

- Multiple events from same subject
- Events from same subject may be correlated
- Clusters are formed representing each subject
  - Different subjects do not share frailty
  - Observations from same subject share frailty

Shared frailty models can also be applied to recurrent event data. It is reasonable to expect that multiple events occurring over follow-up from the same individual would be correlated. To handle within-subject correlation, clusters are formed, each containing observations from the same subject. In this setting, it is not the case that different subjects share the same frailty. Rather, multiple observations representing the same subject share the same frailty.

Recurrent events:

- Topic of next chapter (Chapter 8)

Survival analyses on recurrent events are the focus of the next chapter (Chapter 8) of this text. An example of a Weibull model with shared frailty applied to recurrent event data is presented in the next chapter.

XIII. Summary

Parametric Models

- Assume distribution for survival time
- Distribution specified in terms of parameters
- Parameters estimated from data

In this chapter we presented parametric survival models as an alternative to the Cox model. They are called parametric models because the distribution of the time-to-event variable is specified in terms of unknown parameters, which are estimated from the data. Distributions that are commonly utilized are the exponential, the Weibull, the log-logistic, the lognormal, and the generalized gamma.

$f(t)$ specified $\Rightarrow$ corresponding $S(t)$, $h(t)$ also determined

Moreover,
Specifying one of f(t), S(t), or h(t) determines all three functions

More precisely, for parametric survival models, it is the probability density function f(t) of the distribution that is specified in terms of the parameters. Once f(t) is specified, the corresponding survival and hazard functions S(t) and h(t) can also be determined. Moreover, specifying any one of the probability density function, survival function, or hazard function allows the other two functions to be determined.

Parametric models

- Need not be PH models
- Many are AFT models

The proportional hazards (PH) assumption is the underlying assumption for a Cox PH model. However, parametric survival models need not be proportional hazards models. Many parametric models are acceleration failure time (AFT) models rather than proportional hazards models.

Acceleration factor (γ)

- Key measure of association in AFT models
- Describes stretching or contraction of S(t)

The acceleration factor (γ) is the key measure of association obtained in an AFT model. It describes the "stretching out" or contraction of survival functions when comparing one group to another. If $S_1(t)$ and $S_2(t)$ are the survival functions for Group 1 and Group 2, respectively, then the AFT assumption can be expressed as $S_2(t) = S_1(\gamma t)$.

AFT assumption

$$S_2(t) = S_1(\gamma t)$$
$$\uparrow \quad \uparrow$$
Group 2 Group 1

Detailed examples presented:

- Exponential model
- Weibull model
- Log-logistic model

We presented detailed examples of the exponential, Weibull, and log-logistic model using the remission dataset.

Exponential Model

- h(t) = λ (constant hazard)
- Special case of Weibull model

The underlying assumption for an exponential model, a special case of the Weibull model, is that the hazard function is constant over time (i.e., h(t) = λ). The Weibull model is unique in that if the PH assumption holds then the AFT assumption also holds (and vice versa). The log-logistic model does not satisfy the PH assumption. However, if the AFT assumption holds in a log-logistic model, then the proportional odds (PO) assumption also holds (and vice versa).

Weibull Model

- AFT ⇔ PH

Log-logistic Model

- Not a PH model
- AFT ⇔ PO

PO assumption

$$\text{OR} = \frac{S(t,x^*)/[1-S(t,x^*)]}{S(t,x)/[1-S(t,x)]}$$

OR is constant over time

The idea underlying the proportional odds assumption is that the survival (or failure) odds ratio comparing two specifications of covariates remains constant over time.

Graphical Evaluation

Weibull and Exponential

- Plot ln[−ln $\hat{S}$(t)] against ln(t)

Log-logistic:

- Plot $\ln\left[\frac{\hat{S}(t)}{(1-\hat{S}(t))}\right]$ against ln(t).

Check for linearity

We presented graphical approaches for evaluating the appropriateness of the exponential, Weibull, and log-logistic model by plotting a function of the Kaplan–Meier survival estimates $\hat{S}(t)$ against the log of time and then checking for linearity.

For evaluation of the exponential and Weibull assumptions, the ln[−ln $\hat{S}(t)$] is plotted against ln(t) and for evaluation of the log-logistic assumption the log odds of $\hat{S}(t)$ is plotted against ln(t).

Presented other parametric models

- Generalized gamma model
- Lognormal model
- Gompertz model

We briefly discussed other parametric models such as the generalized gamma, lognormal, and Gompertz models and showed additional parametric approaches such as modeling ancillary (shape) parameters as a function of predictor variables.

Contributions to Likelihood

If event at t, contributes f(t)

If censored, integrate over f(t)

$$\int_0^{t_1} f(t)dt : \text{left} - \text{censored at } t_1$$

$$\int_{t_1}^{\infty} f(t)dt : \text{right} - \text{censored at } t_1$$

$$\int_{t_1}^{t_2} f(t)dt : \text{interval} - \text{censored from } t_1 \text{ to } t_2$$

The parametric likelihood was developed and includes a discussion of left-, right-, and interval-censored data. If a subject has an event at time t, then that subject's contribution to the likelihood is f(t). On the other hand, if a subject is censored (i.e., exact time of event unknown), then the subject's contribution to the likelihood is found by integrating over f(t). The integration limits are determined by the time and type of censorship (see left).

Full likelihood (L)

$$L = \prod_{j=1}^{N} L_j \quad \text{j} = 1, 2, \ldots, \text{N}$$

where L_j is the contribution from jth subject and N = # of subject's

Assuming independence among subjects, the full likelihood can be formulated as a product of each subject's contribution.

Binary regression for interval-censored data

- Follow-up divided into intervals
  - Allows for multiple observations per subject
- Binary outcome variable defined
  - Indicates survival or failure over each interval

We showed how binary regression could be applied to interval-censored data by defining a dichotomous outcome variable indicating subjects' survival or failure over each interval of their follow-up. The data layout for this type of analysis allows multiple observations per subject, representing intervals of survival prior to failure (or censorship).

Binary regression for discrete survival analysis

- Analogous to interval-censored data
  - Discrete outcome–subjects survive discrete units of time
  - Interval outcomes–subjects survive intervals of time

Binary regression can also be used for discrete survival analysis in which the "time-to-event" variable is considered discrete rather than continuous. The data layout is similar to that for interval-censored data except subjects are conceptualized as surviving discrete units of time rather than continuous intervals of time.

Frailty, α

$$h(t|\alpha) = \alpha h(t)$$

multiplicative from effect on h(t)
mean = 1, varience = θ

θ estimated from data

We concluded with a discussion of frailty models. The frailty α is a multiplicative random effect on the hazard designed to account for individual-level unobserved factors that add an extra layer of variability beyond what has already been specified in the model. The frailty is generally assumed to follow a distribution with mean equal to 1 and is typically parameterized in terms of the variance θ which is estimated from the data.

Chapters

1. Introduction to Survival Analysis
2. Kaplan–Meier Curves and the Log-Rank Test
3. The Cox Proportional Hazard Model
4. Evaluating the Proportional Hazards Assumption
5. The Stratified Cox Procedure
6. Extension of the Cox Proportional Hazards Model for Time-Dependent Covariates
- ✓7. Parametric Survival Models

Next:

8. Recurrent Event Survival Analysis

The presentation is now complete. The reader can review the detailed outline that follows and then answer the practice exercises and test.

In the next chapter (8) entitled "Recurrent Event Survival Analysis," we consider approaches for analyzing data in which individuals may have more than one event over the course of their follow-up.

Detailed Outline

I. Overview (pages 292–294)

A. Parametric Survival Models

i. Outcome assumed to follow specified distribution

ii. Weibull, exponential (a special case of the Weibull), log-logistic, lognormal, and generalized gamma are supported with popular software (SAS and Stata)

iii. Contrasts with Cox model in which baseline hazard and survival functions are not specified

II. Probability Density Function in Relation to the Hazard and Survival Function (pages 294–295)

A. If any one of the hazard h(t), survival S(t), or probability density f(t) functions is known then the other two functions can be determined.

B. If f(t) is specified, then $S(t) = \int_t^\infty f(u)du$

C. If S(t) is specified, then $h(t) = (-d[S(t)]/dt)/S(t)$ and $f(t) = (-d[S(t)])/dt$

D. If h(t) is specified, then $S(t) = \exp\left(-\int_0^t h(u)du\right)$ and $f(t) = h(t)S(t)$

III. Exponential Example (pages 295–297)

A. Hazard is constant (i.e., not a function of time) in an exponential model

i. Stronger assumption than the PH assumption that the HR is constant

B. Exponential PH model (one predictor X_1)

i. In terms of the hazard: $h(t) = \lambda$ where $\lambda = \exp(\beta_0 + \beta_1 X_1)$

ii. Hazard ratio: HR ($X_1 = 1$ vs. $X_1 = 0$) = $\exp(\beta_1)$

IV. Accelerated Failure Time Assumption (pages 298–300)

A. Underlying assumptions

i. AFT — effect of covariates is multiplicative with respect to survival time

ii. PH — effect of covariates is multiplicative with respect to the hazard

B. The acceleration factor (γ) is the key measure of association in an AFT
   i. Acceleration factor is a ratio of survival times corresponding to any fixed value of S(t); that is, t_A/t_B where A and B denote two individuals for which $S(t_A) = S(t_B)$
   ii. $S_2(t) = S_1(\gamma t)$, survival function for Group 1, $S_1(t)$ is stretched (or contracted) by a factor of γ compared to survival function for Group 2, $S_2(t)$

C. AFT illustration
   i. Dogs are said to grow older 7 times faster than humans, $S_D(t) = S_H(7t)$

V. Exponential Example Revisited (pages 300–304)

A. Exponential AFT model (one predictor X_1)
   i. In terms of survival: $S(t) = \exp(-\lambda t)$ where $\lambda = \exp[-(\alpha_0 + \alpha_1 X_1)]$
   ii. In terms of time: $t = [-\ln(S(t)] \times \exp(\alpha_0 + \alpha_1 X_1)$
   iii. Acceleration factor ($X_1 = 1$ vs. $X_1 = 0$), $\gamma = \exp(\alpha_1)$

B. An exponential PH model is an exponential AFT model (but uses different parameterization)
   i. $\beta_j = -\alpha_j$, where β_j and $-\alpha_j$ are PH and AFT parameterization for the jth covariate
   ii. $\alpha > 1$ for ($X_1 = 1$ vs. $X_1 = 0$) implies effect of $X_1 = 1$ is beneficial to survival
   iii. HR > 1 for ($X_1 = 1$ vs. $X_1 = 0$) implies effect of $X_1 = 1$ is harmful to survival

C. Exponential model is a special case of a Weibull model
   i. Graphical approach for evaluating appropriateness of exponential model is described in the section on the Weibull example

VI. Weibull Example (pages 304–309)

A. PH form of the Weibull model (one predictor X_1)
   i. In terms of the hazard: $h(t) = \lambda p t^{p-1}$ where $\lambda = \exp(\beta_0 + \beta_1 X_1)$
   ii. Hazard ratio: HR ($X_1 = 1$ vs. $X_1 = 0$) = $\exp(\beta_1)$

iii. Weibull hazard is monotonic with its direction determined by the value of the shape parameter p
   a. $p > 1$ hazard increases over time
   b. $p = 1$ constant hazard (exponential model)
   c. $p < 1$ hazard decreases over time

A. Graphical approach for evaluating appropriateness of Weibull model
   i. Plot the log negative log of the Kaplan–Meier survival estimates against the log of time for each pattern of covariates
      a. If Weibull assumption is correct then plots should be straight lines of slope p
      b. If exponential assumption is correct then plots should be straight lines with slope equal to one ($p = 1$)
      c. If plots are parallel straight lines then Weibull PH and AFT assumptions are reasonable

B. AFT form of the Weibull model (one predictor X_1)
   i. In terms of survival:
   $S(t) = \exp(-\lambda t^p) = \exp[-(\lambda^{1/p}t)^p]$ where $\lambda^{1/p} = \exp[-(\alpha_0 + \alpha_1 X_1)]$
   ii. In terms of time:
   $t = [-\ln(S(t)]^{1/p} \times \exp(\alpha_0 + \alpha_1 X_1)$
   iii. Acceleration factor ($X_1 = 1$ vs. $X_1 = 0$), $\gamma = \exp(\alpha_1)$

C. A Weibull PH model is a Weibull AFT model (but uses different parameterization)
   i. Unique property of Weibull model (exponential is special case, $p = 1$)
   ii. $\beta_j = -\alpha_j p$ where β_j and α_j are PH and AFT parameterization, respectively, for the jth covariate

VII. Log-Logistic Example (pages 309–314)

A. Log-logistic hazard function:
   $h(t) = \lambda p t^{p-1}/(1 + \lambda t^p)$.
   i. $p \leq 1$ hazard decreases over time
   ii. $p > 1$ hazard first increases and then decreases over time (unimodal)

B. Graphical approach for evaluating appropriateness of log-logistic model
   i. Plot the log of the survival odds (using KM estimates) against the log of time for each pattern of covariates
      a. if log-logistic assumption is correct then plots should be straight line of slope $-p$
      b. If plots are parallel straight lines then log-logistic proportional odds (PO) and AFT assumptions are reasonable

C. Log-logistic AFT model (one predictor X_1):
   i. In terms of survival:
   $S(t) = 1/(1+\lambda t^P) = 1/(1+(\lambda t^{1/P} t)^P)$ where $\lambda^{1/p} = \exp(-(\alpha_0 + \alpha_1 X_1))$
   ii. In terms of time:

$$t = \left[\frac{1}{S(t)} - 1\right]^{1/p} \times \exp(\alpha_0 + \alpha_1 X_1)$$

   iii. Acceleration factor ($X_1 = 1$ vs. $X_1 = 0$), $\gamma = \exp(\alpha_1)$

D. Log-logistic proportional odds (PO) model (one predictor X_1)
   i. In terms of survival: $S(t) = 1/(1 + \lambda t^p)$ where $\lambda = \exp(\beta_0 + \beta_1 X_1)$
   ii. Odds of an event (failure odds) by time t: $(1 - S(t))/S(t) = \lambda t^p$
   iii. Odds of surviving event (survival odds) beyond t: $S(t)/(1 - S(t)) = 1/\lambda t^p$
   iv. Failure odds ratio:
   HR ($X_1 = 1$ vs. $X_1 = 0$) $= \exp(\beta_1)$
      a. PO assumption is that the odds ratio is constant over time
   v. Survival odds ratio:
   HR ($X_1 = 1$ vs. $X_1 = 0$) $= \exp(-\beta_1)$
      a. Survival odds ratio is reciprocal of failure odds ratio

E. A log-logistic AFT model is a log-logistic PO model (but uses different parameterization)
   i. Log-logistic model is not a proportional hazards (PH) model
   ii. $\beta_j = -\alpha_j p$ where β_j and α_j are PO and AFT parameterization for the jth covariate
      a. Shape parameter with Stata is parameterized as gamma $= 1/p$

VIII. A More General Form of the AFT Model (pages 314–316)

A. General form with one predictor (X_1):
$\ln(T) = \alpha_0 + \alpha_1 X_1 + \epsilon$

B. Include additional parameter, σ:
$\ln(T) = \alpha_0 + \alpha_1 X_1 + \sigma\epsilon$

C. Let $\sigma = 1/p \Rightarrow \ln(T) = \alpha_0 + \alpha_1 X_1 + (1/p)\,\epsilon$

D. Additive in terms of ln(T) but multiplicative in terms of T:

$$T = \exp\left(\alpha_0 + \alpha_1 X_1 + \frac{1}{p}\epsilon\right)$$
$$= \exp[\alpha_0 + \alpha_1 X_1] \times \exp\left(\frac{1}{p}\epsilon\right)$$

E. Collapse α_0 into baseline term,let
$T_0 = \exp(\alpha_0)\exp\left(\frac{1}{p} \in\right)$:
so $T = \exp(\alpha_1 X_1) \times T_0$

IX. Other Parametric Models (pages 316–318)

A. Generalized gamma model
   i. Additional shape parameters give flexibility in distribution shape
   ii. Weibull and lognormal are special cases

B. Lognormal model
   i. ln(T) follows a normal distribution
   ii. Accommodates AFT model

C. Gompertz model
   i. PH model, not AFT model

D. Modeling failure time as an additive model
   i. Additive model with one predictor:
   $T = \alpha_0 + \alpha_1 TRT + \epsilon$ (no log link)

E. Modeling ancillary parameters
   i. Typically shape parameter p is considered a fixed constant
   ii. Can reparameterize shape parameter in terms of predictor variables and regression coefficients

X. The Parametric Likelihood (pages 318–321)

A. Product of each subject contribution (assuming independence)

B. Subject's contribution uses probability density function f(t)
   i. Subject contributes f(t) if event is observed at time t

ii. Integrate over f(t) if subject is censored
   a. Integrate from 0 to t if subject is left-censored at t
   b. Integrate from t to infinity if subject is right-censored at t
   c. Integrate over interval of censorship if subject is interval-censored

XI. Interval-Censored Data (pages 321–326)

A. Binary regression is alternative approach if data are interval-censored
B. Binary outcome variable represents survival or failure over each subinterval of subject's follow-up
C. Specify a link function when using binary regression
   i. Logit link for logistic regression
   ii. Complementary log–log link is an alternative to logistic regression
D. Discrete survival analysis
   i. Time-to-event variable is discrete
   ii. Binary regression can be applied in a similar manner to that of interval-censored data

XII. Frailty Models (pages 326–340)

A. The frailty α is an unobserved multiplicative effect on the hazard function
   i. Hazard, conditioned on the frailty, $h(t|\alpha) = \alpha h(t)$
   ii. Survival, conditioned on the frailty, $S(t|\alpha) = S(t)^{\alpha}$
B. Frailty assumed to follow some distribution $g(\alpha)$ of mean 1 and variance θ
   i. The variance θ is a parameter estimated by the data
   ii. Gamma distribution offered by Stata and R software
C. Designed to account for unobserved individual-level factors that influence survival
   i. Distinction is made between the individual-level and population-level hazards. PH assumption may hold on individual level but not on population level

D. Shared frailty models allow individuals to share the same frailty
   i. Play similar role as adding a random effect to a linear regression
   ii. Can account for within-group correlation.

XIII. Summary (pages 341–344)

Practice Exercises

Answer questions 1 to 5 as true or false (circle T or F)

T F 1. The acceleration factor comparing exposed and unexposed subjects, (E = 1 vs. E = 0), is a ratio of their median survival times (time to S(t) = 0.5), or more generally the ratio of their times to any fixed value of S(t) = q.

T F 2. Let $S_0(t)$ be the survival function for unexposed subjects (E = 0) and let $S_1(t)$ be the survival function for exposed subjects (E = 1). If $S_0(t) = S_1(3t)$ then the median survival time for the unexposed subjects is 3 times longer than the median survival time for the exposed subjects.

T F 3. The Cox proportional hazards model is a parametric model.

T F 4. if the acceleration failure time (AFT) assumption holds in a Weibull model then the proportional hazards assumption also holds.

T F 5. The hazard is assumed constant in a log-logistic model.

Questions 6 and 7 make use of the output (copied below) presented in Sections III and V containing an example of the exponential model. This example used the remission data with treatment status (coded TRT = 1 for the experimental treatment and TRT = 0 for the placebo). The exponential survival and hazard functions are, respectively, S(t) = exp(−λt) and h(t) = λ where $\lambda = \exp[-(\alpha_0 + \alpha_1 TRT)]$ for the AFT parameterization and $\lambda = \exp(\beta_0 + \beta_1 TRT)$ for the PH parameterization. The output for both the AFT and PH forms of the model are presented.

Exponential regression accelerated failure-time form $\lambda = \exp[-(\alpha_0 + \alpha_1 TRT)]$

_t	Coef.	Std. Err.	z	p>|z|
trt	1.527	.398	3.83	0.00
_cons	2.159	.218	9.90	0.00

Exponential regression log relative-hazard form $\lambda = \exp(\beta_0 + \beta_1 TRT)$

_t	Coef.	Std. Err.	z	p>|z|
trt	−1.527	.398	3.83	0.00
_cons	−2.159	.218	−9.90	0.00

6. In this chapter it was shown in an exponential model that the time to event is $t = [-\log(S(t)] \times (1/\lambda)$ given a fixed value of S(t). Use the output from the AFT form of the model to estimate the median survival time (in weeks) for the treated group (TRT = 1) and the placebo group (TRT = 0).

7. Use the output from the PH form of the model to estimate the median survival time for the treated group (TRT = 1) and the placebo group (TRT = 0). Notice the answers from Questions 6 and 7 are the same, illustrating that the AFT and PH forms of the exponential model are just different parameterizations of the same model.

Questions 8 to 11 refer to a log-logistic AFT model using the data from the Veteran's Administration Lung Cancer Trial. The exposure of interest is treatment status TX (standard = 1, test = 2). The control variables are performance status (PERF), disease duration (DD), AGE, and prior therapy (PRIORTX). These predictors are used in the section on frailty models. The outcome is time to death (in days). The output is shown below.

Log-logistic regression — accelerated failure-time form

Log likelihood = −200.196		LR chi2(5) = 61.31 Prob > chi2 = 0.0000		
_t	Coef.	Std. Err.	z	p>|z|
tx	−.054087	.1863349	−0.29	0.772
perf	.0401825	.0046188	8.70	0.000
dd	.0042271	.0095831	0.44	0.659
age	.0086776	.0092693	0.94	0.349
priortx	.0032806	.0225789	0.15	0.884
_cons	1.347464	.6964462	1.93	0.053
/ln_gam	−.4831864	.0743015	−6.50	0.000
gamma	.6168149	.0458303		

8. State the AFT log-logistic model in terms of S(t) (note gamma = $1/p$).

9. Estimate the acceleration factor γ with a 95% confidence interval comparing the test and standard treatment (TX = 2 vs. TX = 1). Interpret your answer.

10. The AFT log-logistic model is also a proportional odds model. Use the output to estimate the odds ratio (odds of death) comparing the test and standard treatment. Also estimate the survival odds ratio comparing the test and standard treatment.

11. The Akaike Information Criterion (AIC) is a method designed to compare the fit of different models. For this question, three models are compared using the same 5 predictors:
    1. A Weibull model without frailty (presented as Model 1 in the section on frailty models);
    2. A Weibull model containing a frailty component (presented as Model 2 in the section on frailty models); and
    3. The log-logistic model presented above.

Below is a table containing the log likelihood statistic for each model.

Model	Frailty	Number of parameters	Log likelihood
1. Weibull	No	7	−206.204
2. Weibull	Yes	8	−200.193
3. Log-logistic	No	7	−200.196

The goal for this question is to calculate the AIC statistic for each model and select the model based on this criterion. **The AIC statistic is calculated as: −2 log likelihood + 2p** (where p is the number of parameters in the model). A smaller AIC statistic suggests a better fit. The addition of 2 times p can be thought of as a penalty if nonpredictive parameters are added to the model. Each model contains the 5 predictors, an intercept, and a shape parameter. Model 2 contains an additional variance parameter (theta) because a frailty component is included in the model. The log likelihood was unchanged when a frailty component was added to the log-logistic model (not shown in table).

Note that if we are just comparing Models 1 and 2 we could use the likelihood ratio test because Model 1 is nested (contained) in Model 2. The likelihood ratio test is considered a superior method to the AIC for comparing models but cannot be used to compare the log-logistic model to the other two, because that model uses a different distribution.

Which of the three models should be selected based on the AIC?

Questions 12 to 14 refer to a generalized gamma model using the Veterans' data with the same five predictor variables that were used in the model for Questions 8 to 10. The generalized gamma distribution contains two shape parameters (kappa and sigma) that allow great flexibility in the shape of the hazard. If kappa = 1, the model reduces to a Weibull distribution with p = 1/sigma. If kappa = 0 the model reduces to a lognormal distribution. The output is shown below.

Gamma regression — accelerated failure-time form

Log likelihood = −200.626		LR chi2(5) = 52.86 Prob > chi2 = 0.0000		
_t	Coef.	Std. Err.	z	p>|z|
tx	−.131	.1908	−0.69	0.491
perf	.039	.0051	7.77	0.000
dd	.0004	.0097	0.04	0.965
age	.008	.0095	0.89	0.376
priortx	.004	.0229	0.17	0.864
_cons	1.665	.7725	2.16	0.031
/ln_sig	.0859	.0654	1.31	0.189
/kappa	.2376	.2193	1.08	0.279
sigma	1.0898	.0714		

12. Estimate the acceleration factor γ with a 95% confidence interval comparing the test and standard treatment (TX = 2 vs. TX = 1).

13. Use the output to test the null hypothesis that a lognormal distribution is appropriate for this model.

14. A lognormal model was run with the same five predictors (output not shown) and yielded very similar parameter estimates to those obtained from the generalized gamma model shown above. The value of the log likelihood for the lognormal model was −201.210. Compare the AIC of the generalized gamma model, the lognormal model, and the log-logistic model from Question 11 and select a model based on that criterion. Note: each model contains an intercept and five predictors. The generalized gamma distribution contains two additional shape parameters and the log-logistic and lognormal distributions each contain one additional shape parameter (see Question 11 for further details on the AIC).

Questions 15 to 17 refer to a Weibull model using the remission data with treatment as the only predictor (coded TRT = 1 for the test treatment and TRT = 0 for the placebo). In this model both λ and p are modeled as functions of the predictor TRT. The model can be stated in terms of the hazard function: $h(t) = \lambda p t^{p-1}$ where $\lambda = \exp(\beta_0 + \beta_1 \text{ TRT})$ and $p = \exp(\delta_0 + \delta_1 \text{ TRT})$. Typically, the shape parameter in a Weibull model is assumed constant (i.e., $\delta_1 = 0$) across levels of covariates. This model is discussed in the section of this chapter called "Other Parametric Models." The output obtained using Stata is shown below.

Weibull regression — log relative-hazard form

				LR chi2(1) = 1.69	
Log likelihood = −47.063396				Prob > chi2 = 0.1941	
	_t	Coef.	Std. Err.	z	p>|z|
_t					
	trt	−1.682	1.374	−1.22	0.221
	_cons	−3.083	.646	−4.77	0.000
ln_p					
	trt	−.012	.328	−0.04	0.970
	_cons	.315	.174	1.82	0.069

15. Even though λ is parameterized similarly to that in a PH Weibull model, this model is not a PH model because the shape parameter p varies across treatment groups. Show the PH assumption is violated in this model by estimating the hazard ratios for TRT = 0 vs. TRT = 1 after 10 weeks and after 20 weeks of follow-up.
16. Perform a statistical test on the hypothesis $\delta_1 = 0$ (the coefficient for the treatment term for $\ln(p)$). Note: if we assume $\delta_1 = 0$, then the model reduces to the example of the Weibull PH model presented in Section VI of this chapter.
17. Consider the plot of the log negative log of the Kaplan–Meier survival estimates against the log of time for TRT = 1 and TRT = 0. How should the graph look if $\delta_1 = 0$?

Test

Answer the following true or false questions (circle T or F).

T F 1. The accelerated failure time model and proportional hazards model are both additive models.

T F 2. If the survival function is known then the hazard function can be ascertained (and vice versa).

T F 3. If survival time follows a Weibull distribution then a plot of the ln[−ln S(t)] against ln(t) should be a straight line.

T F 4. If the acceleration failure time (AFT) assumption holds in a log-logistic model then the proportional hazards assumption also holds.

T F 5. If the acceleration factor for the effect of an exposure (exposed vs. unexposed) is greater than one, then the exposure is harmful to survival.

T F 6. Let $S_0(t)$ be the survival function for unexposed subjects (E = 0) and let $S_1(t)$ be the survival function for exposed subjects (E = 1). If γ is the acceleration factor comparing E = 1 vs. E = 0 then $S_0(t) = S_1(\gamma t)$.

T F 7. Frailty models are designed to provide an approach to account for unobserved individual-level characteristics.

T F 8. If you include a gamma distributed frailty component to the model, then you will see an additional parameter estimate for the variance of the frailty in the model output.

T F 9. If survival time T follows a Weibull distribution, then ln(T) also follows a Weibull distribution.

T F 10. If a subject is lost to follow-up after 5 years, then the subject is left-censored.

Questions 11 to 17 refer to a Weibull model run with the "addicts" dataset. The predictor of interest is CLINIC (coded 1 or 2) for two methadone clinics for heroin addicts. Covariates include DOSE (continuous) for methadone dose (mg/day), PRISON (coded 1 if patient has a prison record and 0 if not), and a prison–dose product term (called PRISDOSE). The outcome is time (in days) until the person dropped out of the clinic or was censored. The Weibull survival and hazard functions are, respectively, $S(t) = \exp(-\lambda t^p)$ and $h(t) = \lambda p t^{p-1}$ where $\lambda^{1/p} = \exp[-(\alpha_0 + \alpha_1 \text{CLINIC} + \alpha_2 \text{PRISON} + \alpha_3 \text{DOSE} + \alpha_4 \text{PRISDOSE})]$ for the AFT parameterization and $\lambda = \exp[\beta_0 + \beta_1 \text{CLINIC} + \beta_2 \text{PRISON} + \beta_3 \text{DOSE} + \beta_4 \text{PRISDOSE}]$ for the PH parameterization. The Stata output for both the AFT and PH forms of the model are presented as follows:

Weibull regression
accelerated failure-time form

Log likelihood = −260.74854

_t	Coef.	Std. Err.	z	P > |z|
clinic	.698	.158	4.42	0.000
prison	.145	.558	0.26	0.795
dose	.027	.006	4.60	0.000
prisdose	−.006	.009	−0.69	0.492
_cons	3.977	.376	10.58	0.000
/ln_p	.315	.068	4.67	0.000
p	1.370467			
1/p	.729678			

Weibull regression
log relative-hazard form

Log likelihood = −260.74854

_t	Coef.	Std. Err.	z	P > |z|
clinic	−.957	.213	−4.49	0.000
prison	−.198	.765	−0.26	0.795
dose	−.037	.008	−4.63	0.000
prisdose	.009	.013	0.69	0.491
_cons	−5.450	.702	−7.76	0.000
/ln.p	.315	.068	4.67	0.000
p	1.370467			
1/p	.729678			

11. Estimate the acceleration factor with a 95% confidence interval comparing CLINIC = 2 vs. CLINIC = 1. Interpret this result.
12. Estimate the hazard ratio with a 95% confidence interval comparing CLINIC = 2 vs. CLINIC = 1. Interpret this result.
13. Estimate the coefficient for CLINIC in the PH Weibull model using the results reported in the output from the AFT form of the model. Hint: the coefficients for a Weibull PH and AFT model are related $\beta_j = -\alpha_j p$ for the jth covariate.
14. Is the product term PRISDOSE included in the model to account for potential interaction or potential confounding of the effect of CLINIC on survival?
15. Use the output to estimate the median survival time for a patient from CLINIC = 2 who has a prison record and receives a methadone dose of 50 mg/day. Hint: use the relationship that $t = [-\ln S(t)]^{1/p} \times (1/\lambda^{1/p})$ for a Weibull model.
16. Use the output to estimate the median survival time for a patient from CLINIC = 1 who has a prison record and receives a methadone dose of 50 mg/day.
17. What is the ratio of your answers from Questions 15 and 16 and how does this ratio relate to the acceleration factor?

Questions 18 and 19 refer to the Weibull model (in AFT form) that was used for the previous set of questions (Questions 11 to 17). The only difference is that a frailty component is now included in the model. A gamma distribution of mean 1 and variance theta is assumed for the frailty. The output shown on in the following contains one additional parameter estimate (for theta).

Weibull regression
accelerated failure-time form
Gamma frailty

Log likelihood = −260.74854

_t	Coef.	Std. Err.	z	P > |z|
clinic	.698	.158	4.42	0.000
prison	.145	.558	0.26	0.795
dose	.027	.006	4.60	0.000
prisdose	−.006	.009	−0.69	0.492
_cons	3.977	.376	10.58	0.000
/ln_p	.315	.068	4.67	0.000
p	1.370467			
1/p	.729678			
theta	.00000002		.0000262	

Likelihood ratio test of theta=0:
chibar2(01) = 0.00
Prob>=chibar2 = 1.000

18. Did the addition of the frailty component change any of the other parameter estimates (besides theta)? Did it change the log likelihood?
19. A likelihood ratio test for the hypothesis H_0: theta = 0 yields a p-value of 1.0 (bottom of the output). The parameter estimate for theta is essentially zero. What does it mean if theta = 0?

Answers to Practice Exercises

1. T
2. F: The median survival time for the unexposed is 1/3 of the median survival time for the exposed.
3. F: The Cox model is a semiparametric model. The distribution of survival time is unspecified in a Cox model.
4. T
5. F: The hazard is assumed constant in an exponential model.
6. $t = [-\log(S(t)] \times (1/\lambda)$, where $S(t) = 0.5$, and $1/\lambda = \exp(\alpha_0 + \alpha_1 \text{ TRT})$.
   For TRT = 0: estimated median survival = $[-\ln(0.5)] \exp(2.159) = 6.0$ weeks.
   For TRT = 1: estimated median survival = $[-\ln(0.5)] \exp(2.159 + 1.527) = 27.6$ weeks.
7. $t = [-\log(S(t)] (1/\lambda)$, where $S(t) = 0.5$, and $\lambda = \exp(\beta_0 + \beta_1 \text{ TRT}) \Rightarrow 1/\lambda = \exp[-(\beta_0 + \beta_1 \text{TRT})]$.
   For TRT = 0: estimated median survival = $[-\ln(0.5)] \exp[-(-2.159)] = 6.0$ weeks.
   For TRT = 1: estimated median survival = $[-\ln(0.5)] \exp[-(-2.159 - 1.527)] = 27.6$ weeks.
8. $S(t) = 1/(1 + \lambda t^p)$ where $\lambda^{1/p} = \exp[-(\alpha_0 + \alpha_1 \text{ TX} + \alpha_2 \text{ PERF} + \alpha_3 \text{DD} + \alpha_4 \text{AGE} + \alpha_5 \text{PRIORTX})]$.
9. $$\gamma = \frac{\exp[\alpha_0 + \alpha_1(2) + \alpha_2 PERF + \alpha_3 DD + \alpha_4 AGE + \alpha_5 PRIORTX]}{\exp[\ \alpha_0 + \ \alpha_1(1) + \ \alpha_2 PERF + \ \alpha_3 DD + \ \alpha_4 AGE + \ \alpha_5 PRIORTX]}$$
   $$= \exp(\ \alpha_1)$$
   $$\hat{\gamma} = \exp(-0.054087) = 0.95$$
   $$95\%\ \text{CI} = \exp[-0.054087 \pm 1.96(0.1863349)] = (0.66,\ 1.36)$$
   The point estimate along with the 95% CI suggests a null result.
10. The coefficients for a log-logistic proportional odds (PO) and AFT model are related $\beta_1 = -\alpha_1\ p = -\beta_1/\text{gamma}$, where β_1 is the coefficient for TX in a PO model.
    OR = $\exp(-\alpha_1/\text{gamma})$
    estimated OR = $\exp(-0.054087/0.6168149) = 0.92$
    estimated survival OR = $1/[\exp(-0.054087/0.6168149)] = 1.09$.

11. The AIC statistic is calculated as -2 log likelihood $+2$p (where p is the number of parameters in the model). A smaller AIC statistic suggests a better fit. The AIC statistic is shown below for each of the three models.

Model	Frailty	Number of parameters	Log likelihood	AIC
1. Weibull	No	7	−206.204	426.408
2. Weibull	Yes	8	−200.193	416.386
3. Log-logistic	No	7	−200.196	414.392

Based on the AIC, the log-logistic model is selected yielding the smallest AIC statistic at 414.392.

12. $$\gamma = \frac{\exp[\alpha_0 + \alpha_1(2) + \alpha_2 PERF + \alpha_3 DD + \alpha_4 AGE + \alpha_5 PRIORTX]}{\exp[\ \alpha_0 + \ \alpha_1(1) + \ \alpha_2 PERF + \ \alpha_3 DD + \ \alpha_4 AGE + \ \alpha_5 PRIORTX]}$$
$$= \exp(\ \alpha_1)$$
$$\hat{\gamma} = \exp(-0.131) = 0.88$$
$$95\%\ \text{CI} = \exp[(-0.131 \pm 1.96(0.1908)] = (0.60,\ 1.28)$$

13. The generalized gamma distribution reduces to a lognormal distribution if kappa $= 0$.

$H_0:$ kappa $= 0$

Wald test statistic: $z = \dfrac{0.2376}{0.2193} = 1.08$ (from output)

p-value: 0.279 (from output)

Conclusion: p-value not significant at a significance level of 0.05. Not enough evidence to reject H_0. The lognormal distribution may be appropriate.

14. The AIC statistic is shown below for the generalized gamma, lognormal, and log-logistic models.

Model	Number of parameters	Log likelihood	AIC
Generalized Gamma	8	−200.626	417.252
Lognormal	7	−201.210	416.420
Log-logistic	7	200.196	414.392

As in Question 11, the log-logistic model is selected yielding the smallest AIC at 414.392.

15. $h(t) = \lambda p t^{p-1}$ where $\lambda = \exp(\beta_0 + \beta_1 TRT)$ and $p = \exp(\delta_0 + \delta_1 TRT)$
let $\lambda_0 = \exp[\beta_0 + \beta_1(0)]$, $\lambda_1 = \exp[\beta_0 + \beta_1(1)]$let $p_0 = \exp[\delta_0 + \delta_1(0)]$, $p_1 = \exp[\delta_0 + \delta_1(1)]\hat{\lambda}_0 = 0.0458$, $\hat{\lambda}_1 = 0.0085$, $\hat{p}_0 = 1.3703$, $\hat{p}_1 = 1.3539$ (calculated using output)

$$\text{HR (TRT} = 0 \text{ vs. TRT} = 1) = \frac{\lambda_0 p_0 t^{p_0 - 1}}{\lambda_1 p_1 t^{p_1 - 1}}$$

$$\widehat{HR}(\text{as a function of t}) = \frac{(0.0458)(1.3703)t^{0.3703}}{(0.0085)(1.3539)t^{0.3539}}$$

$$\widehat{HR}(\text{t} = 10) = \frac{(0.0458)(1.3703)(10^{0.3703})}{(0.0085)(1.3539)(10^{0.3539})} = 5.66$$

$$\widehat{HR}(\text{t} = 20) = \frac{(0.0458)(1.3703)(20^{0.3703})}{(0.0085)(1.3539)(20^{0.3539})} = 5.73$$

The estimated hazard ratios for RX at 10 weeks and at 20 weeks are different, demonstrating that the hazards are not constrained to be proportional in this model. However, the estimated hazard ratios are just slightly different, suggesting that the PH assumption is probably reasonable.

16. H_0: $\delta_1 = 0$

$$\text{Wald test statistic : } z = \frac{-0.0123083}{0.328174} = -0.04 \text{ (from output)}$$

p-value: 0.970 (from output)

Conclusion: p-value is not significant. No evidence to reject H_0. The PH assumption is reasonable.

17. If the Weibull assumption is met, then the plots should be straight lines with slope p. If $\delta_1 = 0$, then the slope p is the same for TRT = 1 and TRT = 0 and the lines are parallel.